Contents

The underwriting cycles under Solvency II .......................... 1
Anthony Derien

Empirical pricing kernels and investor preferences .................. 19
Kai Detlefsen, Wolfgang Karl Härle and Rouslan A. Moro

Calibration of a multiscale stochastic volatility model .............. 49
Lorella Fatone, Francesca Mariani, Maria Cristina Recchioni and
Francesco Zirilli

A reserve risk model for a non-life insurance company ............. 63
Salvatore Forte, Matteo Ialenti and Marco Pirra

Parameter estimation for differential equations using
fractal-based methods and applications to economics and
finance ............................................................................. 79
Herb Kunze, Davide La Torre, Cinzia Colapinto, Jelena Loncar and
Matteo Fini

Jump telegraph processes and a volatility smile ..................... 93
Nikita Ratanov

Estimating the true embedded risk management cost of total
return strategies ...................................................................... 113
Francesco Sandrini, David Edelman and Dermot Ryan
Preface

This is the first of two special issues of *Mathematical Methods in Economics and Finance* devoted to the *International Conference MAF 2008 – Mathematical and Statistical Methods for Actuarial Sciences and Finance* held in Venice (Italy) from March 26 to 28, 2008.

The conference has been the first international edition of a biennial national series begun at 2004, which was born by a brilliant belief of the colleagues – and friends – of the Department of Economics and Statistical Sciences of the University of Salerno: the idea following which the cooperation between mathematicians and statisticians in working in actuarial sciences, in insurance and in finance can improve research on these topics. The proof of the goodness of this belief has consisted in the wide participation to these events.

This issue collects a series of original papers freely submitted to the journal by Contributors of the conference and, following the usual praxis, each peer reviewed by at least two anonymous referees.

Guest Editors (and co-Chairs of the conference):

Marco Corazza and Claudio Pizzi
The underwriting cycles under Solvency II*

Anthony Derien
Laboratoire de Science Actuarielle et Financière, Université de Lyon, Université Lyon 1, 50 av. Tony Garnier, 69366 Lyon Cedex 07, France
anthony.derien@etu.univ-lyon1.fr

Abstract. The presence of the underwriting cycle in non-life insurance is well established. In recent years this subject has been renewed with the focus on the new European Directive, Solvency II, which supports the need to hold capital due to the uncertainty about the insurance premium. The standard formula used ignores the existence of two regimes, one where the premiums increase (a hard market) followed by another where the premiums decrease (a soft market). We use a regime switching model to demonstrate the dynamic change of the variables and how it could be used in an internal model. Our analysis has been performed on the Motor Liability, Motor Damage and Property lines in the French market. The results support the main theories of the underwriting cycle and the need to allow a different behavior of the variables across the cycle. We suggest another specification for the premium risk in the standard formula.

Keywords. Underwriting cycle, regime-switching, Markov chain.

M.S.C. classification. 91B30, 60J10, 62J02.


1 Introduction

Since 1997 the European Commission and the Committee of European Insurance and Occupational Pension Supervisors (CEIOPS) have worked on the new rules of solvency for the European insurance industry. These rules, named Solvency II, have been implemented to improve risk management in the insurance industry by identifying different types of risks and allowing insurance companies to use an internal model to estimate their capital (see the article 110 of [5]). The different classes of risk proposed by the CEIOPS are: market risk, default risk, non-life risk, life risk, health risk and operational risk. Each of these risks are split into modules. For the non-life risk class, two types of risks are identified: underwriting risk and catastrophe risk. Following the definition, TS.XIII.A.1 in

* The Author is grateful to Jean Paul Laurent and Stéphane Loisel for their detailed comments and Emmanuel Le Floc’h for helpful discussion. This research has been supported by Aon Re France.
The underwriting risk is the specific risk arising from insurance contracts. It relates to the uncertainty about results of the insurer’s underwriting. This risk concerns both premium and reserve risk.

The premium risk describes in the standard formula used volatility by line of business, an assumption of a lognormal distribution and the estimated earned premium to assess the need of capital. This approach disregards the underwriting cycle’s behavior which can be separated in two regimes. A soft market is characterized by an extensive supply of insurance contracts and a decrease in the premium, and a hard market with a low supply of contracts and an increase in insurance premium.

Five main problems arise with the CEIOPS’s methodology:

- The use of loss ratio (claims divided by the premium) to analyze the uncertainty about the premium, see TS.XIII.B.12 in [2]. A high loss ratio could be due to high claims or low premium, and a low premium could be due to a decrease in the number of policyholders or a premium’s decrease. We cannot identify the cause of the loss ratio evolution.
- Using historic loss ratios based on nominal value, though the future rules of the European directive requires companies to estimate the losses in economic value.
- The CEIOPS applies the volatility estimated across the cycle, ie the volatility inter regime.
- The premium risk is applied without distinguishing the regime. In a hard market, the capital requirement would increase due to the premium increase.
- No investigation of the causes of the underwriting cycle.

To overcome these problems, we investigate the main theories in the fields of the underwriting cycle to model its behavior in an internal model and add some features to show that the required capital calculation in the standard formula results in an inverse output to the one originally planned.

- The capacity constraint hypothesis (supported by [11], [12], [13], [17], [22], [23], [24]) is based on the fact that an insurer has to hold an amount of capital to underwrite policies. In case of the occurrence of a shock which decreases the capital, an insurance company has to increase its premium to form its capital. Because it would be more expensive to raise capital through financial markets due to the information asymmetry.
- The financial pricing model is based on discounted cash flows. The insurance premiums are equal to policy expenses plus the expected value of future claims. Many papers have been written on this subject by [7], [14], [15], [19].
- In the actuarial model, the premium equals the expected present loss plus a risk loading, see [1].
- A time series analysis is performed on the behavior of the loss ratio to determine the evolution of the underwriting cycle, see [21].

We based our approach on the existence of two regimes (a hard and a soft market). It lead us to use a regime switching model in a Markovian environment.
Regression techniques are used to estimate the coefficient of the variables (loss, outstanding, lagged value of the premium evolution, surplus, investment income).

This analysis has been performed with different distribution functions of the regression to check the Solvency II’s assumption of the lognormal distribution. In order to test and estimate the regime switching model, we worked on a panel of 19 French insurance companies for the three main lines of business: Motor Liability, Motor Damage, Property. We can distinguish the volume and price effect on our panel from the premium earned by the insurer.

Our work shows the existence of different behavior for the variables in each regime (for example: lower volatility in the soft market than in the hard market, and the coefficient of the capital is significant in the hard market but not in the soft market). The assumption of a lognormal distribution for Motor Damage and Property is supported, but the normal distribution is more suitable for Motor Liability.

The Markovian environment used allows us to calculate the average length of the cycles. This length vary from 5.6 years for Motor Damage, 6 years for Motor Liability and 7.5 years for Property. The soft market is longer than the hard market for the Motor Liability lines, while the hard market is longer than the soft market for Motor Damage and Property Lines. This model can be easily implement in an internal model to derive the regulatory capital.

An alternative to the formula provide by the CEIOPS is suggested to better grasp the behavior of the underwriting cycle. Our proposition takes into account the regime and is in keeping with the methodology of Solvency II.

The remainder of this paper is organized as follows: The first section presents the hypothesis of the behavior of the underwriting cycle. Data, models and their outputs are then described in the next section. Finally, we summarize and conclude the study.

