Long-term decision making in the presence of financial disasters

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: https://dspace.lboro.ac.uk/2134/32644

Publisher: © Ilias Chondrogiannis

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
Long-term decision making in the presence of financial disasters

By

Ilias Chondrogiannis

School of Business and Economics
LOUGHBOROUGH UNIVERSITY

A dissertation submitted to Loughborough University in accordance with the requirements of the degree of Doctor of Philosophy in Finance.

SEPTEMBER 2017
This thesis is dedicated to my parents, Konstantinos Chondrogiannis and Maria Chondrogianni, and my sister, Anthi Chondrogianni.

I would not be here without their boundless support, understanding and patience through all these years and they have an enormous share in anything I have accomplished, and will in the future. This thesis would not have been completed without your love, help and encouragement.
Author’s declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University’s Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate’s own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Long Term Decision Making in the Presence of Financial Disasters by
Ilias Chondrogiannis
Thesis
Submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy of Loughborough University
(date)
© by Ilias Chondrogiannis (2017)

SIGNED: .................................................... DATE: ............................................
Acknowledgements

I would like to thank my supervisors, Prof. Mark Freeman and Prof. Andrew Vivian, for their constant and unconditional support and trust during my PhD. They have been excellent supervisors and scientists, and ideal for my character. I am deeply thankful for the opportunity they gave me and very fortunate to have worked with them.

I would also like to thank my professors from the Department of Economics at the University of Athens. In particular, I would like to thank Profs. Nicholas Theocarakis, Manos Athanasiou, Antonis Bartzokas, Manos Kountouris, Nikos Petralias and Alexandros Paneuthimitakis, for their advise and academic support during and long after my undergraduate degree. Their teaching has proven to be unparalleled and of truly rare quality, it has shaped my way of thinking, and I carry it to this day. In addition, I would like to thank Profs. Dennie Oude-Nijhuis and Willem-Bastiaan (Bas) Van Bockel, for their academic support but most importantly for their spirit and friendship. The warmest thanks go to Prof. Giorgos Deligiannidis, to whom I owe more than I can ever repay. This PhD would not have even started, let alone continued and ended, without his help.

Special thanks go to my friends Jade Brooks, Sabrina Heimler, Higor dos-Reis-Leite, Ed Sykes, Thanos Goulas, Ben Wallis and Gaby Wollerink (in random order). We have been here for each other, and we did it together. Also, to my friends in Greece and around the world - you know who you are, you know what I owe you and you are too many to mention. Thank you for making time and distance vanish.

Finally, I would like to thank anyone who has taught me, coached me or crossed blades with me in the Loughborough Students Fencing Club during the past four years. You have been a essential part of that time and you have made my life here much fuller. Special thanks go to Prof. Graham Stretton, Jackie ”Jaxx” Skelton, Matt Eplett and Simon Blackwell for their teaching and coaching, and to fellow epeeists Elsa Phillipson and Matt Dalton for their patience, good will and countless lessons in humility.

And thanks to Guiness, Lagavulin, Talisker, Ardbeg, Chianti and tsipouro.
Abstract

The research question focuses on three areas. First, what is the most appropriate model and estimation method for studying portfolio optimisation under tail risk with an aim towards managerial incentives. Second, how outcomes differ for investors who take jumps into account compared to those who do not. Third, how managerial incentives in the form of fees and compensation structures create a conflict of interest between investors and funds in the presence of jumps, leading to a need for policy suggestions. To answer those questions the thesis builds up from a CARA single-state model to an SV model to an SVCJ model with jumps in returns and volatility, leverage and heteroskedasticity. The model and its SV only counterpart is estimated via MCMC. A closed-form solution for the portfolio weights is derived and used in subsequent simulations. The results are that the investor always has an incentive to knowingly ignore tail risk in terms of wealth but never in terms of utility, the manager has an incentive in the short- and mid-run to undertake excess risk but not in the long-run, the criteria for the incentive horizon are risk aversion and how investor wealth moves between funds, and policy suggestions are made based on those grounds.
# Table of Contents

<table>
<thead>
<tr>
<th>List of Tables</th>
<th>xiii</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>xv</td>
</tr>
</tbody>
</table>

## 1 Introduction
1.1 Research question ........................................ 1  
1.2 General direction of the thesis ............................. 1  
1.3 Importance of the research question ......................... 2

## 2 Literature review
2.1 Essential literature ....................................... 5  
2.2 The Equity Premium puzzle as an introduction to jumps literature ... 13  
  2.2.1 The Equity Premium Puzzle in more detail ............... 13  
  2.2.2 Deviations from normality and jumps .................... 16  
  2.2.3 Further relevant considerations about jumps .......... 19  
2.3 Jump diffusion models in the literature .................... 20  
  2.3.1 Jump diffusion specifications and their properties .... 21  
  2.3.2 Jumps as a tool, their research scope and interpretative power . 22  
  2.3.3 Time variability in jump frequency ...................... 26  
  2.3.4 Further extensions, comparisons and practical applications . 27  
2.4 Summary ....................................................... 30

## 3 Isolating fat tails as factor
3.1 Research aims and methods .................................. 33  
3.2 The CARA model and its variations .......................... 36  
  3.2.1 Single state optimisation ................................ 36  
  3.2.2 Optimal weights for each case ........................... 37
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.3</td>
<td>Connecting the weight formulas</td>
<td>39</td>
</tr>
<tr>
<td>3.3</td>
<td>The two-state Markov case</td>
<td>40</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Optimal weights</td>
<td>41</td>
</tr>
<tr>
<td>3.4</td>
<td>Simulation setup and results</td>
<td>43</td>
</tr>
<tr>
<td>3.5</td>
<td>Conclusion</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>Estimation method and preliminary results</td>
<td>47</td>
</tr>
<tr>
<td>4.1</td>
<td>The model</td>
<td>47</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Introduction to Bayesian analysis</td>
<td>47</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Gibbs and Metropolis-Hastings sampling</td>
<td>50</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Reversibility and burn-in periods</td>
<td>54</td>
</tr>
<tr>
<td>4.1.4</td>
<td>Advantages, disadvantages and justification of method selection</td>
<td>55</td>
</tr>
<tr>
<td>4.1.5</td>
<td>MCMC and Maximum Likelihood/ General Method of Moments</td>
<td>57</td>
</tr>
<tr>
<td>4.1.6</td>
<td>MCMC and GARCH</td>
<td>58</td>
</tr>
<tr>
<td>4.2</td>
<td>Technical application - Stochastic Volatility</td>
<td>59</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Derivation of posterior distributions</td>
<td>59</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Sampling methodologies</td>
<td>64</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Alterations and expansions</td>
<td>65</td>
</tr>
<tr>
<td>4.3</td>
<td>Results and conclusion</td>
<td>67</td>
</tr>
<tr>
<td>5</td>
<td>Model estimation and optimal portfolio weights</td>
<td>71</td>
</tr>
<tr>
<td>5.1</td>
<td>The model</td>
<td>71</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Basic set-up</td>
<td>71</td>
</tr>
<tr>
<td>5.1.2</td>
<td>The SVCJ model</td>
<td>72</td>
</tr>
<tr>
<td>5.1.3</td>
<td>The SV model without jumps</td>
<td>72</td>
</tr>
<tr>
<td>5.1.4</td>
<td>Further discussion of features and advantages</td>
<td>73</td>
</tr>
<tr>
<td>5.2</td>
<td>Estimation and discussion</td>
<td>74</td>
</tr>
<tr>
<td>5.2.1</td>
<td>The algorithm and the data</td>
<td>74</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Estimation results</td>
<td>75</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Discretisation bias and validity - jumps in very long, low frequency data sets</td>
<td>76</td>
</tr>
<tr>
<td>5.3</td>
<td>Optimal portfolio weights</td>
<td>78</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Utility and stochastic processes</td>
<td>78</td>
</tr>
<tr>
<td>5.3.2</td>
<td>The Bellman equation and optimal weights</td>
<td>79</td>
</tr>
<tr>
<td>5.3.3</td>
<td>The weight result relative to the literature</td>
<td>81</td>
</tr>
<tr>
<td>5.3.4</td>
<td>Characteristics, properties and values of the EJP solution</td>
<td>83</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>5.4</td>
<td>Summary of results</td>
<td>86</td>
</tr>
<tr>
<td>5.5</td>
<td>Conclusion</td>
<td>88</td>
</tr>
<tr>
<td>6</td>
<td>Investor and managerial incentives to ignore jumps</td>
<td>91</td>
</tr>
<tr>
<td>6.1</td>
<td>Mutual and hedge funds literature review</td>
<td>92</td>
</tr>
<tr>
<td>6.1.1</td>
<td>Academic literature</td>
<td>92</td>
</tr>
<tr>
<td>6.1.2</td>
<td>Current prospects according to the industry</td>
<td>97</td>
</tr>
<tr>
<td>6.1.3</td>
<td>Summary</td>
<td>98</td>
</tr>
<tr>
<td>6.2</td>
<td>General setup</td>
<td>99</td>
</tr>
<tr>
<td>6.3</td>
<td>Jumps and the investor</td>
<td>100</td>
</tr>
<tr>
<td>6.4</td>
<td>Jumps and the manager</td>
<td>101</td>
</tr>
<tr>
<td>6.5</td>
<td>Conclusion and summary of results</td>
<td>105</td>
</tr>
<tr>
<td>7</td>
<td>Conclusion and further extensions</td>
<td>109</td>
</tr>
<tr>
<td>7.0.1</td>
<td>Results</td>
<td>109</td>
</tr>
<tr>
<td>7.0.2</td>
<td>Further extensions</td>
<td>112</td>
</tr>
<tr>
<td>Appendix A</td>
<td>Appendix for Chapter 3</td>
<td>115</td>
</tr>
<tr>
<td>Appendix B</td>
<td>Appendix for Chapter 4</td>
<td>123</td>
</tr>
<tr>
<td>B.1</td>
<td>Tables and Figures</td>
<td>123</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Appendix for Chapter 5</td>
<td>131</td>
</tr>
<tr>
<td>C.1</td>
<td>n-dimensional Ito’s lemma in the Bellman equation</td>
<td>131</td>
</tr>
<tr>
<td>C.2</td>
<td>Algebra of the Bellman equation</td>
<td>132</td>
</tr>
<tr>
<td>C.3</td>
<td>SVCJ posteriors</td>
<td>133</td>
</tr>
<tr>
<td>C.4</td>
<td>SV posteriors</td>
<td>140</td>
</tr>
<tr>
<td>C.5</td>
<td>Figures and tables</td>
<td>141</td>
</tr>
<tr>
<td>Appendix D</td>
<td>Appendix for Chapter 6</td>
<td>147</td>
</tr>
<tr>
<td>D.1</td>
<td>Tables 1.1 - 1.4</td>
<td>147</td>
</tr>
<tr>
<td>D.2</td>
<td>Tables 2.1 - 2.4</td>
<td>149</td>
</tr>
<tr>
<td>D.3</td>
<td>Table 3.1 - 3.4</td>
<td>151</td>
</tr>
<tr>
<td>D.4</td>
<td>Table 4.1 - 4.4</td>
<td>153</td>
</tr>
<tr>
<td>D.5</td>
<td>Table 5.1 - 5.4</td>
<td>155</td>
</tr>
<tr>
<td>Bibliography</td>
<td></td>
<td>161</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1 Ex post expected utility across combinations of weights and realised distributions. Bold denotes the winner of the row. Table is read from left to right</td>
<td>118</td>
</tr>
<tr>
<td>B.1 MCMC parameters for the SV model for different horizons and data sets</td>
<td>123</td>
</tr>
<tr>
<td>B.2 Summary statistics</td>
<td>123</td>
</tr>
<tr>
<td>C.1 MCMC parameters for the SVCJ model. Values reported as daily percentages and annual decimals compared to the values of Eraker et al (2003) and Brooks and Prokopszuk (2012). Where necessary, the parameters are converted as described in Chapter 4</td>
<td>141</td>
</tr>
<tr>
<td>C.2 MCMC parameters for the SV model. Values reported as daily percentages and annual decimals compared to the values of Eraker et al (2003) Where necessary, the parameters are converted as described in Chapter 4</td>
<td>142</td>
</tr>
<tr>
<td>C.3 Sample Summary Statistics</td>
<td>142</td>
</tr>
<tr>
<td>C.4 Optimal portfolio weights for the SVCJ and SV models (1980 - 2016) for $r = 2%$ compared to the weights corresponding to the EJP parameters ($r = 4.5%$) and the Liu, Longstaff and Pan (2003) replicated parameters. The LLP weights refer to S&amp;P500 options data between 1-1-1987 and 31-12-1996, for which the methodology, frequency and parameter estimation is incomparable. They are referred here only as a successful replication of an existing result</td>
<td>143</td>
</tr>
</tbody>
</table>
C.5 Replication of existing results in the literature. Branger and Hansis (2012, 2015) transform the EJP parameters from percentage log returns to annual decimals that correspond to the LLP formulation that includes volatility and jumps premia in the returns process and stochastic arrival intensity. 

\[ \lambda_{EJP} = \lambda_{LLP} \bar{V}, \quad \bar{V} = \theta + \mu \lambda/\kappa, \]

ERP = Equity Risk Premium. Common parameters are in Table 1 and only the additional parameters of the LLP version are reported here.

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.1.1</td>
<td>( \gamma = 5 )</td>
<td>147</td>
</tr>
<tr>
<td>D.1.2</td>
<td>( \gamma = 4 )</td>
<td>147</td>
</tr>
<tr>
<td>D.1.3</td>
<td>( \gamma = 3 )</td>
<td>148</td>
</tr>
<tr>
<td>D.1.4</td>
<td>( \gamma = 2 )</td>
<td>148</td>
</tr>
<tr>
<td>D.2.1</td>
<td>( \gamma = 5, 2+20% ) fees, Symmetric wealth transfer function</td>
<td>149</td>
</tr>
<tr>
<td>D.2.2</td>
<td>( \gamma = 5, 1+10% ), Symmetric wealth transfer function</td>
<td>149</td>
</tr>
<tr>
<td>D.2.3</td>
<td>( \gamma = 5, 2+20 ), Asymmetric wealth transfer function</td>
<td>150</td>
</tr>
<tr>
<td>D.2.4</td>
<td>( \gamma = 5, 1+10 ), Asymmetric wealth transfer function</td>
<td>150</td>
</tr>
<tr>
<td>D.3.1</td>
<td>( \gamma = 4, 2+20% ), Symmetric wealth transfer function</td>
<td>151</td>
</tr>
<tr>
<td>D.3.2</td>
<td>( \gamma = 4, 1+10% ), Symmetric wealth transfer function</td>
<td>151</td>
</tr>
<tr>
<td>D.3.3</td>
<td>( \gamma = 4, 2+20 ), Asymmetric wealth transfer function</td>
<td>152</td>
</tr>
<tr>
<td>D.3.4</td>
<td>( \gamma = 4, 1+10 ), Asymmetric wealth transfer function</td>
<td>152</td>
</tr>
<tr>
<td>D.4.1</td>
<td>( \gamma = 3, 2+20 ), Symmetric wealth transfer function</td>
<td>153</td>
</tr>
<tr>
<td>D.4.2</td>
<td>( \gamma = 3, 1+10 ), Symmetric wealth transfer function</td>
<td>153</td>
</tr>
<tr>
<td>D.4.3</td>
<td>( \gamma = 3, 2+20 ), Asymmetric wealth transfer function</td>
<td>154</td>
</tr>
<tr>
<td>D.4.4</td>
<td>( \gamma = 3, 1+10 ), Asymmetric wealth transfer function</td>
<td>154</td>
</tr>
<tr>
<td>D.5.1</td>
<td>( \gamma = 2, 2+20 ), Symmetric wealth transfer function</td>
<td>155</td>
</tr>
<tr>
<td>D.5.2</td>
<td>( \gamma = 2, 1+10 ), Symmetric wealth transfer function</td>
<td>155</td>
</tr>
<tr>
<td>D.5.3</td>
<td>( \gamma = 2, 2+20 ), Asymmetric wealth transfer function</td>
<td>156</td>
</tr>
<tr>
<td>D.5.4</td>
<td>( \gamma = 2, 1+10 ), Asymmetric wealth transfer function</td>
<td>156</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1 Normal distribution terminal wealth for the same series when the jump occurs in different times</td>
<td>115</td>
</tr>
<tr>
<td>A.2 Gamma distribution terminal wealth for the same series when the jump occurs in different times</td>
<td>116</td>
</tr>
<tr>
<td>A.3 Inverse Gaussian distribution terminal wealth for the same series when the jump occurs in different times</td>
<td>117</td>
</tr>
<tr>
<td>A.4 Normal distribution - pdf plot for N(0,1)</td>
<td>119</td>
</tr>
<tr>
<td>A.5 Gamma distribution - pdf plots for different parameters</td>
<td>120</td>
</tr>
<tr>
<td>A.6 Inverse Gaussian distribution - pdf plots for different parameters</td>
<td>121</td>
</tr>
<tr>
<td>B.1 Markov chains for $\mu, \phi, \sigma^2$ (top - bottom), not corrected sample, 1991 - 2007</td>
<td>124</td>
</tr>
<tr>
<td>B.2 Markov chains for $\mu, \phi, \sigma^2$ (top - bottom), corrected sample, 1991 - 2007</td>
<td>125</td>
</tr>
<tr>
<td>B.3 Smoothed volatility states (top) and absolute returns (bottom), not corrected sample, 1991 - 2007</td>
<td>126</td>
</tr>
<tr>
<td>B.4 Smoothed volatility states (top) and absolute returns (bottom), corrected sample, 1991 - 2007</td>
<td>127</td>
</tr>
<tr>
<td>B.5 Smoothed volatility states (top) and absolute returns (bottom), 2005 - 2015</td>
<td>128</td>
</tr>
<tr>
<td>B.6 Markov Chains ($\mu, \phi, \sigma^2$) (top - bottom), 2005 - 2015</td>
<td>129</td>
</tr>
<tr>
<td>C.1 SVCJ Portfolio weight plot for $\gamma = 5, T = 5$ and $T = 9.000$</td>
<td>144</td>
</tr>
<tr>
<td>C.2 MCMC Chains - (Left, top to bottom) $\lambda, \rho, \sigma^2, \mu_V, \mu_Y$, (Right, top to bottom) $\mu, \rho, \sigma^2, \alpha, \beta$</td>
<td>145</td>
</tr>
<tr>
<td>D.1 Symmetric wealth transfer function, cap at $\pm 10%$, $\delta = 0.25$</td>
<td>157</td>
</tr>
<tr>
<td>D.2 Asymmetric wealth transfer function, upper cap at 10% lower cap at $-15%$, $\delta = 0.25, \tau = 0.5$</td>
<td>157</td>
</tr>
<tr>
<td>D.3 Sigmoid wealth transfer function, $\delta = 0.5$</td>
<td>158</td>
</tr>
</tbody>
</table>
List of Figures

D.4 TFUJ histogram, $\gamma = 3, T = 24, 1 + 10\%$ fees, symmetric . . . . . . . . . . . 159
D.5 TFUNJ histogram, $\gamma = 3, T = 24, 1 + 10\%$ fees, symmetric . . . . . . . . . . . 159
Chapter 1

Introduction

1.1 Research question

The research question can be summarised as follows. The overarching question is how managerial incentives in the form of fees and compensation structures create a conflict of interest between investors and funds in the presence of jumps and lead to a motivation for the manager to knowingly undertake excess risk and aim for higher compensation. This theme generates two secondary, yet important, questions. The first one is what is the most appropriate model and estimation method for studying portfolio optimisation under tails with an aim towards managerial incentives. The second one is how outcomes differ for investors who take jumps into account compared to those who do not.

1.2 General direction of the thesis

The path the thesis selects in order to answer the main question is to show that a certain amount of technical complexity is essential and necessary, and is not introduced for its own sake or as a means to impress. The first set of empirical results comes from an attempt to identify the effect of jumps on a CARA investor with a one-period horizon. When that proved to be inadequate, the next attempt focused on the effect of heteroskedasticity and volatility clustering as a way to match a series and generate sufficient tails. This method was also unsuccessful and thus established the need to expand on a continuous time framework that replicates many stylized facts on volatility and returns simultaneously. Model complexity unavoidably called for estimation complexity, since standard maximum likelihood methods were unsuitable or inaccurate. The underlying idea of a Markov switch in the first empirical attempt evolved to MCMC Bayesian estimation in the later
stages and also underlines a wide body of literature dealing with continuous time Markov processes in one form or another.

Although the main contribution of the thesis is not technical, the process necessarily engages with technical deficiencies and improvements. In that front, a new way to tackle a feature of existing solutions is proposed, which leads to a more tractable solution for the model at hand. This translates into a buy-and-hold strategy under the probability of a disaster during the investment period. This paves the way for the manager and investor simulations of the last chapter, where jumps are identified as a generator of diverging motives. Further intuition is added by using a mechanism for investor reactions when the fund they have invested in underperforms. At this stage the thesis has put all the necessary components in place and can focus on investor and managerial incentives over time and across different compensation schemes.

The need to build up is essential also because the route highlights the issues that appear in the final results. The inability of the one-period optimiser to distinguish jumps is reminiscent of the manager’s incentive to ignore jumps in the short-term, reasoning that the shock will not take place ”on average”. Neither one will take them into account. The failure of stochastic volatility to deal with outliers shows the limits of extracting information and the power of exogenous events or ”flash crashes”. The reasoning behind the selection of one model or another highlights the technical limitations and means that must, ultimately, serve the ends. The red thread that connects the chapters of the thesis culminates into a result that leads to clear policy suggestions on how to deal with excess risk taking. The suggestions rely on both the empirical and the theoretical and are solidly founded in the thesis.

1.3 Importance of the research question

The importance of the question is highlighted by the answer itself. The investor and the manager are found to have diverging incentives in the short- and mid-term, while the incentive for the manager dissipates for long horizons. Although the preferences of the investor and the manager are identical, the fact that they draw utility from different amounts (portfolio wealth and managerial fees respectively) creates a moral hazard issue. The source of the divergence is not the level of managerial compensation but the degree of risk aversion and how investors react to positive or negative relative returns. If investors react the same to good or bad results, the incentive is greater for the manager than when they react more viciously to negative than positive results, because more investors
abandon the losing fund and, thus reduce its wealth and consequently the base upon which managerial fees are calculated.

One of the main components of the answer is the isolation and use of jumps as the source of risk which generates those incentives. This is achieved by estimating and comparing a stochastic volatility model without jumps and a stochastic volatility model with jumps in both the return and the volatility process. This particular method of employing and replicating jumps is justified in detail, and the origins of this justification can be detected in Weitzman (2009a), Barro (2006), Martin (2012), Wachter (2013), Julliard and Ghosh (2012) and Jondeau and Rockinger (2012) among others. The basic idea is that an increased probability of jumps, or fat tails in a distribution, leads to an investor being willing to pay an infinite amount today in order to secure against a future catastrophic event. Since observations on tails are rare, their actual shape and the estimation of the related parameters is a matter for debate, leading to very high sensitivity in parameters and model selection. In a financial context, the equity premium puzzle and its possible solutions provide fertile ground for the discussion. Jumps are found to provide an improvement in model accuracy and partially explain certain features of financial markets, but at the same time they are not a panacea and there is still debate about their proper representation. The thesis starts from a very basic setup and builds upwards, since extreme events are notably absent from standard optimisation and decision making methods. This denotes a need to enrich and expand a simple method, as well as select the proper model.

A second component is a set of portfolio allocations and strategies that allow for direct comparisons between cases. This is achieved by deriving a closed-form solution for optimal weights. It is shown that the employment of each model leads to significantly different allocations, which are found to be time invariant. Despite the extensive technical literature and research that focuses on portfolio allocation and model fitting, the aspect of moral hazard and investor/manager choices is absent from the literature. Notably, the factors that affect and lead to the final results are both technical and conceptual. Risk aversion (a parameter of great importance in any model) has an effect at least as important as the reaction and perception of investors to losses and profits. Therefore, the importance of the research question lies not only in conducting a fitting exercise but identifying a wider set of elements that need to be taken into consideration, and consist the environment in which jumps do have a strong effect.
Chapter 2

Literature review

The literature review can be separated in four main areas. The first part discusses a number of key papers that outline the origins and importance of the research question, and provide a guidance to the gaps in the literature that will be addressed by the PhD. The second part discusses the equity premium puzzle and its expansions, as well as portfolio optimisation under extreme events. The third part focuses on jumps-related literature, with a brief reference to other methods of modeling and incorporating tail risk. The fourth part is technical and will discuss finance and mathematics literature exclusively with regards to estimation. The technical part can be found in the third chapter where the estimation methodology is explained, with some additional papers in the fourth chapter.

2.1 Essential literature

The foundation of the research question is two-fold and can be traced in environmental economics and asset pricing or portfolio management. The basis of the former is Weitzman’s ”Dismal Theorem”, which states that tail thickness causes the investor to discount future catastrophic events at a very high rate, even for moderate risk aversion (Weitzman (2009a), Weitzman (2009b), Weitzman (2009c)). The result is that an investor is willing to pay a virtually infinite amount of money today to secure against a long-term disaster of very low probability. The argument is based upon the selection of a utility function and/or a probability distribution function that unavoidably imposes restrictions on how tail risk is perceived, estimated and evaluated, and how different structural assumptions may lead to exploding results under very fat tails. Unperceived or exogenous contemporary changes in parameters may lead to ill specified models and outcomes. The
essence is that uncertainties are linked, their effects are greater than the sum of their parts and attempts to limit those effects by restricting parameter calibration do not (and perhaps cannot) address the issue. Since such restrictions or limitations are very rarely justified a priori, the result is an individual that is infinitely risk averse. The “dismal theorem” states that societies are willing to pay infinite wealth in order to hedge against great unexpected disasters (dominance effect). From a portfolio management point of view, an investor operating in such a world may reverse the argument and be willing to accept a less-than-infinite premium being burdened less in insurance terms), essentially betting that the disaster will not take place during his lifetime.

The discussion expands to areas beyond the scope of the research question, but a certain string of literature that carries on to this day offers insight. Nordhaus (2009) formulates a reply focusing on the paradigm and contextual limitations of the Theorem, highlighting the use of CRRA utility functions, parameter hypersensitivity, combination of factors needed and very fat-tailed distributions as prerequisites for the argument to hold. Weitzman (2009c) argues that the very same uncertainty pointed out by Nordhaus is the same feature that causes Cost-Benefit Analysis to be such a precarious and imprecise alternative, despite its central position in modern economic and environmental analysis.

Keen structural research and constant reassessment of models may relieve pressure on extreme outcomes, but the sheer complexity of the task is so great that it is unlikely to achieve something better than an approximation without eliminating tail risk and effects. The debate expands on the specifics of environmental research. It is important, however, to note some contributions with clear parallels to portfolio optimisation and long-term discounting. Karp (2009) notes that discounting should be seen as hyperbolic, not constant, therefore climate policy should be seen as an intergenerational, not intertemporal optimisation game, and that “the likelihood of intra-generational conflict arising from optimal climate policy may be exaggerated under two circumstances: (i) if we fail to recognize that we can internalize climate damages by changing the composition of investment between man-made and natural capital, or (ii) if we fail to recognize that the correct business as-usual baseline involves investment decisions that internalize neither future nor current damages”. Nordhaus (2012) focuses on policy implementation and finds no evidence that outcomes based on the tails of the distribution have indeed dominated smoother, more normal movements. However, he emphasises on the usefulness of the Dismal Theorem as a word of caution against cases of high uncertainty, the catastrophic events they imply and the differences between frameworks. Horowitz and Lange (2014) introduce available technological and policy options to deal with climate change, in the
elementary form of a term that generates a sure unit of consumption in the future using a
unit of today’s consumption as input. When future output is not certain, the range of risk
aversion that generates the same results as Weitzman’s is narrower. Cost-Benefit Analysis
must, therefore, employ not just a stochastic discount factor but a joint probability
between consumption and investment. Under that structure, society is rarely willing to
proceed into the infinite exchange between present and future consumption implied by
the Dismal Theorem.

The entire debate around the Dismal Theorem is underlined by (no) policy imple-
mentation, the potential effects and the appropriate course of action. This very limited
selection of papers is used to highlight the tangents with the economic and financial
concepts of allocation between generations, long-term discounting and the importance
of selecting a suitable discount factor, correctly assessing the magnitude and frequency
(shape of tails) of extreme events and how to cover against disasters. As it will become
apparent later on, the discussion is reminiscent of the parallel debate in finance about
the use and nature of jumps and calibration/ estimation problems. The importance of
tail risk and the message of the Theorem is that downside risk and catastrophic losses
(with catastrophe in need of a definition) are certainly non-negligible.

Although tail thickness in asset prices has been observed since Mandelbrot (1963),
it is only in the last decades that it has received extensive modeling attention and has
been used as a tool to address a number of empirical issues. The second foundation
can be identified in the Equity Premium Puzzle (Mehra and Prescott (1985)) and its
mirror image, the Risk-Free Rate Puzzle (Weil (1989)). Both papers will be discussed in
detail in the next part, but as a very brief introduction the former records the failure
of existing asset pricing models to capture the gap between stocks and bonds returns
(estimated at roughly 6% at the time) and the latter shows that if a model is estimated
by using historical stocks returns as benchmark, the result is a risk-free rate much
higher than the one observed. The conclusion underlined the significant limitations of
the standard models in the literature, and a reconciliation was proposed by Rietz (1988).
The inclusion of an additional disastrous state in the Mehra and Prescott setup could
be able to generate excessive equity premia. This additional state was, in essence, the
introduction of a fatter tail in the returns distribution contrary to the staple normal
distribution. However, as Mehra and Prescott (1988) note, the solution falls short in a
number of ways. The suggested disasters were of unprecedented magnitude (50%, 75%,
100% of GDP), risk aversion was very high, unanticipated inflation and its effect on
nominal and real bill rates was ignored and historical support was lacking. However,
beyond this critique, a fundamental issue both sides faced was the smooth consumption and GDP time series used as data. It is a well-observed fact that such series cannot generate enough variance and volatility for jumps to be generated and estimated in a sensible matter, leading to a need for implausible assumptions and findings.

The argument of Rietz was brought to the forefront by Barro (2006), who addresses the Equity Premium Puzzle again but tries to adjust for the arbitrary values of Rietz. A model that includes jumps is calibrated according to historical data on disaster frequency and size, using a time additive power utility function and a random walk process with drift and stationary variance for consumption (or future equity claims) where jumps are i.i.d. The dividend return rate is set equal to the equity return rate. The data is annualised, which adds up the multi-year effect of a disaster and treats it as instantaneous, leading to an increased equity premium. The asset process is the simplest form that can include jumps and the Lucas framework puts the model in the same category as the Mehra and Prescott approach. The results show that jumps can improve on the equity premium puzzle estimations and provide a potential explanation. Gabaix (2012) follows the Barro line of reasoning by moving on to a set of puzzles in Macro-Finance and expands to a time variable probability of disasters, contrary to the constant calibration of the above. He expands the framework to incorporate time variability in disaster probability and Epstein-Zin preferences instead of power utility. The findings on ten puzzles in finance support the idea that including disaster risk improves the results in many cases. However, the model is calibrated, not estimated, and certain features of asset prices are not included due to its parsimonious nature. An important comment is the effect of time variability and how a (constant or stochastic) probability of disaster can be calibrated properly. The discussion expands slightly on political measures of disaster risk and disaster risk measured by tail behaviour. Both these factors have a part in the literature review, with the latter receiving particular attention in the technical section.

Julliard and Ghosh (2012) contradict Barro’s solution on the basis of the C-CAPM rejection under empirical likelihood, loss of information due to annualised data, the fact that disasters should occur much more often and the execution of two estimations with international data that cannot generate neither disasters of sufficient magnitude nor equity premia. A distribution that would rationalize the equity premium puzzle should have a fatter left tail and be skewed to the left (assigning higher probabilities to disasters and implying low consumption and low stock returns). It is true that the paper records a number of reverse-engineering failures. On the other hand, the data manipulation and fitting in these methods seems to be extensive. In addition, the intuition behind
mean-variance optimization methods such as the CAPM is quite standard, since it ignores higher moments and additional measures, so jump information is likely not included in the covariance matrix no matter the data set or the assets selected. The basic intuition behind the data set, however, and how disaster data should be treated (annualized or not) is valid and a matter of consideration. In that context it could be said that the equity premium is not necessarily generated by a fear for disasters but from a fear for recessions. However, if disasters are considered as exogenous/unpredictable while recessions as tractable from observing fundamentals, a question arises about both investor behaviour in the bonds and stock markets respectively, as well as illusion issues. Agent heterogeneity might not be enough of an explanation.

Some additional space needs to be dedicated to Barro and Ursúa (2008) as well as the critique and commentary by Blanchard and Constantinides (2008), since it contains the dataset in question which has been used in subsequent research. The paper expands on the Barro (2006) results and data but keeps the same model. The critique focuses on the inability of the Lucas framework to cater for such a research (Blanchard) since the process includes a very precise, even counterfactual, calibration of data. This "peak to trough" calibration is applied on the first year of the disaster, not the last, and is multiplied by the first year rate of return (Constantinides) ending up in an amplification of the disaster effects and a double-counting of the total effect.

Wachter (2013) uses time variability in disaster risk as well, causing the equity premium to follow the same pattern. Epstein-Zin preferences are used to disentangle time preference and risk aversion and thus a well known limitation of the power utility function is avoided (a high price-dividend ratio implies high excess returns). There are two uncorrelated stochastic processes, one for the asset price and one for the arrival intensity, with a single source of risk. The constant mean and variance of the former means that during normal periods without jumps volatility is constant, but the occurrence of a jump adds an additional time varying term on the equity premium. However variance is not explicitly modelled, with the standard deviation term in the process being constant. Compared to the other papers of this section, the model and methodology for deriving optimal portfolio is the closest to the one adopted in the thesis but the aims are vastly different. The focus is on time-varying equity premia, dividend yields, matching stylised facts and generating sufficient volatility, under the very limiting assumption of constant volatility in an attempt to isolate the effects of jumps and their stochastic arrival intensity. The model is again calibrated by using Barro (2006) and Barro and Ursúa (2008) parameters. Its importance lies on being a complete example of the scope
CHAPTER 2. LITERATURE REVIEW

and structure of models with jumps, the existence of closed form solutions subject to the setup, its successes and limitations.

Martin (2012) focuses on an issue highlighted by all of the above papers, particularly Weitzman (2009a): the effect of calibration and parameter sensitivity. The introduction of jumps in any form imposes tail thickness and in essence calls for the calibration or estimation of a probability area where very few observations are found. From a statistical point of view, an assumption on the shape of the distribution (if a known fat-tailed distribution is assumed) or the way tails are thickened (by a Poisson or Levy jump process) may or may not be justified, with both outcomes being on uncertain grounds. Untractable tail degeneration or the effect of fat tails on the hump of the distribution (both issues discussed in great detail in Kemp (2011)) are existing issues that may lead to erroneous results and have no clear solution, since the selection of a distribution or a process immediately imposes underlying assumptions on the interpretation of the data at hand and the nature of the phenomena in question. Martin (2012) uses Epstein - Zin preferences, like Wachter (2013) and an i.i.d consumption growth process, like Barro (2006) but focuses on the resulting distribution and higher moments. Contradicting Nordhaus (2012) and his environmental framework and data, Martin finds strong evidence that tails have a significant impact on asset prices when the focus is on stocks prices and that parameter sensitivity is huge. Therefore, slight deviations can have a great impact on results and model validity. A final aspect of using tailed instead of (log)normal distributions is the latent information that can be extracted by studying higher moments.

The starting point of Jondeau and Rockinger (2006),Jondeau and Rockinger (2009),Jondeau and Rockinger (2012) is the fact that mean-variance optimisation a la Markowitz under Constant Relative Risk Aversion utility functions becomes more unreliable the more the asset distribution deviates from normality, and skewness/ kurtosis become more apparent. Stochastic higher moments add to the complexity but apparently contain additional economic importance, leading to decision making by using distribution timing as well as market and volatility timing: predictions of the underlying features of the returns distribution and its overall shape (skewness and kurtosis) opposed to predictions based on expected returns (mean) and expected volatility. The paper offers an example of how a GARCH-DCC model structure with unexpected returns innovations coming from an asymmetric, fat-tailed skewed t distribution can be used in extreme events analysis. Two debatable points are the assumption of constant expected returns, to allow focus on the higher moments, and the sample period, which might not contain a sufficient number of jumps. Although jumps are not explicitly modelled, the tail
thickness is an accurate representation. The GARCH structure of the covariance matrix also allows for volatility clustering and dynamic correlations. The maximisation cases are a joint normal distribution (corresponding to mean-variance optimisation), a Skewed-t distribution (forecasting both the time-varying covariance matrix and the conditional distribution of asset returns), and a benchmark case of mean-variance strategy with constant weights estimated through sample moments throughout the period, and the results are clearly in favour of the latter. The joint normal case is estimated by a standard two-step GARCH-DCC process, while for the t case the parameters in the first and second moments must be estimated jointly with those in the joint distribution. For that purpose a Michaud (1998) resampling methodology is used. The comparison reveals that bad news affect volatility more than good news and large negative shocks are followed by a decrease in kurtosis, while positive shocks are followed by increased kurtosis and volatility timing provides an increase in utility.

Foster and Young (2012) discuss the reliability of downside risk tests in portfolio returns, when the manager has the option to use derivatives or leverage in order to cover highly risky positions. They test whether compound excess returns have been generated by a martingale process with zero conditional expectation or have a non-positive and positive at some points expectation (they are the result of a superior strategy). The intuition places accidental returns against returns of leveraged positions through options that cover downside excess risk. The test does not use assumptions about distributions and is strategy-proof. However, this may lead to quite lax confidence intervals, which accounts for unobserved tail risk and attempts to improve confidence lead to implausibly high positive returns. In other words, in a non-transparent world a manager needs to exhibit an enormous size of returns in order to pass the proposed test and prove that he is indeed skillful. More transparency and more information about tail risk would lower that threshold. Two interesting observations is that the leveraged version of the test demonstrates the least power loss when lognormality is assumed, while it is asymptotically as strong as the t-test.

This diverse and contradicting literature manages to underline both the key themes of the thesis and the areas of advancement as well as, via their absence, the gaps that this research manages to cover. Weitzman (2009a) emphasises the profound effect rare extreme events can have in decision making and recourse (portfolio) allocation, and the challenges they pose in selecting appropriate instruments. Rietz (1988) and Barro (2006) bring this line of reasoning in the contemporary context of mean-variance optimisation in an attempt to improve theoretical deficiencies, and despite their shortcomings they
show that rare events can provide an explanation. Julliard and Ghosh (2012) note the
importance of selecting a dataset with appropriate properties and plausible calibration
values (jump frequency in particular). Wachter (2013) moves away from assuming
that jumps arrive at a constant rate and provides a modelling improvement, that has
been supplemented with other structures. Another important theme is when and how
time variability in jump frequency matters and what features of asset prices are (not)
included by selecting a certain structure. Although the selected model fits the need for a
time varying equity premium, it may be misleading for other purposes. Martin (2012)
highlights the great effect of uncertainty in estimation which is inherent in tail studies due
to the low number of observations, which translates into imprecise estimates. Jondeau
and Rockinger (2012) put portfolio optimisation to the forefront with and without fat
tails, reporting a significant improvement over the mean-variance optimiser, although
in a higher moments framework. It must be noted how the selection of a particular
distribution contradicts both Weitzman’s and Martin’s intuition on the effects of such
a choice. Finally, Foster and Young (2012) is the only paper that refers to managerial
incentives, despite the vast literature in all the above areas. It also provides a statistical
test to identify managerial behaviour that entails excess risk, a key theme in the thesis.

What has not been pursued, and in fact has seen very little attention, is the effect
of managerial incentives on portfolio allocation. In all the models presented above,
there is no reference to the manager and how his incentives differ from those of the
investor. The manager is either absent or is treated as identical to the investor, in typical
methodological fashion, with the only technical paper to identify a disparity being Foster
and Young (2012). In the thesis the manager is treated as a different entity that receives
utility by a managerial fee structure contrary to the investor who is concerned about
utility derived from portfolio wealth. The mean-variance optimiser who uses a power
utility function and a normal distribution of returns is called ”naive” in the literature,
contrary to a better informed optimiser who takes jumps into account. The research will
reverse the convention and, following the Weitzman and Nordhaus debate, will argue
that instead of being naive, the investor (manager) knowingly ignores the probability
of rare events effectively betting that the chance of occurrence during the investment
horizon is negligible and can be ignored. By consequence, such an agent will knowingly
undertake higher risks in hope of higher returns, hoping that there will not be a disaster
as long as the investment is still active. This change of perspective and context will be
the base of expansions in all the aforementioned areas and lead to a series of novel results.
More than a retelling, the research will expand existing methodologies on new ground
and see if the different incentives of managers and investors lead to a conflict of interest, if the results can be attributed to technical reasons and make policy suggestions.