2 Framework of the underwriting cycle

2.1 French underwriting cycle

We consider three lines of business: Motor Damage, Motor Liability and Property for the non-life French insurance market from a panel of 19 companies, which represents 30% of the market. We focus on these three lines for different reasons. We need the number of contracts for these lines in order to distinguish between the volume and price effect of the premium earned. With this distinction, we are able to determine if the growth of the premium earned stems from an increase in the number of contracts or if it is due to the underwriting cycle. The earned premiums are scaled by this exposure.

Our panel is mainly constituted of mutual companies. They underwrite principally personal lines. These three lines account for approximately 60% of the gross earned premiums in the non-life market. In order to match the definition provided by the CEIOPS, the property line include personal, professional and agricultural property and natural catastrophe (specific treatment in France, see
calculating the mean value for each year, we obtain the following output:

We clearly see a pattern of decrease followed by an increase.

The Motor Liability shows a decrease from 1997 to 1999, then an increase until 2003. A new period of decrease has occurred since 2003. Motor Damage shows a decrease until 1999, followed by a strong increase (more than 5pts) continuing until 2005. In 2006, a new decreasing trend appeared for this line.

The Property line shows a similar evolution (decreases until 1999, a strong increase in 2000. This increase sped up in 2003, then this increase slow down). The hard market observed for the property lines is relatively longer than the Motor Damage and Motor Liability one.

There are many reasons for this. We used the definition of the CEIOPS’s line of business. As we mentioned, we aggregate different lines, it can introduce a bias in the length of the cycle as mentioned in [10].

Over the last decade, many natural events hit France:

- 1999: Two major windstorms
- 2002-2003: Several floods
- 2003: Drought
- 2006: Windstorm

These evolutions justify the need to hold an amount of capital in the soft market due to the underpricing, which increases the insolvency of insurance companies.
2.2 Underlying theories of the underwriting cycle

Underwriting cycle phenomenon is a major issue for the insurance industry when assessing the insurance pricing method and the regulation constraints.

Indeed a soft market may contribute to insolvencies if insurers price and underwrite too aggressively, while high prices and restrictive underwriting in the hard market may contribute to reduce the flows of goods and services.

No consensus appears in the literature. Many fields have been explored to obtain a stochastic model for the behavior of the premium.

Some analysis are based on the behavior of the loss ratio to determine the evolution of the underwriting cycle (see the papers of [7], [21]). They used time series analysis to show the cyclical pattern stem from regulatory and accounting lags.

Most attention has been focused on the financial pricing model and capacity constraint hypothesis.

The financial pricing models are based on discounted cash flows. The insurance premiums are equal to policy expenses plus the expected value of future claims. The insurance company is risk neutral and has rational expectation with respect to the claims. Many authors, such as [7], [14], [15], [19] use times series analysis to confirm financial pricing assumptions. This theory implies a positive coefficient for the claims outstanding, the paid losses and the premium’s evolution lagged when regressing them with the premium’s evolution. The coefficient of the interest rate has to be negatively linked with the premium’s evolution.

The capacity constraint hypothesis is based on the fact that an insurer has to hold sufficient capital to meet its liabilities. In the case of an occurrence of a shock to the institution, an insurance company would have difficulties in raising capital in financial markets. The external capital would be more expensive than the internal one due to the information asymmetry between the manager and the investor that exists in financial markets. Regarding the increase of capital, investor does not see in a good perspective. If this were the case, it would increase its capital by self-financing. The investor would ask for a higher risk premium, so the company should increase its premium to form its capital. This approach involves a negative relationship of the capital’s evolution between the premium’s evolution. The capacity constraint is supported by [4], [11], [12], [13], [17], [22], [23], [24].

[6] proposed an extension to the capacity constraint hypothesis in adding the assumption that the policyholders agree to pay more if the insurer has a lower probability of default than its competitors.

The most commonly used pricing method is in practice the actuarial model. Premiums are equal to the expected present loss plus a risk loading. The risk loading is a buffer for having an acceptable ruin probability. This model is linked positively with the variance of losses and negatively on the capital. Actuarial model has been chosen by the CEIOPS to assess insurer solvency level, see [1].

All the previous papers, except [17], use the same statistical approach (regression, time series analysis) to model the underwriting cycle. Those techniques imply a time invariant parameter of the variables. It seems difficult to believe...
that the variables have the same behavior across the cycle. For example, with the capacity constraint hypothesis, when the capital decreases, the premium should increase, but the explanation for premium decrease is more uncertain. The use of a regime switching model allows us to estimate the parameters of the variables in each phase of the cycle. We extend the work of [17] by allowing the standard deviation to be different in each state.

3 Empirical analysis

3.1 Model

We follow the approach proposed by [16] and [18]. We assume that the regime shifts are exogenous with respect to all realizations of the regression vector. We also assume that the exogenous variable is the realization of the two-states Markov chain with:

\[ Pr(s_t = j | s_{t-1} = i, s_{t-2} = k, y_{nt-1}, y_{nt-2}, \ldots) = Pr(s_t = j | s_{t-1}) = P_{ij}. \]  

(1)

Where

\( s_t \) is the variable representing the two state of the world.
\( y_{nt} \) is the premium’s growth for the company \( n \) in year \( t \).

The probability \( P_{11} \) is the probability to stay in the state 1, which is the state where an insurance company increases its premium. While \( P_{22} \) is the probability to stay in the state 2, which is the state where an insurance company decreases its premium. Since we have only two state and the Markov Chain is ergodic, then \( P_{12} = 1 - P_{11} \) and \( P_{21} = 1 - P_{22} \). We presume the situation of \( s_t \) only through the behavior of \( y_{nt} \).

We define the information set as:

\[ F_{t-1} = (y_{nt-1}, y_{nt-2}, \ldots, y_{(n-k)(t-1)}, y_{(n-k)(t-2)}, \ldots). \]  

(2)

We use the following regression model for each line, based on panel data.

\[ y_{nt}^{S_t} = \beta_0^{S_t} + \alpha_n^{S_t} + \sum_{j=1}^{5} \beta_j^{S_t} x_{jn}^{S_t} + \epsilon_t. \]  

(3)

for \( n = 1, \ldots, 19 \) and \( s_t = 1, 2 \).
\( y_{nt}^{S_t} \) is the premium growth for the firm \( n \) in the state \( s_t \).
\( \alpha_n^{S_t} \) is a constant specific to the company \( n \) in the state \( s_t \).
\( x_{jn}^{S_t} \) It represents the growth for the \( j^{th} \) variable (interest rate, outstanding losses, losses paid, capital, lagged value of premium growth) for the company \( n \) in the state \( s_t \).
\( \beta_j^{S_t} \) are the regression coefficients to be estimated in the state \( s_t \).
\( \beta_0^{S_t} \) is a constant term in the state \( s_t \).
The underwriting cycles under Solvency II

\( \epsilon_s \) is the error term in the state \( s_t \).

We have:

\[
\mu(y_s^n) = E(y_s^n | s_t = j, F_{t-1}) = \beta_0^S_t + \alpha^S_t + \sum_{j=1}^{5} \beta_j^S_t \epsilon_{jn}^S
\]

\[
\sigma^2(y_s^n) = Var(y_s^n | s_t = j, F_{t-1}).
\]

As \( s_t \) is exogeneous with respect to \( \epsilon_s \), we have:

\[
E(\epsilon_s^n | s_t = j, F_{t-1}) = 0.
\]

In the literature, the common assumption for the conditional distribution function of \( (y_s^n | s_t = j, F_{t-1}) \) is the Gaussian distribution.

The actual rules for determining the premium risk in Solvency II are based on the assumption that the underlying risk follows a lognormal distribution.

To avoid restricting to these two cases, we investigate two other distributions: Gamma and Weibull.