To conclude, the thesis will justify the selection of a specific model that includes jumps based on the features it needs to contain, proceed into an estimation in order to include a period as representative as possible, proceed to a closed- or semi-closed form solution for portfolio weights for direct comparison between cases that will have value on its own accord, and conduct a set of comparative simulations between an investor that considers jumps in the optimal portfolio selection and one that does not. In addition, managerial fees will also be included so that managerial incentives and utility can be compared to those of the corresponding investor. The research question can, thus, be summarised as follows. First, what is the most appropriate model and estimation method for studying portfolio optimisation under tails with an aim towards managerial incentives. Second, how outcomes differ for investors who take jumps into account compared to those who do not. Third, how managerial incentives in the form of fees and compensation structures create a conflict of interest between investors and funds in the presence of jumps, leading to a need for policy suggestions.

2.2 The Equity Premium puzzle as an introduction to jumps literature

The second part will analyse in detail the literature starting from the equity premium puzzle and focuses on portfolio optimisation under different utility functions and asset price distributions. The advantage of the equity premium puzzle as starting point is that it encapsulates a great part of the problematique and the analytical challenges addressed or still unresolved by potentially introducing jumps. The discussion will show how the attempt to solve the puzzle has led to the inclusion of different sets of preferences and assumptions on investor behaviour, and how tail risk (not only in the form of jumps but also in higher moment analysis) has led to technical and theoretical advancements in modelling and optimal portfolio solutions.

2.2.1 The Equity Premium Puzzle in more detail

The starting point is Mehra and Prescott (1985), where the Equity Premium Puzzle is analysed in detail as the product of the inability of the mainstream framework of CRRA utility functions and/or normality to generate empirically verifiable equity premia. The
focus is the historically persistent higher yield of stocks over bonds in the US which was not accounted for by standard models under plausible calibrations of the risk aversion parameter and was estimated at 6% at that time. The illustration of this inability came through a representative agent economy with constant elasticity of substitution, time additive expected utility preferences and complete markets. More specifically, the model employed an Arrow - Debreux frictionless economy with a Lucas consumption model and CRRA utility while output followed an ergodic Markov process. The conclusion was that the observed return rates could not be replicated in such a limiting setup and, consequently, parameterisation in the literature and/or model assumptions were out of place.

Around this focal point many authors have argued around a variety of topics which are closely related, if not central, to the thesis. Coherent summaries of the debate and assessments of contemporary results can be found in Kocherlakota (1996), Mehra and Prescott (2003), Van Ewijk et al. (2012) and Constantinides (2002), which identify the main axes of research: improving the assumptions of identical agents (homogeneous preferences), complete markets (perfect insurance), violations of normality (as discussed in the previous section) and no transaction costs. The assumptions, methods and analytical successes and failures the following papers report are relevant in the more specialised area of disaster risk. The first issue to be discussed is advances in the type of preferences and utility functions, market/agent types and investor behaviour in the standard model.

Weil (1989) showed that disentangling elasticity of substitution from the risk aversion parameter through Kreps-Porteus preferences, thus smoothing consumption over states and over time, does not provide better results and, moreover, a second puzzle named the risk-free rate puzzle emerges: the risk-free rates generated are too high compared to the real ones. When these processes are connected, either the risk free rate or the equity premium can be replicated but not simultaneously. When they are independent they can be arbitrarily replicated but with very high risk aversion. In the first case the equity premium predicted by the model is too low compared to empirical values (equity premium puzzle) while in the second the risk-free rate is too high (risk-free rate puzzle). Epstein and Zin (1991) are less successful in their attempt to answer the puzzle through Epstein – Zin preferences (very similar to Kreps - Porteus), since they still have the equity premium puzzle appearing and their results are only marginally accepted. Duffie and Epstein (1992) expand the recursive utility model to continuous time and produce a stochastic differential formulation, which is able to produce a Bellman optimal solution. Although in this setup it is possible to distinguish between different types of utility
function, when the decision criterion is not expected utility this distinction is no longer possible (e.g. between a certainty equivalent based on expected utility and smooth local expected utility). The papers are theoretical but form the basis for the type of solutions to be used in stochastic models for optimal weights.

Constantinides (2002) is in favour of relaxing the assumptions of the basic model but still maintains the need to remain in a rational paradigm, while Detemple (2014) provides an expansive and comprehensive review on portfolio selection literature. Habit formation in consumption in Campbell and Cochrane (1999), where changes in consumption have low (high) effect when consumption is high (low) and where negative states appear as consumption shocks, yields better results compared to focusing on time (in)consistency as above. The authors use the term “fat tail” to describe the skewed distribution of the surplus consumption ratio, which is an explicit way to bring forward a deviation from normality. However, the host of positive results can be attributed to reverse engineering. Technically, it is very fortunate that the coefficient of relative risk aversion is multiplied by a very small quantity and therefore its size has very little effect on the risk-free rate. Also, average risk aversion over time can be high with high variation but the risk-free rate remains low and stable. Heterogeneous consumers and incomplete securities markets are another promising assumption that reappears later. In Constantinides and Duffie (1996) investors have the same preferences but different initial endowments and they suffer from idiosyncratic employment shocks that are very difficult to hedge, especially when they are inaccurately represented through a log-normal distribution. Lack of heterogeneity would lead to an over- or underestimation of the subjective discount rate and the risk aversion coefficient. This finding is related to the counter-cyclicality of the equity premium, since in periods with high job insecurity people are less willing to invest in stocks as they are a poor substitute for income loss.

Freeman (2004) focuses on market incompleteness as a solution to the puzzle. Uninsurable risk present in times of contraction, an idea similar to downside risk/ loss aversion faced by an investor facing tail risk, plays a considerable role in explaining the price when a disaster state is introduced, contrary to the unrealistic parameters and results in the standard Mehra and Prescott setup. Jacobs et al. (2013) also discuss market incompleteness and note how its inclusion lowers the risk aversion necessary to generate estimates of the equity premium closer to empirical values. However the data sets are incomplete or suffer from inaccuracies. Gourio (2008) introduces time variability in disaster risk combined with Epstein - Zin and CRRA preferences and discrete time model, thus differentiating from Wachter (2013). Disaster risk is i.i.d and the state variable of
CHAPTER 2. LITERATURE REVIEW

the equity premium is independent of dividend and consumption growth. While CRRA utility fails to match empirical stock returns, Epstein-Zin preferences match the pattern but are imprecise in the magnitude of estimation.

Additional contributions focused mainly on the points mentioned above and reported varying results. Bach and Møller (2011) follow Constantinides and Duffie (1996) but divide consumers into asset holders and non-asset holders under a non-constant risk free rate. They estimate for each group separately and for the aggregate, finding a coefficient of risk aversion close to 8 for asset holders and problematic fit for the risk free yields when testing for non-asset holders. Compared with the aggregate, the results for asset holders are better, implying better information for this group. The explanation given is the high volatility of consumption for asset groups that allows for a lower CRRA. Guvenen (2009) uses a similar model in which non-stockholders do not consume wealth (empirically wealth is very unevenly distributed between the two categories) but receive stochastic labour income (inelastic and elastic markets) and wish to smooth consumption through the bond market. In negative events, stockholders practically pay non-stockholders countercyclically. The result is that differentiation (low EIS of non-stockholders) generates high countercyclical volatility and equity premia because of the high skewness of the non-stockholders utility curve, which translates to very low IES. Auer (2011) finds evidence of the risk aversion coefficient being higher than normally assumed. He conducts cross-checks between the habit formation model and the standard power utility model (both estimated via GMM) benchmarking through the Hansen-Jagannathan nonparametric method using German investment funds as a benchmark. The models, under all circumstances, demonstrate high risk aversion but cannot be rejected. It is interesting to see how the coefficient for Campbell and Cochrane (1999) is even higher than the original model. As a critical comment, Carroll (2001) provides a technical, estimation-based critique and arguments against the use of (log-linear) Euler consumption functions and their second-order Taylor approximations, stating that micro data are inadequate in the estimation of risk aversion and the close link between consumption and predictable income growth leads to false test rejections.

2.2.2 Deviations from normality and jumps

The second issue is how deviations from normality were gradually introduced in the equity premium discussion as an improvement of the standard model. This, in essence, includes introducing jumps and non-normal distributions. The calibration of jump frequency, or equivalently a certain type of jumps modelling, is directly related to the essential
2.2. THE EQUITY PREMIUM PUZZLE AS AN INTRODUCTION TO JUMPS

behaviour and feature the model is supposed to replicate, bringing jump factors to the forefront. There is an ongoing debate on what consists a jump, not just in terms of magnitude but also in terms of source and reaction. The modelling approach is covered by a line of mathematical and finance literature focusing on the types of processes and their properties, which will be covered extensively later on. The calibration approach brings forward the elements of definition and data mining.

Guidolin and Timmermann (2008) take an international point of view by bringing skewed and kurtotic distributions of returns in international portfolios when home bias is present. The analysis uses the ICAPM model but with emphasis on higher moments under bull and bear regimes and time variability. Fat tails are also present and investors are found to show an aversion towards them (kurtosis) but a preference towards positive skewness. Although such a perspective is beyond the scope of the thesis, the paper is a good example on how tails affect decision making and the severe limitations of ignoring fat tails either by assuming normality or by limiting optimal solutions to the mean-variance criterion. Kole et al. (2006) also focus on international decision making under crises, but their setup is fundamentally different. Instead of treating jumps as one-off events, they consider crises to be systemic, lingering events that change the state of the economy. In a continuous time Markov switching model that resembles a jump-diffusion model (with the difference being in the structure of the jump, which is not a Poisson of generally Levy process) closed-form solutions for portfolio weights are derived. The paper finds strong crisis persistence and major differences between the crisis awareness and crisis ignorance states. The latter is an idea that will be explored in the thesis but in a much different context and from a completely different angle. Finally, a major drawback is the use of monthly data for a period of 30 years, a choice in stark contrast with the continuous time framework that drastically limits the number of observations making the results rather questionable. Azevedo et al. (2014) perform a similar exercise in a deterministic finite investment horizon. Maheu et al. (2013) find further evidence on the existence of a ”skewness premium” and the relationship between time-varying disaster risk and the equity premium, in a GARCH setup with a third-order Taylor approximation for the utility function.

Liu et al. (2004) look for a way to increase consumption volatility in the Mehra - Prescott model with a method close to the main string of the literature. They use a jump – diffusion process for the share of dividends in consumption as well as a different process for consumption itself. The risk premium consists of the standard consumption effect, the effect of a stock price jump and the correlation between the dividend share
and consumption times risk aversion. However there is a number of debatable points: dividends are assumed to be regular and equal to a constant payout ratio times aggregate corporate earnings, consumption shocks are set to a yearly 10% which is consistent only with Great Depression values. The authors assume countercyclical behaviour of dividends but report empirical evidence of strong procyclicality of the corporate fraction, and dividends as variable are generally considered to have minor effects.

Harvey et al. (2010), as in Jondeau and Rockinger (2012), use the Michaud Bayesian resampling technique in a framework of higher moments and parameter uncertainty, but they highlight the loss of utility stemming from that process that results in a sub-optimal results in terms of expected utility according to Jensen’s inequality. Their proposed alternative is a Bayesian model that relies upon a distribution that is the product of a multivariate normal pdf and a multivariate normal cdf. The estimation process is MCMC (the method of choice of the thesis) and another MCMC step optimises the expected utilities resulting from the distribution drawings. Since expected utilities are a function of the weights, optimal weights are also simulated. The weaknesses of the paper are the very limited and selective data set and the lack of any out-of-sampling testing other that the Odds ratio to determine the better model. As a side note, Harvey et al. (2008) discuss the Michaud technique in greater detail.

From a technical point of view, a basic reason for the poor-to-modest results is the fact that consumption time series have too low variance to generate either high risky asset returns or low risk-free rates. Since these models estimate consumption Euler equations, the independent variable is simply incapable of generating the desired results. The conclusion of this body of literature is that there can be no proper treatment of the behaviour of asset prices, at least in the context of explaining the equity premium puzzle, in such a smooth framework that relies on overly restrictive assumptions on market homogeneity, intertemporal elasticity of substitution, risk aversion and using consumption as the underlying variable. Of particular importance is the failure to consistently improve results when the CRRA class of functions is abandoned, which shows that, despite its well-documented limitations, power utility still has merit as an interpretative tool. Yet, some of these models managed to point later research to a more fruitful direction.

This underlines the importance of jumps as explanatory factor. On the other hand, the inclusion of tails and/ or the study of higher moments produces consistently improved results. Up to this point it has become apparent how the focus has shifted from the equity premium puzzle to the advantages and potential provided by alternative utility functions and, more importantly, tail thickness. The advantage of using Epstein - Zin preferences
are not as conclusive as those of higher moments and disaster risk. As mentioned above, 
the use of power utility is still relevant and addresses the research question properly. 
Also, the major issue of proper modelling and parametrisation of tail risk has become 
obvious and will be discussed in detail in the next part, where a large body of continuous 
time jump-diffusion literature will establish the selection of the model. A final remark 
is that managerial incentives are still absent, despite the rich and fruitful work on the 
elements that affect investor behaviour.

2.2.3 Further relevant considerations about jumps

An important element of the discussion is the nature of jumps and how qualitative, 
behavioural or political considerations can be linked to disaster risk. An important 
contribution comes from Berkman et al. (2011). They use an international political 
crisis database to assign probabilities for disasters, with an aim towards time-variation 
in disasters. A crisis is defined as "a perceived change in the probability of a threat that 
results in the start or end of an international political crisis [...] likely to be closely aligned 
with the news events to which investors might react." The paper highlights a number 
of important facts, among which are: the number of actors/factors in an international 
crisis is 5 of larger in half the cases (crisis spreading - contagion), there is correlation 
between consumption crises and involvement in international crises (with the effects 
being much stronger in the case of wars), stock market volatility is affected by global 
political insecurity (in agreement with Wachter (2013)). The chance of being involved in 
a crisis is 1 every 15 years - not very far away from the 6 - 10 years demand of Julliard 
and Gosh.

The methodology is regressions and GARCH with dummy variables and the findings 
are generally in agreement with time-varying volatility models. The paper fails to 
connect future market returns and crisis risk, but it does find a correlation between 
the earnings–price ratio and the dividend yield. However potential validity issues are 
not avoided. It is not discussed how a political crisis translates into an economic/ 
consumption/ stock market crisis, and a theoretical tool to expand from the international 
to the national level (if one accepts that there are crises without international elements) 
is not provided. In a nutshell, it can be difficult to establish a measurable relationship 
between the political and the economic domain, and this is probably why the authors stay 
at a level of mere inference. This is the most complete study of its kind with quantitative 
applications.

Gabaix et al. (2013) provides a small summary of the literature that focuses on
specific periods, finding strong ties between war related and political risk and asset prices. A link between political risk and the volatility and level of asset prices. The disaster hypothesis is, therefore, relevant and extreme event risk can be theorised further both conceptually and technically. Benartzi and Thaler (1995) introduce the concept of myopic loss aversion and find a positive relationship between evaluation frequency in a period and risk aversion. Observed risk premia are consistent with yearly evaluations, in particular those of pension fund allocations (60% stocks – 40% bonds) and management (since they must produce positive results they prefer a more “certain” but sub-optimal allocation within a time frame of 2 – 5 years rather than an “optimal” allocation in favour of stocks that will pay off in the distant future but may be harmed by shocks).

Jagd and Madsen (2009) expand on those findings by introducing capital gains/losses. They find that prospective value curves for real long bond returns (contrary to short bonds) are extraordinarily low (this amounts to inflation bias, myopic loss aversion or government regulations). This, however, does not explain why people buy long duration bonds (possible explanations: hedging other material or immaterial assets, transaction/time costs of rerolling bonds), a result that is partially explained in Sangvinatsos and Wachter (2005). This behavioural-based approach is at the same time a comment on how more technical, traditionally based literature has stayed away from linking disaster risk and managerial incentives.

### 2.3 Jump diffusion models in the literature

This section consists the technical part of the literature review. It will discuss in detail jump-diffusion literature and the form of solutions for optimal portfolio weights. The topics are the specific characteristics and aims of jumps-related literature, particularly jumps-diffusion, its evolution over time and how its different features are tied to the PhD. The papers used from a purely technical contribution will be discussed in the appropriate chapters of the thesis but will be briefly mentioned here as well. The aim is to provide sufficient technical background to justify the methods selected and classify the research questions along with the answers provided.

In short, jump-diffusion models are a flexible way to introduce fat tails without the limitations of a particular distribution with known pdf. The implementation of different types of jumps with different theoretical and mathematical properties is straightforward. A state-space model allows for heteroskedasticity and different factors of risk, multiple asset frameworks and higher moment analysis. The correlation between the "hidden"
2.3. JUMP DIFFUSION MODELS IN THE LITERATURE

The volatility process and the asset price process gives rise to the leverage effect, while identical, independent or correlated jump factors are straightforward to model and estimate by either calibrating the jumps process or changing it altogether. Volatility clustering is also represented. In continuous time, normal periods are represented by a Brownian motion with drift while jumps can be binary (Poisson-type) occurrences of a certain magnitude or follow a more general specification (e.g. Levy). The versatility of those models allows for many different parametrisations (constant or time varying arrival intensity, correlated sources of risk, multi-asset portfolio, volatility premia in options, short-selling etc) which may or may not lead to closed-form solutions. The nature of the research question calls for a closed or, at the very least, a tractable semi-closed solution, therefore it becomes imperative to study portfolio optimisation across this specific literature. Typically, such a solution comes as an intertemporal optimisation problem and includes a Bellman equation (also referred as HJB or Hamilton - Jacobi - Bellman). For an overview, Runggaldier (2003) and Meinerding (2012) are suitable.

2.3.1 Jump diffusion specifications and their properties

The first topic to be discussed is the different jump-diffusion specifications and the consequences of selecting one or the other. A useful introductory resource is Chernov et al. (2003), which contains a thorough presentation and performance comparison among different variations of affine jump-diffusion models. The Stochastic Volatility Jump-Diffusion (SVCJ) class of models is found to dominate all other options, a result upon which Eraker et al. (2003) add that jump independence in returns and volatility (assumes two independent sources of risk) provides only a marginal improvement at the cost of model complexity. The paper serves as a guide for specifications found scattered across the literature, including pure volatility models without jumps (the Heston volatility model with constant mean and square-root standard deviation - or not), an Ornstein-Uhlenbeck process for the mean, and multiple processes for volatility.

Affine jump-diffusion models include, among others, the above specifications with a Poisson jump in returns only and a hidden square-root volatility process, an Ornstein-Uhlenbeck process with jumps in both volatility and prices and log-linear variations. Jump parameters, particularly the mean, may or may not be constrained to zero, while the Poisson arrival rate $\lambda$ is constant and may or may not be common in price and volatility jumps. It must be noted that, especially in formulations with square roots for volatility like the Heston and the Eraker et al. (2003) models, the hidden process depicts variance, not standard deviation, despite being often called "the volatility process". The
square root of the variance is used as standard deviation in the asset price process, hence the characterisation ”square-root”. This terminology will be used throughout the thesis.

With these specifications in mind it is now easier to keep track of the multitude of models used, each with its own properties and disadvantages. The important differences are about constant/ stochastic mean and volatility in the diffusion process, with the latter being of particular importance, processes for jump arrival intensity (constant Poisson parameter, stochastic, following a Cox process etc) and the link between volatility and price jumps (only in one process, independent, correllated, simultaneous etc) and some additional embellishments. Belaygorod et al. (2014) note that affine jump - diffusion models with jumps only in returns in the style of Bates (2000) are the best in terms of estimation and fit flexibility. Otherwise the cost of model complexity makes the process inefficient and forecasting problematic when MCMC is used. That cost in efficiency does exist and will be discussed in the methodology section, but the importance of volatility jumps remains and has been emphasised by further literature.

2.3.2 Jumps as a tool, their research scope and interpretative power

A closely related topic is estimation challenges and identifying the effect of jumps. The goal is to highlight estimation deficiencies, how jumps alter results on volatility and equity premia and how those issues relate to the nature of the data. The main themes are suitability but also financial implications. The use of jumps affects much more than just the fit of a model and touches directly upon equity premia, volatility, portfolio allocations and a host of other topics. Methodological suitability brings forth the unavoidable interplay between the model and the research question. Therefore, the aim of this section is to guide through the jumps literature and see how certain methods have been applied in different contexts and what their results are. A further issue is appropriate estimation, which will be discussed first.

Aıt-Sahalia (2004) isolates the effect of jumps from that of the diffusion process. This allows for independent pricing of various premia and can be expanded from Poisson to Cauchy and Levy processes corresponding to a large number of small jumps. Volatility is not modelled and estimation is via GMM and Maximum Likelihood. The distinction of the diffusion component from the jump component had no effect on the estimation of variance under ML, a result that holds under other processes as well. The explanation provided is that the more jumps implied in the process the closer it can be approximated by the
Brownian motion and thus its effect can be picked up. However, this highlights more a problem in the (GMM) ML estimation of such models rather than a conclusive argument, and in addition volatility is treated as constant. Such a parametric result contradicts findings in parametric equity premium literature because it essentially concludes that the parameters and therefore premia will be similar with present and absent jumps. Along similar lines, Li et al. (2006) report better fitting results for Levy-type jumps rather than Poisson jumps, but that paper will be discussed in great depth later on.

Aıt-Sahalia et al. (2009) and Aıt-Sahalia and Matthys (2015) expand to portfolio selection under jumps. In the former they perform an orthogonal decomposition of jumps and diffusion and provide a closed-form solution and a discussion of the position such solutions imply. However volatilities still remain constant and the solution is, therefore, a variation of the well-known Merton portfolio. In the latter they report closed-form robust optimal portfolio solutions for a model with Levy jumps and constant volatility and semi-closed solutions for Poisson jumps, and note that when jumps are symmetric around zero, deviations in the mean play a greater role than in the jump sizes. When jumps are only negative misspecification problems arise, while under a Poisson process the overall sensitivity of the solution is decreased as a higher compensation is already included in expected return. The results highlight the need for both a more complex setup as well as improved estimation techniques.

Methods like those used in Chernov et al. (2003), Chacko and Viceira (2003) or Pan (2002), who focus on parameter estimation on different variations of jump models, come with important limitations and deficiencies. Besides the apparent need to include the recent financial crisis in the dataset, GMM and ML techniques in continuous time have given way to MCMC and GARCH estimation, with good reason (which will be discussed further in the thesis, but the findings of Aıt-Sahalia (2004) provide sufficient intuition). The need to include the recent crisis goes deeper than simply including additional data with more or more frequent jumps.

Bollerslev and Todorov (2011), in line with aforementioned literature, show that equity and variance risk premia largely include compensation for tail risk and that such worries are time-varying. The effect of jumps on total variance in a model only a returns process with jumps can be up to 7%, biased downwards (Huang and Tauchen (2005)). Bandi and Renò (2016) see jumps in both volatility and prices as a way to explain sudden surge in volatility, as in Eraker et al. (2003), and look at the implications for price and volatility premia. The key assumption here is co-dependence of jumps in volatility and jumps in prices, a relationship that is under debate and largely depends on whether options or
price data are used. The established ways of modelling jumps (either independent or simultaneous, but without the in-between option), modelling tails and filtering volatility under low frequency returns data may be potential reasons for the failure to "relate volatility measures unaffected by risk premia to sudden price changes". They find that when co-jumps take place there is a very strong leverage effect that depends solely upon the price dynamics. Rodrigues and Schlag (2009) disassemble the index data commonly used as the risky asset and focus on jumps of individual stocks compared to those of S&P500. The index jumps do not coincide with jumps in the majority of stocks. Also, many individual returns turn negative, which could mean that a factor that drives jumps in the index may lie in the correlations. Individual volatility is greater and the leverage effect is smaller compared to that of the index. Through principal component analysis a multitude of individual volatility factor is found, with at least one being a significant common factor that could be interpreted as market drive. Another question that arises when jumps are used as an interpretative tool is if jumps actually exist in the data or they are the result of data frequency. Limiting the discussion to indices with sufficient quantity of observations over time and for very high frequency intraday data, the price path may be smooth enough to attribute all sudden movements to volatility and thus rare events might have no explanatory power, or be so rare that they become truly one-off events. On the other hand, for very long periods of low frequency, jumps may also not manifest. For high frequency data like those used in Wang et al. (2000), it is common to observe constant returns for a period, a sudden jump at a different level and constant returns until another jump takes place. Whether regime switching or jumps is a more appropriate approach to model such behaviour is a matter for debate. Nevertheless, when an index is traded continuously such movements are more smooth, and reducing discretisation usually smoothens the series (minute or second data are much smoother than millisecond data). Intuitively, when jumps are very frequent they can correspond to high volatility just as well. For the purposes of the argument, since the thesis uses index data, the early statement holds.

Bollerslev et al. (2008) use high frequency data and a test based on cross-covariance to study the phenomenon based on a selection of stocks individually and combined in an index. As in the previous paper, the index jumped less often than individual stocks due to diversification of idiosyncratic risk and jumps being largely unrelated. In the case of co-jumps, however, there is strong evidence of timing as stocks tend to move together at the time of daily announcements. Therefore, the same patterns in the ultra-high frequency and the daily frequency levels are reported. The thesis will also discuss, albeit
suggesting caution, the relationship between daily and higher frequencies. On the same subject, Moreno et al. (2011) use a short noise function in the jump process to account for the effect of unexpected news on the parameters. This stochastic process is related to the jump time of a Poisson jump and acts as a decay function when a jump takes place. The diffusive mean and standard deviation are constant. A link between the short noise function and autocorrelation in the data is identified, a finding that is in favour of jumps having a lingering effect on prices. Another way to implement jump persistence comes from Todorov (2011) who models volatility as a function of previous jumps, a method stemming from Barndorff-Nielsen and Shephard (2001). Instead of a separate, additive jump (Poisson) term, the time-varying parameters (in this case different variance factors) are modelled as moving average functions of past jumps without a specified jump component in the space or the state process. In contrast, a common ARMA process for the state variables (variances $\sigma_{1,2}$) is modelled based on a Poisson measure and which is later scaled in the price process to attribute to diffusive volatility (small movements) and ”jumps” (large volatility movements). By ”switching off” component 1 or 2 allows movement between different models. The paper reports an overall improvement of model performance based on salient price features.

Cartea and Karyampas (2016) also use the number of jumps as a variable for volatility combined with high frequency data. Although that type of data will be of interest only in terms of discretisation bias and whether jumps truly exist or are a perceived illusion due to differences in data frequency, the finding that they have meaning in the context of a Poisson large jump or an ”infinite-jump” Levy process using minute-by-minute log-returns is indicative of their widespread application. The number of jumps is shown to have more explanatory power for daily volatility than other commonly used factors and play an important role in its overall level. It also increases forecasting performance for AR models and disseminates information not captured in the VIX index. Bäuerle and Rieder (2007) also address the issue of a hidden stochastic Markov process for jumps and provide a optimal solutions for a jump-diffusion model with constant mean and volatility in prices (and therefore quite limiting). It can be seen as a model similar to Wachter (2013), in which the investor cannot observe any information about a jump but knows they exist. Jeanblanc et al. (2010) pursue the same idea in a multi-asset model with stochastic (not modelled) mean and standard deviation where the unobservable Poisson process is also a learning process, and find similar solutions.

Jump-diffusion models have also been used to study the effect of market structure and investor features. Bellamy (2001) uses a jump diffusion model with a jump-diffusion
process for the asset price without stated expressions for the mean and the standard deviation. Market incompleteness is introduced as discontinuity in prices due to the Poisson jumps, so the result expands on the standard Merton solution and can be seen as a preamble of the optimal weight solutions in this part and the thesis. nun use different information in a jump diffusion model with insiders and outsiders. An insider has better information than the outsider on the occurrence of a future event, an idea that is expressed by a wider filtered probability space or, more intuitively, flow of future information. Under looser assumptions than the main string of the literature, they derive optimality conditions for the market where both insiders and outsiders participate. Epstein and Ji (2013) focus on ambiguity (model uncertainty) aversion for both the drift and the volatility of the process. The mathematical background of the paper deviates much from the existing framework, but very briefly the assumption of a single measure defines null events is dropped, making it impossible to apply the Girsanov theorem.

Feunou et al. (2012) focus on downside risk volatility, where investors welcome positive jumps only worry about crashes. Strong evidence of downside risk is found, there is a positive relationship between that risk and the conditional mode of returns, skewness can be priced and, finally evidence of time variation in structural parameters of disappointment aversion preferences is found. Liu and Loewenstein (2007) discuss the effect of proportional transactions costs when the risky asset follows a jump-diffusion process with constant drift and volatility. The optimal strategy is for the weight on the risky asset to lie between two boundaries. When a jump forces the portfolio to violate a boundary, the investor will revert to the nearest weight allowed and the transaction cost increases. An increase in transaction costs may increase trading frequency while jumps have a great negative impact on the amount of the risky asset and frequency. Das and Uppal (2004) use a multivariate setup, international data and method of moments estimation systemic risk reduces only slightly the gains from international diversification implied by the standard portfolio models.

2.3.3 Time variability in jump frequency

The choice of time-varying or constant arrival intensity for jumps underlines a large part of the literature and deserves a separate section. Although some of the papers cited s positive results for certain specifications, it is important to identify when and where time-varying disaster risk poses an advantage rather than a hindrance, and what are the consequences of such a choice. As Eraker et al. (2003) mention, in an SVCJ setup there is evidence of misspecification for stochastic arrival intensity. In addition,
the modelling approaches of Liu et al. (2003) or Eraker (2004) are problematic in their own way. An attempt to replicate the results of Eraker (2004) failed due to unresolved software errors. It must be stressed that, contrary to the Pan (2002) parameters in options, who have become very popular in the literature, the Eraker (2004) parameters have been ignored. Belaygorod et al. (2014) provide a formulation almost identical to Liu et al. (2004) and Eraker (2004) that includes volatility and jump premia in the drift (mean) but there is a key difference: the stochastic arrival intensity is constant instead of $\lambda V_t$, so their parameters do not correspond exactly. Also, the Cox process for the jump arrival intensity was adopted by Liu et al. (2003) for reasons of reverse-engineering and convenience in the solution for weights. As a specification, it only provides a linear link between volatility and jump frequency. In addition, the implied parameters are very similar to the constant case. For time variation to have real merit in that framework, a different process like a Hawkes process needs to be employed contrary to those commonly found.

A crucial point is that the vast majority of the papers reporting successes under time-varying arrival intensity (e.g. Bäuerle and Rieder (2007), Wachter (2013) among others) do so without modelling volatility. Setting estimation and misspecification problems aside, the improvement attributes to stochastic arrival intensity can also be attributed in a more intuitive way to stochastic volatility. A greater or smaller number (frequency) of jumps in asset prices can be naturally tied to volatility following a stochastic path and the leverage effect. This line of reasoning is supported by the marginal improvement under stochastic $\lambda$ for Poisson processes, the general lack of estimation for the Liu et al. (2003) specification (Pan (2002) estimates are still used) and the almost identical results between correlated and independent co-jumps.

### 2.3.4 Further extensions, comparisons and practical applications

The last points for discussion are similarities to existing literature either in the research question or in empirics. Although it is very clear that the ground covered by the thesis is novel and the research questions have not been addressed before, there is existing work that is adjacent to some elements. These are applications in certain markets, uses of similar concepts and modelling improvements. An example is studies in pension funds and how they allocate their assets. Pension funds are long-term investments with very specific goals and conservative nature, so they are a potential area of application for the
outcomes of the thesis.

Menoncin (2005) introduces an investor that wants to maximise the expected utility of terminal wealth under the usual conditions. Two new exogenous variables that express notions from outside the financial market are introduced. They are described as "two sets of variables that do not affect the asset prices but directly affect the level (exogenous level variables) and the growth ratio of the investor’s wealth (exogenous ratio variables)". The interest lies in the interpretation rather than the methodology, which uses what is in essence an SV model. Those variables are introduced not in the asset processes but directly on the wealth process and can be labour income (level variable), exchange rate, inflation (ratio variables) and, most importantly, entry and exit of policy holders in a pension fund or indemnity payments in an insurance company. This bears a resemblance to investors leaving or entering a fund, an idea that is used in the thesis when managerial incentives are introduced.

However the similarities end here. The structure of the model is entirely different, the solution in Menoncin (2005) is approximate while in the thesis is as close to closed-form as possible, expected utility of wealth is optimised under a much different process and the research question is, ultimately, of a much different kind. This paper uses an alternative, technical way to introduce a notion that is utilised in the thesis in a much more practical manner and in a different context. However, as far as investor incentives are concerned, it poses a fruitful question that can be combined with active versus passive investment strategies, and which one performs best.

Staying on the relevant topic of pension funds and how such conservative investments can hedge against rare events, Josa-Fombellida and Rincón-Zapatero (2012) study the asset allocation of defined corporate benefit pension plans. Since wages are linked with benefits, a jump-diffusion model applying to the former also affects the latter. An Hamilton-Jacobi-Bellman equation is solved for optimal strategies and contributions, and linear relationships between the optimal supplementary cost and the optimal investment strategy, and between this strategy and the optimal fund are found. They also find that "it is possible to select the technical rate of interest such that the optimal contribution does not depend on the parameters of the benefit process, getting a spread amortization and the stability and security of the plan in the long term". Ngwira and Gerrard (2007) also consider a defined benefit scheme with a jump-diffusion process for the risky asset, and find that the optimisation results hold for both constant and stochastic benefits. A main result is that the jump size has a positive relationship with the allocation of the risk-free asset and a negative relationship with the risky asset.
The Eraker et al. (2003) SVCJ model, with contemporaneous jumps in volatility and returns and leverage effect, is the method of choice for the thesis and belongs to the class of affine models. It is now useful to debate those models with their non-affine counterparts and highlight some strengths and weaknesses. Ignatieva et al. (2015) verifies the now trivial factor that models with jumps outperform models without jumps. The paper also compares non-affine specifications to affine ones. The differentiation lies in the variance process, where the drift, the standard deviation or both are polynomials of different types, and therefore by using a general specification it is straightforward to switch certain parameters off and move between models. The derivation process for the MCMC posteriors is explained in sufficient detail and can be of further use, since the technique has become popular in this literature. Although the results validate a slight supremacy of non-affine specifications, the disadvantages become instantly obvious.

As mentioned above, the trade-off between model and estimation performance is enormous. The burn-in periods, estimation times and sampling methodologies are unacceptably high and simply not worth the marginal improvement in fit. The added model complexity of non-affine models is a detrimental factor, makes the process very time consuming and computationally intractable (C++ coding is for most intents and purposes out of scope) and very similar results can be achieved with more parsimonious methods. The SVCJ specification in particular is found to dominate the performance of the equivalent non-affine formulation. Another issue relevant to the aims of the PhD is whether non-affine models can produce closed-form or at least tractable solutions for portfolio weights. Affine specifications do, and are therefore suitable. Non-affine specifications pose unnecessary complexities in that regard. Finally, the paper improves upon the results of Ignatieva et al. (2009) which report a preference of jump affine models over non-jump non-affine models, and in addition raise the issue of economically unrealistic parameter estimation for the latter.

Chourdakis and Dotsis (2009) demonstrate the ability of non-affine models to capture market timing. However they do not include leverage in their model and assume continuous asset returns, a choice that biases the comparison across models. Also, the improvement over affine models is small. In addition to the above literature that provides optimal portfolio solutions for further purposes and under different assumptions, there is a vast literature whose focus is the mathematical derivation and existence of such solutions for each individual model. Kurmann (2009) and Guo and Xu (2004) provide examples, He and Meng (2012) introduce Knightian uncertainty in asset prices, Hong and Jin (2016) establish a multi-asset solution for an SVCJ variation, Jin and Zhang
(2012) and Jin and Zhang (2013) decompose optimal portfolios and derive a solution under investment constraints. Finally, Wu (2003), Wu (2013) are further examples in a Markov setup.

2.4 Summary

The main challenges of the thesis can be aptly identified in each of the three parts. The first part highlights the key literature and sets the three main questions of the PhD; i) from the multitude of estimation and modelling methods, which one best fits the purpose of studying managerial incentives; ii) how investors are affected by knowingly ignoring jumps; iii) whether managers have an incentive to disregard jumps, given that excess risk is tied to higher returns and manager compensations are related to portfolio performance. Based on this literature, the literature review described the main considerations in further detail.

The second part takes the viewpoint of utility, its embedded characteristics and how tail risk challenges the standard way of reaching an optimal solution. The argument is in favour of taking event risk into account, both in the face of increasing evidence and in front of the practical issue of, first, treating assets as less risky than they really are, and second, extracting information from time series that was previously inaccessible. Agent actions and presuppositions also took a prominent role, since ultimately the contribution of the thesis is not technical in nature but focuses on potential conflict of interest and a motivation to knowingly undertake excess risk in exchange for higher returns. The term "behavioural" has been deliberately left unused because there is no such pretext or any radical deviations from the standard framework of rationality, the homo economicus and complete markets. In such a setup it would be better to use the verb "act" instead of "behave" with regard to the agent to avoid any confusion with behavioural economics. The aim here was to see how disaster risk led to advancements from the equity premium puzzle point of view, how that intuition expanded to considerations about proper calibration, jump frequency, the nature of data and how it can be an instrument to introduce or expand considerations on investor behaviour and market structure, such as downside loss aversion.

The third part focuses on the performance and features of different specifications of jump-diffusion as well as portfolio optimisation, in an attempt to map a diverse field of research and provide further intuition on the analytical and estimation challenges posed in both parameter estimation and optimisation. In addition, the aim was to see the effect
of different models in different contexts and establish a rationale that justifies the model selected. The advantages and disadvantages of various alternatives were highlighted and it was shown that in every single case a jumps setup would dominate a no jumps setup. From that point on, the focus moved to balancing the tradeoffs between mathematical and statistical complexity, desired features given the research question and a reasonable representation of stylised facts. Another crucial element is the existence of closed-form or at the very least tractable semi-closed form solutions for portfolio weights. Models without the ability to produce such an intertemporal solution for the wealth function are inappropriate, because the comparisons between investors and managers under different parametrisations need to be as precise and replicable as possible. Also, the multitude of models means a multitude of parameters in the literature under datasets of varying indices, assets and lengths. Not all models have received the same attention and not all estimation techniques perform well. It is a necessity to benchmark results with existing literature, and a model that has not been thoroughly studied and parametrised could be hazardous and misleading to use.

There is an unfortunate interplay between different formulations of jump models and stated aims of papers. A prime example is the relationship between asset price volatility, which is a component on the diffusive part, and jumps (size, type, rate) and depending on the structure certain features may be attributed to either jumps or variance. Model selection has an important role in positive or negative findings and conflicting papers, while reverse engineering and preempting results ranges from the very subtle to the thinly disguised. While that may sound as a triviality and acceptable parsimony of science, it has widespread implications in this particular line of research due to the technical difficulties in disseminating the properties of rare events exactly because of their rarity. Moreover, certain types of models may be overcomplicated, misspecified or disregard important effects like leverage.

The thesis achieves a solution for its selected class of models that is not closed-form only because it is the product of a rational and an exponential function, and in that respect it improves upon existing mathematical results. The major gap in the literature is the managerial attitude on event risk. Only a handful of papers discuss either investor or managerial incentives in the presence of jumps, and of those that do only Foster and Young (2012) show an explicit methodological focus on excess risk undertaken by the manager. Also, with one exception, there is no mention in the literature of fees at an employee level. Traders are generally not present, only hedge/mutual fund managers and fee structures for the entire fund. The literature on conflict of interest, misreporting and
managerial incentives tied to compensation packages is long, but has not connected the issue of rare events (jumps) to the motivations of a manager so far.
Chapter 3

Isolating fat tails as factor

3.1 Research aims and methods

Chapters 3 and 4 have the dual role of framing the different technical areas of the thesis and recording learning prowess. Chapter 3 shows why the isolated introduction of disaster risk is not sufficient to explain differences in portfolio allocation, neither from a fitting nor from a portfolio point of view. It establishes the need to move to a more complex setup, by applying at a very basic level portfolio optimisation and simulations. The learning goal was to demonstrate how simulations can be used to answer such a research question and make the first step towards a more complex application. Also, preliminary empirical results show that simply introducing fat tails in a parsimonious setup is no panacea and an investor is practically unable to differentiate between different outcomes. In more detail, the idea is as follows.

The stated goal is not to perform a fitting exercise and make a technical contribution but study investor and managerial incentives. A reasonable argument is parsimony; why proceed into a complicated continuous-time framework when fat tails can be introduced directly by well-known distributions? In addition, since the focus is on tail risk, the mere introduction of fat tails may cause the agents to discriminate between cases, reducing the need for complexity. A comparison between an agent who takes jumps (fat tails) into consideration and one who does not can be straightforward and it may be possible to show that the agent is able to differentiate between the cases of different tail forms, or even across states. This call for parsimony is tempting but shown to be woefully inadequate.

The need in this chapter is to focus on the fundamental idea stated above in an easy to follow framework with tractable mathematical formulations. In that spirit, the
simplest type of investor is one with a CARA utility function who allocates his wealth over an investment horizon to a risk-free asset of known, constant return and a risky asset that follows a distribution. The investor makes an assumption about which distribution is followed in reality, calculates optimal weights and then the real distribution is revealed, the investment takes place and the effect on utility is measured. The distributions are a Normal, a Gamma and an Inverse Gaussian representing a standard case and two different cases of fatter tails. In addition, they were selected because of their known probability density functions which is very helpful analytically, their ability to produce (almost) closed-form solutions and their mathematical properties. Also, CARA utility was selected over CRRA utility because it keeps the focus solely on the effect of tail risk, not on the investor’s changing attitude towards it, plus the mathematical convenience of working with a negative exponential utility function. In short, parsimony, simplicity, tractability and clarity are the main motivators in model selection in order to isolate the effect of extreme events as much as possible.

Constant Absolute Risk Aversion (CARA) utility assigns a level of risk aversion that is the same regardless of individual wealth. Since attitudes towards risk are irrelevant to wealth (no wealth effects), an investor would treat an investment (its potential risk, losses and effect in utility) in the same way regardless of having 10 or 10,000 pounds to invest. In other words, such an investor would always be willing to pay the same amount in order to avoid a fair bet (the investor would always risk the same amount as wealth increases). The Arrow - Pratt coefficient of absolute risk aversion is defined as 

\[ RA(W) = \frac{U''(W)}{U'(W)}, \]

which is constant and equal to \( \gamma \) for an exponential utility function. In contrast, Constant Relative Risk Aversion (CRRA) utility expresses risk aversion in terms of relative wealth. The relative risk aversion coefficient is defined as 

\[ RR(W) = -W \frac{U''(W)}{U'(W)}. \]

It follows that if CRRA is constant for a utility function such as power utility, then CARA is decreasing as wealth increases (as wealth increases, the investor is willing to risk more). For CRRA functions, wealth effects appear - an investor will weigh risk according to its current level of wealth. CARA represents reaction to absolute changes in wealth while CRRA to proportional (percentage) changes in wealth. A CRRA investor would always risk the same proportion of wealth as wealth increases.

The methodological foundations of the chapter rely on Markowitz-type optimisations. For a complete treatise, Cochrane (2009) is a typical reference while for a more modern approach Pennacchi (2008) provides a thorough presentation accompanied by a host of variations. CARA utility functions are not as widely used as their CRRA counterparts, but have their own place in the portfolio optimisation literature. The brief summary of
literature does not intend to be exhaustive, and merely intends to show relevance of the CARA assumption, that it is not obsolete but still used in research.

Vigna (2014) compares CARA and CRRA inefficiency under mean-variance optimisation. Palczewski et al. (2015) discuss CARA dynamic optimisation under under state-dependent drift and transaction costs. Naive and less risk-averse investors incur the most losses in utility, while reduced trading causes half of the total loss from transaction costs. Bodnar et al. (2013) compare three quadratic optimization problems (the Markowitz mean–variance problem, the mean–variance utility function which, under normal distributions, coincides with CARA utility, and the quadratic utility problem) and their conditions for efficiency. The problems are found to be mathematically equivalent but this equivalence does not hold from a stochastic point of view. They also derive the probabilities that the estimated solutions of the Markowitz problem and the quadratic utility problem are mean–variance efficient. de Palma and Prigent (2009) discuss the investor choice under CARA and HARA utility among standardized portfolios which are based on cash, bond and stock indexes. The choice relies on market performance and the type of investor. For the utility functions envisaged, they calculate the losses from not having access to a customized portfolio and show that these losses may be severe. Broadie and Shen (2016) provides a continuous time multi-asset example, where CARA utility is in a multi-asset setup where each asset follows a geometric Brownian motion. Becherer (2003) distinguishes between tradable and non-tradable sources of risk for a rational CARA investor. The hedging strategy is analysed in a general semi-martingale market framework. The result is a computation scheme, valuation bounds, and a discussion on diversification and information effects.

CARA and jumps have received less attention in the literature, partly because of the appealing properties of CRRA and the built-in limitations of CARA. Aït-Sahalia et al. (2009) discusses CARA along with jumps in asset prices. Zhou (2009) combines CARA utility with Levy jumps and derives closed-form solutions to the maximization problem under the Martingale approach. Zou and Cadenillas (2014) derive explicit solutions for an insurer’s investment and risk control strategies under HARA and CARA utility. The agent wants to maximise expected utility of terminal wealth where the risk process is modeled by a jump-diffusion process and is negatively correlated with the capital gains in the financial markets. Liu (2004) and Jondeau and Rockinger (2006) use CARA utility in a modern financial setup on portfolio allocation. CARA utility seems to be more popular for insurance markets, and the examples where jumps are discussed in its context are very scarce. Given the need for simplicity, it is prudent to focus on a specific argument.
and employ the simplest methods available. CARA utility neglects wealth effects but allows for tractable closed-form solutions. In addition, the use of a Moment Generating Function (MGF) instead of a stochastic process allows straightforward comparisons.

3.2 The CARA model and its variations

In greater detail, an investor with a CARA utility function allocates wealth between a risky and a risk-free asset over a single period. The investor maximises utility and determines how much wealth will be allocated in each asset. Returns are received ex post in a Monte Carlo simulation that produces average terminal utilities across the sample. The return of the risk-free asset is constant while the return of the risky asset follows a Normal, a Gamma or an Inverse Gaussian distribution with the same mean and variance in all cases. The most important property is the well-defined moment generating functions of all distributions that allows for simple optimisation. The investor selects a distribution to use for the derivation of portfolio weights in logical time, then the investment begins where the real distribution of the risky asset is revealed and the focus is on the difference in expected utility between the cases of a correct and a false prediction. Therefore, two different cases will be discussed.

3.2.1 Single state optimisation

The investor faces a single state optimisation problem and is free to maximize under whatever distribution he desires, but the risky asset may follow a different distribution. More formally, the utility function is of the form

$$U(W) = -e^{-\alpha W}$$

with wealth $W$ and risk aversion coefficient $\alpha$. Total initial wealth $W_0$ is split between the amount initially invested on the risky asset, $W_{M0}$, and that on the risk-free asset, $W_{f0}$. Trivially, terminal wealth is derived by substituting the 0 subscript with T. During numerical calculations, $r_m$ stands for gross instead of net returns for simpler substitutions and to avoid $r_m$ turning negative if it represented net returns, since that would cause problems with some Moment Generating Functions (MGFs). Some of the moment generating functions have support only in $R^+$. A net return of $-$% falls outside that support, while a gross return of 0.99 does not and is still usable.
In terms of utility, for the single state case,
\[ U(W_T) = -e^{-aW_T} = -e^{-aW_T - aW_M T} = -e^{-aW_{f0}(1+r_f) - aW_{Mr}m} \]

Since \( r_f \) is a known constant and only \( r_m \) is stochastic, taking the utility expectation for time \( T \) yields
\[ E[U(W)] = E[-e^{-aW_{f0}(1+r_f) - aW_{Mr}m}] = -e^{-aW_{f0}(1+r_f)}E[e^{-aW_{Mr}m}] \tag{2.1} \]

with the expectation being replaced by the moment generating function of each distribution. The mean and the variance in each case are the same, while gross returns \( r_m = \mu = \frac{E[W_{Mr}]}{W_{Mr}} \). When \( r_m \) stands for net returns, the equality can be written as \( 1 + r_m = \mu \) with the corresponding alterations in the formulas above. All parameters have been calculated for gross returns, however for completeness and comparison the second notation is also used in the formulas. The graphs for each distribution can be found in the Appendix.

### 3.2.2 Optimal weights for each case

The next part describes the maximisation process for each single state case. In the normal case, for known mean \( \mu \) and standard deviation \( \sigma \), the MGF is of the form \( e^{\mu + \frac{1}{2}a^2\sigma^2} \) for \( t=1 \), with scaling properties \( xN(\mu, \sigma^2) = N(x\mu, x^2\sigma^2) \). The probability density function (pdf) is of the general form \( f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{x(\mu)^2} \). The normal distribution is a symmetric leptokurtic distribution with all moments beyond the first two (mean and variance) being zero. It assigns equal probabilities to positive and negative outcomes, it contains more than 98% of the probability mass within a distance of three standard deviations and, thus, is not very suitable for replicating extreme outcomes or outliers. The optimal allocation of wealth is the first derivative of expected utility function w.r.t. \( W_{M0} \). After plugging the MGF of a normal distribution in (2.1),
\[ \max_{W_M}(-e^{-aW_{fT}e^{-aW_{M0}\mu + \frac{1}{2}a^2W_{M0}^2\sigma^2}}) \]
\[ = \max_{W_M}(-e^{-a(1+r_f)(W_0-W_{M0})-aW_{M0}\mu + \frac{1}{2}a^2W_{M0}^2\sigma^2}) \]
\[ = \max_{W_M}(-e^{-a(1+r_f)W_0-aW_{M0}(\mu-1-r_f)+\frac{1}{2}a^2W_{M0}^2\sigma^2}) \]

And differentiating yields
\[ \frac{\partial}{\partial W_{M0}} = -e^{(c)}[a(1 + r_f - \mu) + a^2\sigma^2W_{M0}] = 0 \iff \]

37
\[ 1 + r_f - \mu + a\sigma^2 W_{M0} = 0 \iff W_{M0} = \frac{\mu - 1 - r_f}{a\sigma^2} = \frac{1 + r_M - 1 - r_f}{a\sigma^2} \iff W_{M0} = \frac{r_M - r_f}{a\sigma^2} \tag{2.2} \]

which represents the monetary amount to be allocated on the risky asset.

For the Gamma case, the moment generating function is of the form \((1 - \theta t)^{-k}\), where \(k\) is a shape parameter and \(\theta\) is a scale parameter. The scaling properties are \(xf(\kappa, \theta) = f(x\kappa, \theta)\) with \(f\) being the gamma probability density function. For the mean and variance, \(\mu = k\theta = (1 + r_M), \sigma^2 = k\theta^2\). The form of the pdf used here is of the general form \(f(x; k, \theta) = \frac{1}{\Gamma(k)} \theta^k x^{k-1} \exp(-x/\theta)\). Its support is positive, it demonstrates a very elongated right tail, although with proper calibration it can resemble a normal distribution with low skewness and kurtosis. However, the left tail is shorter and steeper than the right tail for most parametrisations. Consequently, (2.1) becomes

\[ E[U(W)] = -e^{-aW_{M0}r_f} E[e^{-aW_{M0}(1+r_f)(1+aW_{M0})}] = -e^{-a(W_0-W_{M0})(1+r_f)(1+aW_{M0})} \]

The first order condition is

\[ \frac{\partial E(U)}{\partial W_M} = -e^{(-)}a(1 + r_f)(1 + aW_{M0})^{-k} - e^{(-)}(-k)(1 + aW_{M0})^{-1-k}a\theta = 0 \iff \]

\[ a(1 + r_f)(1 + aW_{M0}) = k\theta \iff 1 + r_f + aW_{M0} = k\theta \iff \]

\[ W_{M0} = \frac{k\theta - (1 + r_f)}{a\theta(1 + r_f)} = \frac{(1 + r_M) - (1 + r_f)}{a\theta(1 + r_f)} \]

Since \(\frac{k\theta}{k\theta^2} = \frac{\mu}{\sigma^2}, \quad \theta = \frac{\sigma^2}{\mu} = \frac{\sigma^2}{1 + r_M} \)

Substituting in the optimal weight yields

\[ W_{M0} = \frac{r_M - r_f}{a\frac{1+r_f}{1+r_M}} = \frac{(1 + r_M)(r_M - r_f)}{a\sigma^2(1 + r_f)} \tag{2.3} \]

For the Inverse Gaussian case, the moment generating function is of the form \(e^{(\frac{\lambda}{x})(1-\sqrt{1-2x^2})}\). Here, the scaling factor is \(-aW_{M0}\) and the properties are \(X \sim IG(\mu, \lambda) \Rightarrow tX \sim IG(t\mu, t\lambda)\). The mean and variance are \(\mu\) and \(\sigma^2 = \frac{\mu^3}{\lambda}\). The pdf is \(f(x; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp(-\frac{(\lambda(x-\mu)^2}{2\mu^2})\). The Inverse Gaussian distribution has a similar shape to the Gamma distribution but shares many similarities with the Normal distribution. That is
the reason why, as $\lambda$ increases, the Inverse Gaussian distribution approximates a Normal distribution. Expression (2.1) then becomes

$$E(U(W)) = -e^{-a(W_0 - W_M_0)(1 + r_f)} + \frac{\lambda}{\mu} \sqrt{1 + \frac{2\mu^2 a W_M_0}{\lambda}}$$

The first order condition is

$$\frac{\partial E(U)}{\partial W_M} = -e^{(\cdot)}(a(1 + r_f) - \lambda \frac{1}{\mu^2} \frac{1}{\sqrt{1 + \frac{2\mu^2 a W_M_0}{\lambda}}} (2\mu^2 a W_{M0}) = 0$$

$$\iff \lambda \frac{\mu^2}{2a\mu^2} \frac{(1 + r_f)^2}{(1 + r_f)^2} = \frac{\mu^2}{2a\mu^2} \frac{(1 + r_f)^2}{(1 + r_f)^2} = W_{M0} \iff W_{M0} = \frac{\lambda}{2a\mu^2} \frac{\mu^2}{(1 + r_f)^2}$$

Substituting the definitions for the mean and variance and simplifying,

$$W_{M0} = \frac{\mu}{2a\sigma^2} \frac{(1 + r_M)^2 - (1 + r_f)^2}{(1 + r_f)^2} = \frac{1 + r_m}{2a\sigma^2} \frac{(1 + r_m)^2 - (1 + r_f)^2}{(1 + r_f)^2}$$

(2.4)

### 3.2.3 Connecting the weight formulas

It is helpful to collect expressions (2.2), (2.3) and (2.4) of the optimal weights and see how they are related.

For the Normal distribution, (2.2) is

$$W_{M0}^N = \frac{r_M - r_f}{a\sigma^2}$$

For the Gamma distribution, plugging (2.2) into (2.3) yields

$$W_{M0}^G = \frac{(1 + r_M)(r_M - r_f)}{a\sigma^2(1 + r_f)} = \frac{r_M - r_f}{a\sigma^2} \frac{1 + r_M}{1 + r_f} = W_{M0}^N \frac{1 + r_M}{1 + r_f}$$

For the Inverse Gaussian distribution, plugging (2.3) into (2.4) yields

$$W_{M0}^{IG} = \frac{1 + r_m}{2a\sigma^2} \frac{(1 + r_m)^2 - (1 + r_f)^2}{(1 + r_f)^2}$$

$$= \frac{1 + r_m}{2a\sigma^2} \frac{1 + 2r_M + r_M^2 - 1 - 2r_f - r_f^2}{(1 + r_f)^2} = \frac{1 + r_m}{2a\sigma^2} \frac{1 + r_M^2 - r_f^2 + 2(r_m - r_f)}{(1 + r_f)^2}$$
\[ \frac{1 + r_m (r_M + r_f)(r_M - r_f) + 2(r_M - r_f)}{2a\sigma^2} (1 + r_f)^2 = \frac{1 + r_M r_M - r_f}{1 + r_f} \]

\[ = W_{M0}^G \frac{2 + r_M + r_f}{1 + r_f} \]

It is interesting to notice how all weights are scaled versions of each other. Since the optimal allocations of wealth in each case have been derived and depend only upon the return of the risky asset, they can be directly used in simulations.

### 3.3 The two-state Markov case

The setup is the same, with the same assumptions on utility and distributions, but the optimisation question becomes a two-state Markov problem. A high probability “normal” state and a low probability “disaster” state with lower mean but the same variance are introduced and the agent is asked to optimise his utility.

The ”disastrous” state is characterised by considerably lower returns (distribution means) in the risky asset compared to the ”good” high-probability state. In practice, this is a very simple regime switching framework with the additional feature of fat tails both in the good and the bad state. For simplicity, the variance between the two cases is assumed to be the same. Dropping the \( m \) subscript for simplicity, this reduces to assuming \( r_1, r_2 \) random variables for returns that follow Normal/ Gamma / Inverse Gaussian distributions with the same variance \( \sigma \) but different means \( \mu_1, \mu_2 \). Expressions for the conditional means are also given further below, since in order to maintain a mean across states that is equal to the single state \( \mu_1, \mu_2 \) need to be slightly adjusted.

The probability between the two states is \( \pi = 0.98 \) for 1, \( 1-\pi=0.02 \) for 2 and the utility function takes the form

\[ U(W) = -e^{-aW} = -e^{-aW_0 r_M T} = -e^{-a(1+r_f)(W_0-W_M)-aW_0 r_{M,i}} \]

\( i \) takes values 1 or 2, according to the state of the world, and \( r_f = 2\% \). Utility across states is, therefore,

\[ U(W) = -\pi * e^{-a(1+r_f)(W_0-W_M)-aW_0 r_{M1}} - (1 - \pi) e^{-a(1+r_f)(W_0-W_M)-aW_0 r_{M2}} \]

\[ = -e^{-a(1+r_f)(W_0-W_M)} (\pi * e^{-aW_0 r_{M1}} + (1 - \pi) e^{-aW_0 r_{M2}}) \]
3.3. THE TWO-STATE MARKOV CASE

This is the terminal utility form used in the weight simulations. However, for the weight formulas, it is more useful to use the Law of Iterated Expectations.

\[
\max E(U) = E(U|\mu_1)\pi + (1-\pi)E(U|\mu_2)
\]

\[
\frac{\partial (E(U))}{\partial W_M} = 0 \iff \pi[E(U|\mu_1)]' + (1-\pi)[E(U|\mu_2)]' = 0 \iff \frac{(E(U|\mu_1))'}{(E(U|\mu_2))'} = \frac{\pi - 1}{\pi}
\]

where the utility forms have already been given in the previous section.

3.3.1 Optimal weights

For the Normal case, the calculations are as follows.

\[
E(U) = -e^{-a(1+r_f)(W_0-W_M)-a\mu_i W_M + \frac{1}{2}a^2 W_M^2 \sigma^2}, i = 1, 2
\]

\[
\frac{\partial E(U)}{\partial W_M} - e^{(+)}(a(1+r_f) - a\mu_i + a^2 \sigma^2 W_M), i = 1, 2
\]

\[
\frac{(E(U|\mu_1))'}{(E(U|\mu_2))'} = \frac{-e^{(+)}(a(1+r_f) - a\mu_1 + a^2 \sigma^2 W_M)}{-e^{(+)}(a(1+r_f) - a\mu_2 + a^2 \sigma^2 W_M)} = \frac{\pi - 1}{\pi}
\]

Working with the exponents, the identical terms are cancelled out and what remains is

\[
-e^{-a(1+r_f)(W_0-W_M)-a\mu_1 W_M + \frac{1}{2}a^2 W_M^2 \sigma^2} - e^{-a(1+r_f)(W_0-W_M)-a\mu_2 W_M + \frac{1}{2}a^2 W_M^2 \sigma^2} = e^{-aW_M(\mu_1-\mu_2)}
\]

Concluding,

\[
e^{-aW_M(\mu_1-\mu_2)} \frac{a(1+r_f) - a\mu_1 + a^2 \sigma^2 W_M}{a(1+r_f) - a\mu_2 + a^2 \sigma^2 W_M} = \frac{\pi - 1}{\pi}
\]

and in terms of the random variables

\[
e^{-aW_M(r_1-r_2)} \frac{(1+r_f) - r_1 + a\sigma^2 W_M}{(1+r_f) - r_2 + a\sigma^2 W_M} = \frac{\pi - 1}{\pi} = -0.02
\]

(2.6)

For the Gamma case,
\[ E(U) = -e^{-a(W_0 - W_{M0})(1 + r_f)}(1 + aW_{M0}\theta_i)^{-k_i}, \ i = 1, 2 \]

And
\[
\frac{(E(U|\mu_1))'}{(E(U|\mu_2))'} = \frac{-e^{(\cdot)}a(1 + r_f)(1 + aW_{M0}\theta_1)^{-k_1} - e^{(\cdot)}(-k_1)(1 + aW_{M0}\theta_1)^{-k_1}\frac{1}{a\theta_1}}{-e^{(\cdot)}a(1 + r_f)(1 + aW_{M0}\theta_2)^{-k_2} - e^{(\cdot)}(-k_2)(1 + aW_{M0}\theta_2)^{-k_2}\frac{1}{a\theta_2}} \]
\[
= \frac{-e^{(\cdot)}(1 + aW_{M0}\theta_1)^{-k_1-1}((1 + r_f)(1 + aW_{M0}\theta_1) - \mu_1)}{-e^{(\cdot)}(1 + aW_{M0}\theta_2)^{-k_2-1}((1 + r_f)(1 + aW_{M0}\theta_2) - \mu_2)} \leftrightarrow \frac{(1 + aW_{M0}\theta_1)^{-k_1-1}((1 + r_f)(1 + aW_{M0}\theta_1) - \mu_1)}{(1 + aW_{M0}\theta_2)^{-k_2-1}((1 + r_f)(1 + aW_{M0}\theta_2) - \mu_2)} = \frac{\pi - 1}{\pi} \quad (2.7) \]

For the Inverse Gaussian case, for \( i = 1, 2 \)
\[ E(U) = -e^{-a(W_0 - W_{M0})(1 + r_f)} e^{\frac{\lambda_i}{\mu_i}(1 - \sqrt{1 - \frac{2\mu^2(a\theta_i)}{\lambda_i}})} = -e^{-a(W_0 - W_{M0})(1 + r_f)} \]
\[ e^{\frac{\lambda_i}{\mu_i}(1 - \sqrt{1 + \frac{2\mu^2(a\theta_i)}{\lambda_i}})} \]

And differentiating yields
\[
\frac{\partial E(U)}{\partial W_M} = -e^{(\cdot)}(a(1 + r_f) - \frac{\lambda_i}{\mu_i} 2\sqrt{1 + \frac{2\mu^2aW_{M0}}{\lambda_i}}) = -e^{(\cdot)}(a(1 + r_f) - \frac{a\mu_i}{\sqrt{1 + \frac{2\mu^2aW_{M0}}{\lambda_i}}}) \]

Finally
\[
\frac{(E(U|\mu_1))'}{(E(U|\mu_2))'} = \frac{-e^{(\cdot)}a(1 + r_f) - \frac{a\mu_1}{\sqrt{1 + \frac{2\mu^2aW_{M0}}{\lambda_1}}}}{-e^{(\cdot)}a(1 + r_f) - \frac{a\mu_2}{\sqrt{1 + \frac{2\mu^2aW_{M0}}{\lambda_2}}}} \]
\[
= \frac{(1 + r_f) - \frac{\mu_1}{\sqrt{1 + \frac{2\mu^2aW_{M0}}{\lambda_1}}} e^{(\cdot)-(\cdot)} \leftrightarrow \frac{(1 + r_f) - \frac{\mu_1}{\sqrt{1 + \frac{2\mu^2aW_{M0}}{\lambda_1}}}}{\sqrt{1 + \frac{2\mu^2aW_{M0}}{\lambda_2}}} e^{(\cdot)-(\cdot)}}{\frac{(1 + r_f) - \frac{\mu_2}{\sqrt{1 + \frac{2\mu^2aW_{M0}}{\lambda_2}}}}{\sqrt{1 + \frac{2\mu^2aW_{M0}}{\lambda_2}}} e^{(\cdot)-(\cdot)}} \]
\[
= \frac{\frac{2\mu_1}{\lambda_1}(1 - \sqrt{1 + \frac{2\mu^2aW_{M0}}{\lambda_1}})}{\frac{2\mu_2}{\lambda_2}(1 - \sqrt{1 + \frac{2\mu^2aW_{M0}}{\lambda_2}})} = \frac{\pi - 1}{\pi} \quad (2.8) \]
It is obvious that there is no closed-form solution in these cases, since the product of an exponential and a rational function appears. Nevertheless, the numerical solution from the simulations proves to be sufficiently accurate.

### 3.4 Simulation setup and results

The coefficient values are presented in Table A.1, also accounting for the conditional mean in the Markov case. The table is constructed around the realisation of the distribution and must be read horizontally. An alternative way is to take a distribution and see which one of the optimisers (the one using normal weights, the one using gamma weights or the one using inverse gaussian weights) has fared better. The type of the distribution is the same for both states but the investor does not know which one will be realised eventually. SS denotes the single state case, $\alpha$ denotes the risk aversion parameter set to 0.0001, $\mu, \mu_i$ and $\sigma, \sigma_i$ denote the means and standard deviations for the single and Markov cases. In order to have a conditional mean equal between the two, the mean for the bad state is calibrated, with the effect of pushing the mean of the good state slightly higher. The change in standard deviation due to conditioning proves to be negligible compared to the SS case, so the Markov standard deviation is kept the same. 100,000 returns were drawn from each distribution for the simulations, with the Markov joint sample being split into 98,000 drawings from the good state and 2,000 drawings from the bad state. That is the ratio that corresponds to the probability of each state. The means (standard deviations) were 5%(16%), 8%(24%), 14%(32%) for the single or good state and −15% for the Markov bad state, creating 9 different combinations for each case and risk aversion parameter.

The weights used in each case are the closed form solutions (which correspond to the simulation results) and therefore the table can also be read as cross-examining different combinations of optimised weights and distributions across cases. The investor picks a distribution to use prior to investment and calculates the appropriate amounts, and then the real distribution is realised. The cases where the prediction was correct and the optimal weights were used lie on the diagonal. Across all parameters, the difference in expected terminal utility is practically non-existent and statistically insignificant. The simulations were repeated many times for verification and inconsistencies were present, which are interpreted statistically as overlapping confidence intervals and economically as the outcomes being practically equivalent and interchangeable. The outcome that appeared
CHAPTER 3. ISOLATING FAT TAILS AS FACTOR

the most times was the one recorded, however. Higher variance (which corresponds to a higher realisation of extreme values) or level of risk aversion do not change the result. The conclusion is that a difference between ex ante and ex post returns translated to a different probability for abnormal returns does not have a tangible effect on investor utility. However, even marginally, a correct prediction leads to a higher utility as shown by the table values, and the difference can be significant. This result is in accordance to previous literature, and it agrees with the theoretical assumption that the highest average utility would be achieved when the assumed distribution is the same as the occurring distribution. This outcome is sample dependent: the difference in a test of means (or simply the certainty equivalent) is statistically insignificant and numerically very small. For a tangible difference to take place, one must look beyond a one period game. For the single period variation discussed here, the player that predicts correctly does have a benefit in utility but that benefit is neither strong nor conclusive. Therefore, there are other factors beyond fat tails that enhance extreme risk, such as stochastic volatility, leverage or time variation in equity premia, jump occurrences etc.

Moving on to the Markov setup, the probability of a negative result is now more pronounced. It is directly expressed by the low probability state and the tails of the distributions. However the previous result still holds. Utility is slightly lower than in the single state case but the probability of a jump has a significant effect on neither the weights (they are somewhat lower in the two-state problem) nor the differences between cases. This depicts a pattern very similar to the one above. The investor does not appear to be able to distinguish between states (distributions) and even if he did he does not have a tangible incentive to alter his behaviour. In other words the mere existence of a different, thicker or longer, tail is not enough for a shift in portfolio weights very much. Thus tail risk does not manifest in a single period choice model with or without state uncertainty.

Another point of consideration is how a myopic investor that used the unconditional parameters (weights) instead of the conditional ones would perform. The results proven to be identical so they are omitted.

An additional question examined is how disaster timing affects terminal wealth during multiple periods. Here the investor is assumed to have an investment horizon of 30 periods and his behaviour to be myopic: he is allowed to maximise only for the next period. One of the 30 periods is supposed to be a disaster period, where the return is set to 0.75. The period when this return is realised is, in turn, the 6th (an early disaster), the 17th (in the middle) and the 28th (late). The return sample is the same in all three
cases for comparison reasons. The result is that the earlier the jump occurs the higher the negative impact on terminal wealth is, although the numerical difference is again small. Due to the initial assumptions this is essentially a compounding problem but even so a disaster effect still appears. The results are depicted in figures A.1 - A.3. It is important to notice that the effect on terminal wealth is visible, yet each returns series follows the same pattern. Even a very late jump is not enough to make the investor worse off than a middling or early jump.

3.5 Conclusion

Of course, the setup is too limiting to provide safe conclusions. Constant risk aversion is a questionable assumption since such an individual always allocates a constant sum of wealth in each asset. In addition, wealth effects are non-existent due to the CARA assumption. The investor has the same attitude towards risk no matter how much wealth he has accumulated. - he will always pay the same amount in order to avoid taking a fair gamble, regardless of his wealth. This is a counter-intuitive assumption, especially when one considers that absolute risk aversion has merit when it is decreasing, not constant. A natural expansion is introducing relative risk aversion followed by multiple periods. This could lead to a well-known environment of intertemporal optimization but with the introduction of a bonus structure.

Another limitation is the investment horizon. In a single-period setup, rare events have a very small effect on expected utility, which is calculated "on average". The investor can be fairly confident that a disastrous return will not be realised within the investment period. Even if it does, as shown in figures A.1 -A.3, the CARA investor has time to recover even if the jump takes place late in the investment (it must be noted again that the weights used are the one-period weights, which remain constant for the duration of the investment). With a simple buy-and-hold strategy where the same allocation is selected each year, the investor will have time to make up for losses. The time of the jump does not seem to matter in that context. The limitation of the setup makes for a limiting and counter-intuitive result - while a resilient investor might be willing to make up for early losses, an investor incurring losses late in his investment might be much less tolerant.

The fact that both in the single state and the Markov state the investor is unable to differentiate between outcomes calls for a more complicated setup. No level of average expected returns (mean), risk (variance) or skewness/ kurtosis (distributional shape
containing additional information) seems to be able to have a significant impact. In the more promising Markov setup, which contains tails in both the good and the bad state, the result is the same. The limitation here is not the preselected distributions but the arbitrary frequencies and sizes of the bad outcome. The setup does not allow some finer points to materialise, such as the effect of longer horizons or heteroskedasticity, in an intertemporal framework. Unavoidably, everything is churned into an average terminal utility indicator that eliminates any significant disaster risk.

The most important outcome is that disaster risk, either in the form of fat-tails or different states, is not enough on its own to cause consideration to the investor. A jump structure on its own is not enough to cause differentiation in investor decisions or give rise to tangible distortions and losses under mean-variance optimisation, always "on average". It is, thus, important to introduce a more complicated structure that takes into account more stylised facts of asset returns, such as volatility clustering. The simulations approach will be maintained both in inference (under Bayesian methods) and in the last part of the thesis.

Lastly, it is worth noting that the result is also a nod to the limitations discussed in the literature review when known tailed distributions are used. Although the Gamma and Inverse Gaussian distributions have similar shapes and a skew to the right, the Inverse Gaussian can have a very long tail. In order to achieve that, the peak of the distribution must strangely be rather high, mid-to low segments need to be rather low and the remaining mass creates a very long tail. The Gamma distribution, on the other hand, can be seen as a nesting or parent distribution for the $\chi$, t or even Poisson distributions given the correct parametrisation. Lastly, the difference in the tails does not only imply a change in the frequency of extreme returns but a change in "normal" returns as well. Since the integral of a pdf must always be 1, fatter tails imply lower humps in the distribution, often combined with differences in skewness and kurtosis. This point will be discussed in more detail in Chapter 4, where higher kurtosis is related to the financial crisis.
Chapter 4

Estimation method and preliminary results

4.1 The model

In Chapter 3 the goal was to isolate fat tail effects in a simplistic, tractable setup of mathematical convenience and apply optimisation solutions and simulations in order to derive some preliminary results. Tail risk on its own was not enough to make a difference, and also the use of prespecified known distributions may be inflexible and subject to parametrisation considerations. This chapter will isolate the effect of stochastic volatility instead and study to what extend such a model can replicate a price behaviour similar to the one expected under jumps. In addition, the primary estimation method of the thesis will be discussed along with the relevant literature and an introductory model will be estimated. The aim is to replicate the results of a paper in order to show methodological aptitude and a functional application of the method, to present the method in detail and focus on the effect of stochastic volatility. A description and critique of the method will be provided and its selection will be justified by comparing its advantages and disadvantages to those of other options. The method is Markov Chain Monte Carlo, or MCMC.

4.1.1 Introduction to Bayesian analysis

The central idea of Bayesian analysis, and Markov Chain Monte Carlo in particular, can be best described by a pseudo-historical anecdote. Suppose a person is recovering from an illness, with no company and little entertainment apart from a deck of cards. That person also happens to be highly proficient in statistics. Since the recovery is long, the patient keeps laying solitaires one after the other, sometimes solving the game and others failing - but in the meantime he becomes so experienced that he very rarely
makes a "wrong" move, to the best of his knowledge and ability. At some point the focus changes from solving a laid down solitaire to calculating the probability that this particular solitaire (or any other random one) can be solved. In other words, what is the probability of a combination of cards to occur that can lead to a solvable solitaire? In more Bayesian terms, the goal was to calculate a very complicated conditional probability (or probability density function) regarding the position of every one of the 52 cards in the deck relative to the position of the other 51, and disseminate the probability of a solution. Suffice to say, this problem was analytically untractable and impossible to solve, like many similarly complex problems in statistics, physics and chemistry. However, the patient realised he did not really have to solve the problem. He only had to attend the outcome. In a Law of Large Numbers approach, and given his solitaire prowess that prevented mistakes, he only had to keep on playing solitaire and count how many times he got a solution. The probability of a certain outcome could be approximated given certain conditions (in his case, no cheating) and a large enough number of repetitions that would allow the number of occurrences to approximate the true value.

So a taxing mathematical problem becomes a repeated experiment that provides the desired information a posteriori. The material for the discussion that follows relies heavily on Gilks et al. (1995b) "Markov Chain Monte Carlo in practice", Lynch (2007) "Introduction to Applied Bayesian Statistics for Social Scientists" and Robert and Casella (2005) "Monte Carlo Statistical Methods" along with Gilks and Wild (1992), Henneke et al. (2006) and Gilks et al. (1995a). In order to avoid constant citations that would make the text unreadable, these books and papers should be treated as constant references. They provide an excellent discussion of MCMC from the very basics to elaborate topics.

In more formal terms, let D denote the data of a model and \( \theta \) denote a vector of model parameters and/or variables (for MCMC there is no conceptual difference between the two, as they are treated as unknown quantities). The question asked is "what parameter values are most likely to have been generated (match) by the data set at hand". We define a prior distribution for \( \theta \) which provides potential parameter values. The prior and its calibration provides the source of potential values and is a way to impose existing knowledge and perceptions on the parameter to be estimated. If for example a certain range of values is expected, such as positive, then the prior can be truncated. Having a completely uninformative prior is also a valid option, if there is no reason to impose restrictions or conditions. Therefore, by using Bayes, the joint probability distribution over all quantities of the model is \( P(D, \theta) = P(D|\theta)P(\theta) \) and the distribution of \( \theta \) conditional on D (representing the question) is
4.1. THE MODEL

\[ P(\theta|D) = \frac{P(\theta)P(D|\theta)}{\int P(\theta)P(D|\theta)d\theta} \]

\( P(\theta|D) \) is called the posterior distribution of \( \theta \) and the answer to the question is its expectation. The denominator is the normalising constant and can remain undefined since it simply scales the distribution, turning the equality to proportionality. This is the often unknown distribution of the data, and integrating that term had been the main analytical obstacle in Bayesian inference until the development of MCMC. \( P(\theta) \) is the prior distribution. \( P(D|\theta) \) is the likelihood function for the data at hand, or some sort of distributional form for a model.

The posterior expectation of a function \( f(\theta) \) is

\[ E[f(\theta)|D] = \frac{\int f(\theta)P(\theta)P(D|\theta)d\theta}{\int P(\theta)P(D|\theta)d\theta} \]

MCMC evaluates this by taking sample drawings \( X_t \) from the posterior distributions and approximating this as \( E[f(X)] = \frac{1}{n} \sum_{t=1}^{n} f(X_t) \).

Simplifying notation, let us assume that a vector \( X \) of random quantities (parameters in this context, each following its own prior distribution, although in Bayesian inference there is no conceptual difference between variables and parameters as they are both unknown random quantities) follows a distribution \( \pi(.) \). Then, \( E[f(X)] = \frac{\int f(x)\pi(x)dx}{\int \pi(x)dx} \).

It is important to note that the distribution of \( X \) is allowed to be known only up to the normalization constant (which can be unknown), expressed here by the denominator. This normalisation constant simply scales the distribution and leads to a proportionality. If its general form is known (the quantity within the integral corresponds to a normal, beta etc), this practically means that it is proportional to a familiar pdf of some kind, allowing for direct (Gibbs) sampling of values. If it is not known, other methods such as Metropolis - Hastings sampling need to be applied. The key elements now are ”approximation” and ”convergence”. A Markov chain of \( X \) is a process where the distribution of \( X_t \) depends only on the previous value \( X_{t-1} \). In order to have a meaningful average above, this distribution must converge towards a stationary distribution which must be also distribution \( \pi(.) \). The way to achieve this match is to create an appropriate Markov chain with three properties.
1) irreducibility (the chain has a positive probability to reach every possible value - it can go everywhere).

2) aperiodicity (the chain must not visit states in a pattern or order).

3) positive recurrence (a stationary distribution \( \pi(.) \) exists such that if the initial value of \( X \) is sampled from it then all subsequent \( X \)s must be distributed according to \( \pi \).

When these conditions hold, the chain is called ergodic and converges to the target/posterior distribution \( \pi(.) \). The drawings of the algorithm will eventually converge to the drawings we would get from the posterior distribution (ergodic averages converge to their expectations under the stationary distribution). It can be shown that for a sampler like those discussed in this chapter these conditions always hold. Unfortunately it is a very taxing, complicated and ultimately fruitless exercise that offers no additional intuition. It is telling that, aside from theoretical research, such a check is never performed in applied Bayesian estimation.

Regarding construction, an important but not necessary property is reversibility. Loosely, this allows the chain to move back and forth in time seamlessly. The mathematical requirements are for the chain to be positive recurrent with a stationary distribution \( \pi \) with transition probabilities \( \pi_iP_{ij} = \pi_jP_{ji} \). If we move forwards across states (time) and then backwards, then the transition probabilities \( P \) must be the same. In other words, if we select two random spots in the chain the chance to reach one from the other is the same. If we have many variables, this means that the order of updating them in each iteration must not follow a pattern.

### 4.1.2 Gibbs and Metropolis-Hastings sampling

The next topic is the sampling methodology, or how to construct a chain where its stationary distribution is exactly the posterior distribution in question. The main focus of the discussion is on Gibbs, Metropolis - Hastings and Accept- Reject sampling because these are the methods used in the PhD. However, to provide context, other methods will also be discussed.

Gibbs sampling can be seen as a special case of Metropolis - Hastings and is employed when the conditional distribution of a variable is known in form (or differently up to the normalising quantity) and direct drawing is possible. Assuming more than one variables, so that conditioning has meaning, the algorithm takes drawings of a variable conditional on all other variables and updates the previous value. If in the previous example \( \theta = (\theta_1, \theta_2, ..., \theta_j) \), the algorithm can be summarised as
1) Set initial values for all parameters, or \( \theta^0 = (\theta^0_1, \theta^0_2, \ldots, \theta^0_j) \). Also, specify the order of parameter updating and number of iterations (T).

2) For \( t < T \), sample \( \theta^t_1 \) from \( p(\theta^t_1|\theta^{t-1}_2, \theta^{t-1}_3, ..., \theta^{t-1}_{j-1}) \)

3) Sample \( \theta^t_2 \) from \( p(\theta^t_2|\theta^t_1, \theta^{t-1}_3, ..., \theta^{t-1}_{j-1}) \)

4) Sample \( \theta^t_3 \) from \( p(\theta^t_3|\theta^t_1, \theta^t_2, \theta^{t-1}_4, ..., \theta^{t-1}_{j-1}) \)

5) Continue until \( j \), so all \( \theta_j \) have been updated.

6) Repeat for \( t + 1 < T \) until \( T \) is reached.

This is where the usefulness of conjugate priors becomes apparent. In some cases, it is possible to select a prior distribution such that the product of the prior distribution and the likelihood function will yield the same type of posterior distribution as that of the prior, or at least an identifiable functional form. For example, the product of two normal distributions is a normal distribution, or the product of an exponential and a normal distribution is also a normal distribution. Sampling from a pdf known up to the normalising constant is possible and convenient, and allows for Gibbs sampling in the manner above. Gibbs sampling is useful when sampling from the joint posterior of the parameters is not feasible, but sampling from each individual conditional posterior (or blocks of them) is. The methodology of conjugate priors and conditioning pdfs will be shown in practice when the stochastic volatility model is introduced.