We estimate the parameters by the maximum likelihood. The likelihood function is given by:

\[
L = \prod_{n=1}^{N} \prod_{t=1}^{T} ln(f(y_s^n | F_{t-1})).
\]

where

\[
f(y_s^n | F_{t-1}) = \sum_{s_t} \sum_{s_{t-1}} f(y_s^n | s_t, s_{t-1}, F_{t-1})
\]

\[
= \sum_{s_t} \sum_{s_{t-1}} f(y_s^n | s_t, s_{t-1}, F_{t-1}) Pr(s_t, s_{t-1} | F_{t-1})
\]

\[
= \sum_{s_t} \sum_{s_{t-1}} f(y_s^n | s_t, s_{t-1}, F_{t-1}) Pr(s_t | s_{t-1}) Pr(s_{t-1} | F_{t-1}).
\]

By definition \( \sum_{i=1,2} Pr(s_{t-1} = i | F_{t-1}) = 1. \)

We proceed iteratively to solve this algorithm via the recursive filter [16]:

\[
Pr(s_{t-1} = i | F_{t-1}) = \sum_{s_t} Pr(s_t = i, s_{t-1} | F_{t-1})
\]

\[
= \sum_{s_t} \frac{f(y_s^n | s_t = i, F_{t-1}) Pr(s_t = i | s_{t-1}) Pr(s_{t-1} | F_{t-1})}{f(y_s^n | F_{t-1})}.
\]

3.2 Data

We used the data from the financial reports of insurance companies published in French GAAP (nominal values). These data are similar to those used by rating agencies to estimate a rating based on public information, see [9], [20].
We focus on the three main lines of business for the French Market: Motor Liability, Motor Damage and Property. Our sample has been restricted to allow us to distinguish the premium’s growth between the volume and price effect. Even if this panel is limited, our purpose is illustrative and we believe that if more data were available the results could be generalized. The data used in the regression are for each line:

**Losses** The losses occurring in the accident year and paid in the year

**Premium** The earned premiums in the accident year

**Allocated investment income** The investment income for each line of business

**Outstanding** The reserve of the losses in the calendar year (the sum of accident year outstanding plus the variation of previous year’s reserve).

**Capital** The capital at the company level.

These variables are scaled by the number of contracts.

In the data used, no reference to interest rate appears, although it is present in the tested theories. In reviewing the results of the last two “Quantitative Impact Study” (QIS) carry out by the CEIOPS, the interest rate has an impact. The CEIOPS provided two different term structures (see the graph 2) to estimate the present value of losses.

![Fig. 2. Term Structure of the Second and Third Quantitative Impact Study](image)

The following table presents the ratio of the expected discounted value of the outstanding losses with respect to the accounting value of the outstanding losses, for the French market for the second and third “Quantitative Impact Study”.

---

**Anthony Derien**
The underwriting cycles under Solvency II

### Table 1. Ratio of the Discounted Value of the Cash Flow on the Accounting Value

<table>
<thead>
<tr>
<th>LoB</th>
<th>QIS2</th>
<th>QIS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor (Damage + Liability)</td>
<td>85%</td>
<td>78%</td>
</tr>
<tr>
<td>Property</td>
<td>82%</td>
<td>81%</td>
</tr>
</tbody>
</table>

So with an increase of 30% in the interest rates, Motor (property + liability) line shows a decrease of 8% for the Best Estimate. For the Property line, no difference appears for the expected value. Clearly the difference in payment pattern appears between these lines.

Unfortunately we cannot specifically test the interest rate. Indeed, to estimate the discounted value of the cash flow of the outstanding losses, we need cash flow information on paid claims. It is not available. We could use the proxy defined by [24] but we should apply it for all the companies. It is difficult to believe that each company should have the same cash flow. The proxy works when we use times series, not panel data, as did [4], [24].

The same reason can be given for the use of interest rates in the explanatory variables. Moreover a methodological problem arises with the use on the interest rate in the discounted value of cash flow and at the same time the use of the interest rate as the explanatory variables.

Nevertheless, to verify the impact of the time value of the money, we include in the regression a proxy: The investment income allocated by line of business. We do this in order to see if a relationship appears with the premium, which should prove that the economic environment impacts the premium.

#### 3.3 Results

**Distribution function** We test the four distributions, and we look at which gives us the maximum likelihood function.

### Table 2. Results for the maximum likelihood function

<table>
<thead>
<tr>
<th></th>
<th>Motor Liability</th>
<th>Motor Damage</th>
<th>Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma</td>
<td>136.72</td>
<td>171.43</td>
<td>141.78</td>
</tr>
<tr>
<td>Gaussian</td>
<td>145.73</td>
<td>181.40</td>
<td>150.34</td>
</tr>
<tr>
<td>Lognormal</td>
<td>134.51</td>
<td>184.43</td>
<td>165.62</td>
</tr>
<tr>
<td>Weibull</td>
<td>126.09</td>
<td>153.15</td>
<td>120.25</td>
</tr>
</tbody>
</table>

The lognormal distribution is the best distribution to represent the premium’s growth for Motor Damage and Property. These results support the assumption of the CEIOPS, but we reject it for Motor Liability which exhibits a smaller tail than the lognormal.
As we work on panel data, we test for each line if the presence of the specific company constant (the parameter $\alpha_{St}$ in equation 3) increases the prediction capacity of the model by the AIC criteria. The AIC criteria is defined by:

$$AIC = \ln \left( \frac{\sum_{i=1}^{n} e_i^2}{n} \right) + \frac{2k}{n}$$

Where $e_i$ is the residual of the regression, $k$ is the number of regression parameters and $n$ the size of our sample.

Table 3. AIC results

<table>
<thead>
<tr>
<th></th>
<th>Motor Liability</th>
<th>Motor Damage</th>
<th>Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>With company effect</td>
<td>-6.14</td>
<td>-6.74</td>
<td>-6.28</td>
</tr>
<tr>
<td>Without company effect</td>
<td>-6.63</td>
<td>-6.79</td>
<td>-5.88</td>
</tr>
</tbody>
</table>

Specific company constant increases the performance of the model just for Property; for Motor Liability and Motor Damage, this effect can be ignored. One explanation for the presence of the specific company constant term in the Property line could stem from the fact that we include in the property line different types of risk (personal, agricultural, professional goods and natural catastrophic line). Even if the last three categories represent a small part of the earned premium it could affect the premium’s evolution.

Regression estimation Following are the regression coefficients for each line, the standard errors are indicated in brackets:

Table 4. Regression Coefficients for the Motor Damage Line (*Significant at 5%)

<table>
<thead>
<tr>
<th>Motor Damage</th>
<th>Soft Market</th>
<th>Hard Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-3.99%*</td>
<td>3.52%*</td>
</tr>
<tr>
<td>Investment income</td>
<td>-0.22%</td>
<td>-0.35%</td>
</tr>
<tr>
<td>Outstanding</td>
<td>2.57%*</td>
<td>2.70%*</td>
</tr>
<tr>
<td>Losses</td>
<td>2.29%*</td>
<td>0.85%</td>
</tr>
<tr>
<td>Capital</td>
<td>1.26%</td>
<td>-4.40%*</td>
</tr>
<tr>
<td>Premium N-1</td>
<td>-0.72%</td>
<td>23.57%*</td>
</tr>
<tr>
<td>Volatility</td>
<td>2.02%*</td>
<td>3.47%*</td>
</tr>
</tbody>
</table>

All the variables have the expected signs, when they are significant. We observe that the claims paid and the outstanding losses are positively linked to the premium’s growth. The coefficients of the outstanding losses are quite similar between the two phases of the cycle. We have a confirmation of the
financial pricing theory and the actuarial model, because we have a premium adjustment to the claims. If we have an increase in the loss in year $n - 1$, in year $n$ we would have an increase in the premium.