Metropolis-Hastings sampling is most useful when direct sampling is not feasible. Assume state \( X_t \) as current and \( X_{t+1} \) as next. We employ a candidate (proposal) distribution \( q(\cdot|X_t) \) of known, suitable form that provides proposal points for the chain (values that could have come from the posterior, or simply \( Y = X_{t+1} \)). The form of this proposal can be anything - it does not affect the posterior \( \pi(\cdot) \) at all. The candidate point \( Y \) is accepted with probability \( \alpha(X,Y) = \min (1, \frac{\pi(Y)q(X|Y)}{\pi(X)q(Y|X)}) \).

In more detail, the Metropolis-Hastings step essentially asks the question which value is more likely to have come from the posterior distribution \( \pi(\cdot) \), the existing \( X_t \) or the newly proposed \( X_{t+1} \). To answer that, the ratio of the posteriors times the ratio of the proposals, each evaluated at the existing and proposed values. However, the reversion must be noted. If the fraction is split, then the ratio of the posteriors is loosely read as "candidate over existing" values, while the ratio of the proposals is read as "existing over candidate". After \( \alpha(X,Y) \) is calculated, it is compared with a random drawing from a \( U(0,1) \) distribution representing the probability of acceptance. If \( U < \alpha \), the candidate value is accepted. If \( U > \alpha \), the existing value is kept. If accepted, then indeed \( X_{t+1} \) takes the value \( Y \) and the chain moves. If rejected, \( X_t \) remains unchanged.
(or \(X_t = X_{t+1}\)) and the chain does not move. An important element is the shape of
the proposal distribution. Although its nature does not affect the result, a symmetric
proposal like a normal or a uniform distribution allows for \(q(.|.)\) to be simplified. This
makes the calculation of the probability even easier. As a word of caution, for numerical
reasons the calculation of ratios should be done using logarithms. If a symmetric proposal
such as a uniform or a (standard) normal is used, the method is called random-walk
Metropolis - Hastings. Finally, Metropolis - Hastings can be applied to both univariate
and multivariate distributions, contrary to inversion sampling, and does not require an
assumption on, or the construction of, a proposal like rejection sampling and its variants,
where an envelope function is required.

A formal way to illustrate the method can be described as

1) Set the initial value of parameter \(X^0\) and the number of iterations \(T\). Also, select
a proposal density \(q(X^{t+1}|X^t)\) from which new candidate values \(Y\) are sampled.

2) For \(t < T\), calculate \(\alpha(X^t, Y) = \min (1, \frac{\pi(Y)q(X^t|Y)}{\pi(X^t)q(Y|X^t)})\). If \(q\) is symmetric, then
\(\alpha(X^t, Y) = \min (1, \frac{\pi(Y)}{\pi(X^t)})\).

3) Sample \(U(0, 1)\). If \(U < \alpha(X^t, Y)\) set \(X^{t+1} = Y\). If \(U > \alpha(X^t, Y)\) set \(X^{t+1} = X^t\).

The selection of the proposal, like that of the priors, is a matter of consideration.
A ”bold” fat-tailed proposal generates extreme values more frequently, and therefore if
the actual shape of the posterior is not very close to the shape of the proposal it will
take less time to find the area of convergence. On the other hand, these sudden changes
and jumps in the chain may lead the algorithm to explore irrelevant areas. A more
conservative proposal will move more slowly but will show a high number of acceptances,
contrary to the more erratic behaviour of a bold proposal. Also, it might be the case
that after a number of iterations a recalibration of the proposal can improve efficiency.
A common case is in random walks, where the variance of a normal distribution can
be recalibrated after a number of iterations if the ratio of acceptances/rejections is
too high or too low. There is no given rule on the way (Koop et al 2007), but as a
general guide a flat increase (decrease) in variance can improve the ratio and lead to
faster convergence. On the issue of prior selection, conjugacy is a welcome property for
obvious reasons. When that is not feasible, random walks are a common choice due to
being symmetric and non-informative. If bivariate normal distributions are involved,
Inverse Wishart (equivalent to Inverse Gamma) priors and Jeffreys priors are com-
monly used to condition the variance term. The latter is of particular importance, as it
is scale invariant, non-informative and is not affected by reparametrisation. It is defined as

\[ p_J(\theta) \propto I(\theta)^{\frac{1}{2}} \]

where \( X \sim \pi(x|\theta) \) and \( I(\theta) = -E(\frac{d^2}{d\theta^2} \log \pi(X|\theta)) \) is the Fisher information matrix.

The Jeffreys prior may perform poorly in practice for multivariate cases but is very useful for univariate ones. When applied, it breaks down to \( \frac{1}{\sigma^2} \) (although Jeffreys proposes using \( \frac{1}{\sigma} \) due to better convergence properties), and for multivariate normal densities it is \( 1/ \Sigma^{1/2} \) for the covariance matrix.

Accept - Reject (A-R), or Rejection Sampling, is similar to Metropolis - Hastings. The difference is that if \( Y \) is rejected, \( X_{t+1} \) does not take the value \( Y \) but new values for \( Y \) are drawn until there is an acceptance. A prerequisite for that is for the proposal to always lie above the posterior. This property can be represented as \( \pi(\theta) < M \ast q(\theta) \), or a constant \( M \) acting as an upper bound for the ratio of the proposal and the posterior. In that context, the final sample is "exact" - all its elements have been filtered and are more likely to have come from the posterior distribution. However there is a computational drawback. If the chain gets stuck in an area where updating is rare, then the algorithm becomes very slow and inefficient. Metropolis - Hastings will not hinder the entire algorithm but will simply use the existing values, contrary to A-R. Another similar issue is that the Accept - Reject performs worse in multi-dimensional cases, where the probability of a rejection increases exponentially and \( M \) is large. The overall performance of the algorithm depends greatly on how similar the proposal is to the posterior, or having an \( M \) as close to 1 as possible. For multidimensional and multivariate distributions, or abnormally shaped (very peaked, two-pronged etc) this can be very challenging and render Rejection Sampling very inefficient.

Gilks and Wild (1992) improve upon this issue by proposed Adaptive Rejection Sampling (ARS). Instead of using a predefined proposal that may be inaccurate at some regions, they construct a piecewise proposal (in the paper a piecewise exponential but a piecewise linear works as well) that acts as an envelope function and allows for direct sampling. Each time there is a rejection for a proposed value, this value is used in order to define a new piece in the proposal thus updating it and making it more similar to the posterior. The process continues until there is an acceptance, in which case the envelope function is constructed from the beginning in the next iteration. The pieces can be defined both as tangents and as secants. The method works best for log densities and can only be applied to concave functions. Gilks et al. (1995a) expand ARS to include non-log-concave functions. For non-concave, the piecewise envelope may find itself below
CHAPTER 4. ESTIMATION METHOD AND PRELIMINARY RESULTS

the posterior for a piece. To accommodate for such cases, a Metropolis - Hastings step is applied when the ARS step accepts a proposed value $Y$. If $h(Y) = \min(q(Y), \pi(Y))$, then the Metropolis - Hastings step is $U < \min(1, \frac{\pi(Y)h(X_t)}{\pi(X_t)h(Y)})$ where $X_t$ is the existing value.

When the piecewise $q(.)$ lies above the posterior, the algorithm works as before. When it lies before, the extra step ensures that the existing, more preferable value will be retained if a sub-par value is accepted in the ARS step. This methodology is called Adaptive Rejection Metropolis Sampling (ARMS). Martino et al. (2012) point out that in that process the parts below the posterior may never be updated, which means that the envelope will never improve for those regions, and also that the initial use of exponential pieces leads to great numerical problems. The proposed solution is to include an additional accept-reject step when the Metropolis-Hastings step is completed, that will determine if a support point will be added to the piecewise. Another method is to consider past drawings in the construction of the proposal.

For reasons of completeness, it is useful to mention Independence sampling, Slice sampling and Hamiltonian sampling. These are variants of the Metropolis and Metropolis - Hasting algorithms that can be useful under certain circumstances, but go beyond the scope of the thesis.

4.1.3 Reversibility and burn-in periods

Gibbs sampling provides a good basis to explain the importance of scanning during one iteration of the algorithm. Three indicative strategies, that are also applied to M - H, rejection sampling and hybrid algorithms, are Deterministic Scan (DSGS), Random Permutation Scan (RPGS) and Random Scan (RSGS) (Johnson (2009) PhD Thesis). In the first case, the order of updating is set: variable A will be updated first, then variable B (using the updated value of A and the non-updated values of the other variables), then C (using updates As and Bs etc) etc until all variables are updated. This form is non-reversible. In the second case, all variables are updated during the iteration but at random order. For example, among variables A, B and C, B is selected, then between A and C, A is selected, and C is the last to be updated. This version is reversible if the probability of selection is the same for all variables. Here it is 1/3 at first and 1/2 afterwards, with the remaining one being trivially 1, so reversibility holds. In the third case, only one variable is selected for updating at random during one iteration. This is also reversible. Rejection sampling and Metropolis - Hastings are reversible by construction.
One last technical matter is setting initial values and burn-in periods. In order to initiate the algorithm some initial values must be set for the chain to begin, which must not affect the final result of convergence. Unavoidably, some starting points are better than others but there is no way of knowing. To avoid a bias, the first iterations (burn-in period) are discarded so the effect of the starting points is neutralised when the averages are calculated. Again, there is no rule on the length of the burn-in period.

4.1.4 Advantages, disadvantages and justification of method selection

This section will discuss the relative advantages and disadvantages of MCMC, both in theory and in practice. In addition, MCMC will be compared to other methods and an explanation will be provided on why it is more advantageous to be used in the context of the thesis.

The first thing to note about MCMC is that, by construction, it is suitable to deal with the simultaneous estimation of many parameters and variables. This makes it particularly efficient when complicated models are involved, such as SVCJ. Because Bayesian inference relies on quantifying and assessing probability and does not focus on explicit parameter estimation, it is better suited to address highly complex models. Also, despite the variations in sampling and other technical sides, the basis is always the Bayes rule and how to move from a prior distribution to a posterior through likelihood. It is not required to know the exact shape or form of the joint posterior, or impose exogenous assumptions, but the problem can be broken down to individual bits estimated separately. This is not an advantage shared by the General Method of Moments (GMM), Maximum Likelihood Estimation (MLE) or GARCH estimation and makes MCMC very versatile. In addition, MCMC is agnostic. It only requires the three properties of aperiodicity, positive recurrence and irreducibility to hold, and it does not depend on the initial values or the posteriors used. In fact, a proper algorithm always converges no matter what the priors or the initial values are, and the chains will explore first and converge after even if the true parameter values are used to initialise the process. That was the case (and a performance test, among others) for the PhD. Third, in-depth knowledge of distributional properties is not required apart from the form of the likelihood. The exact shape can be unknown and the functional form can be multidimensional or very non-standard.

The problems of MCMC, particularly in the context of the PhD, are empirical. First, although different priors are mathematically equivalent and will all lead to convergence
CHAPTER 4. ESTIMATION METHOD AND PRELIMINARY RESULTS

as long as the fundamental conditions hold, in practice they may affect the speed of convergence, overall performance and may even cause ”prior bias”, or limiting the chain to a region. In many cases this will lead to a software crash because the parameters will violate some condition somewhere, or to an erratic chain and the existence of a problem will become obvious. If the problem is detected, one needs to simply use a different prior and recondition. Second, in practice MCMC can be time-consuming. The long burn-in periods in MCMC literature with jumps are indicative of the issue, because a lot of iterations need to be discarded before it is safe to assume convergence. A poor selection of priors (e.g. in a M - H step with many rejections) may result to slow mixing of the chains, slow convergence or ”prior bias”, causing the algorithm to get stuck in a region. An example is the problem of a ”two-pronged distribution” phenomenon where only one peak is covered. Those problems can be addressed by selecting different priors or implementing a Metropolis step and using ARMS, respectively. Discretization bias is a concern but in the papers that do discuss it (most notably Eraker et al. (2003), it is found to be negligible with high frequency data.

Some more technical issues are the slow mixing of the chain if the proposal is too conservative or too bold, the production of correlated samples. The first issue is easily solved by readjusting the proposal. The second issue can be dealt with sampling in blocks or from joint posteriors. If there is reason to believe that two or more parameters are correlated, it is possible to construct a bivariate or multivariate posterior for them and sample in pairs or groups. This does not exclude prior conjugacy, as in many cases multivariate conjugate priors can also be used (e.g. normal). From another point of view, output is still expressed in ergodic averages, so in essence MCMC is an elaborate method to estimate means and variances ”on average”. This average can have very low standard deviation but it still is an estimate over an entire sample that may vary a lot, contain structural breaks etc. Therefore, critique on the use of such parameters in forecasting rare events cannot be avoided. However, both MCMC and SVCJ are better equipped to address tail risk than other counterparts.

Despite all the advantages mentioned above, MCMC is still a numerical technique. This makes estimation errors unavoidable, as in all methods, but in addition it is from difficult to impossible to know how large that error is. The best guide is the standard deviation of the realisations of the chain, but the equivalent of a t-test cannot be used. Still, measures like the Odds ratio to compare between models are available.
4.1.5 MCMC and Maximum Likelihood/ General Method of Moments

Maximum Likelihood can, in a sense, include both GMM and GARCH estimation since is it based on the idea of choosing parameters that maximise the likelihood function. This assumes the existence of a true parameter value which is estimated according to a confidence interval. On the opposite, Bayesian analysis does not discriminate between variables and parameters, treating them as unknown quantities, threatens them both as random variables, and focuses on the uncertainty surrounding them. It calculates their expectation based on probability, rather than trying to find an optimal value. The difference between classical statistics and Bayesian inference can also be described as follows. In statistics, ML for example, one is concerned with estimating a parameter with as much precision as possible. In Bayesian inference, the uncertainty of a system is expressed, inferred and measured by probability, given a set of observations. These is not a hidden true value to be approximated, only an amount of uncertainty that can be analysed through the prior - posterior path and the Bayes rule. In an even cruder way, it relies on empirical observation of results (ex post) rather than a hidden distributional law (ex ante). Where ML needs to maximise a very complicated, non-standard, high-dimensional multivariate distribution (a task ranging from difficult to impossible), MCMC only needs to condition that likelihood and approximate the posteriors of each parameter Paap (2002) for a comprehensive discussion on the subject). This is both feasible and faster.

In addition, the very nature of ML estimation assumes some underlying features for the form of the likelihood, such as being differentiable and at least some moments to exist. This becomes very obvious in the case of the General Method of Moments and the Cauchy-Lorentz distribution (Student’s t with one degree of freedom). GMM cannot be applied because that distribution simply does not have a mean. Because we are dealing with unknown distributions of unknown shape and properties, the caveat of imposing false underlying assumptions via the adoption of a methodology is always present, or very simply put estimating things that are not there. The existence of (higher) moments is a very strong assumption for finance, because one can be fairly certain on the existence of the mean and somehow certain on the existence of variance, but for skewness and kurtosis there is very little empirical evidence.

Another advantage of MCMC in jumps estimation is that if a possible parameter value does not appear in the sample its ML estimation will be zero. That is not the case
for MCMC, where the Bayesian estimator will reach zero only when forced by a prior. In
the case of rare observations, such as jumps, Maximum Likelihood can be problematic.
It is indicative that when ML is used along with jumps the model is fairly simple, e.g. as
in Aıt-Sahalia (2004) or Bates-like models with only one process. This is telling of the
practical limitations in application.

Finally, Bayesian analysis allows the introduction of beliefs via the priors, a feature
not shared by Maximum Likelihood. For a proper prior and sufficiently large sample,
ML and Bayesian estimators will converge. Also, in practice ML estimates are more
affected by noise in the data. Maximum likelihood does not allow for prior information
and its interpretation relies on having a tractable pdf and an idea about the distribution,
leading to conclusions on something that might not be there.

4.1.6 MCMC and GARCH

Compared to GARCH, MCMC has a series of advantages. In a nutshell, GARCH
maximises the likelihood function given a set of initial parameters and operates as a black
box: convergence to the true value cannot be monitored and is often dependent upon the
initial value of the parameters. For some parametrisations the GARCH estimator may
not converge at all or perform badly, while for others it might converge. MCMC on the
other hand is exactly the opposite. It is completely indifferent to the selection of initial
parameters and despite some being more reasonable than others the only difference is the
speed of convergence. In addition, the performance of the algorithm can be visualised
by plotting the chains. A realisation of the parameter Markov chains can reveal if the
chain has degenerated (e.g. create a ”funnel” due to increasing variance), sufficiently
explored its region and, most importantly, converged. A converged chain demonstrates
very low variance around a centre value and looks like a ”hairy caterpillar” rather than
a ”random walk”. That type of empirical validation is not possible when using GARCH,
which relies on a set of difficult to track tests and in addition may be subject to structural
issues (proper number of lags). A discussion on formal convergence diagnostics, along
with their potential failures, can be found in Cowles and Carlin (1996). A complete
and excellent discussion on MCMC techniques in ARMA-GARCH models is given by
Henneke et al. (2006). As a final note, when (G)ARCH models are not estimated by
MCMC, they often assume a Student’s t distribution as in Bollerslev (1987) or Jondeau
and Rockinger (2009, 2012). This highlights again the matter of imposing distributional
features, where MCMC along with jump-diffusion allows for a more independent, flexible
structure.
4.2 Technical application - Stochastic Volatility

4.2.1 Derivation of posterior distributions

This section focuses on the paper to be replicated. It is Hautsch and Ou (2008), which uses a simple Auto Regressive Stochastic Volatility (AR SV) model. The paper relies heavily on Kim et al. (1998), one of the first papers to apply MCMC estimation in SV models, but differentiates by using DAX, Dow Jones and the GBP/USD rate as data instead of only the latter and for a much longer period. As a contribution, an advancement in reversibility will be introduced. The order of parameter updating in Hautsch and Ou (2008) is fixed and thus non-reversible. The MH steps are reversible by construction. However there is no theoretical ground to impose a certain order (e.g. why should the mean be updated before the mean reversion parameter?) despite purely technical justifications. Therefore a Random Permutation Gibbs Scan (all parameters updated in random order during each iteration) will be applied which will ensure reversibility.

The two equations describing the model are

\[ y_t = \exp(h_t/2)u_t, \quad u_t \sim N(0,1) \]  \hspace{1cm} (4.1)

\[ h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \quad \eta_t \sim N(0, \sigma^2) \]  \hspace{1cm} (4.2)

where \( y_t \) is log returns at time \( t \), \( h_t \) is log volatility modelled by a stationary AR process with parameter \(|\phi| < 1\). \( u_t, \eta_t \) are error terms and the unconditional distribution of \( h_t \) is \( h_t \sim N(\mu, \frac{\sigma^2}{1-\phi^2}) \). The returns formulation is the equivalent of the Heston square-root volatility model without drift.

Setting \( \theta = (\mu, \phi, \sigma^2) \) to denote a vector of all the parameters and \( h = (h_1, ..., h_T) \), by Bayes’ theorem we have

\[ p(\theta, h|y) \propto p(y|\theta, h) * p(h|\theta) * p(\theta) \]

with the LHS being the posterior distribution, which is proportional to the likelihood function (first term) specified by (4.1), the conditional distribution of the volatility states (second term specified by (4.2) and the prior distribution (third term).

Specifying prior distributions for the parameters, \( p(\mu) = N(\alpha_\mu, \beta^2_\mu) \), \( p(\phi) = N(\alpha_\phi, \beta^2_\phi) \) truncated between -1 and 1, and \( p(\sigma^2) = IG(\alpha_\sigma, \beta_\sigma) \). The parameters of the priors (hyper-parameters) are set by the researcher and the distributions are normal, normal and inverse gamma. They all yield conjugate priors for the respective terms, and the
inverse gamma (or Wishart) prior is a common option for conditioning variance. Their values may be completely agnostic, denoting no particular knowledge of the values/ sign of the parameters (e.g zero mean and huge variance) or may impose some prior belief on the value of the parameter. In this case they are agnostic.

The conditional posteriors of each parameter are

\[ p(\mu | y, h, \phi, \sigma^2) \propto p(y | h, \mu, \phi, \sigma^2) \times p(h | \mu, \phi, \sigma^2) \times p(\mu) \]  
\[ (4.3) \]

\[ p(\phi | y, h, \mu, \sigma^2) \propto p(y | h, \mu, \phi, \sigma^2) \times p(h | \mu, \phi, \sigma^2) \times p(\phi) \]  
\[ (4.4) \]

\[ p(\sigma^2 | y, h, \phi, \mu) \propto p(y | h, \mu, \phi, \sigma^2) \times p(h | \mu, \phi, \sigma^2) \times p(\sigma^2) \]  
\[ (4.5) \]

The first RHS term is the full conditional likelihood which is a constant with respect to the model parameters and can thus be omitted. The only variable that affects \( y \) is \( h \), so in the conditionals for the parameters \( \mu, \phi, \sigma \) \( p(y | h, \mu, \phi, \sigma^2) \) does not play any role. \( h \) subsumes all information about the other parameters. We now focus on the second term, which includes conditioning every element in the vector of vol.states according to the other parameters. This yields

\[ p(h | \mu, \phi, \sigma^2) = p(h_1 | \mu, \phi, \sigma^2) \prod_{t=1}^{T-1} p(h_{t+1} | h_t, \mu, \phi, \sigma^2) \]

The separation of \( p(h_1 | .) \) happens because for \( h_1 \) there is no prior state \( h_{t-1} \) so the unconditional mean and variance need to be used, instead of the conditional ones used for all other \( t \). Also, manipulating (2) yields \( h_t \sim N(\mu + \phi(h_{t-1} - \mu), \sigma^2) \). The posterior for \( \mu \) and \( \phi \) are, therefore, the product of normal probability density functions. In what follows the normalising constant (generally \( 1/\sqrt{2\pi \sigma^2} \)) is omitted because it simply scales the results, and the focus is on the exponential part of the normal pdf. Restating, (4.3) becomes

\[ p(\mu | y, h, \phi, \sigma^2) \propto p(h | \mu, \phi, \sigma^2) \times p(\mu) = p(h_1 | \mu, \phi, \sigma^2) \prod_{t=1}^{T-1} p(h_{t+1} | h_t, \mu, \phi, \sigma^2)p(\mu) \]

\[ = \exp(-\frac{(h_1 - \mu)^2}{2\sigma^2/(1-\phi^2)}) \exp(-\frac{\sum_{t=1}^{T-1}(h_t - \mu - \phi(h_{t-1} - \mu))^2}{2\sigma^2}) \exp(-\frac{(\mu - \alpha\mu)^2}{2\beta^2\mu}) \]  
\[ (4.6) \]
4.2. TECHNICAL APPLICATION - STOCHASTIC VOLATILITY

\[ \exp\left(-\frac{1}{2} \frac{h_1 - 2 \mu h_1 + \mu^2}{\sigma_h^2/(1 - \phi^2)}\right) = \exp\left(-\frac{1}{2} \frac{(\mu^2 - 2 \mu h_1)(1 - \phi^2)}{\sigma_h^2}\right) \]

\[ = \exp\left(-\frac{1}{2} \frac{\mu^2(1 - \phi^2)}{\sigma_h^2} - \frac{1}{2} \frac{-2 \mu h_1(1 - \phi^2)}{\sigma_h^2}\right) \] (4.7)

This seemingly awkward way to write the equation will be helpful to understand the manipulations that lead to a normal pdf later on. It is also illustrative of the process, hence the unnecessary detail. With a similar algebraic treatment, the middle exponential part of (4.6) becomes

\[ \exp\left(-\frac{1}{2} \frac{\sum_{t=1}^{T-1} (h_t - \phi h_{t-1} - \mu(1 - \phi))^2}{2\sigma_h^2}\right) \]

\[ = \exp\left(-\frac{1}{2} \frac{\mu^2(1 - \phi)^2(T - 1) - \sum_{t=1}^{T-1} 2\mu(1 - \phi)(h_t - \phi h_{t-1})}{\sigma_h^2}\right) \] (4.8)

The point of interest is how the sum across T created the (T-1) multiplication with the quadratic term. The sum applies only to volatilities, and since \( \mu \) does not depend upon time it will be multiplied by the length of the sum. The separation of the \( \mu^2 \) and \( 2\mu \) terms can be omitted. The third term of (4.6), the prior, is simply

\[ \exp\left(-\frac{\mu^2 - 2\mu \alpha + \alpha^2}{2\beta^2}\right) = \exp\left(-\frac{\mu^2 - 2\mu \alpha}{2\beta^2}\right) \] (4.9)

Collecting the results together, the posterior is proportional to the product of the three exponential parts, so summing the powers in (4.7) - (4.9) and collecting terms leads to

\[ \exp\left[-\frac{1}{2} \left(\frac{(T - 1)(1 - \phi)^2}{\sigma_h^2} + \frac{1 - \phi^2}{\sigma_h^2} + \frac{1}{\beta^2}\right)\mu^2 \right. \]

\[ - \frac{1}{2} \left(\frac{h_1(1 - \phi^2)}{\sigma_h^2} + \sum_{t=1}^{T-1} (1 - \phi)(h_t - \phi h_{t-1})\right) \left(\frac{\alpha}{\beta^2}\right)(-2\mu) \right] \]

\[ = \exp\left[-\frac{1}{2} \left(A\mu^2 - 2B\mu\right)\right] = \exp\left[-\frac{A}{2}(\mu^2 - 2\frac{B}{A}\mu)\right] \]

This formulation is almost the exponential part of a normal pdf. The variance is \( 1/A \) in the \( 2/A \) term while the \( B/A \) term is the mean if the square is completed in the parenthesis. The square can be completed, seemingly out of nowhere, because the additional \((B/A)^2\) term needed does not contain \( \mu \) so it does not affect the proportionality
- it is simply a constant term than can be ignored. After completing the square, the posterior can be written as

\[ p(\mu | y, h, \phi, \sigma^2_\eta) \propto \exp\left( -\frac{(\mu - B/A)^2}{2A} \right) \]

so \[ p(\mu | y, h, \phi, \sigma^2_\eta) \propto N(B/A, 1/A) \]

These calculations are standard process when using normal distributions. In the same manner, the posterior for \( \phi \) is also shown to be a normal. The posterior for \( \sigma^2_\eta \) is of slightly more interest. In a more succinct manner and substituting directly for (4.4),

\[
p(\sigma^2_\eta | y, h, \phi, \mu) \propto p(h_1 | \mu, \phi, \sigma^2_\eta) * \prod_{t=1}^{T-1} p(h_{t+1} | h_t, \mu, \phi, \sigma^2_\eta) * IG(\sigma^2_\eta | \alpha, \beta) \\
\propto \left( \frac{1}{\sqrt{2\pi\sigma^2_\eta}} \right)^T \exp\left( -\frac{(h_1 - \mu)^2(1 - \phi^2)}{2\sigma^2_\eta} - \frac{\sum_{t=1}^{T-1} (h_t - \mu - \phi(h_{t-1} - \mu)^2)}{2\sigma^2_\eta} \right) \times \frac{\beta^\alpha_\sigma}{\Gamma(\alpha_\sigma)} (\sigma^2_\eta)^{\alpha_\sigma+1} \exp\left( -\frac{\beta^\alpha_\sigma}{\sigma^2_\eta} \right)
\]

Because \( \sigma^2_\eta \) appears in the denominator, no terms are excluded in the exponential parts and it also appears in the root quantity to the left. The T power appears from the product of the normal distributions. Ignoring all the constants and the gamma function and collecting terms,

\[
p(\sigma^2_\eta | y, h, \phi, \mu) \propto (\frac{1}{\sigma^2_\eta})^{T/2} (\frac{1}{\sigma^2_\eta})^{\alpha_\sigma+1} \exp\left( -\frac{2\beta_\alpha}{\sigma^2_\eta} + (h_1 - \mu)^2(1 - \phi^2) + \sum_{t=1}^{T-1} (h_t - \mu - \phi(h_{t-1} - \mu)^2) \right) \Leftrightarrow \\
(\frac{1}{\sigma^2_\eta})^{D} \exp - \frac{C}{\sigma^2_\eta}
\]

which is proportional to an inverse gamma function with parameters \( IG(D, C) \). The derivation also illustrates the reason why an inverse gamma function is needed to produce a conjugate prior.

To sum up, after substituting the prior distributions in the conditional posteriors, we create conjugate priors proportional to the priors but with different parameters specified as follows.

\[ p(\sigma^2_\eta | y, h, \phi, \mu) \propto IG(\hat{\alpha}_\sigma, \hat{\beta}_\sigma) \]
4.2. TECHNICAL APPLICATION - STOCHASTIC VOLATILITY

\[ p(\phi | y, h, \mu, \sigma^2_n) \propto N(\hat{\alpha}_\phi, \hat{\beta}_\phi), \text{ truncated between -1 and 1} \]

\[ p(\mu | y, h, \phi, \sigma^2_n) \propto N(\hat{\alpha}_\mu, \hat{\beta}_\mu) \]

with the corresponding formulas for each parameter being

\[ \hat{\alpha}_\sigma = \alpha_\sigma + \frac{T}{2} \]

\[ \hat{\beta}_\sigma = \beta_\sigma + \frac{1}{2} \left\{ \sum_{t=1}^{T-1} (h_{t+1} - \mu - \phi(h_t - \mu))^2 + (h_1 + \mu)^2 (1 - \phi^2) \right\} \]

\[ \hat{\alpha}_\phi = \hat{\beta}_\phi \left\{ \frac{\sum_{t=1}^{T-1} (h_{t+1} - \mu)(h_t - \mu)}{\sigma^2_n} \right\} + \frac{\alpha_\phi}{\hat{\beta}_\phi} \]

\[ \hat{\beta}_\phi = \left\{ \frac{\sum_{t=1}^{T-1} (h_t - \mu)^2 - (h_1 - \mu)^2}{\sigma^2_n} + \frac{1}{\hat{\beta}_\phi^2} \right\}^{-1} \]

\[ \hat{\alpha}_\mu = \hat{\beta}_\mu \left( \frac{h_1 (1 - \phi^2) + (1 - \phi) \sum_{t=1}^{T-1} (h_{t+1} - \phi * h_t)}{\sigma^2_n} \right) + \frac{\alpha_\mu}{\hat{\beta}_\mu^2} \]

\[ \hat{\beta}_\mu^2 = \left\{ \frac{1 - \phi^2 + (T - 1)(1 - \phi)^2}{\sigma^2_n} + \frac{1}{\hat{\beta}_\mu^2} \right\}^{-1} \]

Each posterior can be sampled directly and each drawing will be conditional upon the values at hand of all the other parameters. As the chains progress, the values of each parameter will start concentrating around an average area, ”limiting” each other. The time and speed of convergence does not have to be the same for all.

On the contrary, the full conditional posterior for \( h_t \) is non-standard and we can not obtain any direct drawings. For this step, an Accept - Reject step must be implemented. From Bayes’ theorem we get

\[ p(h_t | y, h_{-t}, \theta) \propto p(y_t | h_t, \theta) * p(h_t | h_{-t}, \theta) = \frac{1}{\sqrt{2\pi e^{\exp(h_t)}}} e^{\exp\left(-\frac{y_t^2}{2 e^{\exp(h_t)}}\right)} * p(h_t | h_{-t}, \theta) \]

\[ = f^*(y_t, h_t, \theta) * p(h_t | h_{-t}, \theta) \]  \hspace{1cm} \text{(4.11)}

where we see that this posterior does not belong to a known family of pdfs (\( h_{-t} \) denotes all elements of \( h \) excluding the currently selected \( h_t \)). In order to progress, we need a suitable proposal distribution - we can derive one through a Taylor approximation on (4.11). Here, formula (4.1) is also used because \( h_t \) appears in it. The whole point is how to construct an appropriate envelope function. We notice that the AR model has
a Markov structure where the current observation depends upon the previous and the forthcoming one in the chain, so

\[ p(h_t|h_{t-1}, \theta) = p(h_t|h_{t+1}, h_{t-1}, \theta) = p_N(h_t|\alpha_t, \beta^2) \quad (4.12) \]

where \( \alpha_t = \mu + \frac{\phi(h_{t+1} - \mu) + h_{t-1} - \mu}{1 + \phi^2} \) and \( \beta^2 = \frac{\sigma^2}{1 + \phi^2} \) are the parameters of the normal density function \( p_N(\cdot|\cdot) \).

This result comes from taking two successive values \( h_t, h_{t+1} \) from (4.2). After tedious algebra that follows earlier methodology (excluding all terms not containing \( h_t \)), it is shown that

\[ p(h_t|h_{t-1}, \theta) \propto \exp\left(-\frac{1}{2} \frac{(h_t - (\mu(1 + \phi) + \phi h_{t-1}))^2}{\sigma^2} \right) \exp\left(-\frac{1}{2} \frac{(h_{t+1} - (\mu(1 + \phi) + \phi h_t))^2}{\sigma^2} \right) \]

which corresponds to (4.12). As a reminder, the errors in (4.2) are normal.

In practical terms, for each repetition there is a "cursoring" across all elements of vector \( h \) that draws a new vector of \( h \), with \( h_1 \) drawn from its unconditional distribution. The Accept - Reject step uses \( \exp(-h_t) \) being bounded by a linear function in \( h_t \) so we expand (4.11) around \( \alpha_t \) to get

\[ \log f^*(y_t, h_t, \theta) \leq -\frac{1}{2} \log(2\pi) - \frac{1}{2} h_t - \frac{y^2_t}{2} \exp(-\alpha_t)(1 + \alpha_t - h_t) = \log^*(y_t, h_t, \theta) \]

Since \( p(h_t|h_{t-1}, \theta) = p_N(h_t|\alpha_t, \beta^2) \) it follows that

\[ p(h_t|h_{t-1}, \theta) * f^*(y_t, h_t, \theta) \leq p_N(h_t|\alpha_t, \beta^2) * g^*(y_t, h_t, \theta) \]

after removing the logs, so the RHS can be written as \( k * p_N(\alpha^*_t, \beta^2) \) with \( k \) constant. The parameters are \( \alpha^*_t = \alpha_t + \frac{\beta^2}{2} (y^2_t \exp(-\alpha_t) - 1), \beta^2 = \) the same as before. This means that up to a constant \( k \) the RHS is bound by \( p_N(\alpha^*_t, \beta^2) \) which is the proposal envelope distribution.

### 4.2.2 Sampling methodologies

The algorithm for sampling each parameter is as follows

1) set initial values \( h^{(0)}, \mu^{(0)}, \phi^{(0)}, \sigma^2_{\eta} \)

2) For a number of steps \( s = 1,...,S \), sample each parameter conditional upon the value of all the other parameters in the previous step, or if they have already been updated from the current step. In Hautsch and Ou (2008) this happens in a set sequence:
first sample vector \( h \), then \( \sigma_\eta^2 \), then \( \phi \), then \( \mu \). More formally,

*Sample \( h_t^{(s)} \) from \( p(h_t | y, h_{t-1}^{(s)}, h_{t+1}^{(s-1)}, \mu^{(s-1)}, \phi^{(s-1)}, \sigma_\eta^2) \)

*Sample \( \sigma_\eta^2 \) from \( p(\sigma_\eta^2 | y, h^{(s)}, \mu^{(s-1)}, \phi^{(s-1)}) \)

*Sample \( \phi^{(s)} \) from \( p(\phi | y, h^{(s)}, \sigma_\eta^2, \mu^{(s-1)}) \)

*Sample \( \mu^{(s)} \) from \( p(\mu | y, h^{(s)}, \phi^{(s)}, \sigma_\eta^2) \)

The introduced alteration is updating \( h, \sigma_\eta^2, \phi \) and \( \mu \) in random order during the step. This makes the algorithm reversible. \( \mu, \phi \) and \( \sigma \) are sampled directly (Gibbs sampling) while the Accept - Reject algorithm for \( h_t \) is

for \( t=1,\ldots,T \)

1) Draw candidate \( h_t^* \) from \( p_N(h_t | \alpha^*_t, \beta^2) \)

2) Draw \( U \) from uniform distribution \( U[0,1] \)

3) If \( U \leq \frac{f(y, h_t^*, \theta)}{g(y, h_t, \theta)} \) set \( h_t = h_t^* \) otherwise return to 1 and take a new drawing of \( h \).

4) Repeat until acceptance.

The total number of runs is 10.000, of which the first 5.000 are discarded as burn-in period. Volatility states are computed according to

\[
hh_t = \frac{1}{G} \sum_{g=1}^{G} \exp \frac{h_t^{(g)}}{2}
\]

current where \( G \) is the remaining repetitions after discarding the burn-in and \( g \) the realisation of the chain.

### 4.2.3 Alterations and expansions

A number of expansions and alterations are introduced, for various reasons.

1) Data set issues

Hautsch and Ou (2008) use daily log returns of three indices (DAX, Dow Jones, GBP/USD FX rate) from 1-1-1991 to 21-3-2007. Only the Dow Jones series was used for replication. They report summary statistics and a sample size of 4.231 observations. Using the DJINDUS series from Datastream, the number of observations between these two dates is 4.232, with the very first log return for 1-1-1991 being zero and a non-trading
day (and naturally excluded). However, all other trading days have not been excluded from the 4.231-sized sample. After doing so, the actual number of trading days for that period is 4.079. In practice this does not affect the results, because all log returns are very close to zero and similar to arithmetic returns, since daily data is used. The sample statistics and results are only slightly affected because log returns are very close to zero by construction. The corrected values are as reported. In addition, a sample between 3-1-2005 and 23-7-2015 which includes the financial crisis of 2007-2009 was used. The data set comes from Datastream (tagged DJINDUS, 2656 observations excl. non-trading days) and includes daily log-returns. Comparative statistics of the samples are reported in Table B.2.

2) Change of Metropolis step

Although Hautsch and Ou (2008) set up an Accept - Reject step, during the replication it proved to be problematic. When Accept - Reject sampling was used, the code would get stuck very early (70 - 200 repetitions), meaning that there no values could be drawn from a posterior given the most recent parameter values and distributional forms. The problematic step was the volatility Accept - Reject. The mathematical derivation is completely accurate but the implementation appears to be less than perfect. When the Metropolis - Hastings step was employed instead, the results of the paper could be fully replicated and are stated here. The proposal and posterior are exactly the same as discussed above, with the only difference being in step 3 which becomes

3) If $U < \frac{f^*(y_t, h_t^*, \theta)}{g^*(y_t, h_t^*, \theta)}$ set $h_t = h_t^*$ otherwise keep existing value.

To find the reason for this discrepancy the authors were contacted. Access was given to the original code but, unfortunately, it was written in C++ and had very few comments. After reverse-engineering it, it became apparent that the step was coded as $\log U = \frac{f}{g}$. Since $0 < U < 1$, $\log U < 0$ and obviously all drawings are accepted. There was no part of the code where $f^*$ or $g^*$ were specified or turned into logarithms. There was also no part of the code where anything else was defined or turned into logarithm either. Mechanically, this error most likely causes all drawings to be accepted and therefore the chain does not get "stuck". This is a very cautious result, due to lack of proficiency in C++. It is very difficult to follow the entire process so there is always the possibility of misinterpreting something. However, since the mathematical derivation is correct, the aim of demonstrating the construction of a proposal and how an Accept-Reject step
works is achieved.

3) Reversibility

As already discussed, the algorithm was changed to reversible from non-reversible. Also, a mistake in the Taylor expansion for \( \log f \) was corrected. It contained an additional erroneous \( \exp(h_t) \) term in Hautsch and Ou (2008) but is correctly stated in Kim et al. (1998).

4.3 Results and conclusion

The expansions are 1) the application of a Metropolis - Hastings step instead of an acceptance - rejection step 2) turning the algorithm from deterministic to reversible 3) running an additional simulation with more recent data and more violent movements in returns. The step change along with reversibility still produce results very close to the original paper, showing that the more flexible method is still efficient. There is no theoretical ground to impose one order of updating or another. Intuitively, the agent conducting this repeated experiment has no reason to update the mean before volatility etc. The success of reversibility shows that an artificial construct is not needed, and that a reversible structure is equally effective.

In addition, convergence is largely achieved after 70 - 200 repetitions. This was a useful observation since for a full replication with 25,000 repetitions the duration would be more than 30 hours, which imposed time restrictions. This was conducted only once, to verify that everything worked as intended, and all other operations were conducted with 10,000 repetitions of which 5,000 were burn-in period. As seen from the realisations of the chains, the loss of accuracy was practically non-existent. This is not the case for \( \sigma_\eta \) which is the slowest to converge and remains somehow inaccurate, quite possibly due to the low number of iterations. Even that, however, does not deviate enormously.

There are two elements for replication. The first one is the table of parameters, and the second one the comparison between smoothed volatility states and absolute log returns. Table B.1 shows almost exact replication of the paper results, excluding the mean of \( \sigma \) which is slightly higher. Still, the values obtained by the corrected and non-corrected data are almost the same, while its SDs are identical for all cases. Unit root behaviour is also verified. Comparing the 1991 - 2007 and the 2005 - 2015 period, the mean of returns is the same but with a higher standard deviation, while \( \sigma \) is also
higher and more volatile as one would expect. The most important finding is that the model is still able to identify volatility clustering with precision, even during a period with severe jumps. This is demonstrated by Figures B.3, B.4 and B.5 where smoothed volatility estimates closely mimic absolute log-returns.

The main aim of the chapter is to demonstrate technical aptitude and validate the application of an estimation process via replication. It is useful to discuss the estimation results in further detail, although they are almost identical to the base paper. Unavoidably, the conclusions are similar with the exception of standard deviation and kurtosis. The value of $\phi$ (0.9876) shows high persistence for the autoregressive process. This is almost a unit root, implying that the process is getting close to being non-stationary. This is a common finding for returns series and is also connected to volatility clustering and stochastic volatility (heteroskedasticity), like in the model at hand. Equation (4.2) can be rewritten as

$$h_t = (1 - \phi)\mu + \phi h_{t-1} + \eta_t$$

where, after substituting the numerical values of the replication, the first term is negative and equal to -0.119. The standard deviations of all parameters are very small and almost identical to the original estimates. The difference is in $\sigma$, whose value is 0.127236 at the replication compared to the original 0.087. The standard deviation of both estimates is 0.01, so they are both accurate. Both under the original and the corrected sample, the replication of $\sigma$ is virtually the same. The most important change is in the Accept - Reject step. If that was indeed the reason for the discrepancy, then all other parameters would be inaccurate to some extent. This is not the case, and there are two more elements in favour of the replication outcome contrary to the original value. The first is the coding deficiencies discussed above, which cast doubts on the validity of the original results. The second is the comparable values of the corrected sample and the 2005 - 2015 sample. This shows that the code operates consistently and accurately.

The 2005 - 2015 sample tests parameter accuracy for a very volatile period, such as the financial crisis. The parameter estimates are almost identical to those of the corrected sample but with much higher standard deviations. This is to be expected, since the model is ill equipped to deal with so sudden movements in prices and estimation is unavoidably less accurate. Nevertheless, the impact is limited to the standard deviations of $\mu$ and $\phi$. $\sigma$ is up at 0.1751 from 0.1276, which shows an attempt to accommodate for the large returns movements. Its standard deviation is also higher but still comparable to the original set. The reason for these increases is depicted in Figure B.5 in the Appendix.
Although volatility is able to track absolute returns rather closely, it is unable to track outliers such as those observed to the left and right of the main masses on the returns graph. This is a strong visual representation to introduce jumps and account for those movements, because they cannot be replicated by an SV setup and volatility clustering has its limitations. Smoothed volatility states cannot represent such high deviations. To a lesser extent, this behaviour can be observed in figures B.3 and B.4 for the original and corrected samples. Absolute or squared returns can be seen as a reasonable proxy for volatility, but if the data set is known to contain outliers it is a very precarious choice.

The fact that the model is unable to capture the distributional properties of returns is also evident in kurtosis. Its formula can be derived directly by using the second and fourth moments of (4.1), which yields

\[
Kurt(y_t) = \frac{E(y_t^4)}{E(y_t^2)^2} = 3exp(\frac{\sigma_n^2}{1 - \phi^2})
\]

Using the authors’ parameters for Dow Jones, estimated kurtosis is found to be 4.3884 compared to 8.276 calculated by the sample. The estimates of the corrected sample give kurtosis of 5.78506 compared to 7.99061 in the sample. This is expected given the different \(\sigma_n\). Kurtosis for the 2005 - 2015 sample is, according to estimated parameters, 8.02188. As expected, in the higher volatility period the difference between kurtosis as calculated by the data (14.1445) and kurtosis implied by the model is even higher. A stochastic volatility model is ill suited to replicate this phenomenon, especially in more unstable periods. This highlights the need to include jumps in both the returns and the volatility process. Also, despite the improvement by clearing the sample, the sample kurtosis remains higher than the estimated kurtosis in all cases. This is another element in favour of the replication parameters, since the resulting kurtosis is closer to that of the sample. Hautsch and Ou (2008) provide a Jarque-Bera test and a QQ plot which show the inability of the model to account for excess kurtosis. This applies to the replication parameters as well, since they are very similar. Finally, the explanation for the high kurtosis in the 2005 - 2015 period is the much fatter tails and the high number of extreme returns in the set. While the SV model is able to account for that increase to some extent, it falls very short on adequately dealing with kurtosis. The value in Table B.2 is 14.1445 while in Table B.1 it is 8.022, while the corresponding values for \(\sigma\) are 0.175 and 0.127 respectively. This shows that higher variance for the error term is not sufficient to capture such effects. Clearly, jumps need to be explicitly introduced in an SV setup to adequately capture the nature and effect of jumps and improve model fit.
CHAPTER 4. ESTIMATION METHOD AND PRELIMINARY RESULTS

After cleaning the 1991 - 2007 data, the results were practically left unaltered. Daily log-returns are typically numbers very close to zero so the change had a noticeable but marginal effect on results. When comparing the two different periods, the increase in the standard deviation of parameters is clear. This demonstrates the difficulty of the model to capture both outliers and volatility dynamics, as seen in parameter parsimony. According to the paper, the normal innovations in the volatility process are not sufficient for capturing such behaviour. The model is completely unable to account for excess kurtosis and jumps. This highlights the importance of introducing jumps and that stochastic volatility on its own is not enough in a AR / GARCH framework. Even an elementary jump structure will improve model fit, as discussed earlier in the literature and shown here.

Nevertheless the main goal, which was replication and expansion, was achieved. The method was successfully replicated and analysed, the intuition and results were found to hold over both samples and the parameters of the paper were estimated with sufficient accuracy. This was an important step due to the complexity of MCMC inference. It is always good practice to start from a basic model with few parameters and build up, because in that way both efficiency and tractability are improved and possible problems can be corrected at an early stage. If, for example, a Metropolis-type step is not efficient of has a low rate of acceptance, it is unlikely that its performance will increase when more parameters are introduced and a better solution must be found. From a learning point of view, the code used in this chapter provides the basis for the code of the SVCJ model, with correlation between the volatility and returns processes, additional parameters and jumps in both volatility and returns. From a contribution point of view, the inability of the SV framework to accommodate tail risk sufficiently was established. Excess kurtosis was detected in both a "calm" and a "tumultuous" period, and in addition outliers were completely ignored. This poses enough justification for an expansion towards including jumps. Admittedly, the derivation included a lot of detail. The intention was to avoid similar explanations during the much more complicated conditioning of the SVCJ model in the next chapter.

70
Chapter 5

Model estimation and optimal portfolio weights

5.1 The model

5.1.1 Basic set-up

In the previous two chapters it was established that neither jumps nor heteroskedasticity are sufficient on their own to capture extreme event risk. An investor faced with uncertainty about tail risk does not see a difference in terms of average terminal utility and is therefore not concerned. Stochastic volatility is not able to fully capture the stylised facts of asset price movements or the distribution of returns and is consistently outperformed by even the simplest models with jumps. A combination of the two, however, is promising and has produced significant literature in both portfolio selection and model fitting. This section will introduce and discuss the main model of the thesis, its parameter estimation via MCMC and their comparison to existing literature, and the derivation of the optimal portfolio solution in detail. A comparison between previous estimation results and other types of solutions is a key part of this section, as the first large extension of the thesis relate on the existence of a closed-form solution for this class of models. In addition, the discussion of a corpus of literature will take place on a strict base of usefulness. before it is formally presented and discussed. The simplest form of a jump model is a Brownian motion for returns, with or without drift, with constant standard deviation plus a Poisson jump with constant parameter. From chapters 3 and 4, jumps and volatility clustering need to be introduced. This immediately introduces a stochastic process for volatility, which should be mean-reverting. Additionally, correlation between returns and volatility can be introduced in order to allow for leverage effects. This leads to the Heston square volatility model. The introduction of jumps gives different types of stochastic volatility models: jumps only in returns, jumps only in volatility,
independent jumps in both, simultaneous jumps in both (SVJ models). The performance and use of this general class of models has been discussed in chapter 2. The leverage effect and stochastic volatility are crucial, while jumps in both processes tackle discontinuities in variance and "spikes" observed, among others, by Bates (2000) and are a meaningful enhancement. In terms of performance, it improves as we move from the SV to the SVJ model with returns only jumps to an SVJ version with double jumps. The need is to select a model that combines flexibility, a good performance record and a level of complexity that allows multiple features (especially those discussed above) to be captured simultaneously. The model that fits the prerequisites is the Stochastic Volatility model with Correlated Jumps (SVCJ). It is arguably the best performing affine jump-diffusion model with a large number of research applications. Proof of that lies in the literature review, and some of the papers estimating the model are discussed later in this chapter.

5.1.2 The SVCJ model

The model of choice is the Eraker et al (2003) continuous time SVCJ model (EJP)

\[
\begin{align*}
    dY_t &= \mu dt + \sqrt{V_{t-}}dW_t^Y + \xi_t^Y dN_t \\
    dV_t &= (\kappa\theta - \kappa V_{t-}) dt + \sigma_V \sqrt{V_{t-}}dW_t^V + \xi_t^V dN_t
\end{align*}
\] (5.1)

\[
\begin{align*}
    dY_t &= \mu dt + \sqrt{V_{t-}}dW_t^Y + \xi_t^Y dN_t \\
    dV_t &= (\kappa\theta - \kappa V_{t-}) dt + \sigma_V \sqrt{V_{t-}}dW_t^V + \xi_t^V dN_t
\end{align*}
\] (5.2)

\[dY_t\] is log returns where \( Y_t \) is the log price, \( V_t \) is volatility and \( V_{t-} \) the left limit of \( V_t \) (the point in time closest to it). \( dW_t^Y, dW_t^V \) are Brownian motions with correlation \( \rho, dN_t \) is a Poisson process with constant arrival intensity \( \lambda \) that is common in the two processes, \( \mu \) is diffusive mean returns and is constant, \( \xi_t^Y, \xi_t^V \) are jump sizes of returns and volatility with correlation \( \rho_j \). \( \sigma_V \) is the ”volatility of volatility” parameter or the standard deviation of \( V_t \), \( \kappa \) is the speed of mean reversion for \( V_t \) and \( \theta \) is the diffusive long-run volatility mean. The shorthands \( \alpha = \kappa\theta, \beta = -\kappa \) are also used in some formulations. The returns jumps follow a normal distribution \( N(\mu_Y + \rho_j \xi_V, \sigma_Y^2) \) and the volatility jumps follow an exponential distribution \( \exp(\mu_V) \) which guarantees positivity of volatility.

5.1.3 The SV model without jumps

The no jumps case is the Heston square-root stochastic volatility model

\[
\begin{align*}
    dY_t &= \mu dt + \sqrt{V_{t-}}dW_t^Y \\
    dV_t &= \kappa(\theta - V_{t-}) dt + \sigma_V \sqrt{V_{t-}}dW_t^V
\end{align*}
\] (5.3)

\[
\begin{align*}
    dY_t &= \mu dt + \sqrt{V_{t-}}dW_t^Y \\
    dV_t &= \kappa(\theta - V_{t-}) dt + \sigma_V \sqrt{V_{t-}}dW_t^V
\end{align*}
\] (5.4)
where the diffusion term in the volatility process has been factored purely for illustration purposes. In order for the models to be estimated, an Euler discretisation needs to be applied. The discretised versions are stated in the Appendix. Euler discretisation is a standard mathematical process and will not be described in full here, since the final form is stated in every cited paper that performs MCMC estimation of an SVCJ model. For completeness, a sufficient description can be found in Johannes and Polson (2009). Trivially, the Brownian motions correspond to drawings from \( \epsilon_t^{Y,V} \sim N(0, 1) \) while the Poisson process is discretised to a Bernoulli where the occurrence of a jump \( J \sim Ber(\lambda) \) times its magnitude \( \xi_t^{Y,V} \).

5.1.4 Further discussion of features and advantages

The main advantages of the model are its ability to nest many different variations, its ability to capture leverage effects (the negative relationship between returns and volatility) by assuming correlation, and its popularity that allows for a more widespread comparison of performance and parameters. Any affine formulation can be nested in the SVCJ model by removing correlations, jumps or assuming separate \( \lambda_Y, \lambda_V \) for independent jumps. Stochastic equity premia or arrival intensities can also be introduced by modelling \( \mu \) and \( \lambda \) with a stochastic process. Additional equity premium terms can be introduced linearly in the mean of the returns process. It is important to stress that \( \mu \) is the average return of the diffusive part and does not take the effect of jumps into account. Briefly, if one assumes a risk-free rate \( r \), then \( \mu \) is the sum of \( r \) and an arbitrary risk premium, or diffusive equity premium EP. A "volatility premium" term \( \eta V_t \) appears in some variations but both Eraker et al and authors literature find it to be negligible for equity returns, and is very often relaxed. A "jumps premium" can be included as \( \mu_Y \lambda \) (or \( \mu_Y \lambda V_t \) in further specifications), and in the EJP variation the drift would be \( \mu - r - \mu_Y \lambda \), where \( \mu_Y \) is estimated to be negative. The topic will be discussed in further detail in the Appendix, but for now it is important to identify how the drift terms correspond to different quantities. The EJP model is well documented and arguably the best performing of its class. It performs as well as SVIJ (independent jumps) and on par with non-affine models as discussed in the literature review.

The first paper to present and estimate the model via MCMC was Eraker et al. (2003). Johannes and Polson (2009) provide a full theoretical and applied discussion of MCMC applications for a variety of models and is a standard reference. Other papers include, but are not limited to, Raggi (2004), Asgharian and Bengtsson (2006), Li et al. (2006), Brooks and Prokopczuk (2013), Witzany (2013). Their common trait is that
they both estimate the model as well as derive the posteriors, and they often provide expansions and comparisons across different frameworks. Witzany (2013) expands to a bivariate specification, Li et al. (2006) compare with an infinite-jump Levy process among others, Asgharian and Bengtsson (2006) look into contagion of jumps in international markets, Brooks and Prokopczuk (2013) discuss jumps in commodity markets. Each author has his own preferences in sampling and setting priors, which leads to different posterior distributions. The reasons for that are, apart from personal convenience, the computational efficiency of the algorithm (directly related to coding and the software at hand) and the avoidance of numerical errors and precision issues.

Such examples are joint normal posteriors for \((\xi_Y, \xi_V)\) or \((\kappa, \theta)\), sampling \((\alpha, \beta)\) and transforming back to the initial model. The most important alternative concerns the parameters \((\rho, \sigma^2)\). In this form, the model leads to Metropolis - Hastings sampling because no conjugate priors can be derived. If the volatility error term in (5.2) is rewritten as \(\epsilon_t^V = \rho \ast \epsilon_t^Y + \sqrt{1 - \rho^2} \ast \zeta_t\), where \(\zeta\) is a random normal variable independent of \(\epsilon_t^Y\), and \(\omega = \sigma^2_V (1 - \rho^2)\) and \(\phi = \sigma_V \ast \rho\) are defined, then conjugate priors exist (an inverse gamma and a normal, respectively) and direct Gibbs sampling of \(\omega\) and \(\phi\) is possible. Trivially, \(\rho = \frac{\phi}{\sigma_V}, \sigma^2_V = \omega + \phi^2\). This is a standard statistical transformation and is used in this content by Jacquier et al. (2004), among others.

None of the complete set of posteriors available elsewhere fully fits the thesis because some are simply unnecessary, some cause software issues and some can be simplified. All posterior distributions have been derived from first principles and the full process can be found in the Appendix of the chapter. Briefly, conjugate priors have been used for all variables and parameters except \(V_t\), which is unavoidably non-standard, all quantities are sampled individually (no joint sampling or multivariate distributions) and the transformations used are for \(\alpha, \beta, \omega, \phi\) as described above. For completeness, posteriors for joint sampling and Metropolis - Hastings steps are provided, explaining why they were not applied.

5.2 Estimation and discussion

5.2.1 The algorithm and the data

The algorithm is as follows

1) Set initial values for all parameters \(\tau = (\mu, \alpha, \beta, \omega, \rho_j, \phi, \mu_Y, \mu_V, \sigma^2_Y, \lambda)\) and vector
5.2. ESTIMATION AND DISCUSSION

$V_t$, number of iterations M and burn-in period of G.

For iteration m between 1 and M,

1) Sample $p(\tau_t^{(m)} | \tau_t^{(m-1)}, V_t^{(m-1)}, \xi_t^{V,(m-1)}, \xi_t^{V,(m-1)})$,
2) Sample $p(\tau_t^{(m)} | \tau_t^{(m-1)}, V_t^{(m-1)}, \xi_t^{V,(m-1)}, \xi_t^{V,(m-1)}, J_t)$,
3) Sample $p(\xi_t^{V,(m)} | V_t^{(m-1)}, V_t^{(m-1)}-1, ..., \tau_m) \sim Bernoulli$
4) Sample $p(V_t^{(m)} | \tau_t^{(m)}), \xi_t^{V,(m)}-1, ..., \tau_m) \sim N$
5) Sample $p(V_t^{(m)} | \tau_t^{(m)}), \xi_t^{V,(m)}-1, ..., \tau_m) \sim N$
6) Sample $p(V_t^{(m)} | \tau_t^{(m)}), \xi_t^{V,(m)}-1, ..., \tau_m) \sim N$

The data consists of 9132 daily percentage log-returns ($\left(\log S_t - \log S_{t-1}\right) \times 100$) of S&P from 2-1-1980 till 29-3-2016. Compared to other literature, the dataset has the advantage of including the financial crisis of 2007 - 2009 and the following relatively more tranquil period, which nevertheless features a number of shocks. Descriptive statistics and estimation results are provided in Tables C.3 and C.1. The results are presented together with parameters from Eraker et al. (2003), which do not include the crisis and is close to Li et al. (2006), and Brooks and Prokopczuk (2013) which do. In addition, a shorter estimation focusing on the crisis and its aftermath is conducted. The parameters are presented in daily percentages and annual decimals. The methodology for annualisation is described in the Appendix of Branger and Hansis (2015) and is mentioned here in brief. Broadie et al. (2007), who also perform a similar exercise, do not provide a full description. All datasets are S&P daily log-returns of different but comparable length. Convergence and the realisation of the chains is reported in Figure 2. The number of repetitions M is 90.000, the burn-in period G is 45.000 and convergence is completed after about 12.000 repetitions.

5.2.2 Estimation results

Complete results can be found in Table C.1 of the Appendix.

The first observation addresses the nature of the time series and how the inclusion of the crisis affects estimation accuracy. Most of the literature avoids taking the crisis into account, considering it too distorting. For the SVCJ model, the parameters for $\lambda, \kappa, \theta$ show that the model has the tendency to trade jump frequency for volatility. The volatility related parameters are higher when the crisis is included but jump frequency drops. On an annual basis 1.66 jumps are expected each year for the pre-crisis period but only 1 to 1.4 jumps when the crisis is included. $\theta$, on the other hand, increases from 0.0135 annually to roughly 0.02. For the crisis-only period, between 2007 and 2016, the
same pattern is identified with 1.3 jumps per year but a $\theta$ of 0.0343. This shows that the model does not necessarily need high jump frequency or size to match volatility - it is perfectly capable of doing so with increased volatility parameters. The feature is particularly fitting when many small- and medium-size jumps are present in a period, but does not bode well when the aim is the perfect fit of the data at hand. Since the latter is not the goal of the thesis and the parameters are comparable to the rest of the literature, the point will not be explored further and Li et al. (2006) will be cited, who discuss how jumps can be better captured by Levy processes.

The standard deviations are very low and similar across the papers, showing that the crisis does not significantly affect accuracy. The reason for that is the inclusion of a sufficiently long period before and after 2007-2009. Even if the pre-2007 period is considered too calm, the post-2009 period is volatile enough to ensure an adequate number of sizable jumps. This establishes that extreme price movements are not overly rare in the sample. Another remark is about the correlation of jump sizes, $\rho_J$. This parameter is notoriously difficult to capture due to the rarity of jumps, and previous estimates are very inaccurate as shown in the table. The result of the thesis is a value of zero with very low standard deviation. This allows for the assumption to be dropped (e.g. Broadie et al. (2007) do so on estimation difficulty grounds). Although the literature tends to follow this path, the thesis provides a tangible reason to do so other than estimation difficulty and potential overfitting. The leverage effect proves to be significant, with an $\rho$ value of $-0.67$, and as expected the jump sizes and returns jump variance demonstrate high standard deviations. This is understandable due to the rarity of jumps and their much different magnitude, leading to a relative lack of precision. Still, however, the range is satisfactory. The negative sign of the returns jump follows the literature and the Compared to the SV parameters, the SVCJ delivers a higher diffusive mean but lower volatility related estimates. This is in agreement with the intuition that jumps, especially small and medium sized, can be captured to a very limited extent by SV models due to an increase in estimated volatility. Where the model fails is for large movements and outliers which are captured by the SVCJ model.

5.2.3 Discretisation bias and validity - jumps in very long, low frequency data sets

A lingering question is whether the improvement by including jumps is factual rather than an arithmetic phenomenon. The main counter-argument comes from studies that
employ high-frequency data, where by construction the time series is very smooth and sudden returns and volatility spikes are more rare. From that point of view, what is perceived as jumps is just a result of daily discretisation. This interpretation lack rigour, however, because estimation with high-frequency data can not cover a time frame as wide as with daily data due to sheer computational constraints. On the other hand, with a low frequency and greater differences in returns, jumps might be irrelevant because their effect can be attributed to volatility.

An often disregarded caveat in continuous time research is discretisation bias. In order to estimate such a model an Euler discretisation is unavoidable, which gives rise to some degree of distortion. Eraker et al. (2003) are referenced, if not for any other reason, at least for stating that for daily data the discretisation bias in their model is negligible. However, beyond that point, almost no further discussion exists on how estimation, particularly Bayesian, is affected. A thorough discussion on data issues in the estimation of continuous time models can be found in Bergstrom (1988), while a first treatment of the issue in Bartlett (1946). Bartlett (1946) focuses on estimating continuous time parameters from discreet time data states. In essence, he highlights that the parameter bias does not shrink along with the observation (discretisation) period but will always be present when his method is followed, and in addition there is no way to get non-biased estimates on a second order SDE from the first two sample autocorrelation coefficients. The important feature is the always present discretisation bias, which Eraker et al. (2003) mention as negligible for daily data and which all subsequent literature takes as granted. Apart from the great difficulties of the task at hand, Bergstrom (1988) refers to the interplay between the restrictions of the discrete distribution of the data and the continuity of the model. For regression methods, the bias was found to be of the same order of smallness as the square of the unit observation period. A method to outweigh that effect was to take into account the a priori restrictions, whose effect covers the bias.

This discussion does not cover Bayesian inference but aims to highlight an often overlooked statistical issue. Under the assumption that the relationship between bias and frequency holds, low-frequency estimations are not advisable and, even then, they will not be free of bias. The thesis will rely on the Eraker et al. (2003) result but raise a point that should be explored further and could have important implications for SV and SVCJ models, especially under ultra-high frequency and big data estimation. The second point raised, that of the stability of the optimal portfolios derived by such models, is discussed in detail in Korn and Kraft (2004) and is a signal for caution. Another significant resource on the same topic is Tunaru (2017) that analyses a list of empirical
CHAPTER 5. MODEL ESTIMATION AND OPTIMAL PORTFOLIO WEIGHTS

and theoretical caveats in model risk and continuous time finance.

Expanding the EJP statement, however, implies that for weekly or monthly frequencies it is not. In addition, the question is posed whether using logarithmic instead of arithmetic returns will lead to discrepancies or, in other words, if the month-to-month return varies so much in percentage terms that log returns are erroneous. While the danger is identified, the choice is daily logarithmic returns in order to keep consistency with other results. From another point of view, the economic interpretation of the data, the market and the parameters is debatable. The US economy and the S&P index have changed dramatically over the period and coming up with uniform, average parameters poses validity issues. The results using monthly data must be read with caution and must be read only as a fitting exercise fueled by equal parts of curiosity and defense, as they may be technically unreliable and scientifically misleading. The point is to show that a $\lambda$ parameter is still relevant.

To counter this critique, a dataset of 1745 monthly returns for S&P500 was used in estimation, covering the full length of the index life time. Robert Shiller’s web page (http://www.econ.yale.edu/shiller/data.htm) provides the data, which cover the period between January 1872 and May 2017. The standard deviation of the parameters is significantly larger, which is to be expected given the nature of the data. The results are presented in Table C.3. The annualised arrival rate is 0.20 with a very low standard deviation and amounts to one jump every 5 years. This is in accord with empirical evidence and a plausible estimate. It should not, however, be directly compared to the annualised values of the other estimates since monthly estimations are vastly different in magnitude than daily ones. The jump size mean is -10% with a standard deviation of 2.2 and the standard deviations are also significantly smaller than the SV case. Therefore, even with low frequency data, jumps still have an important role to play both in terms of arrival rate and in terms of magnitude.

5.3 Optimal portfolio weights

5.3.1 Utility and stochastic processes

The next step after model estimation is calculating the portfolio weights selected by an investor. The two types of investors are one that takes jumps into account and optimises using the SVCJ model to replicate the behaviour of the risky asset and one that chooses to ignore them by using the SV model. Each investor allocates wealth to one risk-free asset
5.3. OPTIMAL PORTFOLIO WEIGHTS

with a constant annual return of 2% and one risky asset represented by the SVCJ model. The investment horizon extends up to 30 years, and for that reason parameters used in the optimisation process are annual decimals instead of daily percentage parameters. That also allows for easy comparison with literature that focuses on portfolio optimisation rather than model performance, since the parameters there are usually annual. Finally, the investor has a power utility function \( U(W) = \frac{W^{1-\gamma}}{1-\gamma} \).

The first thing needed is the wealth process. For wealth \( W_t \), return of the risky asset \( Y_t \) and asset weight \( \phi \) denoting the percentage of wealth allocated to the risky asset, in general

\[
W_t = W_{t-1}(1 - \phi)r + \phi W_{t-1}Y_t \tag{5.5}
\]

and stochastically, after substituting \( Y_t \) by its process,

\[
\frac{dW_t}{W_t} = (1 - \phi)rdt + \phi \mu dt + \phi \sqrt{V_t}dZ_t^Y + \phi E(\xi_t^Y) dN_t \tag{5.6}
\]

where \( dZ_t \) denotes the Brownian motion to differentiate from the wealth notation and \( dN_t \) is the Poisson process. The diffusive mean \( \mu \) is equal to the risk-free rate plus any risk premium captured by the model, or \( \mu = r + EP \). This allows for substitution in (5.6), leading to

\[
\frac{dW_t}{W_t} = (r - \phi r + \phi r + \phi EP)dt + \phi \sqrt{V_t}dZ_t^Y + \phi E(\xi_t^Y) dN_t
\]

\[
= (r + \phi EP)dt + \phi \sqrt{V_t}dZ_t^Y + \phi E(\xi_t^Y) dN_t
\]

or

\[
dW_t = (r + \phi EP)W_t dt + \phi W_t \sqrt{V_t}dZ_t^Y + \phi W_t E(\xi_t^Y) dN_t \tag{5.7}
\]

and volatility still following (5.2)

\[
dV_t = \kappa(\theta - V_t)dt + \sigma_V \sqrt{V_t}dW_t^V
\]

5.3.2 The Bellman equation and optimal weights

Equations (5.7) and (5.2) create a two-dimensional state-space model, as before, only this time the goal is to solve for optimal weights. This requires the construction of a Hamilton-Jacobi-Bellman equation. Although it is a well-known result in the literature, the derivation and full solution will be presented in order to highlight a difference with
existing formulas which will lead to the first technical contribution of the thesis.

An indirect utility function $F$ of unknown form is assumed. The goal is to solve

$$0 = F_t + \max_{\phi}[L(F)]$$ (5.8)

where $L(F)$ comes from applying an n-dimensional Ito’s lemma on the correlated stochastic processes of wealth and volatility (see this chapter’s Appendix) and is

$$L(F) = (r + \phi E)WF_W + \frac{1}{2}\phi^2 W^2VF_{WW} + \kappa(\theta - V)F_V + \frac{1}{2}\sigma_V^2 VF_{VV} + \sigma_V\phi W V F_{WV} + \lambda E[F(W(1 + \phi E(\xi^Y)), V + E(\xi^V), t) - F]$$

The next step is to assume a functional form for $F$. The solution method is to conjecture and then verify that $F(W_t, V_t, t)$ is of a certain form, namely

$$F(W_t, V_t, t) = \frac{W^{1-\gamma}}{1-\gamma} \exp(A(t) + B(t)V)$$

where $A(t), B(t)$ depend only upon $t$ but not $W$ and $V$.

The optimal weight $\phi$ is the solution to the following system.

$$\phi = \frac{EP}{\gamma V} + \frac{\rho \sigma V B(t)}{\gamma} + \frac{\lambda E(\xi^V)(1 - \phi E(\xi^Y))^{-\gamma}}{\gamma V} \exp(B(t)E(\xi^V))$$ (5.9)

where $B(t)$ solves the differential equation

$$B'(t) - \frac{1}{2}\gamma \phi^2 (1-\gamma) + \frac{1}{2}\sigma_V^2 B^2(t) + (\sigma_V \phi (1-\gamma) - \kappa)B(t) = 0$$ (5.10)

with initial conditions

$$A(T) = 0, B(T) = 0$$

The full solution can be found in this chapter’s Appendix. The solution for the SV model follows the same pattern and is

$$\phi = \frac{EP}{\gamma V} + \frac{\rho \sigma V B(t)}{\gamma}$$ (5.11)

under the same differential equation (5.10) and conditions as above.
5.3.3 The weight result relative to the literature

It is now necessary to discuss how this solution is different from other known results and why it was necessary to derive it from the beginning. The same methodology is described in Liu et al. (2003) for a variation of the SVCJ model. They parametrise by adapting the Pan (2002) values for stocks and options. In fact the model is widely used in the options related literature, while papers using equity data rely on the EJP model.

\[
\frac{dS_t}{S_t} = (r + \eta V_t - \mu \lambda V_t) dt + \sqrt{V_t} dW_t^Y + \Xi_t^Y dN_t \tag{5.12}
\]

\[
dV_t = (\alpha' - \beta' V_t - \kappa' \lambda V_t) dt + \sigma_V \sqrt{V_t} dW_t^V + \Xi_t^V dN_t \tag{5.13}
\]

where \(S_t\) is the price, not the return or the log price as previously, and \(V_t\) is volatility, \(r\) is the constant risk-free rate, \(\eta V_t\) is the volatility premium which is insignificant for stock prices (Eraker et al. (2003)), \(\Xi^{S,V}_t\) are jump sizes with means \(\mu\) and \(\kappa\) respectively, stochastic arrival intensity is \(\lambda V_t\) of Poisson process \(N\), \(\mu \lambda V_t\) is returns jump premium and \(\kappa' \lambda V_t\) is the volatility jump premium.

The most important difference is the use of a stochastic arrival intensity that is linearly related to volatility. The process for the Poisson parameter is a Cox process of the general form \(\lambda = \lambda_0 + \lambda_1 V_t\). Setting \(\lambda_0 = 0\) leads to the specification in (5.12) and (5.13) while setting \(\lambda_1 = 0\) leads to the EJP specification of constant rate in (5.1) and (5.2). A less important difference is the inclusion of the volatility premium, which has some importance for options, and the correction in the mean for the jump premium \(\mu \lambda V_t\). The drift of the diffusion thus includes one additional term that compensates for jumps. Since these changes are introduced linearly, the solution is not greatly affected.

The solution for \(\phi\) is

\[
\phi = \frac{\eta - \mu \lambda}{\gamma} + \frac{\rho \sigma_V B(t)}{\gamma} + \frac{\lambda E(\xi^Y)(1 - \phi E(\xi^Y))}{\gamma} - \frac{\gamma}{e^{\gamma \exp\left(B(t)E(\xi^V)\right)}} \tag{5.14}
\]

where \(B(t)\) solves the differential equation

\[
B'(t) - \frac{1}{2} \gamma \phi^2(1 - \gamma) + \frac{1}{2} \sigma_V^2 B^2(t) + (\sigma_V \phi (1 - \gamma) - \kappa \lambda - \beta)B(t)
\]

\[+(\eta - \mu \lambda)(1 - \gamma)\phi + \lambda E(\xi^Y)(1 - \phi E(\xi^Y))^{-\gamma} e^{\gamma \exp\left(B(t)E(\xi^V)\right)} = 0 \tag{5.15}\]

with initial conditions

\[A(T) = 0, B(T) = 0\]

The reason why stochastic arrival intensity and a Cox process were selected is because they allow the time-varying \(V\) terms to be eliminated from the denominators. An
additional term $\lambda V_t$ appears in the numerators of each fraction which leads to (5.14) and (5.15). The ODE for $B(t)$ is not a Riccati equation and can only be solved numerically. Setting $\lambda_1 = 0$ on the other hand (or equivalently defining the arrival rate as $\lambda$) turns the ODE for $B(t)$ into a Riccati equation which can be solved in closed form Branger and Hansis (2015). However, in the EJP model $V$ can not be eliminated: the absence of the jump premium in the drift excludes $\lambda V_t$ from appearing and the last jumps related term is multiplied by $\lambda$ only.

The Liu et al model comes with a set of drawbacks. The selection of a Cox process does, indeed, produce a numerical solution for portfolio weights and provides a time variable jump arrival intensity. Nevertheless, estimations of that model in the literature prove to be extremely rare. Apart from the GMM estimation of Pan (2002), the only other paper to the knowledge of the thesis that provides parameters for that exact model is Eraker (2004) using MCMC estimation largely taken from Eraker et al. (2003). In all the following literature (Branger, Brag, Schneider Hansis etc) the Pan parameters are the only one used (based on a dataset ending in 1999), with the Eraker parameters not only having been ignored but also having never been updated. During the PhD an attempt was made to adapt the code to fit to its stochastic arrival intensity counterpart and estimate it, but after 100 – 200 repetitions the software would crash due to memory overload issues. On the contrary, the EJP model has been estimated multiple times in different markets and over different periods, thus providing a verified and tested history of application and results. In addition, the parameters in Eraker et al. (2003) and Broadie et al. (2007) have been calibrated in numerous cases (e.g Branger and Hansis (2012, 2015)) to fit the Liu et al. model when needed. At the very least, this demonstrates a difficulty in getting parameter estimates via MCMC for the stochastic arrival intensity variation of SVCJ and, given the need to take the financial crisis into account, existing estimates are inadequate.

A second consideration is the Cox process itself. Even in the early jumps literature there was evidence for misspecification under stochastic arrival intensity with only marginal improvements in accuracy at the cost of complexity (Bates (2000)). The main advancements took place with a constant $\lambda$ and where time variability was introduced the model tended to be simpler (e.g constant volatility, as in Wachter (2013). These remarks highlight analytical and estimation based issues. Also, the (seemingly positive) relationship between jump frequency and volatility may be intuitive but not necessarily linear. It is included in the Liu et al model because it leads to a tractable solution for portfolio weights combined with an ODE for a numerical solution, which translates into
reverse-engineering convenience. No further argument is offered. Other processes that can provide similar, or more intuitive, structures are infinite-jump Levy processes, who are better at capturing small jumps Li et al. (2006) or Hawkes processes (Fičura and Witzany (2015)), who introduce a self-exciting (-feeding) component. The latter is also related to persistent jumps whose effects dissipate over time. Cox processes are still instantaneous, are able to attribute a higher jump probability when volatility is higher but pose estimation difficulties and, most importantly, do not perform better than models with constant arrival intensity.

5.3.4 Characteristics, properties and values of the EJP solution

This section discusses the key features of the portfolio weight. The weights are time invariant, are derived by annualised log parameters and to reach a closed form solution a substitution is required.

Two important observations about the optimal solution defined by (5.9) and (5.10) are, first, that the optimal weight is expressed in log terms and, second, that the V term in the denominator is stochastic. On the first remark, the "log" parameters are estimated by log returns leading to "log" weights. This is the reason why there are no Ito terms in the expressions that contain the drift parameters. The estimation issues discussed in the previous section eliminate the option of using arithmetic returns, and therefore the only alternatives are either to apply Ito’s lemma and turn the "log" parameters to their arithmetic counterparts or use the parameters and weights as they are, calculate portfolio total log returns at a later stage and apply an exponential part to remove the logarithm. Here, the second method is used due to its simplicity and intuition, especially since the actual equity premia are of no interest. For completeness, the Ito terms increase both SVCJ and SV weights while keeping their difference the same. Since the results are shown to depend on the difference of the weights rather their absolute values, using log or arithmetic processes is of no particular consequence.

In order to circumvent the second issue and derive a proper solution, the thesis follows the approach in Branger and Hansis (2012), where the EJP and Broadie et al. (2007) parameters are adapted from the EJP to the Liu et al formulation. This is achieved first by annualisation, and second by setting \( V_t \) equal to its long-run average. That can be \( V = \theta + \mu V \lambda / \kappa \), for the annualised parameters. This allows the elimination of the denominator terms and leads to a tractable closed-form solution under constant arrival intensity.
INTENSITY. It must be emphasised that Branger and Hansis (2012) use that substitution to transform the estimated value of $\lambda$ in (5.1 – 5.2) into a value that would correspond to (5.11 – 5.12). The idea, and novelty, to use $\bar{V}$ instead of $V_t$ in the optimal weights solution was conceived independently and, in the process, that application was discovered which now acts as a more solid methodological foundation. A set of simulations showed that the long-run value is a reasonable approximation for $V$. It also agrees with similar long-run volatilities used and estimated in the cited literature. Finally, the same literature does not mention any distortion from the approximation when portfolio weights or equity premia are discussed. The tests of the thesis were not exhaustive but suggest that using long-run volatility is realistic.

As an example, for $\gamma = 5$ and the SVCJ model, $EP = 0.0858 - 0.02 = 0.0658$, $\bar{V} = 0.027$ and $EP\bar{V} = 2.52$. All calculations are done with annualised parameters and decimals are rounded, so slight differences are to be expected. The calculation of the diffusive equity premium follows Liu et al. (2003), where the parameter values are substituted directly into the differential equations for the solution (in essence $EP/V$ corresponds to $\eta - \mu\lambda$, following the Liu notation). As discussed above, the Ito term $0,5V_t$ in the equity premium does not appear (Branger and Hansis (2012) and Broadie et al. (2007) for a more detailed discussion on an arithmetic returns - prices - specified process with options data, where parameters estimated under log prices (returns) are transformed to their arithmetic returns (flat prices) counterparts, hence an Ito term appears). Here, $EP$ contains a number of separate terms in other specifications and expresses a generic premium in log returns terms. There are also marginal differences if $EP$ is first calculated on the daily parameters and then annualised.

Now that the first two characteristics have been discussed, the straightforward formula for $\phi$ calls for substituting (5.14) in (5.15), which has a closed-form solution. Unfortunately, the exponential bit remains impossible to circumvent. The resulting expression is very complicated and does not provide additional intuition, but is the closest to a closed-form solution this model can have. The only obstacle is a product of a real and an exponential part that cannot be solved algebraically. However, this is only a minor failure and the solution presented here is, for all intents and purposes, a closed form solution.

A very important feature is the fact that the solution is time invariant. The investment horizon $T$ does not affect the optimal weight at any level, as shown by the plots. It does not appear in the formulas and it is only used in the terminal conditions $B(T)=0$, $A(T)=0$. Moreover, with horizons spanning from 5 to 9000 the weights were exactly
5.3. OPTIMAL PORTFOLIO WEIGHTS

the same until a direct vertical drop to 0 when t reaches T in the solution plot. The results are in line with intuition and demonstrate a clearly more conservative stance of the investor operating under the SVCJ model (the ”jumps investor”) compared to the investor operating under the SV model (the ”no jumps investor”). This is the first time in the thesis that a clear difference between thin (normal) and fat tails emerges, which validates the needs for a sufficiently complex setup in order to capture the effect for investors. The stylised facts at play are heteroskedasticity, volatility clustering, rapid mean reversion of volatility, fat tails and leverage effect. For risk aversion $\gamma = (1, 2, 3, 4, 5)$ the gap keeps expanding even when $\phi$ is allowed to be above 1. The strategy is obviously buy-and-hold, since the numerical value of $\phi$ does not change over time and portfolio rebalancing is not modelled, allowed, considered or predicted. The pronounced differences between the weights are the generator of the results in the next chapter.