We obtain a negative relationship between the capital and the premium’s evolution in the hard market but the correlation is not significant in the soft market. As expected when the capital decreases, we observe an increase in the premium but not link appears between the premium’s evolution and the capital in the soft market. This is a validation of the capacity constraint hypothesis for the hard market. This result confirms the use of a regime switching model.

As we could see, the coefficient of the investment income is negative, as expected, but not significant. Due to the large coefficient for the outstanding losses variable, a more accurate approach to treat the outstanding losses in a discounting value should be investigated to better reflect the impact of the economical environment.

We observe a strong correlation between the lagged value of the premium’s evolution and the premium evolution in the hard market for the motor lines and in the soft market for the property line.

As the lagged value of the premium is not significant in soft market for the motor lines, it shows that when the loss decreases, the insurance companies reflect it on the insurance premium quickly. Inversely, we observe an inertia phenomenon in the hard market, we can conclude that due to the competition in the insurance market, the insurance company cannot reflect the whole increase of the loss in the premium, but it spreads it over many years.

---

Table 5. Regression Coefficients for the Motor Liability Line (*Significant at 5%)

<table>
<thead>
<tr>
<th>Motor Liability</th>
<th>Soft Market</th>
<th>Hard Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>89.53%* (5.41%)</td>
<td>87.00%* (3.26%)</td>
</tr>
<tr>
<td>Investment income</td>
<td>-0.29% (0.50%)</td>
<td>0.06% (0.45%)</td>
</tr>
<tr>
<td>Outstanding</td>
<td>5.82%* (0.76%)</td>
<td>7.08%* (1.01%)</td>
</tr>
<tr>
<td>Losses</td>
<td>1.78%* (0.77%)</td>
<td>0.35% (0.62%)</td>
</tr>
<tr>
<td>Capital</td>
<td>1.64%* (0.84%)</td>
<td>-8.84%* (1.11%)</td>
</tr>
<tr>
<td>Premium N-1</td>
<td>0.49% (0.79%)</td>
<td>32.95%* (6.30%)</td>
</tr>
<tr>
<td>Volatility</td>
<td>3.19%* (1.76%)</td>
<td>4.76%* (2.21%)</td>
</tr>
</tbody>
</table>

Table 6. Regression Coefficients for the Property Line (*Significant at 5%)

<table>
<thead>
<tr>
<th>Property</th>
<th>Soft Market</th>
<th>Hard Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-5.41%* (2.90%)</td>
<td>5.90%* (2.49%)</td>
</tr>
<tr>
<td>Investment income</td>
<td>-0.21% (0.67%)</td>
<td>-0.06% (0.72%)</td>
</tr>
<tr>
<td>Outstanding</td>
<td>2.84%* (1.63%)</td>
<td>1.07% (1.33%)</td>
</tr>
<tr>
<td>Losses</td>
<td>-0.19% (1.14%)</td>
<td>2.04%* (1.22%)</td>
</tr>
<tr>
<td>Capital</td>
<td>1.32% (1.57%)</td>
<td>-5.26%* (2.98%)</td>
</tr>
<tr>
<td>Premium N-1</td>
<td>6.67%* (3.87%)</td>
<td>0.38% (1.88%)</td>
</tr>
<tr>
<td>Volatility</td>
<td>3.46%* (1.28%)</td>
<td>3.98%* (1.73%)</td>
</tr>
</tbody>
</table>
We note an important difference in the volatility between the two regimes. For Motor Liability, in the soft market, the volatility is 40% lower than in the hard market. The same result can be observed on Motor Damage. The difference is smaller in Property but exists.

As there are frictional cost in the insurance market (tacit agreement, cancellation at renewal date) it is necessary for an insurance company to decrease its premium to avoid losing market share, which would be difficult to recover. It can explain why we observe a downgrade pressure on the premiums in the soft market.

The regime switching model allows us to see specifically in which phase of the cycle, the variables produce their effect on the premium’s growth.

We perform the same analysis without distinguishing between the regimes. In doing a Chow Test, we obtain the following results:

Table 7. Chow test for the different lines of business

<table>
<thead>
<tr>
<th></th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor Liability</td>
<td>0.36%</td>
</tr>
<tr>
<td>Motor Damage</td>
<td>0.28%</td>
</tr>
<tr>
<td>Property</td>
<td>0.76%</td>
</tr>
</tbody>
</table>

As the P-value are below 1%, these results tell us that we have to distinguish between the regimes, and confirm the use of the regime switching model.

Transition matrix We present below the probability of the Markov Chain for 1-year.

Table 8. 1-Year Transition Matrix

<table>
<thead>
<tr>
<th>Probability</th>
<th>Motor Liability</th>
<th>Motor Damage</th>
<th>Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of staying in hard market</td>
<td>61.66%</td>
<td>68.83%</td>
<td>42.42%</td>
</tr>
<tr>
<td>Probability of leaving hard market</td>
<td>38.34%</td>
<td>31.17%</td>
<td>57.58%</td>
</tr>
<tr>
<td>Probability of staying in soft market</td>
<td>70.67%</td>
<td>59.10%</td>
<td>82.52%</td>
</tr>
<tr>
<td>Probability of leaving soft market</td>
<td>29.33%</td>
<td>40.90%</td>
<td>17.48%</td>
</tr>
</tbody>
</table>

As we can see in the previous table, the probability of staying in the hard market for 1-year is higher than the probability of staying in the soft market for Motor damage. The reverse is true for the two other lines.

We have to note that the probability of staying in the hard market is two times lower than the probability of staying in the soft market for the Property lines. The aggregation of different types of risk can explain it.
We calculate the average length of a period in the soft market, i.e., a premium decreasing regime, and in the hard market, i.e., a premium increasing regime, with the Markov chain. Since the probability of having a period in regime $i$ which lasts exactly $t$ years is $p_{it}^{t-1}p_{ij}$, we can calculate the average length of a period in regime $i$ as:

$$\sum_{t=1}^{\infty} t p_{it}^{t-1} p_{ij} = \frac{1}{p_{ij}}.$$ \hfill (9)

<table>
<thead>
<tr>
<th>Line of Business</th>
<th>Soft Market</th>
<th>Hard Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor Liability</td>
<td>3.41</td>
<td>2.62</td>
</tr>
<tr>
<td>Motor Damage</td>
<td>2.45</td>
<td>3.21</td>
</tr>
<tr>
<td>Property</td>
<td>1.74</td>
<td>5.72</td>
</tr>
</tbody>
</table>

The length of a soft market is longer than the hard market for Motor liability, and the opposite for the other two. If we add the time spent in the soft market and the time spent in the hard market for each line, we obtain an average length of an underwriting cycle of 6.03 years for Motor Liability, 5.66 for Motor Damage and 7.46 for the Property lines. These results are quite similar to those obtained by [21] on US data. He found cycles of 6.3 years for Motor Liability (respectively, 5.5 for Motor Property and 7.2 for Property).

These results allow us to better grasp the behavior of the premium’s growth and justify the need for an insurance company to develop an internal model which takes into account the underwriting cycle to determine the amount of capital for the premium risk and to better reflect the behavior of the insurance market. We wonder if using a standard formula which integrates into the underwriting cycle we can achieve the same result.