Table C.4 and Figure C.1 in the Appendix present the numerical values for the thesis and the weights implied by the Eraker et al. (2003) and Liu et al. (2003) parameters, as well as a plot of the differential equation solution. Additional replications are presented in Table C.5 of the Appendix. These include combined results of Table 1 in Branger and Hansis (2012) and Table 1 in Branger and Hansis (2015) focusing on the different parameters among the EJP and the LLP version (the common parameters are in Table C.1). In addition, numerical values for average volatility, implied risk-free rates and equity premia and a numerical mistake are reported. Average volatility is reported in Branger and Hansis (2012) to be 0.0154, while the value used to calculate $\lambda$ in Branger and Hansis (2015) is correct and equal to 0.023. The latter value is easily calculated by using the annualised parameters provided in the same tables. Another slight discrepancy is the risk free rate (4 – 4.5%), which can only be inferred, and the equity risk premium (7.32% - 6%). These differences partially explain the erroneous value of average volatility. Another reason is the inclusion of additional terms of the true and risk-neutral measures in the formula for average volatility. However, the information given is inadequate and does not explain the discrepancy between the two papers that use identical models and parameter values.

The need to switch from daily to annual parameters is also practical. Even when summary statistics are discussed, the literature tends to perform some sort of annualisation. It is therefore needed to allow for easy comparison across papers and make the reading of results easier. Since the investment horizons will be rather large, annual parameters are more informative in that context. A reporting case is for $\theta$, which is turned into $\sqrt{252} \times \theta$ in Eraker et al. (2003) or Li et al. (2006) etc and then compared
to that of the sample. Broadie et al. (2007) note a brief but incomplete methodology as "scaling some of the parameters by multiplying $\kappa$ and $\lambda$ by 252, $\sqrt{252} \cdot \theta / 100$ gives the mean volatility and $\sqrt{252} \cdot \mu_V / 100$ gives the mean jump in volatility". Branger and Hansis (2015) provide the clearest treatment of the subject in the Appendix, where they turn the daily percentage (log) elements first into decimals and then into annual values. They work directly on the process, dividing $d\log S_t$ and $dV_t$ by 100 and $100^2$ and then multiplying by 252 and $\sqrt{252}$ respectively. The resulting transformations are $\frac{252}{100}\mu$, $252\kappa$, $\frac{252}{100}\theta$, $\frac{252}{100}\sigma_V$, $\frac{1}{100}\mu_Y$, $\frac{252}{100^2}\mu_V$, $\frac{1}{100}\sigma_Y$, $252\lambda$, $\frac{100}{252}\rho_J$, $\rho$ is not altered. The parameters are turned from daily percentages to annual decimals.

Solutions of the SVCJ model across the same lines can be found in the literature, but the solution proposed here is the first one that does not resolve to a numerical approximation. To be more precise, the solution is not completely closed only in the respect that it contains both real and exponential components in a manner that can not be fully resolved. Still, it is the closest one can get. As a further example in addition to the aforementioned papers, Ruan et al. (2013) provide a solution for equity premia, not portfolio weights, for the SVCJ model with constant arrival intensity and discuss clearly when there is a closed-form solution of this system, when a numerical approximation is needed, how the Bellmann equation is derived and/or Taylor approximated. In the case of the thesis, the numerical value for the portfolio weight comes from a Solve operation in Mathematica.

### 5.4 Summary of results

It is useful to provide a final overview of the results in this chapter. The first result is the very precise estimate of $\rho_J$ in the MCMC estimation, which allows for the assumption of correlated jump sizes to be relaxed. It is estimated to be practically zero with a standard deviation of 0.003. The result holds for the 1872-2016 and the 2007-2016 periods as well. All other known estimates of that parameter in the literature, a brief list of which was given above, are burdened with very high standard deviation. The EJP estimate of -0.6 has a standard deviation of 1, while the estimate of Brooks and Prokopczyk is also close to zero (0.05) but with a much higher standard deviation (0.20) It is the first time such a precise estimate is given, which allows for the empirical practice of dropping the parameter altogether to be justified. The other parameters are comparable to the literature. Leverage effects, according to $\rho$ (-0.6765), are very strong compared to EJP (-0.4838) but comparable to -0.5811 of Brooks and Prokopczyk. Jumps in returns , at
almost $-3\%$, are lower than the 1985 - 2010 estimate of $-4.4\%$ but almost double the non-crisis estimate of EJP. It must be kept in mind that this average jump size is additive to the diffusion part, and it is sufficient enough to generate negative daily movements up to $-12\%$ or $-15\%$ during the simulations presented in the next chapter. As a numerical example, if a jump of $-3\%$ takes place and the diffusive part equals a modest $-2\%$ then the daily return is $-5\%$. If the diffusive part equals $2\%$ only a small negative ”jump” is recorded.

The second major contribution is the solution for optimal portfolio weights, which is almost in closed form. Until now the literature had to choose between a Cox process for jumps, tied to outdated or adapted parameters, and a constant arrival intensity burdened by approximate solutions but with a tractable record of parameter estimates. For all intents and purposes, the suggested solution of substituting $V_t$ with its long-run mean was adopted before a similar approach was encountered in the literature. Branger and Hansis (2012) use the idea in a different context and are interested only in adapting the EJP parameters, not in portfolio weights. The context of the PhD is different and expands the methodology one step before a complete solution (which is impossible, as stated, because of the form of the formula for $\phi$). To the best of the PhD’s knowledge, this is the first time this method has been applied and therefore it may represent an important contribution.

A third result is the clear difference between an investor that optimises his allocation while taking jumps into account and an investor that ignores them. This is an evolution from the previous chapters but hardly a surprising finding, apart from the magnitude. The differences between the two cases are significant, persistent and divergent across a range of reasonable degrees of risk aversion. It is also beyond any reasonable statistical or numerical mistake and transformation, in the sense that a modest difference in equity premia, volatility or parameter estimates will not be enough to reverse the observed difference. The effect of the financial crisis and its inclusion in the sample can also be detected when the SVCJ weights are compared to those derived from the EJP parameter. The weights difference between the SV and SVCJ cases is more pronounced. A final remark is the behaviour of the model when periods of high volatility and/or frequent jumps are introduced. The model tends to trade volatility with jump frequency, as shown in the comparison across papers. In the pre-crisis period $\lambda$ is higher than in the (post-)crisis period while volatility parameters are lower. Jump sizes, on the other hand, are significantly increased. This means that the model does not require a high arrival intensity to generate small and medium sized jumps because volatility is sufficiently high.
for that. More vicious movements, however, are represented by both higher volatility and higher jump means. The final point is the list of tables and results replicated (and mistakes noted), which demonstrates correct application and technical aptitude.

5.5 Conclusion

This chapter has derived the parameters of the main model, compared the estimated parameters with the literature and calculated a first set of estimates. The inclusion of the financial crisis period was shown to push the volatility related parameters upwards and jump frequency downwards, because the model can generate sufficient large movements with fewer jumps on average. Jumps were also found to be greater. The most important result was the zero estimate on jump size correlation, which enables the relaxation of the assumption that volatility and returns jump sizes are related. This is the most accurate estimation in the literature and provides justification for a practice used by other papers, due to the extreme difficulty in estimating that parameter.

The most important contribution of the chapter is the derivation of a closed-form solution for a class of models that, up to now, had only numerical solutions. A discussion behind the mechanics of SVCJ variations showed that, for reverse-engineering purposes, a stochastic jump frequency would lead to a convenient formula for portfolio weights. For that reason, the Liu et al. (2004) specification is popular in jumps portfolio literature despite the outdated parameters commonly used. The thesis provided an alternative for the Eraker et al. (2003) formulation, leading to a closed-form Riccati equation and optimal weights solution. This enabled the calculation of optimal, time-invariant portfolio weights for the SVCJ and SV models that result in a buy-and-hold strategy and show a clear preference for the SV investor for riskier positions. The higher the degree of risk aversion the greater the distance between the two allocations and the amount of wealth invested in the risky asset.

Further discussion dealt with the interpretation and importance of jumps in low-frequency series. Parameter behaviour was comparable to that of daily frequency but with somewhat lower accuracy, while jumps were found to be present at a 20% probability on an annual basis. This shows that jumps can be a powerful interpretative tool regardless of frequency, but great care must be taken when discretization bias is taken into account.

This paves the way for the next chapter, where investor and manager behaviour will be discussed. The approach is similar to that of Chapter 3, where a set of simulations was conducted based on a set of optimal weights. This chapter estimated the model
that will generate the data, derived the optimal weighs and discussed their properties, and established the clear differentiation and robustness of the results. Chapter 6 can now proceed to the main contribution of the thesis. Stochastic models with jumps have already been discussed in the literature, but it may be useful to justify the selected variation
Chapter 6

Investor and managerial incentives to ignore jumps

While the previous chapter laid the technical foundations of the thesis, the present chapter presents the main findings. Chapter 5 provided the parameter estimates for the investor and the optimal portfolio allocation he would select. It provided the mechanics upon which this chapter will study investor and managerial incentives, the motives for both actors and their relative performance. The derivation of a closed-form solution for portfolio weights allows for an easy comparison between each case. Of particular importance is the effect of the investment horizon on incentives, an element not discussed in the literature.

The emphasis is now on investor and managerial performance and incentives with and without jumps, and to identify if there is a motive to do so in the long run. The new element is the mutual or hedge fund manager. Contrary to the investor, who receives utility from investment wealth, the fund manager receives utility from administration and performance fees which rely on portfolio returns. An essential assumption is that the data generating process contains jumps and that fact is known to both the manager and the investor - they are free to select whether to take the jumps into consideration by using the SVCJ weights in their portfolio allocation, or ignoring them and deliberately use the SV weights. The thesis explores the case of potential moral hazard, where the manager is motivated to undertake excess risk by ignoring jumps in order to attain higher fees. This is in contrast to the stated goals of the investor, and the study of such attitude is one of the main contributions in both portfolio optimisation and the hedge/ mutual fund literature. As shown in the literature review, there is no explicit discussion on how jumps affect investor and managerial incentives or how (if) they affect moral hazard issues between the two. The final part of the thesis will isolate jumps as a factor and make a contribution by answering that exact question. The type of managers to be discussed are
mutual and hedge fund managers. Despite their differences, the investment activities they partake are comparable and have a strong presence in bond and equity markets. Also, despite the multiple trading strategies undertaken by hedge funds, it is still intuitive to use a simple fund as example where an amount is invested in bonds and an amount in the market portfolio under an index tracking or market neutral strategy. Another argument is the fact that one cannot be more diversified than the market portfolio. Studying the interplay between momentum or uncovered options trading and jumps risk is beyond the scope of the thesis, although it is a relevant subject. What is of interest is the reward structure of a hedge fund manager, and how an option-type compensation creates a set of incentives directly related and affected by extreme event risk.

Managerial incentives are often absent while in the latter they do not include rare events and jumps in asset prices. Two exceptions are liquidity risk (which, as Liu et al. (2003) discuss, is equivalent to jump risk because both do not allow for an instant recalibration of portfolios) and downside risk, which is discussed in a very different context. The new background element in this section is the hedge and mutual fund literature. The starting point of the thesis was portfolio optimisation, and in order to keep on that trajectory the discussion of the literature will focus specifically on managerial incentives and some technical considerations. In addition, the sole investor case is examined on incentive grounds. The process is a simulation based on the SVCJ parameters and optimal allocations of the previous chapter and the technical aspects are the definition of the measure of win and the payoff structure of the fund manager. All these issues will be discussed in detail.

6.1 Mutual and hedge funds literature review

6.1.1 Academic literature

The first thing to note is that there is no particular need to apply risk disregard at a managerial level. Traders in a firm or in their separate divisions are perfectly capable of supporting the argument. Unfortunately, the literature around trader compensation can be summarised as follows: there is no literature. There is a vast and multifaceted literature on manager compensation and incentives in other industries (services, manufacturing etc), particularly CEO benefits, but with a notable difference. These schemes must be publicly announced, which allows for close monitoring and large numbers of data available. On the contrary, trader contracts are private agreements with varying structures and are
very rarely discussed, announced or disclosed. In the hedge fund industry in particular, the only information available to the public is the fee structure of the fund. Unavoidably, the literature focuses on a fund manager, not at trader, level.

Two thorough resources are the hedge literature reviews by El Kalak et al. (2016a) and El Kalak et al. (2016b), which focus specifically on managerial characteristics and risk management styles. Because the papers provide a very clear and thorough classification and listing of results, only a summary will be provided with focus on managerial incentives. A clear break between optimistic early literature and pessimistic, cautious late research can be found, where doubt is cast on the claim that higher performance is linked to higher incentive fees. When a more complex set of incentives (deltas, watermarks etc) is considered, a positive relationship is established. Reputation costs and self investment in the fund can limit a manager’s risk appetite. Managerial discretion (high lock-ups, fewer restrictions) is also found to be positively related to returns, while experience is negatively correlated as is leads to more conservative investments. A key contribution comes from Schwarz (2007) and Liang and Schwarz (2011), where it is noted that "the higher the investor outflow restrictions the lower is the closure likelihood and greater performance loss over time. High pay-performance deltas are not strong enough to prevent overinvestment. Management and incentive fees are correlated with lock-up periods reducing the cross-sectional fee variation. Negative relationship between funds of hedge fund performance and incentive fees. Large funds charge higher fees and are more likely to raise fees level. Investors do not view fee levels as a signal of future fund performance."

Giannetti and Metzger (2015) highlight that exact difficulty by stating that "this is the first paper to explore the structure of compensation for non-executive employees in the financial sector”. Contract information at a trader/ employee level is simply not available. Similar information on at a managerial level is available at an empirical, industry level. Given the circumstances, the literature review will make a brief reference on tangent issues discussed in contemporary corporate-related papers and will then proceed into a targeted discussion of hedge and mutual fund research. The two pillars are managerial incentives and risk assessment, and it will be shown that there is ample space for both a technical and a conceptual contribution. For CEO compensation, Frydman and Jenter (2010) provide a comprehensive review.

Nikolov and Whited (2014) is a typical example of a study focusing on a manager that is compensated by an equity share and a cash performance fee, a structure that is not applicable to hedge fund managers. The focus on conflict between agents (the
investor and the CEO, or between management and ownership) is well documented in the corporate literature where different payment fees are examined. These include flat bonuses, share ownership, options, performance measures etc. Another fitting string of literature is managerial compensation for risk taking. Leisen (2015) uses a continuous time model (without jumps) to study a risk-averse manager who chooses the level of risk (volatility) of a company share. The amount of risk undertaken is capped and the firm has outstanding debt so the probability of default exists. The manager compensation structure is a fixed amount plus a claim to a bonus payment which is based on terminal asset values. The option type compensation is similar to that of a hedge fund but relies on terminal wealth, not on returns. Another difference is the existence of debt - a corporate feature - and the cap on risk. An argument could be made on trader discipline, but empirical assessments might prove that moot. In addition, risk is not separated to welcome upside and unwelcome downside, where each case justifies increases or reductions. The dynamic setup is a bonus.

Danthine and Donaldson (2015) refer again to the difference in incentives between owners and management, a theme that will also be encountered in the PhD in the form of moral hazard. The corporate setup leaves little space for direct comparisons, but the discussion of the "pay-for-luck" puzzle is relevant: large parts of managerial compensation seem to relate to factors beyond his control, an effect that is stronger when returns are positive. This is a clear allusion to exogenous shocks, or jumps in asset prices, and managerial alphas. Standard incentive theory states that a manager should be compensated only for what is entirely under his control. The aim of the paper is not to find a dynamic general equilibrium, which is not relevant to this chapter, but the conclusion that the manager could be tied to the company in some way ensures that his interests are more aligned to those of the shareholders.

Aivaliotis and Palczewski (2014) is the paper closest to the idea pursued in the thesis, but with some very significant differences. A manager that applies mean-variance optimisation on a continuous time process with constant mean and volatility is introduced. The manager optimises the portfolio allocation according to the mean-variance criterion which depends upon a random amount of compensation. With continuous monitoring of performance, the manager’s compensation is proportional to the cumulative future value of the difference between the portfolio return and the benchmark return. when performance and compensation depend only on terminal wealth, the manager’s compensation is proportional to the difference between the portfolio return and the benchmark return. The main contribution of the paper is a method different than the discretisation and
numerical solution of a Bellman equation, which leads to the conclusion that terminal-based fees lead to more conservative behaviour and better performance compared to constant monitoring. It is one of the very few pieces of research that discuss the behaviour of a fund manager or provide a mathematical background. However, the context is drastically different as there are no empirical links particularly for compensation schemes, jumps are not considered and there is only one agent, the manager, with only one criterion.

Bali et al. (2012) identify the importance of different constituents of systemic risk in hedge fund returns and explicitly refer to tail risk. The methodology is, however, regressions where the explanatory power of skewness and kurtosis is tested. Given the modelling advancements in other areas of finance incorporated here, the approach is severely limited and the conclusion that tail risk has small explanatory power does not hold. Nevertheless, evidence is found that idiosyncratic risk plays an important role in fund returns. In CAPM terms a hedge fund is not market neutral but is affected by the manager’s attitude and skill, expressed by alpha. Morton et al. (2006) discuss downside risk in funds-of-hedge-funds and establish non-normality in returns. Meligkotsidou and Vrontos (2008) test for structural breaks but their dataset ends before the financial crisis, so they rely on isolated cases and bubbles to replicate those breaks. The Bayesian method they employ is able to track previously undetected breaks and provide some insight in the differences across funds in terms of strategies and risk exposure.

Giannikis and Vrontos (2011) expand on non-linear exposures to various risk factors, finding that those exposures are asymmetric and that ignoring certain thresholds may lead to misinterpreting alpha. Fung and Hsieh (2011) report that ”empirical analysis finds persistent net exposures to the spread between small vs large cap stocks in addition to the overall market [for long/short hedge funds]. Together, these factors account for more than 80% of return variation. Additional factors are price momentum and market activity”. Downside risk and rare events are absent from the reasoning of the paper.

For managerial incentives, there is a break in the literature where very positive attitudes towards managerial skill in early ’00, driven by strong hedge fund performance, are substituted by more sceptical work particularly in the aftermath of the financial crisis. Early work includes Carpenter (2000), who discuss the implications of the option-type fees structure. The paper considers Decreasing Absolute Risk Aversion utility in particular where the manager is given a call option on assets under management. Its value is found to end up either deep in or deep out of the money. As the asset value goes to zero, volatility goes to infinity. Risk attitude is not as straightforward, as sometimes the
manager’s optimal choice of risk is lower than if we was individually trading. Bollen and Whaley (2009) bring manager authority to the forefront and how strategy is dictated. The focus is on exposures to risk which are perceived as constant when they are in fact time-varying, a disparity that may lead to incorrect measures and abnormal returns. When the related parameters change significantly, the constant alphas are a misleading indicator of performance.

Patton and Ramadorai (2013) make a modelling contribution based on conditional parameters, where returns follow a CAPM formulation but the beta is stochastic. Various time effects are discussed and high responsiveness to market signals and seasonality in the behaviour of parameters is found. Agarwal et al. (2009) are more specific in the link between incentives and discretion. For incentives they use a mixture of option-type compensation structures, ownership and performance thresholds (watermarks) and find that administrative and managerial liberties and provisions play an important role in performance. The nonlinearity of those incentives, who are typically expressed by alpha, creates a nexus of endogenous incentives that affect both leverage and performance measures. This acts as another verification of the need to look beyond returns, fees structures and standard risk exposures in order to have a better understanding of managerial decision making in hedge funds. When high watermarks are combined with incentives (measured by delta), performance is better. Lim et al. (2016) provide a useful introduction to a mechanism that is used in this chapter, that of capital inflows and outflows that affect both the amount of future fees and the attitude of managers. The importance of a wider array of factors is verified again, but the importance is the link between performance monitoring and immediate reward (punishment) via capital movement, the connection between the labour market and this monitoring and how this scrutiny alters the behaviour of the managers. There is a clear link to signaling and incentives literature but this goes somewhat beyond the scope of the thesis.

Yin (2016) refers the important area of conflict of interest between management and investors. While the same disparity has been well documented in the corporate finance literature, similar effects are found when the size of the hedge fund is taken into consideration. Since fees are a percentage of wealth, a manager has the incentive to increase assets under management even at the expense of performance. This creates a different kind of conflict that is not based on compensation schemes per se.
6.1.2 Current prospects according to the industry

Although there is agreement on the structure of hedge fund fees, the standard ”2+20” scheme has started to change in recent years.

Fortado (2016) reports considerations from fund managers that ”2+20” is no longer sustainable, in light of both lower returns and higher flexibility. Only the biggest funds are now able to maintain these levels, with others offering discounts on rates after a period or amount, more choices on fee schemes or straight reductions. Long-term funds are willing to reduce the flat administration fee, while funds that use algorithm trading provide reductions on both management and performance fees. On average, for established funds the structure is 1.65 + 18% while for new ones it is 1.5 + 17.5%. Apart from competition and low performance across the industry, another reason is a shift in more passive strategies, which challenges the willingness and need for a performance fee. The effect on managerial stance is apparent, with a recorded shift towards transparency and promotion of unique skill and investment strategy.

Further information on the subject has been published in reports from Barclays and Prequin. The Preqin (2017) Hedge Fund Management Outlook (March 2017) contains a sample of 276 hedge fund managers who state a recent improvement of their fees which they would like to see continued, although they also record investor demand for lower fees as both being very high and the second most important business challenge, after performance. Hedge funds are expected to attract less capital in 2017, which would normally increase pressure, yet the industry outlook remains positive. With respect to fee levels, the majority now charges administration (management) fees of 1.50 – 1.99% (41%) contrary to 2% (35%) and the trend is increasing. The industry mean is at 1.56%, where new funds charging 2% have reduced to 30%. A similar pattern is detected in the performance fee, but 20% remains popular as 73% of all funds charge it. The number of funds that apply lower commissions than the older staples has been increasing over the last 3 years. According to the report, the challenge is on how that level of performance fees is implemented and justified, rather than negotiating its height. Ways to compensate or reassure investors include hurdle rates, high watermarks and clawback terms. The first tool is used by 4 out of 5 managers while the third option by only 1 out of 10, with an additional 1 considering it. Some research on the terms and conditions of major hedge funds shows that they still charge 2+20 (e.g. Aberdeen), with one exception being Winton who reduced fees from 1+20 to 0.90+16 for its major funds. Based on market reports, smaller funds are unable to maintain those levels. For the needs of the PhD, this calls for alternative structures which are described in the
CHAPTER 6. INVESTOR AND MANAGERIAL INCENTIVES TO IGNORE JUMPS

remainder of this chapter.

The Barclays (2017) Global Hedge Fund Industry Outlook and Trends (February) holds similar conclusions. 2016 has been the first year of net outflows since the financial crisis, and the trend since 2011 of underperformance compared long-only indices still carries on, with a reduced gap. The premium over LIBOR has been 3.5% for Long-Term funds since 2010. Low industry performance lead to the majority of funds missing their annual targets, where high investor expected returns were combined with average realised volatility being lower than average target volatility. The best performing were those more diversified and with low management clutter (e.g. few acting managers). Smaller funds performed better than large ones and investors generally paid lower fees, with differences between funds and strategies. Equities have been the most pressured, where the report cites reduced demand and an "investor's market" as reason. The most resilient were Quant Equity, Fixed Income Relative Value, and Multi-Strategy hedge funds, due to capacity constraints, higher costs and market dynamics. Larger investors, and those with sufficiently large positions in a fund, are more able to get concessions in the form of reduced fees and preferential terms. Banks and family offices, on the contrary, are much less likely to provide discounts than hedge funds. This attitude is not limited to large funds, but increasingly to small funds as well. Reduced fees for longer locks and/or larger tickets are the most ordinary forms of discounts but hurdles and limits on pass-through expenses have become more common.

6.1.3 Summary

Important observations can be made in the aftermath of the review. From a technical point of view, continuous time specifications are rare in the hedge and mutual fund literature. The research aims in the area are served by different models and when there is a stochastic element it is not necessarily translated to a diffusion, let alone a jump diffusion. This leaves space for a contribution in that particular field just by employing the SVCJ model. It is noteworthy that no study was found where jump diffusion parameters (MCMC based or otherwise) were used or estimated in a hedge and mutual fund context. Given that multi-asset setups are present in the literature, there is an important gap to be covered on modelling grounds.

As a conclusion, managerial incentives and risk factors play a central role in the hedge and mutual fund literature. It was underlined that the option-type compensation creates an environment where the manager is prone to optimise his personal payoff at the expense of the investor. This in turn affects the performance and structure of the fund,
especially if further discretion of the manager are considered. The main interpretative variable is a manager’s alpha and how it affects returns and reputation. This is an important foundation for the research question examined in this chapter, since the direct link between risk taking and managers’ attitudes does not discuss jumps. The exposures to risk factors, linear or not and time-varying or not, ignores rare events and the extent to which they pose a serious consideration for the manager. Rare events are not modelled, used or regarded as an important factor in decision making. This is the gap that the PhD intends to cover, and in a multi-faceted way. Compared to the fields discussed in the previous chapters, hedge fund research uses different (and, when comparable, simpler) tools what do not consider leverage effects or heteroskedasticity, does not often employ stochastic models and looks into risk-taking from a practical, managerial perspective rather than linked to portfolio allocation and asset pricing literature. This is not a deficiency per se but leaves much ground to be covered on how certain factors are introduced.

Thus, the contribution of this chapter is, first, to bring the results and intuition of MCMC estimation in an SVCJ model to fund management, second, connect jump risk to manager payoff and third, establish whether there is an incentive to undertake excess risk and potentially cause moral hazard issues.

6.2 General setup

It is useful to establish a methodology before the simulations are introduced. The process will be two-fold, first focusing on the investor and his incentive to ignore jumps and afterwards on the manager. In the previous chapter it was established that there is an important difference in optimal portfolio weights with and without jumps. The aim now it to show that this difference creates an informed reason to take on additional risk in the hope that a severe negative outcome will not be realised before the end of the investment. In the case of the investor this is stated as comparing the performance of an investor that takes jumps into account to one who does not, under different measures of performance. The case of the manager is more complex. Both he and the investor have the same power utility function, but the manager receives utility from fees measured at regular intervals while the investor from the terminal wealth of the investment. Given the lack of data, an assumption based on corporate grounds is made: both managerial and trader (employee) payments follow the same scheme, which is a flat administrative fee plus a performance fee when the portfolio returns exceed a threshold. This is the
same as the publicly announced fee paid by the investor. The corporate argument is uniformity in payment schemes for discipline, equity and comparison purposes. The manager, in essence, is given two possible choices of portfolio weights from the investor, or equivalently he performs the same maximisation exercise as in the previous chapter on behalf of his client (it must be noted that his fees are not part of the maximisation).

6.3 Jumps and the investor

The simulation for the investor is arranged as follows. A time series of daily log returns is generated by using the SVCJ model and daily percentage parameters. Total returns on the risky asset are then used to calculate portfolio returns of a portfolio consisting of a share and a risk-free asset with return \( r = 2\% \) annually. Jumps are thus present in the path of the risky asset and the portfolio, a fact that is known to the investor. In the first case the portfolio uses the optimal weight from the SVCJ model, \( w_J \), thus taking jumps into account. In the second case, the investor knowingly chooses to ignore them and uses the weight from the SV model, \( w \). The measures of win in the horse race are terminal wealth, average terminal utility and number of wins for each investor. The length of the investment period is 2, 3, 5, 10, 15, 24 and 30 years, risk aversion \( \gamma \) takes values 2, 3, 4, 5, starting wealth is \( W_0 = 100 \) and the utility function is \( U(W) = \frac{W^{1-\gamma}}{1-\gamma} \). For a simulation length of 5.000 the results are accurate, since tests up to 100.000 runs led to the same results.

Tables D.1.1 - D.1.4 contain the simulation outcomes for the two types of investors. The result is that the no-jumps investor always wins in terms of wealth but always loses in terms of utility. This is calculated by taking the difference between average terminal wealth (utility) for both investors and looking at the sign. The result is persistent across investment horizons and degrees of risk aversion. The metrics are the number of times the wealth (utility) of the no-jumps investor is greater than that of the jumps investor and the difference between the average terminal wealth (utility) of the no-jumps and the jumps-investor. The percentage of wins ranges from 62\% to 80\% in favour of the No-Jumps investor as the investment horizon ranges from 2 to 30 years, and is the same across all values of \( \gamma \). The percentage of wealth and utility wins is the same because wealth is calculated at the end of the investment and used to calculate terminal utility. Thus, an investor who is interested in terminal wealth only has a major motive to ignore jumps in portfolio allocation. An investor who is interested in utility has, on the other hand, a motive to take jumps into consideration when allocating wealth.
6.4 Jumps and the manager

The structure for the manager is more complicated. As stated before, the manager knows that jumps exist and decides whether to ignore them or not in order to achieve higher fees. The fee structure consists of a flat administration fee and a performance fee whenever portfolio returns exceed 10%. The administration fee is calculated on annual portfolio wealth (value) and, after it is deducted, the potential performance fee is calculated on portfolio profits and deducted as well. To avoid complexities with average portfolio wealth during the course of each year, only initial and terminal wealth values are used. The remaining wealth is the starting wealth of each period until the end of the investment. Each year’s fee is used to calculate the utility of the manager, who uses exactly the same function as the investor. This ensures that any results cannot be attributed to differences in preferences, risk aversion or the shape of utility. Simulations where average portfolio wealth during each year was used provided very similar results that had only a marginal effect on numerical values so the simpler method was adopted.

The scheme of managerial compensation is 2 + 20 and 1 + 10 in percentage terms. Fees between or beyond these limits, such as 1 + 20 or 0.5 + 10 exhibit similar patterns as the fees of choice because the flat administration fee increases compensation horizontally while the performance fee scales smoothly between 10 and 20%. A fee such as 1+15 will, therefore, be slightly higher than 1+10 and a fee of 2+10 will be slightly lower than 2+20. A set of preliminary simulations was run to establish that conclusion. Since the results were very similar and to avoid repetition, no full results for intermediate schemes are presented, as they are adequately represented by Tables D.2 - D.5. The traditional choice is 2% administration fee + 20% performance fee, which has been the industry staple for years. As the literature review has shown, during the last years there has been a reduction in fees and 2+20 is maintained only by the largest hedge funds. The combinations used are able to accommodate both observed structures and averages across industry as well as fees close to mutual funds. These include funds that have selected to charge low administration fees but maintain a relatively high performance fee, and vice versa. In addition, portfolio wealth is now affected by negative results. When investors face losses, some of them may select to move to a fund that has performed better. On technical grounds this is a way to enhance the trivial result obtained by a straightforward introduction of managerial fees. The manager does not get penalised in a way and will always receive his administration fee, so the differentiating criterion is how many times the fund will exceed the threshold. When portfolio weights that ignore jumps are used,
CHAPTER 6. INVESTOR AND MANAGERIAL INCENTIVES TO IGNORE JUMPS

the exposure to the risky asset is greater and, trivially, the jumps manager will exceed the threshold more times than the no jumps manager. When wealth transfer is introduced, there is a flow of capital from the "losing", worse-performing fund to the "winning", better performing one. One fund’s loss is the other fund’s gain, which is calculated on annual performance.

This calls for the definition of a wealth transfer function that determines the amount of wealth to change funds. This is

\[
f(x) = \begin{cases} 
-10 & f(x) \leq 0 \\
-exp(-\delta_1 x) + 1 & x \leq 0 \\
exp(\delta_2 x) - 1 & x \geq 0 \\
10 & f(x) \geq 10
\end{cases}
\]

where x is the relative performance of the funds and is defined as \(r_{NJ} - r_J\), where \(r_{NJ,J}\) is the percentage return of the each portfolio. The function \(f(x)\) demonstrates positive inflow/ negative outflow from the no-jumps fund into the jumps fund. \(\delta_1,2\) is a parameter that defines the curvature of the function and its economic meaning is the sensitivity of wealth transfers to differences in fund performance. For a symmetric function, \(\delta = \delta_1 = \delta_2\) and for the needs of the thesis it is arbitrarily set equal to 0.25, while in the case of asymmetry \(\delta_1 = 0.5, \delta_2 = 0.25\). This makes investors on the no-jumps fund to be more sensitive to relative bad performance than to good performance. The function is barred at 10% and −10% to prevent extreme changes of value. For the asymmetric case, the upper bound remains the same but the lower bound is set to −15%. This produces a stronger reaction when relative losses are observed. Another possible candidate would be a sigmoid function such as the logistic function \(f(x) = \frac{1}{1 + e^{-\delta x}}\), but properly adapted. A form which would yield a similar shape with different curvature properties would be

\[
f(x) = \alpha \left( \frac{1}{1 + e^{-\delta x}} - 0.5 \right)
\]

for \(\alpha = 20, \delta = 0.5\). The function passes from 0, is asymptotic or at least very slowly increasing, but also symmetric - for asymmetries to be introduced a piecewise sigmoid for different \(\delta\)s is needed. In practice the result is the same. The plots can be found in the Appendix (Figure D.1 - D.3).

The setup of the managerial simulations is as follows. Two funds are considered, one where the manager has chosen to use the SVCJ portfolio weights and one where
the manager uses the SV ones. Since the weights are time invariant, the strategy is buy-and-hold until the end of the investment. The portfolio consists of a risky and a risk-free asset with return $r = 2\%$. A path of daily log returns is simulated by using the SVCJ parameters (since the existence of jumps is known to both managers) and then split into annual segments. End-of-year wealth, profits, fees and transfer of wealth are calculated in that order, which determines the starting wealth of next period. The measures are average total fees for each manager, average aggregate total terminal utility for each investor and average total terminal utility for each manager.

The simulation outcome can be found in the collected Tables D.2,D.3,D.4,D.5. Tables D.2.1 - D.2.4 present the results for $\gamma = 5$, Tables D.3.1 - D.3.4 for $\gamma = 4$, Tables D.4.1 - D.4.4 for $\gamma = 3$ and Tables D.5.1 - D.5.4 for $\gamma = 2$. Odd-numbered tables refer to the 2+20 scheme and even-numbered tables to the 1+10 scheme. The first two tables have a symmetric wealth transfer function, while the last two an asymmetric one. The winner for each horizon is marked by blue. For $\gamma = 5$ and for all other levels of risk aversion, the manager that does not take jumps into account always wins. The measure of winning in that case is the total amount of fees amassed by the end of the investment, and at face value the no-jumps manager can be seen as having an incentive to go after higher total fees. For the same level of risk aversion, and for all others as well, the investor that takes jumps into account always wins. The measure of winning is the average total terminal utility of the investor, which is based on portfolio wealth. This is true for every table between D.2.1 and D.5.4 regardless of considering a symmetric or asymmetric wealth transfer function, a different level of $\gamma$ or the length of the investment horizon. This result is the same as in the previous section.

Tables D.2.1 and D.2.2 show that the manager that does not consider jumps and uses the SV weights wins on average for the 2- and 3-year investment horizons but loses in all subsequent ones. The measure of winning is, again, the average total annual utility for the manager, which is based on fees. Tables D.2.3 and D.2.4, where an asymmetric wealth transfer function is used, show that the manager that ignores jumps never wins on average. For the 5.000 simulations, the manager that considers jumps wins only 36\% of the runs for the 2-year horizon under the 2+20 scheme, with the percentage slightly increasing up to 39.6\% for the 30-year horizon. The result is much different in the asymmetric case of Tables D.2.3 - D.2.4, where the same manager wins 37\% of the runs for the 2-year horizon but gradually increases the rate up to 72 – 75\% for the 30-year horizon. In Tables D.3.1 - D.3.2 the winning horizons for the manager that ignores jumps are the 2- and 3-year ones, with the percentage of wins for the manager
that considers jumps being 33 – 36% across the tables. Tables D.3.3 and D.3.4 show the same pattern as Tables D.2.3 and D.3.4, only now the manager that ignores jumps wins the shortest horizon on average. The patterns are largely the same, with slightly reduced win percentages, for tables D.4.1 - D.4.4. Tables D.5.1 - D.5.4, for the lowest value of $\gamma$, show the longest winning horizons for the manager that ignores jumps in both the symmetric (2 to 10 years) and the asymmetric case (2 to 5 years), as well as the lowest winning percentages for the manager that considers jumps.

The results show a clear incentive for the manager to ignore jumps in terms of fees. Across the same horizons, fee levels, wealth transfer patterns and risk aversion levels as before, the no-jumps manager always amasses a higher amount of fees on average. The investor, on the other hand, follows the same pattern as before, with a clear and constant win of the no-jumps investor in terms of wealth but a loss in terms of utility. The results are consistent across the board. Managerial utility is more complicated. For all levels of $\gamma$ the No Jumps manager wins for short horizons and loses for longer horizons. For decreasing risk aversion and the same wealth transfer function, the winning period ranges from 2 to 10 years. For the same level of risk aversion, a change in the fees structure has no effect on the winning period. However, when the symmetric and asymmetric wealth transfer functions are compared, the winning horizon is shorter for the asymmetric case given the same $\gamma$.

The situation is reversed when the number of wins is considered. Here, the no-jumps manager has a significant chance to win across the table for every investment horizon. Although he loses on average terms in the symmetric case across all tables, the probability of winning a single run (measured as the number of wins for the entire set of simulations) ranges between 60% – 70%, depending on risk aversion. This means that, even if his expected outcome is in favour of taking jumps into account, he still has a very high probability of receiving more benefit for a single run. The reason for this discrepancy lies at the tails of the distribution of terminal utility. Histograms D.4 and D.5 compare the Jumps and the No Jumps utility samples and show that the vast majority of realisations are concentrated close to zero but the long tail dominates. It contains very low values that cannot be covered by the more frequent realisations. The tail exists for all $\gamma$s and does not depend on the level of initial wealth. Simulations conducted with $W=1$ led to the same results and the same distribution for managerial utility. Also, the tail is quite pronounced under the modest 1+10 scheme for a long investment horizon and symmetric wealth transfer function. The shape persists across the various combinations of $T, \gamma$ and fee schemes. When the wealth transfer function is asymmetric, the probability of the
no-jumps manager to win starts off at 60% – 70% for a horizon of 2 years, gradually diminishes and is reversed for longer horizons, where the jumps-investor wins both on average and more frequently. The result is, again, consistent across levels of risk aversion and shows how for longer horizons the No Jumps manager loses both on average and in percentage.

A further result is the effect of assuming a symmetric or asymmetric wealth transfer function and its economic meaning. It appears to have a more significant effect on the manager’s outcome than risk aversion, investment horizons and fee structures. For the same structure and risk aversion, fees make no difference. For the same fees and risk aversion, the manager has a much stronger incentive under symmetric wealth transfer. As risk aversion decreases, the manager has an incentive to ignore jumps for horizons from 2 to 5 years as opposed to the 10-year limit under symmetry. The outcome that the shorter the horizon the greater the motivation is valid, but different fee structures do not appear to magnify or dampen down its strength. What does is how investors react to positive or negative results. Asymmetry in that context implies more violent movements in bad results, which in the case of the no-jumps manager implies greater exposure to risk. Investors with symmetric reactions weigh wins and losses equally, while asymmetric reactions mean higher downside risk aversion (the relevant literature has been discussed in the appropriate chapter of the thesis). This is obvious across all tables D.2 – D.5 but particularly for $\gamma = 2$ (Table D.5), which is combined with the longest horizon and managerial incentive.

6.5 Conclusion and summary of results

The findings show a clear incentive for both the investor and the manager to take excess risk. If the investor is concerned about terminal wealth, he is bound to win more often and on average compared to the more conservative investor that does not ignore jumps. On the other hand, if utility is the measure of win, the jumps-investor always wins on average. The technical reason for that is that the optimisation process was conducted on utility basis, so the weights of the SV model are suboptimal when the time series contains jumps. The economic intuition is that, no matter how long the investment horizon, the investor is always worse-off in terms of utility but may choose, either deliberately or because of myopic decision making, to focus on wealth instead. In any case the motivation is there, and the result holds across all levels of risk aversion, initial wealth and investment horizons.
CHAPTER 6. INVESTOR AND MANAGERIAL INCENTIVES TO IGNORE JUMPS

When a more complex structure and a fund manager is introduced, the same intuition holds for the investor. Here, however, the portfolio allocation decision belongs to the manager who makes a choice based on his fees and utility. Total fees are always in favour of the no-jumps manager, a fact that poses a clear incentive from many sides - absolute compensation, reputation, performance measures etc. The manager’s utility is more ambiguous. While in the short run the no-jumps manager wins both in terms of income and utility, in the mid-to-long run the jumps catch up to him leading to worse performance compared to the more disciplined jumps manager. A manager that focuses on the short-run for a short-term investment has a very clear motivation to act against the investor’s interests and take a more aggressive position. For longer term investments, the manager will choose the same if he only considers the short-run, and even if he takes utility into account he still has roughly 60% chance to achieve a good result. The policy implication is that short-term goals and perspective reduce both individual and client welfare in the long-run. The results are, again, homogeneous across time horizons, risk aversion, initial wealth and, this time, managerial compensation schemes.

The differentiating element is, thus, how wealth transfers between the two funds each time. If investor reaction is symmetrical between wins and losses, the jumps-manager wins fewer times on average and the winning horizon for the no-jumps manager is longer than when investor reaction is asymmetrical. As risk aversion increases, the winning horizon shrinks for both cases and is eliminated in the asymmetric case for $\gamma = 5$. High risk aversion coupled with more severe swings of wealth at losses eliminate any advantages in utility for the no-jumps manager. Unavoidably, this poses the question of what leads an investor to have a more severe reaction for downside risk than upside risk, and what policy suggestions that entails. A first factor is information - an ill-informed or naive investor will have no reason to assess negative outcomes different than positive ones. Also, a symmetric function implies that this assessment will lead to identical reactions, while research focusing on downside risk often finds that investors have stronger reactions for losses than for wins (flight to quality).

A first policy suggestion is, thus, transparency on behalf of the fund and publishing sufficient and factual information on investment strategies, portfolio allocation, exposures to risk factor, products, assets and markets where the fund is currently investing. This information will allow the investor to monitor fund performance more effectively and not solely rely on realised profits, as well as plan future movements better. A form of transparency already applied (Preqin (2017)) is providing investors with a managed account structure that allows them to track the fund manager’s decisions. Expanding
similar tools would align investor and managerial attitudes and help the regulator monitor and supervise the entire industry. Therefore, any policy that makes investors more aware to the actual risk they are facing in an investment will cause them to withdraw in the case of increased losses, limiting assets under management for the fund and reducing the incentive of the manager to undertake and/or conceal such risk.

Such knowledge is directly related to investor protection (e.g. clawbacks) and information on the structure of the risky portfolio, and by consequence exposure to risk factors. The wealth transfer function is a way to import a penalising loss of wealth for the losing fund, a loss that can be seen not only as a flight of investors but as paybacks and penalties paid by the fund. A second policy suggestion, supported by the empirical section of this chapter’s literature review Barclays (2017), is a more thorough integration and embodiment in legislative procedures of investor protection, such as clawbacks, preferential terms, lock-up periods for fees, fee stratas and discounts, and contract terms in general. Although fees do play a role in creating an incentive for the manager, the flipside for the investor is related to preferential terms and cover beyond costs. An investor is inclined to welcome positive shocks and be more averse to negative shocks. When the information he receives is such that he is bound to treat those two elements equally, his reaction will be symmetric. Anything that may cause him to have an asymmetric reaction will also reduce the incentive for the manager. This penalty, translated here as movement of wealth, may take various forms (payment buffers, clawbacks, capital cover etc, regulatory penalties, paybacks) which are tied to regulatory responsibilities and overall market transparency.

From a regulator’s point of view, direct intervention on compensation schemes is unlikely to eliminate the incentive for excess risk and align managers and investors. This includes predetermined quotas or enforced schemes. The effect of such a policy would be to reduce the total compensation of the manager (which can be seen as a way to reduce his incentive) but the winning horizon for which that incentive persists is unchanged. Direct intervention of the policymaker that caps managerial returns is a third policy suggestion, but its scope is very specific. If the goal is to limit excessive fees in the industry to protect the investor and limit the absolute amount of managerial compensation, then there is a reason to apply quotas. If, however, the goal is to remove the moral hazard issue and eliminate the circumstances in which such a motive exists, then the policy will be ineffective and probably distorting. Better monitoring, data collection and stricter obligations on behalf of the fund to provide details on its investment strategies and allocation will be more efficient. In that way, the regulating authority will be in a better
position to identify excess risk in the industry and sources of increased risk (or even a "piling up" of disaster probabilities), which will in turn enable it to take corrective measures and inform investors. Investors, in their turn, will adjust their expectations and attitude accordingly and plan their reactions, thus defining the properties of the wealth transfer function.

On the other hand, wealth transfers can be seen as penalising the manager because they reduce portfolio wealth and, therefore, fee margins. Such penalties, like clawbacks, are more likely to have effects because they are applied directly at a wealth level which touches upon fund profits. The main issue lies at reducing the moral hazard problem arising between the aims of the investor and the aims of the manager. For mid- and long-term investments they are aligned on average, according to terminal utilities. The strongest incentive manifests when high fees and low risk aversion are combined and the weakest under low fees and high risk aversion.

More transparency in the terms, in the portfolio structure and more disclosure on exposure in risky investments (e.g. uncovered derivative positions) will play a crucial role in reducing excess risk. Also, better communication between the regulatory authority, the fund and the investors will help in the assessment of the actual risks involved. Another area of improvement is the regulatory framework and investor protection, as well as the application of preferential terms and clauses. The first important result of the chapter is that jump risk was successfully isolated from other sources of risk. It is very important to highlight that the differences between the investor and the manager are not due to differences in utility or preferences, since they both use the same function and parameter. The difference is that the investor’s utility relies on portfolio wealth while the manager’s on fees. Event risk was found to have a clear, strong effect on managerial appetite for risk and it was found to provide a motive to ignore rare events in favour of short-term profits. As the investment horizon lengthens, the incentive vanishes on expected terms but is retained in percentage terms if investors react the same at wins and losses. If investors react more intensely at relative losses, the winning horizon for the risky manager is shorter both on average and in percentage terms. The managerial incentive is, thus, short- to mid-term and is enhanced by low risk aversion. The crucial factor that causes differentiation is how investors treat relative losses and how they decide to switch funds.
Chapter 7

Conclusion and further extensions

7.0.1 Results

The thesis is in the pleasant position to report a series of new results from both a technical and an empirical point of view.

Chapters 3 and 4 discussed to what extent jumps are a necessary technicality for the research question at hand. Chapter 3 focused on whether fat-tailed distributions for asset prices were enough to make a difference in decision making on behalf of the investor. If the investor considered a distribution different than the real one, would there be a difference in expected utility? The answer was no, both in a single state and a Markov-state setup with a ”good” and a ”bad state”. The advantage of the investor using the correct distribution in his decision making was marginal and statistically insignificant. Therefore, tail risk alone was not enough of a factor to create an incentive to select a riskier position. The limitations of the setup (singel period optimisation and CARA utility) were discussed and do play a role in the outcome. Another contribution was the timing of the jump in a compounding exercise. It was shown that an early jump would cause terminal wealth to be slightly lower than a mid- or end- period jump. The result was that even a jump right before the end of the investment would not be enough to make the investor worse-off, compared to the same jump happening earlier. This demonstrates the ability of an investment to recover over time.

Having established that tail risk alone does not pose a sufficient motive for the investor to undertake excess risk, the attention in chapter 4 turned to stochastic volatility. The goal was to see to what extent jumps and tail risk could be represented by an SV model and how stylised facts of asset prices could be replicated. The model proved unable to match excess kurtosis and outliers in the samples, both for the pre-crisis and the crisis period. Parameter accuracy was smaller in the crisis period but the only significantly
different estimate was the standard deviation of volatility errors. This is tied to higher variance in volatility and has also led to higher kurtosis. This kurtosis was still much lower than the sample kurtosis, and the comparison of absolute returns and smoothed volatility states revealed that outliers and frequent mid-range jumps were not represented by volatility. Therefore, it is necessary to introduce jumps explicitly in a stochastic volatility framework.

Chapter 5 demonstrated the estimation results on the SVCJ model that includes leverage effects, stochastic volatility and simultaneous jumps in both volatility and returns. The main results are based on S&P500 daily data from 1980 to 2016 and secondary results on 2007 - 2017 S&P500 daily data and 1872 - 2016 S&P500 monthly data. The comparison with two papers estimating the same model with the same time series at different time spans showed that the model tends to trade volatility with jump frequency when the financial crisis is introduced in the sample, that leverage effects become stronger and estimation accuracy is slightly affected. A major result is the very accurate estimation of the jump size correlation parameter, $\rho_j$, to be 0. Other literature either provides an estimate with very high standard deviation or drops the parameter altogether claiming estimation problems. The thesis provides a practical reason to drop the assumption of correlated jump sizes and not rely simply on convenience. Another, ambiguous, result was that jumps are also present in low-frequency data as well, with one jump every 5 years on average. The next result is the derivation of closed-form portfolio weights for this class of models and compare the manipulations to that of existing solutions. There was a thorough discussion on how stochastic arrival intensity for jumps is a convenient choice because it leads to tractable closed-form solutions. Nevertheless, this feat comes with great deficiencies in parameter estimation, MCMC in particular. Literature that discusses portfolio weights uses stochastic arrival intensity but literature discussing model fit uses constant intensity, like the PhD. The gap was bridged by setting volatility equal to its long-run average, a fact that turned an ODE to a Riccatti equation which, in sequence, has a tractable solution. Although a similar idea was used in parallel by other papers to transform parameters, this is the first time it was used for portfolio weights. The optimal allocations for the SVCJ (jumps) and SV (no jumps) models were very different and therefore not very sensitive to small changes in the parameters. An investor using the SV model would put a significantly larger part of his wealth on the risky asset for any level of risk aversion.

Chapter 6 discussed how jumps create incentives for the investor and the manager of a fund to disregard rare events risk and bet that a jump will not take place before the
end of the investment, harming their wealth and utility. The data generating process contained jumps and two types of investor and manager were considered: one who uses the correct SVCJ weights and one who deliberately uses the SV weights. The same utility function for both the manager and the investor was used so that any outcome would not be attributed to individual differences between the two. The manager derives utility from fees while the investor from portfolio wealth. The measures of win were average terminal wealth, average terminal utility, percentage number of wins and total amount of fees. In the first, investor-only simulations, the no-jumps investor would always win in terms of wealth but always lose in terms of utility, both on average (expected) and on percentage terms. In the second, investor and manager only simulations, a movement of wealth at the end of each year from the losing to the winning fund was introduced. The results for the investor remained the same. The no-jumps manager was found to always have an incentive in terms of absolute fees to ignore jumps and a short-to-mid-term incentive in terms of expected utility. The effect was observed across all levels of risk aversion and fee structures for horizons from 2 to 30 years. Altering $\gamma$ would reduce or increase the winning horizon but never eliminate the incentive. The other strong factor was the pattern of wealth movements (symmetric or asymmetric). If wealth movement to and from the fund is the same for positive and negative relative returns, then the incentive is greater. If wealth movement is more severe (a steeper function with a lower cap), then the winning horizon is visibly reduced. This effect is traced for the same level of risk aversion as well as different levels. On the other hand, manager compensation schemes had no interplay with either risk aversion or wealth transfer - for both $2 + 20\%$ and $1 + 10\%$ the winning horizon was the same. Therefore, the incentive for the manager to ignore jumps is affected by his (and the investor’s) risk aversion and how investors react to losses, and is clearly present in the short and mid-run as opposed to the long-run. In the long-run, jumps tend to catch up and harm managerial utility. The issue is directly linked to investor information around how the manager allocates portfolio wealth, how his actions are monitored and what provisions or penalising clauses are set either by the regulator or the investment contract. The list of policy recommendations includes better monitoring of fund management, extensive disclosure of fund structure and trading strategies followed, wider application and legislation on clawbacks, penalties and investor protection and employment of monitoring tools such as investment accounts.

In that way, the research question of the thesis have received a clear answer in all three parts. The present work can be extended in a number of ways discussed in the next section.
7.0.2 Further extensions

The first area of improvement is technical. The modelling of jumps has received a lot of attention and is always an area of advancement. The Poisson structure is a sufficient and tested way to replicate jumps but, in certain frameworks, can be seen as inaccurate. Levy processes (Li et al. (2006)) have been found to provide a more accurate replication of time series features. A more fruitful way is the use of a Hawkes process, as in Fičura and Witzany (2015). This stochastic process introduces a self-exciting component in the jump that allows the monitoring of sequential jumps as well as the dissipation or possible build-up of jump effects. This replaces the need for a stochastic arrival intensity in the style of Liu et al. (2004) and makes up for a more modern and efficient tool.

There is also plenty of ground to be covered by using novel estimation methods, such as particle filtering. Pitt et al. (2014) provides an application on a model with leverage and jumps in returns only, while a thorough discussion can be found in Andrieu et al. (2010) and Andrieu et al. (2003). Particle filtering is superior in terms of efficiency and convergence to MCMC, as shown in the above literature. In addition, an application with jumps in both returns and volatility is currently missing, so there is space for a contribution.

Moving away from the technical aspect, an area where the present results can be greatly expanded is in the wealth transfer function. The form used here is illustrative and serves a specific mechanical purpose, but the idea can be expanded and get much deeper foundations. There is a direct link to herding, the amplification of anxiety and contagion. An asymmetric curve implies asymmetric reactions to a source of risk but also that investors are relatively more hesitant to run towards already realised wins than run away from realised losses: an investor will leave an underperforming fund easier than he will join a winning fund. In addition, jumps are not restricted to being negative - positive jumps can, and do exist. The estimate on returns jump size, however, implies a negative starting point or, alternatively, a negative average size. Linking jumps to negative returns and having investors react viciously when they occur is, therefore, not unjustified. This calls for a better linkage of the wealth transfer function to investor behaviour and patterns, either at an individual or a collective level.

Another possible area of improvement is the use of a different trading strategy. Time invariable jumps may be a mathematical outcome as well as the depiction of a buy-and-hold strategy, but it is interesting to test how incentives would be affected by (time-)varying jumps. The technical obstacles are identified to the nature of the Bellman equation solution, that relies on a very specific power utility function. Any
alterations in either the methodology or the utility function may lead to a solution being unfeasible. Abandoning the properties used in the thesis is not so much of an issue, but the framework is tight and not very prone to manipulation. Time variation in portfolio weights can also be linked to considerations by Korn and Kraft (2004) about the stability of solutions for SV, SVJ and SVCJ models.
Figure A.1. Normal distribution terminal wealth for the same series when the jump occurs in different times
Figure A.2. Gamma distribution terminal wealth for the same series when the jump occurs in different times.
**Inv. Gaussian distribution - disasters**

*Figure A.3. Inverse Gaussian distribution terminal wealth for the same series when the jump occurs in different times*
Table A.1: Ex post expected utility across combinations of weights and realised distributions. Bold denotes the winner of the row. Tables are read from left to right.

<table>
<thead>
<tr>
<th>Normal weights</th>
<th>Gamma weights</th>
<th>Inv. Gauss. weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal dist</td>
<td>-0.354335</td>
<td>-0.35435</td>
</tr>
<tr>
<td>-0.354334</td>
<td>-0.354342</td>
<td>-0.35435</td>
</tr>
<tr>
<td>Markov Normal</td>
<td>-0.354342</td>
<td>-0.35435</td>
</tr>
<tr>
<td>-0.354339</td>
<td>-0.35435</td>
<td>-0.35435</td>
</tr>
<tr>
<td>Markov Gamma</td>
<td>-0.35435</td>
<td>-0.35435</td>
</tr>
<tr>
<td>-0.354333</td>
<td>-0.35435</td>
<td>-0.35435</td>
</tr>
<tr>
<td>Markov Inv. Gauss.</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>-0.354334</td>
<td>-0.354335</td>
<td>-0.354335</td>
</tr>
<tr>
<td>SS Ex post terminal utility</td>
<td>-1.02</td>
<td>-1.02</td>
</tr>
</tbody>
</table>

*(Note: Table is read from left to right)*
Figure A.4. Normal distribution - pdf plot for $N(0,1)$
Figure A.5. Gamma distribution - pdf plots for different parameters
Figure A.6. Inverse Gaussian distribution - pdf plots for different parameters
Appendix B

Appendix for Chapter 4

B.1 Tables and Figures

Table B.1: MCMC parameters for the SV model for different horizons and data sets.

<table>
<thead>
<tr>
<th></th>
<th>$\mu$</th>
<th>$\phi$</th>
<th>$\sigma_\eta$</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hautsch &amp; Ou</td>
<td>-9.471</td>
<td>0.99</td>
<td>0.087</td>
<td>4.3884</td>
</tr>
<tr>
<td></td>
<td>(0.171)</td>
<td>(0.003)</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Replication</td>
<td>-9.62584</td>
<td>0.987623</td>
<td>0.127236</td>
<td>5.79317</td>
</tr>
<tr>
<td></td>
<td>(0.169997)</td>
<td>(0.00307)</td>
<td>(0.0098449)</td>
<td></td>
</tr>
<tr>
<td>Corrected sample</td>
<td>-9.58047</td>
<td>0.987518</td>
<td>0.127635</td>
<td>5.78506</td>
</tr>
<tr>
<td></td>
<td>(0.175654)</td>
<td>(0.00336461)</td>
<td>(0.0108075)</td>
<td></td>
</tr>
<tr>
<td>2005 – 2015</td>
<td>-9.59566</td>
<td>0.984292</td>
<td>0.175091</td>
<td>8.02188</td>
</tr>
<tr>
<td></td>
<td>(0.264911)</td>
<td>(0.0236879)</td>
<td>(0.0189727)</td>
<td></td>
</tr>
</tbody>
</table>

Table B.2: Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Kurtosis</th>
<th>Sample Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hautsch &amp; Ou</td>
<td>0.00036</td>
<td>0.009</td>
<td>8.276</td>
<td>4231</td>
</tr>
<tr>
<td>Corrected sample</td>
<td>0.0003808</td>
<td>0.0097773</td>
<td>7.99061</td>
<td>4079</td>
</tr>
<tr>
<td>2005 – 2015</td>
<td>0.0001873</td>
<td>0.0116215</td>
<td>14.1445</td>
<td>2656</td>
</tr>
</tbody>
</table>
FIGURE B.1. Markov chains for $\mu, \phi, \sigma^2$ (top - bottom), not corrected sample, 1991 - 2007
Figure B.2. Markov chains for $\mu, \phi, \sigma^2$ (top - bottom), corrected sample, 1991 - 2007
Figure B.3. Smoothed volatility states (top) and absolute returns (bottom), not corrected sample, 1991 - 2007
Figure B.4. Smoothed volatility states (top) and absolute returns (bottom), corrected sample, 1991 - 2007
Figure B.5. Smoothed volatility states (top) and absolute returns (bottom), 2005 - 2015
Figure B.6. Markov Chains $\left(\mu, \phi, \sigma^2\right)$ (top - bottom), 2005 - 2015
Appendix C

Appendix for Chapter 5

C.1 n-dimensional Ito’s lemma in the Bellman equation

The formula for the n-dimensional Ito’s lemma is

\[ dZ_t = \frac{\partial f}{\partial t}(t, X_t)dt + \sum \frac{\partial f}{\partial x_i}(t, X_t)dX_t^i + \frac{1}{2} \sum \frac{\partial^2 f}{\partial x_i \partial x_j}(t, X_t)dX_t^i dX_t^j \]  

(C.1)

The example is merely illustrative and is based on diffusions, so it does not contain jump terms. Here, \( f \) is the indirect utility function and corresponds to \( F_t \), \( X_t^{i,j} \) are the processes for \( W_t \) and \( V_t \) respectively, \( dX_t^i \) is the mean times \( dt \) in each process and \( dX_t^i dX_t^j \) is the product of the two correlated Brownian motions, which according to stochastic calculus yields the product of their standard deviation terms, the correlation and \( dt \). (C.1) is applied to processes (4.7) and (4.2) with an additional jumps term generated. Subscript \( t \) is dropped for convenience. The result is

\[ L(F) = (r + \phi EP)WF_W + \kappa(\theta - V)F_V + \frac{1}{2} \phi^2 W^2 VF_WW + \frac{1}{2} \sigma_V^2 V F_{VV} + \sigma_V \phi W V \rho F_{VV} \]

\[ + \lambda E[F(W(1 + \phi E(\xi^V)), V + E(\xi^V), t) - F] \]  

(C.2)

The last term comes from the Ito’s lemma ability to break jump-diffusion down to its components. Simply put, the expected effect of simultaneous jumps in returns and volatility on indirect utility is the difference between \( F(W,V,t) \) and \( F \) times the arrival intensity of the Poisson. In other words, \( W \) and \( V \) in \( F(W,V,t) \) are placeholders and represent the increase in wealth and volatility by the expected jump size. Some simple algebra around the Poisson function provides the final result.
C.2 Algebra of the Bellman equation

The solution method is to conjecture and then verify that $F(W_t, V_t, t)$ is of a certain form, namely

$$F(W_t, V_t, t) = \frac{W^{1-\gamma}}{1-\gamma} \exp(A(t) + B(t)V) \tag{C.3}$$

where $A(t), B(t)$ depend only upon $t$ but not $W$ and $V$. This leads to the following partial derivatives.

$$F_W = W^{-\gamma} \exp(A(t) + B(t)V) \tag{C.4}$$
$$F_{WW} = -\gamma W^{-1-\gamma} \exp(A(t) + B(t)V) \tag{C.5}$$
$$F_t = \frac{W^{1-\gamma}}{1-\gamma} (A'(t) + B'(t)V) \exp(A(t) + B(t)V) \tag{C.6}$$
$$F_V = \frac{W^{1-\gamma}}{1-\gamma} B(t) \exp(A(t) + B(t)V) \tag{C.7}$$
$$F_{WV} = B(t) W^{-\gamma} \exp(A(t) + B(t)V) \tag{C.8}$$
$$F_{VV} = \frac{W^{1-\gamma}}{1-\gamma} B^2(t) \exp(A(t) + B(t)V) \tag{C.9}$$

For the jumps related term in (C.2)

$$F(W(1 + \phi E(\xi^Y)), V + E(\xi^Y), t) = \frac{W^{1-\gamma}(1 + \phi E(\xi^Y))}{1-\gamma}^{1-\gamma} \exp(A(t) + B(t)(V + E(\xi^Y)))$$

which, after differentiating, yields

$$\frac{\partial F(\ldots)}{\partial \phi} = E(\xi^Y) W^{1-\gamma}(1 + \phi E(\xi^Y)) \exp(A(t) + B(t)V + B(t)E(\xi^Y)) = K \tag{C.10}$$

Substituting (C.2) into the Bellmann equation $0 = F_t + \max_{\phi}[L(F)]$ (4.8) and differentiating with respect to $\phi$ instantly removes some expressions. What remains is

$$EP^* W F_W + \phi W^2 V J_{WW} + \rho \sigma_V W V J_{WV} + \lambda K = 0 \tag{C.11}$$

and with the appropriate substitutions from (C.4 - C.10) into (C.11)

$$EP^* W W^{-\gamma} \exp(A(t) + B(t)V) - \phi W^2 V \gamma W^{-1-\gamma} \exp(A(t) + B(t)V) + \rho \sigma_V W V B(t) W^{-\gamma} \exp(A(t) + B(t)V) + \lambda E(\xi^Y) W^{1-\gamma}(1 + \phi E(\xi^Y)) \exp(A(t) + B(t)V + B(t)E(\xi^Y)) = 0 \leftrightarrow$$

$$EP^* W^{1-\gamma} - \phi \gamma V W^{1-\gamma} + \rho \sigma_V W^{1-\gamma} V B(t) + \lambda E(\xi^Y) W^{1-\gamma}(1 - \phi E(\xi^Y))^{-\gamma} \exp(B(t)E(\xi^Y)) = 0 \leftrightarrow$$
The last expression is equation (4.9) for the optimal portfolio weight.

The final step is to derive the ordinary differential equations for B(t) and A(t) for which the assumed form of the indirect utility function is indeed a solution. In order for that assumption to hold, the solution for \( \phi \) needs to validate the Hamilton-Jacobi-Bellman equation and set it equal to zero. After substituting \( \phi \) (4.9) and \( F(C.3) \) into (5.8), it is possible to eliminate \( W_1 - \gamma \) from all terms. This allows us to separate the terms that contain \( V \) from those who do not. The general form is thus \( D + H^*V = 0 \) where both \( D \) and \( H \) need to be 0, or else

\[
H = B'(t) - \frac{1}{2} \gamma \phi^2(1 - \gamma) + \frac{1}{2} \sigma_V^2 B^2(t) + (\sigma_V \phi \rho(1 - \gamma) - \kappa) B(t) = 0
\]

and the expression for \( A(t) \) is omitted as it does not affect the system of equations. It must be noted that the expression for \( A(t) \) contains the jump term.

To sum up, the optimal weight \( \phi \) is the solution to the following system.

\[
\phi = \frac{EP}{\gamma V} + \frac{\rho \sigma_V B(t)}{\gamma} + \frac{\lambda E(\xi^V)(1 - \phi E(\xi^V))^{-\gamma}}{\gamma V} exp(B(t) E(\xi^V))
\]

where \( B(t) \) solves the differential equation

\[
B'(t) - \frac{1}{2} \gamma \phi^2(1 - \gamma) + \frac{1}{2} \sigma_V^2 B^2(t) + (\sigma_V \phi \rho(1 - \gamma) - \kappa) B(t) = 0
\]

with initial conditions

\[ A(T) = 0, B(T) = 0 \]

C.3 SVCJ posteriors

The discretised version of the SVCJ model described by (4.1 - 4.2) is

\[
Y_t = \mu + \sqrt{V_{t-1}^c} \epsilon_t^Y + \xi_t^Y J_t \quad (C.12)
\]

\[
\Delta V_t = V_t - V_{t-1} = \alpha + \beta V_{t-1} + \sqrt{V_{t-1}^c} \epsilon_t^V + \xi_t^V J_t \quad (C.13)
\]

where \( \alpha = \kappa \theta, \beta = -\kappa, \epsilon^Y, \epsilon^V \sim N(0,1) \) with correlation \( \rho \), log returns \( Y_t = \log(S_t/S_{t-1}) \), \( J_t \sim Ber(\lambda), \xi^Y \sim \exp(\mu_v), \xi^Y \sim N(\mu_Y + \rho_j \xi^V, \sigma^2) \) jump sizes with
correlation \( \rho_j \). The model can be discretised with either \( \alpha = \kappa \theta \) and \( \beta = -\kappa \) or directly \( \kappa, \theta \). Here the first method is used. In this form, Metropolis - Hastings sampling is needed for \( V_t, \rho, \sigma_V \). If the volatility error term is rewritten as \( \epsilon_t^V = \rho * \epsilon_t^Y + \sqrt{1 - \rho^2} * \zeta_t \), where \( \zeta_t \sim N(0, 1) \) independent of \( \epsilon_t^Y \), and \( \omega = \sigma_V^2(1 - \rho^2), \phi = \sigma_V * \rho \) are defined, then they can be sampled directly from the resulting posteriors due to conjugacy and get \( \rho = \frac{\phi}{\omega}, \sigma_V^2 = \omega + \phi^2 \).

The parameters to be sampled are \( \theta = (\mu, \alpha, \beta, \rho, \rho_j, \sigma_V^2, \mu_Y, \mu_V, \sigma_Y^2) \), the vectors (sets) to be sampled are \( (V_t, J_t, \xi_t^Y, \xi_t^V) \) and the notation (...) denotes all other quantities.

It is useful to write down the model likelihood function. It is a bivariate normal distribution of \( Y_t, V_t \). Swapping sides at the discretised model, (A.12) becomes

\[
Y_t - \mu - \xi_t^Y J_t = \sqrt{V_{t-1}} \epsilon_t^Y
\]

and let the LHS = \( A_t \).

\[
V_t - V_{t-1} - \alpha - \beta V_{t-1} - \xi_t^V J_t = \sigma_V \sqrt{V_{t-1}} \epsilon_t^V
\]

Similarly for the volatility process (C.13), let the LHS = \( B_t \). These consist a bivariate normal

\[
p(Y_t, \Delta V_t|V_{t-1}, ..., Y_{t+1}) = \frac{1}{2\pi \sigma_V V_{t-1} \sqrt{1 - \rho^2}} * \exp[-\frac{1}{2(1 - \rho^2)} \left( \frac{A_t^2}{\sigma_V^2 V_{t-1}} + \frac{B_t^2}{\sigma_V^2 V_{t-1}} - 2 \rho A_t B_t \right)]
\]

(C.14)

The methodology for conditioning has been explained in Chapter 3 and will not be repeated here. The constant term outside the exponent will always be ignored.

Volatility yields the most complex posterior.

\[
p(V_t|V_{t-1}, V_{t+1}, Y_t, ...) \propto \frac{1}{V_t} * \exp[-\frac{1}{2(1 - \rho^2)} \left( \frac{\sigma_Y^2 A_t^2 + B_t^2 - 2 \rho \sigma_V A_t B_t}{\sigma_Y^2 V_{t-1}} + \frac{\sigma_Y^2 A_{t+1}^2 + B_{t+1}^2 - 2 \rho \sigma_V A_{t+1} B_{t+1}}{\sigma_Y^2 V_t} \right)]
\]

for the neighbouring values \( V_{t-1}, V_{t+1} \) of \( V_t \) where the Markov property is again exploited. After expanding the squares, conditioning and factoring, the most compact analytical form is
Due to proportionality, any other formulation would suffice but this one is the neatest. This creates a cursoring over the existing volatility vector that keeps or substitutes the current value according to a Metropolis - Hastings step. The proposal is a random walk $e$ that follows $N(0, \sigma^2)$ centered on the previous value, so $N_{\text{prop}} = V_{\text{old}} + e$. Setting $\sigma = 0.05$ works well in practice. It has the advantage of being completely agnostic and with a pace $\sigma$ that can be easily calibrated. Li et al. (2006) provide an insightful comment on the performance of different sampling techniques, which were validated by the thesis. The construction of an appropriate proposal for Accept - Reject is very difficult, Kalman filtering cannot be applied due to non-normality and ARMS is at least comparable in performance to random walk Metropolis - Hastings. A difference in performance was noted without a significant gain in precision, and the random walk approach was selected.

The posteriors for $\rho, \sigma^2_v$ are non-standard. For $\rho$ it is almost identical to the likelihood function (C.14) and a sound Metropolis - Hastings sampler would entail a $U(1−, 1)$ or truncated $N(0,1)$ proposal. For $\sigma$, the form is very similar to an Inverse Gamma distribution. The transformation $\omega = \sigma^2_v (1 - \rho^2)$, $\phi = \sigma_v \rho$ allows the elimination of those terms and separate direct sampling due to conjugacy for $N(0, \frac{1}{\omega})$ for $\omega$ and $\text{IG}(2, 200)$ for $\omega$ as priors. Formally, $p(\phi|V_t, \omega, ...) \propto p(Y_t, V_t|V_{t-1}, \omega, ...)p(\phi|\omega)$ and $p(\omega|V_t, \phi, ...) \propto p(Y_t, V_t|V_{t-1}, \phi, ...)p(\omega)$. The results are an Inverse Gamma posterior for $\omega$ with parameters $\text{IG}(D,C)$

\[
D = \frac{T}{2} + 2
\]

\[
C = \sum \frac{1}{2} \left( \frac{V_t - V_{t-1} - \alpha - \beta V_{t-1} - J_t \xi_t^V)^2}{V_{t-1}} + \frac{1}{200} \right.
\]

\[
- \frac{1}{V_{t-1}} \left( Y_t - \mu - J_t \xi_t^V \right) (V_t - V_{t-1} - \alpha - \beta V_{t-1} - J_t \xi_t^V)^2
\]

\[
\frac{1}{2} \sum \frac{1}{V_{t-1}} (Y_t - \mu - J_t \xi_t^V)^2 + 2
\]

and a Normal $(Z,X)$ posterior for $\phi$ with mean
APPENDIX C. APPENDIX FOR CHAPTER 5

\[ Z = \frac{\sum \frac{1}{V_{t-1}} (Y_t - \mu - J_t \xi_t^Y)(V_t - V_{t-1} - \alpha - \beta V_{t-1} - J_t \xi_t^V)}{\sum \frac{1}{V_{t-1}} (Y_t - \mu - J_t \xi_t^Y)^2 + 2} \]

and variance

\[ X = \frac{\omega}{\sum \frac{1}{V_{t-1}} (Y_t - \mu - J_t \xi_t^Y)^2 + 2} \]

For completeness, the non-conjugate posteriors for \( \rho, \sigma^2 \) are noted.

\[ p(\rho | V_t, ...) \propto \left( \frac{1}{\sqrt{1 - \rho^2}} \right)^T \exp \left[ -\frac{1}{2(1 - \rho^2)} \left( \frac{A^2}{\sigma_V^2 V_{t-1}} + \frac{B^2}{\sigma^2 V_{t-1}} - \frac{2\rho A_t B_t}{\sigma_V V_{t-1}} \right) \right] \]

\[ p(\sigma^2 | V_t, ...) = \left( \frac{1}{\sigma_V^2} \right)^{T/2} \exp \left[ -\frac{1}{\sigma_V^2} \left( \sum \frac{B_i^2}{2(1 - \rho^2)} - \frac{2\rho \sigma_V A_t B_t}{\sigma_V V_{t-1}} \right) \right] \ast p(\sigma_V^2) \]

This is almost inverse Gamma with parameters \( \alpha = T/2 - 1 \) and \( \beta = \) the sum term, but the \( \sigma_V \) in \( 2\rho A_t B_t \sigma_V \) negates conjugacy. Jeffrey's prior or Inverse Gamma priors such as IG(c,C) have no effect. Therefore Independence or Metropolis - Hastings sampling must be used for a suitable proposal. That would be a Uniform or an Inverse Gamma if the problematic \( \sigma \) was ignored, causing the terms to be

\[ IG(c + T/2, C + \frac{1}{2} \sum (V_t - V_{t-1} - \alpha - \beta V_{t-1} - J_t \xi_t^V)^2) \]

This, however, does not work because the posterior proves to be extremely peaked so a proposal with a very calibrated step needs to be used. Raggi (2005) proposes ARMS while Ashgarian and Bengtsson (2006) the approach above.

\[ \left( \frac{1}{\sigma_V^2} \right)^{T/2} \exp \left[ \frac{1}{2} \sum B_i^2 \right] \ast \left( \frac{1}{\sigma^2} \right)^{c+1} \exp \left[ -\frac{C}{\sigma^2} \right] \]

\[ \left( \frac{1}{\sigma_V^2} \right)^{T/2+c+1} \exp \left[ \frac{1}{2} \sum B_i^2 + C \right] \]

The posterior for \( J_t \) produces a vector and is one not encountered before. The prior is a Bernoulli \( Ber(\lambda) \) and the two possible states of \( J \) are 0 and 1. Therefore,

\[ P(1) = P(J = 1|Y_t, V_t, ...) = \lambda \ast p(V_t, Y_t | J = 1, ...) \]

136
C.3. SVCJ POSTERIORS

\[ P(0) = P(J = 0|Y_t, V_t, \ldots) = (1 - \lambda) * p(V_t, Y_t|J = 0, \ldots) \]

and the resulting posterior is \( p(J_t|V_t, Y_t, \ldots) \propto Ber(q), \) where \( q = \frac{P(1)}{P(1) + P(0)} \)

The remaining posteriors are mostly Normal or Inverse Gamma and have been discussed extensively, so only their final parameters will be mentioned. With lax notation of \( X, Z \) corresponding to \( N(\text{mean}, \text{variance}) \) and \( \text{IG}(\text{shape}, \text{scale}) \)

\[ p(\mu|V_t, \ldots) \sim N(X, Z) \]

with prior \( N(k = 2, K = 40) \) for mean and variance, and

\[ Z = \left( \sum \frac{1}{(1 - \rho)^2 V_{t-1} + \frac{1}{K}} \right)^{-1} \]

and

\[ X = \left( \sum Y_t - J_t \xi_t \gamma - \frac{\mu}{\sigma_V}(V_{t-1} + \alpha + \beta V_{t-1} + J_t \xi_t \gamma) \right) (1 - \rho)^2 V_{t-1} + K \]

For \( \alpha \) and \( \beta \), Eraker et al. (2003) and Asgharian and Bengtsson (2006) suggest joint sampling from a bivariate normal distribution due to high correlation. However, it is straightforward to derive separate conditional posteriors. Both of them are conjugate normals, truncated in \((-\infty, 0]\) for \( \beta \) and \((0, \infty]\) for \( \alpha \). The result proves to be identical in practice and it is easier to set up individual samplers. In the case of joint sampling, the posterior is \( N(b^*, B^*) \) where

\[ b^* = (B^{-1} b + \frac{1}{(1 - \rho^2) \sigma_V^2} W' Q) * B \]

\[ B^* = (B^{-1} + \frac{1}{(1 - \rho^2) \sigma_V^2} W' W)^{-1} \]

\( b \) is a vector of ones and \( B \) is the identity matrix, both of appropriate dimension, and are the hyper-parameters of the prior. The expressions for \( Q \) and \( W \) are

\[ Q = \begin{bmatrix} \frac{V_1 - V_0 - J_1 \xi_1 \gamma}{\sqrt{\sigma_V}} \\ \frac{V_2 - V_1 - J_1 \xi_1 \gamma}{\sqrt{\sigma_V}} \\ \cdots \\ \frac{V_T - V_{T-1} - J_{T-1} \xi_{T-1} \gamma}{\sqrt{\sigma_V}} \end{bmatrix} \]

... until \( T \)
APPENDIX C. APPENDIX FOR CHAPTER 5

\[
W = \begin{bmatrix}
\frac{1}{\sqrt{V_0}} & \sqrt{V_0} \\
\frac{1}{\sqrt{V_1}} & \sqrt{V_1} \\
\vdots & \vdots \\
1 & \sqrt{V_T} \\
\end{bmatrix}
\]

The posteriors used in the thesis are the separate ones. For \( \alpha, p(\alpha|V_t, ... \propto (N, Z) \) where

\[
X = \left( \sum \frac{V_t - (1 + \beta)V_{t-1} - J_t\xi_t^V - \rho \sigma_Y(Y_t - \mu - J_t\xi_t^V)}{\sigma_Y^2(1 - \rho^2)} \right) * Z
\]

and

\[
Z = \sum \frac{1}{\sigma_Y^2(1 - \rho^2)V_{t-1}}
\]

For \( \beta, p(\beta|V_t, ... \propto (N, Z) \) where

\[
X = \left( \sum \frac{V_t - V_{t-1} - \alpha - J_t\xi_t^V - \rho \sigma_Y(Y_t - \mu - J_t\xi_t^V)}{V_{t-1} \sigma_Y^2(1 - \rho^2)} \right) * Z
\]

and

\[
Z = \frac{\sum V_{t-1}}{\sigma_Y^2(1 - \rho^2)} + 1
\]

For \( \lambda \) the prior is a Beta \((k, K), \) so \( \lambda \sim Beta(X, Z) \) with parameters

\[
X^* = k + \sum J_t \\
Z^* = K + T - \sum J_t
\]

For \( \sigma_Y^2, p(\sigma_Y^2|V_t, ... \propto p(\xi_t^Y|...)p(\sigma_Y^2) \)

\[
\propto IG\left(\frac{1}{2}T + \epsilon, \frac{1}{2} \sum (\xi_t^Y - \rho J_t\xi_t^V - \mu_Y)^2 + E\right)
\]

with \( IG(e=10, E=40) \) as prior.

For \( \mu_Y, p(\mu_Y|V_t, ...) \propto p(\xi_t^Y|...)p(\mu_Y) \)

with prior \( N(z=0, 100) \), which yields Normal distribution \( N(X, Z) \) with variance

\[
Z = \left( \frac{T}{\sigma_Y^2} + \frac{1}{100} \right)^{-1}
\]
and

\[ X = \left( \frac{\sum (\xi_t^Y - \rho J \xi_t^V)}{\sigma_Y^2} + \frac{z}{100} \right) \ast Z \]

For \( \mu_V \), the pdf of an exponential distribution with mean \( \mu_V \) is

\[ \frac{1}{\mu_V} \text{Exp} \left[ -\frac{\xi_t^V}{\mu_V} \right] \]

since the general form is \( \lambda \text{Exp} \left[ -\lambda x \right] \) and mean \( \lambda^{-1} \)

With an Inverse Gamma (d=10, D=20) as prior and ignoring constants, \( p(\mu_V|...) \propto p(\xi_t^V)p(\mu_V) \)

\[ \propto \left( \frac{1}{\mu_V} \right)^T \text{Exp} \left[ -\sum \frac{\xi_t^V}{\mu_V} \right] (\mu_V)^{-d-1} \text{Exp} \left[ -\frac{D}{\mu_V} \right] \]

\[ \propto IG(T + d, \sum \xi_t^V + D) \]

For \( \xi_t^Y \), the posterior is

\[ p(\xi_t^Y | \Delta V_t, Y, J = 1, ...) \propto p(Y_t, \Delta V_t | \xi_t^Y, J = 1, ...)p(\xi_t^Y) \text{ which leads to a Normal } (X, Z) \]

with

\[ Z = \left( \frac{1}{(1 - \rho^2)V_{t-1}} + \frac{1}{\sigma_Y^2} \right)^{-1} \]

and

\[ X = \left( \frac{Y_t - \mu - \rho B_t}{(1 - \rho^2)V_{t-1}} + \frac{\mu_Y - \rho J \xi_t^V}{\sigma_Y^2} \right) \ast Z \]

When \( J=0 \), the drawing of \( \xi_t^Y \) comes from the unconditional distribution \( \xi_t^Y \sim N(\mu_Y + \rho J \xi_t^V, \sigma_Y^2) \)

For \( \xi_t^V \), the posterior is again standard.

\[ p(\xi_t^V | \Delta V_t, Y_T, J = 1...) \propto p(Y_t, \Delta V_t | \xi_t^V, J = 1)p(\xi_t^V | \xi_t^Y, ...)p(\xi_t^V) \text{. Ignoring the constants, this can be written as} \]

\[ \propto \text{Exp} \left[ -\frac{1}{2(1 - \rho^2)} \left( \frac{\sigma_Y^2 A_t^2 + B_t^2 - 2\rho \sigma_Y A_t B_t}{\sigma_Y V_{t-1}} \right) \right] \text{Exp} \left[ -\frac{(\xi_t^Y - \mu_Y - \rho J \xi_t^V)^2}{2\sigma_Y^2} \right] \text{Exp} \left[ -\frac{\xi_t^V}{\mu_V} \right] \]
leading to \( N(X, Z) \) where

\[
Z = \left( \frac{1}{\sigma_V^2(1 - \rho^2)V_{t-1}} + \frac{\rho_J^2}{\sigma_Y^2} \right)^{-1}
\]

and

\[
X = \left( \frac{(V_t - V_{t-1} - \alpha - \beta V_{t-1}) - \rho \sigma_V(Y_t - \mu - \xi_Y)}{\sigma_V^2(1 - \rho^2)V_{t-1}} + \frac{\rho_J(\xi_Y - \mu_Y)}{\sigma_Y^2} - \frac{1}{\mu_V} \right) Z
\]

When \( J=0 \), the drawing of \( \xi^V \sim \exp(\mu_v) \) which is again the unconditional distribution.