**Alternative model** The previous model explains how to build a stochastic model to estimate the premium risk for the companies which have an internal model. As it is costly to develop an internal model for an insurance company (see [3]), we are interested in evaluating a standard formula which takes into account the underwriting cycle.

As we mentioned in the introduction, the CEIOPS applies the volatility inter regime to the forthcoming earned premium which ever regime we are in, while it would be more realistic to apply the volatility intra regime and the premium’s growth according to the regime.

On the graph 3, we can show the approach followed by the CEIOPS with the letters “A” and “B”, they are applicable regardless of which cycle we are in. In the hard market (premium increase), the model of the CEIOPS requires more capital because the premium risk is calculated by means of a percentage.
applied to the premium earned. But if the company increases its premiums, the
dolvency risk should decrease.

In our alternative, we identify the regime. The distribution function repre-
sented by “C” has a smaller tail, representing the distribution function in a soft
market. We apply the volatility specific to the soft market and the premium’s
growth.

Fig. 3. Alternative model for the Premium Risk

More precisely, according to the current state of the cycle, we estimate the
expected value for the premium risk:

\[
PR = PE[\pi_{s_t,H}(\alpha_H + \rho(X_H)(1 + \alpha_H)) + \pi_{s_t,S}(\alpha_S + \rho(X_S)(1 + \alpha_S))].
\]  
(10)

Where:

\(PR\) is the premium risk.
\(PE\) is the premium earned.

\(\alpha_H, \alpha_S\) represent, respectively, the premium’s growth in the hard market and
soft market (for example: 3% and -2%).

\(\pi_{s_t,H}, \pi_{s_t,S}\) represent, respectively, the probability of moving from the state \(s_t\)
to the hard market, and the probability of moving from the state \(s_t\) to the
soft market. \(s_t\) is the current state, hard market or soft market.

\(\rho(X_H), \rho(X_S)\) correspond, respectively, to the 99.5\textsuperscript{th} percentile of the distribu-
tion function (lognormal) in the hard market and the soft market.

These parameters would be estimated by the local regulatory authority.
The underwriting cycles under Solvency II

Table 10. Parameters for the alternative model

<table>
<thead>
<tr>
<th>Line</th>
<th>$\alpha_S$</th>
<th>$\rho(X_S)$</th>
<th>$\alpha_H$</th>
<th>$\rho(X_H)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor Liability</td>
<td>-4.2%</td>
<td>-12.7%</td>
<td>4.8%</td>
<td>-9.3%</td>
</tr>
<tr>
<td>Motor Damage</td>
<td>-2.9%</td>
<td>-7.9%</td>
<td>4.5%</td>
<td>-5.3%</td>
</tr>
<tr>
<td>Property</td>
<td>-4.2%</td>
<td>-15.3%</td>
<td>5.5%</td>
<td>-6.3%</td>
</tr>
</tbody>
</table>

If we use our estimates (given in table 10) to calculate the premium risk with the same risk measure, we obtain the following percentages to apply to the earned premium:

Table 11. Comparison of the Premium Risk

<table>
<thead>
<tr>
<th>Current State</th>
<th>Motor Liability</th>
<th>Motor Damage</th>
<th>Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soft Market</td>
<td>13%</td>
<td>7%</td>
<td>9%</td>
</tr>
<tr>
<td>Hard Market</td>
<td>9%</td>
<td>4%</td>
<td>4%</td>
</tr>
<tr>
<td>Solvency II</td>
<td>29%</td>
<td>29%</td>
<td>29%</td>
</tr>
</tbody>
</table>

Different explanations can be given to justify the spread.

- Using inter regime volatility, as CEIOPS does, rather than intra regime volatility, leads to a significantly higher capital requirement than in our model.
- The difference stems from the fact that the CEIOPS estimates the premium risk based on loss-ratios. While we have estimated the premium risk on premium scaled by the number of contracts.

In order to compare the impact of the underwriting cycle on the standard formula, we perform the same analysis using the loss ratio on two samples. One is the same as before (Panel A) and the other is an extended sample over 25 companies (Panel B) with a mean period of 9.4 years.

Table 12. Comparison of the Premium Risk with Loss Ratio

<table>
<thead>
<tr>
<th>Sample</th>
<th>Current State</th>
<th>Motor Liability</th>
<th>Motor Damage</th>
<th>Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Soft Market</td>
<td>22%</td>
<td>15%</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>Hard Market</td>
<td>21%</td>
<td>6%</td>
<td>26%</td>
</tr>
<tr>
<td></td>
<td>Solvency II</td>
<td>29%</td>
<td>29%</td>
<td>29%</td>
</tr>
<tr>
<td>Panel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Soft Market</td>
<td>22%</td>
<td>16%</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>Hard Market</td>
<td>20%</td>
<td>6%</td>
<td>26%</td>
</tr>
<tr>
<td></td>
<td>Solvency II</td>
<td>29%</td>
<td>29%</td>
<td>29%</td>
</tr>
</tbody>
</table>
We work on a calendar year basis. It means that the volatility and the growth in the loss ratio that we observe come from the uncertainty from the reserve and the premium. So our estimate corresponds to the premium risk and a part of the reserve risk. The inclusion of the underwriting cycle in the CEIOPS methodology could reduce the capital requirement and better reflect the behavior of the insurance market.

4 Conclusion

The CEIOPS recognizes the fluctuation of the insurance premium, and require insurance companies to hold an amount of capital to protect themselves against the uncertainty around the premium. Unfortunately, no consideration of the underwriting cycle is being made and nor is the current state of the cycle.

We tested the main theories of this subject on a sample of French companies. We extended them by allowing the variables to have a different behavior across the cycle. We used a regime switching model to integrate the two regimes of the underwriting cycle: a soft market characterized by an extensive supply of insurance contracts and a decrease in the premium; and a hard market with a low supply and an increase in the premium.

It is possible to isolate the theories which are applicable according to the regime (for example: negative link between the capital and the premium’s evolution in the hard market only). Our work supported the capacity constraint hypothesis, the actuarial model and the pricing model theory. We show that assuming a constant volatility of the premium growth across the cycle, as CEIOPS does, is a misrepresentation of reality. The volatility of the underwriting cycle is different in each regime, the hard market has a stronger volatility than the soft market.

We suggest another approach to estimate the premium risk using the same methodology, which integrates the underwriting cycle and determines the capital requirement according to the regime.

The use of the loss ratios to estimate the premium risk rather than the use of the premium scaled, leads to an increase in the capital requirement because it includes other elements than the premium uncertainty (for example: loss uncertainty).

Further research is advisable on the impact of the introduction of historic economic loss ratio rather than nominal loss ratio for the estimate of the standard deviation to a certain whether it could create a pro-cyclicility or counter-cyclicility effect.

References

Empirical pricing kernels and investor preferences

Kai Detlefsen¹, Wolfgang Karl Härdle¹, and Rouslan A. Moro²

¹ CASE – Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany {detlefsen, haerdle}@wiwi.hu-berlin.de
² Department of Economics and Finance, School of Social Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, UK Russ.Moro@brunel.ac.uk

Abstract. This paper analyzes empirical market utility functions and pricing kernels derived from the DAX and DAX option data for three market regimes. A consistent parametric framework of stochastic volatility is used. All empirical market utility functions show a region of risk proclivity that is reproduced by adopting the hypothesis of heterogeneous individual investors whose utility functions have a switching point between bullish and bearish attitudes. The inverse problem of finding the distribution of individual switching points is formulated in the space of stock returns by discretization as a quadratic optimization problem. The resulting distributions vary over time and correspond to different market regimes.