For \( \rho_J \) the posterior is a Normal distribution with and the prior is \( N(0,4) \)

\[
p(\rho_J | \xi^V_t, ...) \propto p(\xi^V_t | \xi^V_t, ...)p(\rho_J)
\]

which yields \( N(X,Z) \)

\[
Z = \left( \frac{\sum \xi_{V,t}^2}{\sigma_Y^2} + \frac{1}{4} \right)^{-1}
\]

and

\[
X = \frac{\sum \xi_t^V(\xi_Y - \mu_Y)}{\sigma_Y^2} Z
\]

### C.4 SV posteriors

The easiest way to get the posteriors for the SV model is to take the SVCJ expressions and set the missing parameters equal to 0. This yields exactly the same result as conditioning from the beginning. As a word of caution, this holds only for the model at hand and should not be used in general. In order to verify, the posteriors were properly derived and then compared to the SVCJ formulas when setting the missing parameters equal to 0. They are much simpler than their SVCJ counterparts so to save space this explanation is sufficient.
### C.5 Figures and tables

Table C.1: MCMC parameters for the SVCJ model. Values reported as daily percentages and annual decimals compared to the values of Eraker et al (2003) and Brooks and Prokopszuk (2012). Where necessary, the parameters are converted as described in Chapter 4.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda )</td>
<td>0.0055</td>
<td>1.3919</td>
<td>0.0174</td>
<td>0.2091</td>
<td>0.0053</td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0994)</td>
<td>(0.0029)</td>
<td>(0.0012)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>( \rho_j )</td>
<td>0.0039</td>
<td>0.0030</td>
<td>-0.0009</td>
<td>-0.0009</td>
<td>-0.0014</td>
</tr>
<tr>
<td></td>
<td>(0.0329)</td>
<td>(0.0674)</td>
<td>(0.0578)</td>
<td>(0.0018)</td>
<td>(0.0012)</td>
</tr>
<tr>
<td>( \sigma_Y )</td>
<td>2.6554</td>
<td>0.0266</td>
<td>2.3308</td>
<td>0.0233</td>
<td>4.4246</td>
</tr>
<tr>
<td></td>
<td>(0.3902)</td>
<td>(0.4342)</td>
<td>(1.3707)</td>
<td>(0.5679)</td>
<td>(5.251)</td>
</tr>
<tr>
<td>( \mu_V )</td>
<td>1.0808</td>
<td>0.0254</td>
<td>1.0088</td>
<td>0.0012</td>
<td>1.0555</td>
</tr>
<tr>
<td></td>
<td>(0.1265)</td>
<td>(0.1534)</td>
<td>(0.3722)</td>
<td>(0.3404)</td>
<td>(0.3404)</td>
</tr>
<tr>
<td>( \mu_V )</td>
<td>-2.9251</td>
<td>-0.0293</td>
<td>-10.9183</td>
<td>-1.092</td>
<td>-1.9029</td>
</tr>
<tr>
<td></td>
<td>(0.6197)</td>
<td>(2.1911)</td>
<td>(1.0701)</td>
<td>(1.5566)</td>
<td>(0.9100)</td>
</tr>
<tr>
<td>( \mu )</td>
<td>0.0340</td>
<td>0.0858</td>
<td>0.617</td>
<td>0.074</td>
<td>0.0396</td>
</tr>
<tr>
<td></td>
<td>(0.0081)</td>
<td>(0.0868)</td>
<td>(0.0248)</td>
<td>(0.0112)</td>
<td>(0.0112)</td>
</tr>
<tr>
<td>( \rho )</td>
<td>-0.6757</td>
<td>-0.6757</td>
<td>-0.3739</td>
<td>-0.3739</td>
<td>-0.6875</td>
</tr>
<tr>
<td></td>
<td>(0.0259)</td>
<td>(0.0259)</td>
<td>(0.0504)</td>
<td>(0.0623)</td>
<td>(0.0623)</td>
</tr>
<tr>
<td>( \sigma_V )</td>
<td>0.1429</td>
<td>0.3601</td>
<td>0.6587</td>
<td>0.0791</td>
<td>0.2895</td>
</tr>
<tr>
<td></td>
<td>(0.0055)</td>
<td>(0.0635)</td>
<td>(0.0385)</td>
<td>(0.0074)</td>
<td>(0.0074)</td>
</tr>
<tr>
<td>( \alpha (= \kappa \theta) )</td>
<td>0.0211</td>
<td>0.1337</td>
<td>0.6178</td>
<td>0.0089</td>
<td>0.0517</td>
</tr>
<tr>
<td></td>
<td>(0.0018)</td>
<td>(0.1279)</td>
<td>(0.0065)</td>
<td>(0.0065)</td>
<td>(0.0065)</td>
</tr>
<tr>
<td>( \kappa (- \beta) )</td>
<td>0.0252</td>
<td>6.3561</td>
<td>0.05</td>
<td>-0.6</td>
<td>0.0380</td>
</tr>
<tr>
<td></td>
<td>(0.0211)</td>
<td>(0.0105)</td>
<td>(0.0061)</td>
<td>(0.0041)</td>
<td>(0.0041)</td>
</tr>
<tr>
<td>( \theta )</td>
<td>0.8347</td>
<td>0.0210</td>
<td>12.3554</td>
<td>0.0148</td>
<td>1.3598</td>
</tr>
<tr>
<td></td>
<td>(0.0539)</td>
<td>(0.0539)</td>
<td>(0.0539)</td>
<td>(0.0539)</td>
<td>(0.0539)</td>
</tr>
</tbody>
</table>
Table C.2: MCMC parameters for the SV model. Values reported as daily percentages and annual decimals compared to the values of Eraker et al (2003) Where necessary, the parameters are converted as described in Chapter 4.

<table>
<thead>
<tr>
<th></th>
<th>1980-2016</th>
<th>1872-2016</th>
<th>EJP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Daily</td>
<td>Annualised</td>
<td>Monthly</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.0289</td>
<td>0.0728</td>
<td>0.5505</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td></td>
<td>0.0645</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.6096</td>
<td>-0.6096</td>
<td>-0.203</td>
</tr>
<tr>
<td></td>
<td>(0.0244)</td>
<td></td>
<td>0.04847</td>
</tr>
<tr>
<td>$\sigma_V$</td>
<td>0.1691</td>
<td>0.42617</td>
<td>0.7663</td>
</tr>
<tr>
<td></td>
<td>(0.00857)</td>
<td></td>
<td>0.0199</td>
</tr>
<tr>
<td>$\alpha (= \kappa \theta)$</td>
<td>0.0265</td>
<td>0.0273</td>
<td>1.2924</td>
</tr>
<tr>
<td></td>
<td>(0.0018)</td>
<td></td>
<td>0.1078</td>
</tr>
<tr>
<td>$\kappa (= -\beta)$</td>
<td>0.0245</td>
<td>6.1662</td>
<td>0.1471</td>
</tr>
<tr>
<td></td>
<td>(0.00264)</td>
<td></td>
<td>0.0144</td>
</tr>
<tr>
<td>$\theta$</td>
<td>1.0815</td>
<td>0.0273</td>
<td>8.7887</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table C.3: Sample Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$Mean$</td>
<td>0.000322652</td>
<td>0.000149161</td>
<td>0.36</td>
</tr>
<tr>
<td>$SD$</td>
<td>0.0112694</td>
<td>0.0137424</td>
<td>4.1</td>
</tr>
<tr>
<td>$Length$</td>
<td>9132</td>
<td>2253</td>
<td>1745</td>
</tr>
</tbody>
</table>
Table C.4: Optimal portfolio weights for the SVCJ and SV models (1980 - 2016) for $r = 2\%$ compared to the weights corresponding to the EJP parameters ($r = 4.5\%$) and the Liu, Longstaff and Pan (2003) replicated parameters. The LLP weights refer to S&P500 options data between 1-1-1987 and 31-12-1996, for which the methodology, frequency and parameter estimation is incomparable. They are referred here only as a successful replication of an existing result.

<table>
<thead>
<tr>
<th></th>
<th>1980-2016</th>
<th>EJP</th>
<th>LLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>SVCJ</td>
<td>SV</td>
<td>SVCJ</td>
</tr>
<tr>
<td>5</td>
<td>0.211</td>
<td>0.41</td>
<td>0.57</td>
</tr>
<tr>
<td>4</td>
<td>0.2634</td>
<td>0.52</td>
<td>0.7119</td>
</tr>
<tr>
<td>3</td>
<td>0.3505</td>
<td>0.686</td>
<td>0.9459</td>
</tr>
<tr>
<td>2</td>
<td>0.5235</td>
<td>1.022</td>
<td>1.4096</td>
</tr>
</tbody>
</table>

Table C.5: Replication of existing results in the literature. Branger and Hansis (2012, 2015) transform the EJP parameters from percentage log returns to annual decimals that correspond to the LLP formulation that includes volatility and jumps premia in the returns process and stochastic arrival intensity. $\lambda_{EJP} = \lambda_{LLP}\bar{V}$, $\bar{V} = \theta + \mu V \lambda/\kappa$, ERP = Equity Risk Premium. Common parameters are in Table 1 and only the additional parameters of the LLP version are reported here.

<table>
<thead>
<tr>
<th></th>
<th>$\kappa'$</th>
<th>$\theta'$</th>
<th>$\lambda_{LLP}$</th>
<th>$\lambda_{EJP}$</th>
<th>ERP</th>
<th>$\bar{V}$</th>
<th>ERP/$\bar{V}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVCJ</td>
<td>6.552</td>
<td>0.0135</td>
<td>72.2018</td>
<td>1.6632</td>
<td>0.0732 (6%)</td>
<td>0.0154 (0.023)</td>
<td>4.7532</td>
</tr>
<tr>
<td>SV</td>
<td>5.8212</td>
<td>0.0228</td>
<td></td>
<td></td>
<td>0.0783</td>
<td>0.0228</td>
<td>3.4342</td>
</tr>
</tbody>
</table>
Figure C.1. SVCJ Portfolio weight plot for $\gamma = 5$, $T = 5$ and $T = 9.000$
Figure C.2. MCMC Chains - (Left, top to bottom) $\lambda, \rho_j, \sigma^2_Y, \mu_V, \mu_Y$, (Right, top to bottom) $\mu, \rho, \sigma^2_V, \alpha, \beta$
Appendix D

Appendix for Chapter 6

D.1 Tables 1.1 - 1.4

Investor simulations for $\gamma = 2, 3, 4, 5$. **NJW**: terminal Wealth for the No Jumps investor **NJU**: terminal Utility for the No Jumps investor **WDiff**: difference in average terminal wealths **UDiff**: difference in average terminal utilities

Table D.1.1: $\gamma=5$

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>NJW Wins</td>
<td>61.74%</td>
<td>62.04%</td>
<td>64.06%</td>
<td>67.62%</td>
<td>71.86%</td>
<td>76.04%</td>
<td>79.84%</td>
</tr>
<tr>
<td>NJU Wins</td>
<td>61.74%</td>
<td>62.04%</td>
<td>64.06%</td>
<td>67.62%</td>
<td>71.86%</td>
<td>76.04%</td>
<td>79.84%</td>
</tr>
<tr>
<td>WDiff (NJ - J)</td>
<td>0.02336</td>
<td>0.03261</td>
<td>0.05820</td>
<td>0.13719</td>
<td>0.24676</td>
<td>0.51911</td>
<td>0.82088</td>
</tr>
<tr>
<td>UDiff (NJ - J)</td>
<td>-0.00845</td>
<td>-0.01474</td>
<td>-0.02382</td>
<td>-0.03262</td>
<td>-0.03430</td>
<td>-0.02898</td>
<td>-0.02189</td>
</tr>
</tbody>
</table>

Table D.1.2: $\gamma=4$

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>NJW Wins</td>
<td>60.96%</td>
<td>62.18%</td>
<td>66.26%</td>
<td>69.54%</td>
<td>71.8%</td>
<td>75.52%</td>
<td>77.82%</td>
</tr>
<tr>
<td>NJU Wins</td>
<td>60.96%</td>
<td>62.18%</td>
<td>66.26%</td>
<td>69.54%</td>
<td>71.8%</td>
<td>75.52%</td>
<td>77.82%</td>
</tr>
<tr>
<td>WDiff (NJ - J)</td>
<td>0.02888</td>
<td>0.04526</td>
<td>0.08565</td>
<td>0.19736</td>
<td>0.34603</td>
<td>0.74951</td>
<td>1.13564</td>
</tr>
<tr>
<td>UDiff (NJ - J)</td>
<td>-0.01094</td>
<td>-0.01823</td>
<td>-0.01690</td>
<td>-0.03164</td>
<td>-0.04352</td>
<td>-0.03739</td>
<td>-0.02718</td>
</tr>
</tbody>
</table>

147
### Table D.1.3: $\gamma=3$

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>NJW Wins</td>
<td>62.12%</td>
<td>62.38%</td>
<td>65.28%</td>
<td>70.38%</td>
<td>71.96%</td>
<td>75.84%</td>
<td>79%</td>
</tr>
<tr>
<td>NJU Wins</td>
<td>62.12%</td>
<td>62.38%</td>
<td>65.28%</td>
<td>70.38%</td>
<td>71.96%</td>
<td>75.84%</td>
<td>79%</td>
</tr>
<tr>
<td>WDiff (NJ - J)</td>
<td>0.04603</td>
<td>0.06728</td>
<td>0.12651</td>
<td>0.31298</td>
<td>0.55600</td>
<td>1.2535</td>
<td>2.05842</td>
</tr>
<tr>
<td>UDiff (NJ - J)</td>
<td>-0.00895</td>
<td>-0.01798</td>
<td>-0.02238</td>
<td>-0.03723</td>
<td>-0.05784</td>
<td>-0.05601</td>
<td>-0.08726</td>
</tr>
</tbody>
</table>

### Table D.1.4: $\gamma=2$

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>NJW Wins</td>
<td>60.2%</td>
<td>62.88%</td>
<td>64.62%</td>
<td>69.1%</td>
<td>72.58%</td>
<td>75.5%</td>
<td>77.76%</td>
</tr>
<tr>
<td>NJU Wins</td>
<td>60.2%</td>
<td>62.88%</td>
<td>64.62%</td>
<td>69.1%</td>
<td>72.58%</td>
<td>75.5%</td>
<td>77.76%</td>
</tr>
<tr>
<td>WDiff (NJ - J)</td>
<td>0.07282</td>
<td>0.12097</td>
<td>0.22944</td>
<td>0.63438</td>
<td>1.20383</td>
<td>3.12372</td>
<td>5.57252</td>
</tr>
<tr>
<td>UDiff (NJ - J)</td>
<td>-0.01033</td>
<td>-0.00729</td>
<td>-0.00731</td>
<td>-0.01200</td>
<td>-0.01747</td>
<td>-0.02143</td>
<td>-0.03203</td>
</tr>
</tbody>
</table>
D.2 Tables 2.1 - 2.4

Investor and Manager simulations for $\gamma = 2, 3, 4, 5$. **TFJ**: Total Fees of Manager (Jumps case) **TFNJ**: Total Fees of Manager (No Jumps case) **TUJ**: Average Investor terminal Utility (Jumps case) **TUNJ**: Average Investor terminal Utility (No Jumps case) **TFUJ**: Average Total Annual Utility of Manager from Fees (Jumps case) **TFUNJ**: Average Total Annual Utility of Manager from Fees (No Jumps case). Blue denotes the winning manager.

### Table D.2.1: $\gamma=5$, 2+20% fees, Symmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFJ</td>
<td>4.1134</td>
<td>6.1893</td>
<td>10.278</td>
<td>20.483</td>
<td>30.7218</td>
<td>49.204</td>
<td>61.5978</td>
</tr>
<tr>
<td>TFNJ</td>
<td>5.12877</td>
<td>7.93911</td>
<td>13.1283</td>
<td>26.991</td>
<td>41.9317</td>
<td>71.1001</td>
<td>91.9276</td>
</tr>
<tr>
<td>TUJ ($\times 10^{-9}$)</td>
<td>-2.4962</td>
<td>-2.49619</td>
<td>-2.55146</td>
<td>-2.62194</td>
<td>-2.64366</td>
<td>-2.68481</td>
<td>-78.9905</td>
</tr>
<tr>
<td>TUNJ ($\times 10^{-9}$)</td>
<td>-2.93103</td>
<td>-3.0467</td>
<td>-4.27311</td>
<td>-6.71655</td>
<td>-10.0262</td>
<td>-15.5008</td>
<td>-351.281</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.287002</td>
<td>-0.0430744</td>
<td>-0.07235</td>
<td>-0.146331</td>
<td>-0.204301</td>
<td>-0.362212</td>
<td>-0.453511</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.0280683</td>
<td>-0.0429121</td>
<td>-0.0838553</td>
<td>-0.216396</td>
<td>-0.351235</td>
<td>-0.956448</td>
<td>-1.74013</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>35.96</td>
<td>33.68</td>
<td>33.42</td>
<td>35.54</td>
<td>37.28</td>
<td>39.54</td>
<td>39.6</td>
</tr>
</tbody>
</table>

### Table D.2.2: $\gamma=5$, 1+10%, Symmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFJ</td>
<td>2.07236</td>
<td>3.1236</td>
<td>5.25193</td>
<td>10.7397</td>
<td>16.5157</td>
<td>27.7563</td>
<td>35.8835</td>
</tr>
<tr>
<td>TFNJ</td>
<td>2.65585</td>
<td>4.03675</td>
<td>6.92637</td>
<td>14.6369</td>
<td>23.4346</td>
<td>41.8983</td>
<td>56.3048</td>
</tr>
<tr>
<td>TUJ ($\times 10^{-9}$)</td>
<td>-4.67885</td>
<td>-6.92864</td>
<td>-11.1408</td>
<td>-20.5021</td>
<td>-28.2178</td>
<td>-38.8733</td>
<td>-44.2656</td>
</tr>
<tr>
<td>TUNJ ($\times 10^{-9}$)</td>
<td>-5.07058</td>
<td>-7.8992</td>
<td>-14.2545</td>
<td>-32.729</td>
<td>-58.7133</td>
<td>-121.347</td>
<td>-165.869</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.447582</td>
<td>-0.662913</td>
<td>-1.06606</td>
<td>-1.96226</td>
<td>-2.70095</td>
<td>-3.71829</td>
<td>-4.23515</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.42267</td>
<td>-0.659745</td>
<td>-1.18098</td>
<td>-2.71481</td>
<td>-4.3265</td>
<td>-10.0045</td>
<td>-13.819</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>34.56</td>
<td>33.92</td>
<td>34.08</td>
<td>36.18</td>
<td>35.72</td>
<td>36.88</td>
<td>38.2</td>
</tr>
</tbody>
</table>
## APPENDIX D. APPENDIX FOR CHAPTER 6

Table D.2.3: $\gamma=5$, 2+20, Asymmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFJ</td>
<td>4.14593</td>
<td>6.25992</td>
<td>10.5541</td>
<td>21.7746</td>
<td>33.6008</td>
<td>56.2138</td>
<td>72.4294</td>
</tr>
<tr>
<td>TFNJ</td>
<td>5.14174</td>
<td>7.66955</td>
<td>12.7947</td>
<td>25.6348</td>
<td>38.5557</td>
<td>61.7916</td>
<td>78.3366</td>
</tr>
<tr>
<td>TUNJ ($\times 10^{-9}$)</td>
<td>-5.9979</td>
<td>-10.159</td>
<td>-22.4123</td>
<td>-96.3665</td>
<td>-239.355</td>
<td>-1199.59</td>
<td>-2492.14</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.0279114</td>
<td>-0.0409538</td>
<td>-0.0654901</td>
<td>-0.117992</td>
<td>-0.161075</td>
<td>-0.224718</td>
<td>-0.258033</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.0299182</td>
<td>-0.050489</td>
<td>-0.111401</td>
<td>-0.484775</td>
<td>-1.19765</td>
<td>-5.88703</td>
<td>-12.0652</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>37.36</td>
<td>38.62</td>
<td>45.72</td>
<td>56.3</td>
<td>62.04</td>
<td>71.38</td>
<td>75.44</td>
</tr>
</tbody>
</table>

Table D.2.4: $\gamma=5$, 1+10%, Asymmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFJ</td>
<td>2.08434</td>
<td>3.16859</td>
<td>5.39279</td>
<td>11.4248</td>
<td>18.1162</td>
<td>32.0474</td>
<td>42.6871</td>
</tr>
<tr>
<td>TFNJ</td>
<td>2.62285</td>
<td>3.98125</td>
<td>6.68629</td>
<td>13.8526</td>
<td>21.5614</td>
<td>36.4705</td>
<td>47.2148</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.438135</td>
<td>-0.629298</td>
<td>-0.969264</td>
<td>-1.60384</td>
<td>-2.02792</td>
<td>-2.47643</td>
<td>-2.66441</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.465278</td>
<td>-0.761525</td>
<td>-1.59883</td>
<td>-5.1783</td>
<td>-14.5959</td>
<td>-64.1319</td>
<td>-57.3833</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>37.1</td>
<td>38.76</td>
<td>45.46</td>
<td>54.6</td>
<td>59.78</td>
<td>66.04</td>
<td>72.16</td>
</tr>
</tbody>
</table>
D.3 Table 3.1 - 3.4

Table D.3.1: $\gamma=4$, 2+20%, Symmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFNJ</td>
<td>5.56988</td>
<td>8.59343</td>
<td>14.6642</td>
<td>30.3255</td>
<td>47.1842</td>
<td>79.7479</td>
<td>104.394</td>
</tr>
<tr>
<td>TUJ ($\times 10^{-7}$)</td>
<td>-6.70446</td>
<td>-10.0977</td>
<td>-16.989</td>
<td>-35.1197</td>
<td>-54.0623</td>
<td>-90.8233</td>
<td>-114.948</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.0775296</td>
<td>-0.11669</td>
<td>-0.194751</td>
<td>-0.518134</td>
<td>-0.937875</td>
<td>-2.19226</td>
<td>-4.0686</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.0727639</td>
<td>-0.111609</td>
<td>-0.196224</td>
<td>-0.518134</td>
<td>-0.937875</td>
<td>-2.19226</td>
<td>-4.0686</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>35.36</td>
<td>31.56</td>
<td>30.42</td>
<td>33.62</td>
<td>34.46</td>
<td>35.88</td>
<td>36.22</td>
</tr>
</tbody>
</table>

Table D.3.2: $\gamma=4$, 1+10%, Symmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFJ</td>
<td>2.11209</td>
<td>3.18399</td>
<td>5.33871</td>
<td>10.8649</td>
<td>16.6332</td>
<td>27.8655</td>
<td>35.8588</td>
</tr>
<tr>
<td>TFNJ</td>
<td>2.91586</td>
<td>4.47321</td>
<td>7.69121</td>
<td>16.5574</td>
<td>26.4013</td>
<td>48.9598</td>
<td>65.7202</td>
</tr>
<tr>
<td>TUJ ($\times 10^{-7}$)</td>
<td>-6.38149</td>
<td>-9.48521</td>
<td>-15.5041</td>
<td>-29.63</td>
<td>-42.2345</td>
<td>-60.9076</td>
<td>-72.0084</td>
</tr>
<tr>
<td>TUNJ ($\times 10^{-7}$)</td>
<td>-6.94923</td>
<td>-10.7075</td>
<td>-18.8891</td>
<td>-42.2087</td>
<td>-71.793</td>
<td>-120</td>
<td>-187.89</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.608113</td>
<td>-0.902803</td>
<td>-1.47662</td>
<td>-2.82229</td>
<td>-4.02278</td>
<td>-5.79576</td>
<td>-6.85519</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.553296</td>
<td>-0.851483</td>
<td>-1.49367</td>
<td>-3.33377</td>
<td>-5.66446</td>
<td>-9.51762</td>
<td>-14.7887</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>33.66</td>
<td>31.34</td>
<td>30.64</td>
<td>32.94</td>
<td>34.02</td>
<td>32.84</td>
<td>34.44</td>
</tr>
</tbody>
</table>
### Table D.3.3: $\gamma=4$, 2+20%, Asymmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFJ</td>
<td>4.22939</td>
<td>6.36439</td>
<td>10.7598</td>
<td>22.1103</td>
<td>33.9866</td>
<td>56.6911</td>
<td>72.8306</td>
</tr>
<tr>
<td>TFNJ</td>
<td>5.64367</td>
<td>8.42136</td>
<td>14.186</td>
<td>28.7358</td>
<td>43.1255</td>
<td>70.2579</td>
<td>88.4408</td>
</tr>
<tr>
<td>TUJ ($\times 10^{-7}$)</td>
<td>-6.53198</td>
<td>-9.68427</td>
<td>-15.6483</td>
<td>-29.3506</td>
<td>-41.2985</td>
<td>-60.4752</td>
<td>-71.8269</td>
</tr>
<tr>
<td>TUNJ ($\times 10^{-7}$)</td>
<td>-7.68938</td>
<td>-12.8496</td>
<td>-26.0689</td>
<td>-86.7467</td>
<td>-217.077</td>
<td>-701.23</td>
<td>-147.271</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.0754435</td>
<td>-0.111959</td>
<td>-0.180631</td>
<td>-0.338323</td>
<td>-0.476715</td>
<td>-0.697944</td>
<td>-0.82926</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.0737118</td>
<td>-0.123464</td>
<td>-0.249759</td>
<td>-0.830699</td>
<td>-2.08</td>
<td>-6.64889</td>
<td>-14.4653</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>35.5</td>
<td>37.4</td>
<td>43.04</td>
<td>52.62</td>
<td>58.42</td>
<td>67</td>
<td>70.92</td>
</tr>
</tbody>
</table>

### Table D.3.4: $\gamma=4$, 1+10%, Asymmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFJ</td>
<td>2.12956</td>
<td>3.23453</td>
<td>5.49658</td>
<td>11.6244</td>
<td>16.6495</td>
<td>32.4886</td>
<td>43.277</td>
</tr>
<tr>
<td>TFNJ</td>
<td>2.90262</td>
<td>4.37038</td>
<td>7.39661</td>
<td>15.461</td>
<td>26.4926</td>
<td>41.8518</td>
<td>54.8552</td>
</tr>
<tr>
<td>TUJ ($\times 10^{-7}$)</td>
<td>-6.24131</td>
<td>-9.07638</td>
<td>-14.3224</td>
<td>-24.9325</td>
<td>-42.171</td>
<td>-42.8639</td>
<td>-47.8045</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.595002</td>
<td>-0.864363</td>
<td>-1.36527</td>
<td>-2.37389</td>
<td>-4.01629</td>
<td>-4.08013</td>
<td>-4.55016</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.590602</td>
<td>-0.95773</td>
<td>-1.92217</td>
<td>-5.7049</td>
<td>-6.43929</td>
<td>-33.7565</td>
<td>-49.557</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>35.14</td>
<td>37.34</td>
<td>42.26</td>
<td>51.24</td>
<td>55.08</td>
<td>63.62</td>
<td>67.54</td>
</tr>
</tbody>
</table>
### D.4 Table 4.1 - 4.4

#### Table D.4.1: $\gamma$=3, 2+20, Symmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFNJ</td>
<td>6.42132</td>
<td>9.89045</td>
<td>16.7194</td>
<td>35.6563</td>
<td>56.2634</td>
<td>96.3445</td>
<td>126.559</td>
</tr>
<tr>
<td>TUJ ($\times 10^{-4}$)</td>
<td>-1.00753</td>
<td>-1.52654</td>
<td>-2.5972</td>
<td>-5.40292</td>
<td>-8.4649</td>
<td>-14.4265</td>
<td>-18.8818</td>
</tr>
<tr>
<td>TUNJ ($\times 10^{-4}$)</td>
<td>-1.08821</td>
<td>-1.69373</td>
<td>-3.07705</td>
<td>-7.72644</td>
<td>-12.9465</td>
<td>-26.4833</td>
<td>-53.9419</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.228987</td>
<td>-0.345256</td>
<td>-0.587302</td>
<td>-1.22274</td>
<td>-1.91448</td>
<td>-3.264</td>
<td>-4.26746</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.204424</td>
<td>-0.316268</td>
<td>-0.566061</td>
<td>-2.36253</td>
<td>-3.44285</td>
<td>-6.89118</td>
<td>-9.2923</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>33.12</td>
<td>29.52</td>
<td>30.22</td>
<td>31.76</td>
<td>31.98</td>
<td>31.76</td>
<td>32.8</td>
</tr>
</tbody>
</table>

#### Table D.4.2: $\gamma$=3, 1+10, Symmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFJ</td>
<td>2.25439</td>
<td>3.40396</td>
<td>5.68203</td>
<td>11.4715</td>
<td>17.4741</td>
<td>28.8977</td>
<td>37.208</td>
</tr>
<tr>
<td>TFNJ</td>
<td>3.37101</td>
<td>5.13221</td>
<td>8.8745</td>
<td>19.7625</td>
<td>32.4317</td>
<td>60.3262</td>
<td>83.1845</td>
</tr>
<tr>
<td>TUJ ($\times 10^{-4}$)</td>
<td>-0.974</td>
<td>-1.46171</td>
<td>-2.43214</td>
<td>-4.8063</td>
<td>-7.06255</td>
<td>-10.8066</td>
<td>-13.1058</td>
</tr>
<tr>
<td>TUNJ ($\times 10^{-4}$)</td>
<td>-1.0335</td>
<td>-1.61482</td>
<td>-2.84692</td>
<td>-9.93096</td>
<td>-21.4017</td>
<td>-17.3573</td>
<td>-25.516</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.900758</td>
<td>-1.3465</td>
<td>-2.24105</td>
<td>-4.42816</td>
<td>-6.50014</td>
<td>-9.96095</td>
<td>-12.0838</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.788421</td>
<td>-1.23233</td>
<td>-2.17239</td>
<td>-5.07936</td>
<td>-7.57461</td>
<td>-13.2926</td>
<td>-19.6939</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>32.24</td>
<td>29.98</td>
<td>30.24</td>
<td>31.74</td>
<td>34.02</td>
<td>30.54</td>
<td>30.34</td>
</tr>
</tbody>
</table>
APPENDIX D. APPENDIX FOR CHAPTER 6

Table D.4.3: $\gamma=3, 2+20$, Asymmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFJ</td>
<td>4.5004</td>
<td>6.7658</td>
<td>11.3664</td>
<td>23.2004</td>
<td>35.4803</td>
<td>58.4252</td>
<td>74.7063</td>
</tr>
<tr>
<td>TUJ ($\times 10^{-4}$)</td>
<td>-0.9935</td>
<td>-1.47956</td>
<td>-2.43891</td>
<td>-4.74168</td>
<td>-6.93773</td>
<td>-10.894</td>
<td>-13.4015</td>
</tr>
<tr>
<td>TUNJ ($\times 10^{-4}$)</td>
<td>-1.14786</td>
<td>-1.82251</td>
<td>-3.52386</td>
<td>-9.85826</td>
<td>-19.8829</td>
<td>-62.8707</td>
<td>-89.0543</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.224589</td>
<td>-0.334953</td>
<td>-0.551765</td>
<td>-1.07156</td>
<td>-1.56872</td>
<td>-2.46355</td>
<td>-3.03007</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.215742</td>
<td>-0.346441</td>
<td>-0.655305</td>
<td>-1.83772</td>
<td>-3.68977</td>
<td>-10.349</td>
<td>-15.1948</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>33.74</td>
<td>36.62</td>
<td>38.98</td>
<td>47.76</td>
<td>53.92</td>
<td>60.54</td>
<td>63.56</td>
</tr>
</tbody>
</table>

Table D.4.4: $\gamma=3, 1+10$, Asymmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFJ</td>
<td>2.28253</td>
<td>3.45927</td>
<td>5.88035</td>
<td>12.3192</td>
<td>19.4406</td>
<td>34.0733</td>
<td>45.5048</td>
</tr>
<tr>
<td>TFNJ</td>
<td>3.34766</td>
<td>5.05551</td>
<td>8.66882</td>
<td>18.4966</td>
<td>29.2669</td>
<td>51.7951</td>
<td>70.6413</td>
</tr>
<tr>
<td>TUJ ($\times 10^{-4}$)</td>
<td>-0.959283</td>
<td>-1.41777</td>
<td>-2.28526</td>
<td>-4.21887</td>
<td>-5.86146</td>
<td>-8.29972</td>
<td>-9.59786</td>
</tr>
<tr>
<td>TUNJ ($\times 10^{-4}$)</td>
<td>-1.0859</td>
<td>-1.79289</td>
<td>-3.30339</td>
<td>-9.2767</td>
<td>-15.2494</td>
<td>-32.072</td>
<td>-47.7016</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.884331</td>
<td>-1.30641</td>
<td>-2.10438</td>
<td>-3.88808</td>
<td>-5.40095</td>
<td>-7.65295</td>
<td>-8.84317</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.831653</td>
<td>-1.36319</td>
<td>-2.52766</td>
<td>-7.12084</td>
<td>-11.6281</td>
<td>-24.5588</td>
<td>-35.3786</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>32.58</td>
<td>35.82</td>
<td>38.94</td>
<td>45.58</td>
<td>50.76</td>
<td>57.38</td>
<td>60.26</td>
</tr>
</tbody>
</table>
### Table D.5.1: $\gamma=2, 2+20$, Symmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFJ</td>
<td>5.13</td>
<td>7.64079</td>
<td>12.5097</td>
<td>24.532</td>
<td>35.3789</td>
<td>55.2502</td>
<td>68.652</td>
</tr>
<tr>
<td>TFNJ</td>
<td>7.96</td>
<td>12.1517</td>
<td>20.7963</td>
<td>46.4978</td>
<td>71.0843</td>
<td>126.556</td>
<td>170.177</td>
</tr>
<tr>
<td>TUJ</td>
<td>-0.02018</td>
<td>-0.030532</td>
<td>-0.0518497</td>
<td>-0.107621</td>
<td>-0.169138</td>
<td>-0.286295</td>
<td>-0.368799</td>
</tr>
<tr>
<td>TUNJ</td>
<td>-0.02154</td>
<td>-0.0327491</td>
<td>-0.0567569</td>
<td>-0.13538</td>
<td>-0.196885</td>
<td>-0.350078</td>
<td>-0.543816</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.886234</td>
<td>-1.34045</td>
<td>-2.2789</td>
<td>-4.70585</td>
<td>-7.4273</td>
<td>-12.5276</td>
<td>-16.1619</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.825752</td>
<td>-1.24604</td>
<td>-2.15916</td>
<td>-4.67734</td>
<td>-7.664</td>
<td>-13.3141</td>
<td>-17.2303</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>32.2</td>
<td>29.28</td>
<td>31.2</td>
<td>32.5</td>
<td>32.44</td>
<td>33.02</td>
<td>32.44</td>
</tr>
</tbody>
</table>

### Table D.5.2: $\gamma=2, 1+10$, Symmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFJ</td>
<td>2.63216</td>
<td>3.93694</td>
<td>6.57586</td>
<td>13.0963</td>
<td>19.7262</td>
<td>33.0206</td>
<td>42.4538</td>
</tr>
<tr>
<td>TFNJ</td>
<td>4.14923</td>
<td>6.39352</td>
<td>11.328</td>
<td>25.2517</td>
<td>42.1472</td>
<td>83.0288</td>
<td>115.962</td>
</tr>
<tr>
<td>TUJ</td>
<td>-0.0198272</td>
<td>-0.0297701</td>
<td>-0.0498897</td>
<td>-0.10088</td>
<td>-0.152576</td>
<td>-0.241554</td>
<td>-0.302799</td>
</tr>
<tr>
<td>TUNJ</td>
<td>-0.0207063</td>
<td>-0.0317069</td>
<td>-0.0544408</td>
<td>-0.129646</td>
<td>-0.2346</td>
<td>-0.335995</td>
<td>-0.466738</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>31.7</td>
<td>29.92</td>
<td>28.4</td>
<td>31.58</td>
<td>32.02</td>
<td>30.68</td>
<td>32.14</td>
</tr>
</tbody>
</table>
## APPENDIX D. APPENDIX FOR CHAPTER 6

### Table D.5.3: $\gamma=2, 2+20$, Asymmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFJ</td>
<td>5.17882</td>
<td>7.73102</td>
<td>13.0149</td>
<td>26.2194</td>
<td>39.5972</td>
<td>64.8932</td>
<td>67.7002</td>
</tr>
<tr>
<td>TFNJ</td>
<td>7.9661</td>
<td>11.9224</td>
<td>20.3747</td>
<td>42.7863</td>
<td>66.4353</td>
<td>114.479</td>
<td>166.869</td>
</tr>
<tr>
<td>TUJ</td>
<td>-0.0200331</td>
<td>-0.0300701</td>
<td>-0.0502091</td>
<td>-0.100975</td>
<td>-0.151911</td>
<td>-0.247061</td>
<td>-0.376239</td>
</tr>
<tr>
<td>TUNJ</td>
<td>-0.0221937</td>
<td>-0.0341252</td>
<td>-0.065115</td>
<td>-0.155308</td>
<td>-0.276896</td>
<td>-0.568278</td>
<td>-0.634142</td>
</tr>
<tr>
<td>TFUJ</td>
<td>-0.879143</td>
<td>-1.32388</td>
<td>-2.20046</td>
<td>-4.41868</td>
<td>-6.65744</td>
<td>-10.8331</td>
<td>-16.51</td>
</tr>
<tr>
<td>TFUNJ</td>
<td>-0.849764</td>
<td>-1.30251</td>
<td>-1.91565</td>
<td>-5.63092</td>
<td>-9.54508</td>
<td>-19.4324</td>
<td>-22.0555</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>33.06</td>
<td>34.54</td>
<td>38.02</td>
<td>49.24</td>
<td>49.6</td>
<td>53.38</td>
<td>58.2</td>
</tr>
</tbody>
</table>

### Table D.5.4: $\gamma=2, 1+10$, Asymmetric wealth transfer function

<table>
<thead>
<tr>
<th>Years</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>24</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFNJ</td>
<td>4.06353</td>
<td>6.24395</td>
<td>10.9436</td>
<td>23.6876</td>
<td>39.7276</td>
<td>73.3717</td>
<td>97.8292</td>
</tr>
<tr>
<td>TUJ</td>
<td>-0.0197031</td>
<td>-0.0293917</td>
<td>-0.0483121</td>
<td>-0.09429</td>
<td>-0.137079</td>
<td>-0.209287</td>
<td>-0.258568</td>
</tr>
<tr>
<td>TUNJ</td>
<td>-0.0210771</td>
<td>-0.033172</td>
<td>-0.0552787</td>
<td>-0.157428</td>
<td>-0.237725</td>
<td>-0.428447</td>
<td>-0.604689</td>
</tr>
<tr>
<td>UJ wins (%)</td>
<td>34.12</td>
<td>34.94</td>
<td>37.36</td>
<td>43.14</td>
<td>46.28</td>
<td>50.74</td>
<td>55.76</td>
</tr>
</tbody>
</table>
Figure D.1. Symmetric wealth transfer function, cap at ±10%, δ = 0.25

Figure D.2. Asymmetric wealth transfer function, upper cap at 10% lower cap at −15%, δ = 0.25, τ = 0.5
Figure D.3. Sigmoid wealth transfer function, $\delta = 0.5$
Figure D.4. TFUJ histogram, $\gamma = 3, T = 24, 1 + 10\%$ fees, symmetric

Figure D.5. TFUNJ histogram, $\gamma = 3, T = 24, 1 + 10\%$ fees, symmetric
Bibliography

Optimal portfolio for an insider in a market driven by lévy processes §.


Fortado, L. (22-12-2016). Hedge funds fees take a trim. *Financial Times, available at https://www.ft.com/content/ab1ce98e-c5da-11e6-9043-7e34c07b46ef?mhq5j=e2.*