Keywords. Utility function, pricing kernel, behavioral finance, risk aversion, risk proclivity, Heston model.


1 Introduction

Numerous attempts have been undertaken to describe basic principles on which the behaviour of individuals are based. Expected utility theory was originally proposed by J. Bernoulli in 1738. In his work J. Bernoulli used such terms as

* The research work of R. A. Moro was funded by the German Academic Exchange Service (DAAD). K. Detlefsen was financially supported by Bankhaus Sal. Oppenheim. This work was supported by Deutsche Forschungsgemeinschaft through the SFB 649 “Economic Risk”.

The research work of R. A. Moro was funded by the German Academic Exchange Service (DAAD). K. Detlefsen was financially supported by Bankhaus Sal. Oppenheim. This work was supported by Deutsche Forschungsgemeinschaft through the SFB 649 “Economic Risk”.

The research work of R. A. Moro was funded by the German Academic Exchange Service (DAAD). K. Detlefsen was financially supported by Bankhaus Sal. Oppenheim. This work was supported by Deutsche Forschungsgemeinschaft through the SFB 649 “Economic Risk”.
risk aversion and risk premium and proposed a concave (logarithmic) utility function, see [5]. The utilitarianism theory that emerged in the 18th century considered utility maximization as a principle for the organisation of society. Later the expected utility idea was applied to game theory and formalized by [31]. A utility function relates some observable variable, in most cases consumption, and an unobservable utility level that this consumption delivers. It was suggested that individuals' preferences are based on this unobservable utility: such bundles of goods are preferred that are associated with higher utility levels. It was claimed that three types of utility functions – concave, convex and linear – correspond to three types of individuals – risk averse, risk neutral and risk seeking. A typical economic agent was considered to be risk averse and this was quantified by coefficients of relative or absolute risk aversion. Another important step in the development of utility theory was the prospect theory of [24]. By behavioural experiments they found that people act risk averse above a certain reference point and risk seeking below it. This implies a concave form of the utility function above the reference point and a convex form below it.

Besides these individual utility functions, market utility functions have recently been analyzed in empirical studies by [22], [27] and others. Across different markets, the authors observed a common pattern in market utility functions: There is a reference point near the initial wealth and in a region around this reference point the market utility functions are convex. But for big losses or gains they show a concave form – risk aversion. Such utility functions disagree with the classical utility functions of [31] and also with the findings of [24]. They are however in concordance with the utility function form proposed by [18].

In this paper, we analyze how these market utility functions can be explained by aggregating individual investors’ attitudes. To this end, we first determine empirical pricing kernels from DAX data. Our estimation procedure is based on historical and risk neutral densities and these distributions are derived with stochastic volatility models that are widely used in industry. From these pricing kernels we construct the corresponding market utility functions. Then we describe our method of aggregating individual utility functions to a market utility function. This leads to an inverse problem for the density function that describes how many investors have the utility function of each type. We solve this problem by discrete approximation. In this way, we derive utility functions and their distribution among investors that allow to recover the market utility function. Hence, we explain how (and what) individual utility functions can be used to form the behaviour of the whole market.

The paper is organized as follows: In section 2, we describe the theoretical connection between utility functions and pricing kernels. In section 3, we present a consistent stochastic volatility framework for the estimation of both the historical and the risk neutral density. Moreover, we discuss the empirical pricing kernel implied by the DAX in 2000, 2002 and 2004. In section 4, we explain the utility aggregation method that relates the market utility function and the utility functions of individual investors. This aggregation mechanism leads to
an inverse problem that is analyzed and solved in this section. In section 5, we conclude and discuss related approaches.

2 Pricing kernels and utility functions

In this section, we derive the fundamental relationship between utility functions and pricing kernels. It describes how a representative utility function can be derived from historical and risk-neutral distributions of assets. In the following sections, we estimate the empirical pricing kernel and observe in this way the market utility function.

First, we derive the price of a security in an equilibrium model: we consider an investor with a utility function $U$ who has as initial endowment one share of stock. He can invest into the stock and a bond up to a final time when he can consume. His problem is to choose a strategy that maximizes the expected utility of his initial and terminal wealth. In continuous time, this leads to a well known optimization problem introduced by [26] for stock prices modelled by diffusions. In discrete time, it is a basic optimization problem, see [13].

From this result, we can derive the asset pricing equation

$$P_0 = E^P [\psi(S_T) M_T]$$

for a security on the stock $(S_t)$ with payoff function $\psi$ at maturity $T$. Here, $P_0$ denotes the price of the security at time 0 and $E^P$ is the expectation with respect to the real/historical measure $P$. The stochastic discount factor $M_T$ is given by

$$M_T = \beta U'(S_T)/U'(S_0)$$  \hspace{1cm} (1)

where $\beta$ is a fixed discount factor. This stochastic discount factor is actually the projection of the general stochastic discount factor on the traded asset $(S_t)$. The stochastic discount factor can depend on more variables in general. But as discussed in [13] this projection has the same interpretation for pricing as the general stochastic discount factor.

Besides this equilibrium based approach, [7] derived the price of a security relative to the underlying by constructing a perfect hedge. The resulting continuous delta hedging strategy is equivalent to pricing under a risk neutral measure $Q$ under which the discounted price process of the underlying becomes a martingale. Hence, the price of a security is given by an expected value with respect to a risk neutral measure $Q$:

$$P_0 = E^Q [\exp(-rT)\psi(S_T)]$$
If \( p \) denotes the historical density of \( S_T \) (i.e. \( P(S_T \leq s) = \int_{-\infty}^{s} p(x) \, dx \)) and \( q \) the risk neutral density of \( S_T \) (i.e. \( Q(S_T \leq s) = \int_{-\infty}^{s} q(x) \, dx \)) then we get

\[
P_0 = \exp(-rT) \int \psi(x) q(x) \, dx = \exp(-rT) \int \psi(x) \frac{q(x)}{p(x)} p(x) \, dx = \exp(-rT) \int \exp(-rT) \psi(S_T) \frac{q(S_T)}{p(S_T)} \, dx
\]

(2)

Combining equations (1) and (2) we see

\[
\beta \frac{U'(s)}{U'(S_0)} = \exp(-rT) \frac{q(s)}{p(s)}.
\]

Defining the pricing kernel by \( K = q/p \) we conclude that the form of the market utility function can be derived from the empirical pricing kernel by integration:

\[
U(s) = U(S_0) + \int_{S_0}^{s} U'(S_0) \frac{\exp(-rT) q(x)}{\beta p(x)} \, dx = U(S_0) + \int_{S_0}^{s} U'(S_0) \frac{\exp(-rT)}{\beta} K(x) \, dx
\]

because \( S_0 \) is known.

As an example, we consider the model of [7] where the stock follows a geometric Brownian motion

\[
dS_t/S_t = \mu dt + \sigma dW_t
\]

(3)

Here the historical density \( p \) of \( S_t \) is log-normal, i.e.

\[
p(x) = \frac{1}{x \sqrt{2\pi \sigma^2}} \exp \left\{ -\frac{1}{2} \left( \frac{\log x - \bar{\mu}}{\bar{\sigma}} \right)^2 \right\}, \quad x > 0
\]

where \( \bar{\mu} = (\mu - \sigma^2/2)t + \log S_0 \) and \( \bar{\sigma} = \sigma \sqrt{t} \). Under the risk neutral measure \( Q \) the drift \( \mu \) is replaced by the riskless interest rate \( r \), see e.g. [19]. Thus, also the risk neutral density \( q \) is log-normal. In this way, we can derive the pricing kernel

\[
K(x) = \left( \frac{x}{S_0} \right)^{-\frac{\mu - \gamma}{\sigma^2}} \exp\{ (\mu - r)(\mu + r - \sigma^2)T/(2\sigma^2) \}.
\]

This pricing kernel has the form of a derivative of a power utility

\[
K(x) = \lambda \left( \frac{x}{S_0} \right)^{-\gamma}
\]
where the constants are given by \( \lambda = e^{(\mu - r)(\mu + r - \sigma^2 \gamma)T} \) and \( \gamma = \frac{\mu - r}{\sigma^2} \). This gives a utility function corresponding to the underlying (3)

\[
U(S_T) = (1 - \frac{\mu - r}{\sigma^2})^{-1} S_T^{(1 - \frac{\mu - r}{\sigma^2})}
\]

where we ignored additive and multiplicative constants. In this power utility function the risk aversion is not given by the market price of risk \((\mu - r)/\sigma\). Instead investors take the volatility more into account. The expected return \(\mu - r\) that is adjusted by the riskfree return is related to the variance. This results in a higher relative risk aversion than the market price of risk.

\[\text{Fig. 1.} \] up: Utility function in the Black Scholes model for \(T = 0.5\) years ahead and drift \(\mu = 0.1\), volatility \(\sigma = 0.2\) and interest rate \(r = 0.03\). down: Market utility function on 06/30/2000 for \(T = 0.5\) years ahead.

A utility function corresponding to the Black-Scholes model is shown in the upper panel of figure 1 as a function of returns. In order to make different market situations comparable we consider utility functions as functions of (half year) returns \(R = S_{0.5}/S_0\). We chose the time horizon of half a year ahead for our analysis. Shorter time horizons are interesting economically and moreover the historical density converges to the Dirac measure so that results become trivial (in the end). Longer time horizons are economically more interesting but it is hardly possible to estimate the historical density for a long time ahead. It
neither seems realistic to assume that investors have clear ideas where the DAX will be in e.g. 10 years. For these reasons we use half a year as future horizon. Utility functions $\tilde{U}$ of returns are defined by:

$$\tilde{U}(R) := U(RS_0), \ R > 0$$

where $S_0$ denotes the value of the DAX on the day of estimation. Because of $U' = cK$ for a constant $c$ we have $\tilde{U}'(R) = cK(RS_0)S_0$ and we see that also utility functions of returns are given as integrals of the pricing kernel. The change to returns allows us to compare different market regimes independently of the initial wealth. In the following we denote the utility functions of returns by the original notation $U$. Hence, we suppress in the notation the dependence of the utility function $U$ on the day of estimation $t$.

The utility function corresponding to the model of [7] is a power utility, monotonically increasing and concave. But such classical utility functions are not observed on the market. Parametric and nonparametric models that replicate the option prices all lead to utility functions with a hump around the initial wealth level. This is described in detail later but is shown already in figure 1. The upper panel presents the utility function corresponding to Black-Scholes model with a volatility of 20% and an expected return of 10%. The function is concave and implies a constant relative risk aversion. The utility function estimated on the bullish market in summer 2000 is presented in the lower panel. Here, the hump around the money is clearly visible. The function is no more concave but has a region where investors are risk seeking. This risk proclivity around the money is reflected in a negative relative risk aversion.

3 Estimation

In this section, we start by reviewing some recent approaches for estimating the pricing kernel. Then we describe our method that is based on estimates of the risk neutral and the historical density. The risk neutral density is derived from option prices that are given by an implied volatility surface and the historical density is estimated from the independent data set of historical returns. Finally, we present the empirical pricing kernels and the inferred utility and relative risk aversion functions.

3.1 Estimation approaches for the pricing kernel

There exist several ways and methods to estimate the pricing kernel. Some of these methods assume parametric models while others use nonparametric techniques. Moreover, some methods estimate first the risk neutral and subjective density to infer the pricing kernel. Other approaches estimate directly the pricing kernel.
[1] derive a nonparametric estimator of the risk neutral density based on option prices. In [2], they consider the empirical pricing kernel and the corresponding risk aversion using this estimator. Moreover, they derive asymptotic properties of the estimator that allow e.g. the construction of confidence bands. The estimation procedure consists of two steps: First, the option price function is determined by nonparametric kernel regression and then the risk neutral density is computed by the formula of [8]. Advantages of this approach are the known asymptotic properties of the estimator and the few assumptions necessary.

[22] analyses risk aversion by computing the risk neutral density from option prices and the subjective density from historical data of the underlying. For the risk neutral distribution, he applies a variation of the estimation procedure described in [23]: A smooth volatility function derived from observed option prices gives the risk neutral density by differentiating it twice. The subjective density is approximated by a kernel density computed from historical data. In this method bandwidths have to be chosen as in the method of [1].

[27] use a different approach and estimate the subjective density and directly (the projection of) the pricing kernel. This gives the same information as the estimation of the two densities because the risk neutral density is the product of the pricing kernel and the subjective density. For the pricing kernel, they consider two parametric specifications as power functions and as exponentials of polynomials. The evolution of the underlying is modelled by GARCH processes. As the parametric pricing kernels lead to different results according to the parametric form used this parametric approach appears a bit problematic.

[11] also estimates the pricing kernel without computing the risk neutral and subjective density explicitly. Instead of assuming directly a parametric form of the kernel he starts with a (multidimensional) modified model of [20] and derives an analytic expression for the pricing kernel by the Girsanov theorem, see [10] for details. The kernel is estimated by a simulated method of moments technique from equity, fixed income and commodities data and by reprojection. An advantage of this approach is that the pricing kernel is estimated without assuming an equity index to approximate the whole market portfolio. But the estimation procedure is rather complex and model dependent.

In a recent paper, [3] price options in a GARCH framework allowing the volatility to differ between historical and risk neutral distribution. This approach leads to acceptable calibration errors between the observed option prices and the model prices. They estimate the historical density as a GARCH process and consider the pricing kernel only on one day. This kernel is decreasing which coincides with standard economic theory. But the general approach of changing explicitly the volatility between the historical and risk neutral distribution is not supported by the standard economic theory.

We estimate the pricing kernel in this paper by estimating the risk neutral and the subjective density and then deriving the pricing kernel. This approach does not impose a strict structure on the kernel. Moreover, we use accepted parametric models because nonparametric techniques for the estimation of sec-
ond derivatives depend a lot on the bandwidth selection although they yield the same pricing kernel behaviour over a wide range of bandwidths. For the risk neutral density we use a stochastic volatility model that is popular both in academia and in industry. The historical density is more difficult to estimate because the drift is not fixed. Hence, the estimation depends more on the model and the length of the historical time series. In order to get robust results we consider different (discrete) models and different lengths. In particular, we use a GARCH model that is the discrete version of the continuous model for the risk neutral density. In the following, we describe these models, their estimation and the empirical results.

3.2 Estimation of the risk neutral density

Stochastic volatility models are popular in industry because they replicate the observed smile in the implied volatility surfaces (IVS) rather well and moreover imply rather realistic dynamics of the surfaces. Nonparametric approaches like the local volatility model of [16] allow a perfect fit to observed price surfaces but their dynamics are in general contrary to the market. As [4] points out the dynamics are more important for modern products than a perfect fit. Hence, stochastic volatility models are popular.

We consider the model of [20] for the risk neutral density because it can be interpreted as the limit of GARCH models. The Heston model has been refined further in order to improve the fit, e.g. by jumps in the stock price or by a time varying mean variance level. We use the original Heston model in order to maintain a direct connection to GARCH processes. Although it is possible to estimate the historical density also with the Heston model e.g. by Kalman filter methods we prefer more direct approaches in order to reduce the dependence of the results on the model and the estimation technique.

The stochastic volatility model of [20] is given by the two stochastic differential equations:

\[
\frac{dS_t}{S_t} = r dt + \sqrt{V_t} dW_1^t \\
\frac{dV_t}{V_t} = \xi (\eta - V_t) dt + \theta \sqrt{V_t} dW_2^t
\]

where the variance process is modelled by a square-root process:

\[
dV_t = \xi (\eta - V_t) dt + \theta \sqrt{V_t} dW_2^t
\]

and \( W_1 \) and \( W_2 \) are Wiener processes with correlation \( \rho \) and \( r \) is the risk free interest rate. The first equation models the stock returns by normal innovations with stochastic variance. The second equation models the stochastic variance process as a square-root diffusion.

The parameters of the model all have economic interpretations: \( \eta \) is called the long variance because the process always returns to this level. If the variance \( V_t \) is e.g. below the long variance then \( \eta - V_t \) is positive and the drift drives the variance in the direction of the long variance. \( \xi \) controls the speed at which the
Empirical pricing kernels and investor preferences

variance is driven to the long variance. In calibrations, this parameter changes a lot and makes also the other parameters instable. To avoid this problem, the reversion speed is kept fixed in general. We follow this approach and choose $\xi = 2$ as [4] does. The volatility of variance $\theta$ controls mainly the kurtosis of the distribution of the variance. Moreover, there are the initial variance $V_0$ of the variance process and the correlation $\rho$ between the Brownian motions. This correlation models the leverage effect: When the stock goes down then the variance goes up and vice versa. The parameters also control different aspects of the implied volatility surface. The short (long) variance determines the level of implied volatility for short (long) maturities. The correlation creates the skew effect and the volatility of variance controls the smile.

The variance process remains positive if the volatility of variance $\theta$ is small enough with respect to the product of the mean reversion speed $\xi$ and the long variance level $\eta$ (i.e. $2\xi\eta > \theta^2$). As this constraint leads often to significantly worse fits to implied volatility surfaces it is in general not taken into account and we follow this approach.

The popularity of this model can probably be attributed to the semiclosed form of the prices of plain vanilla options. [9] showed that the price $C(K, T)$ of a European call option with strike $K$ and maturity $T$ is given by

$$C(K, T) = \frac{\exp\{-\alpha \ln(K)\}}{\pi} \int_0^{+\infty} \exp\{-i\ln(K)\} \psi_T(v) dv$$

for a (suitable) damping factor $\alpha > 0$. The function $\psi_T$ is given by

$$\psi_T(v) = \frac{\exp(-rT)\phi_T(v - (\alpha + 1))}{\alpha^2 + a - v^2 + i(2\alpha + 1)v}$$

where $\phi_T$ is the characteristic function of $\log(S_T)$. This characteristic function is given by

$$\phi_T(z) = \exp\left\{ -\frac{(z^2 + iz)V_0}{\gamma(z) \coth \frac{\gamma(z)T}{2} + \xi - i\rho z} \right\} \times \frac{\exp\left\{ \frac{\gamma(T)\gamma(-i\rho z)}{\theta} + izTr + i\gamma\log(S_0) \right\}}{(\cosh \frac{\gamma(z)T}{2} + \xi - i\rho z) \sinh \frac{\gamma(z)T}{2} \frac{2\gamma}{\theta^2}}$$

where $\gamma(z) = \sqrt{\theta^2(z^2 + 1)z} + (\xi - i\rho z)^2$, see e.g. [12].

For the calibration we minimize the absolute error of implied volatilities based on the root mean square error:

$$\text{ASE}_t \overset{\text{def}}{=} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( IV^\text{mod}(t) - IV^\text{mar}(t) \right)^2}$$
where \( \text{mod} \) refers to a model quantity, \( \text{mar} \) to a quantity observed on the market and \( IV(t) \) to an implied volatility on day \( t \). The index \( i \) runs over all \( n \) observations of the surface on day \( t \).

It is essential for the error functional \( \text{ASE}_t \) which observed prices are used for the calibration. As we investigate the pricing kernel for half a year to maturity we use only the prices of options that expire in less than 1.5 years. In order to exclude liquidity problems occurring at expiry we consider for the calibration only options with more than 1 month time to maturity. In the moneyness direction we restrict ourselves to strikes 50% above or below the spot for liquidity reasons.

The risk neutral density is derived by estimation of the model parameters by a least squares approach. This amounts to the minimization of the error functional \( \text{ASE}_t \). [15] provided evidence that such error functionals may have local minima. In order to circumvent this problem we apply a stochastic optimization routine that does not get trapped in a local minimum. To this end, we use the method of differential evolution developed by [30].

Having estimated the model parameters we know the distribution of \( X_T = \log S_T \) in form of the characteristic function \( \phi_T \), see (4). Then the corresponding density \( f \) of \( X_T \) can be recovered by Fourier inversion:

\[
f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \phi_T(t) dt,
\]

see e.g. [6]. This integral can be computed numerically.

Finally, the risk neutral density \( q \) of \( S_T = \exp(X_T) \) is given as a transformed density:

\[
q(x) = \frac{1}{x} f(\log(x)).
\]

This density \( q \) is risk neutral because it is derived from option prices and options are priced under the risk neutral measure. This measure is applied because banks replicate the payoff of options so that no arbitrage conditions determine the option price, see e.g. [28]. An estimated risk neutral density is presented in figure 2. It is estimated from the implied volatility shown in figure 3 for the day 24/03/2000. The distribution is right skewed and its mean is fixed by the martingale property. This implies that the density is low for high profits and high for high losses. Moreover, the distribution is not symmetrical around the neutral point where there are neither profits nor losses. For this and all the following estimations we approximate the risk free interest rates by the EURIBOR. On each trading day we use the yields corresponding to the maturities of the implied volatility surface. As the DAX is a performance index it is adjusted to dividend payments. Thus, we do not have to consider dividend payments explicitly.

### 3.3 Estimation of the historical density

While the risk neutral density is derived from option prices observed on the day of estimation we derive the subjective density from the historical time series of
the index. Hence, the two data sets are independent in the sense that the option prices reflect the future movements and the historical time series the past.

The estimation of the historical density seems more difficult than the estimation of the risk neutral density because the drift is not fixed and it depends in general on the length of the time series. Because of these difficulties we use different models and time horizons for the historical density: First, we estimate a GARCH in mean model for the returns. Returns are generally assumed to be stationary and we confirmed this at least in the time intervals we consider. The mean component in the GARCH model is important to reflect different market regimes. We estimate the GARCH model from the time series of the returns of the last two year because GARCH models require quite long time series for the estimation in order to make the standard error reasonably small. We do not choose longer time period for the estimation because we want to consider special market regimes. Besides this popular model choice we apply a GARCH model that converges in the limit to the Heston model that we used for the risk neutral density. As this model is also hard to estimate we use again the returns of the last 2 years for this model. Moreover, we consider directly the observed returns of the last year. The models and their time period for the estimation are presented in table 1. All these models give by simulation and smoothing the historical density for half a year ahead.

The GARCH estimations are based on the daily log-returns

\[ R_t = \log(S_t) - \log(S_{t-1}) \]
where \((S_t)\) denotes the price process of the underlying and \(t_i, \ i = 1, 2, \ldots\) denote the settlement times of the trading days. Returns of financial assets have been analyzed in numerous studies, see e.g. [14]. A model that has often been successfully applied to financial returns and their stylized facts is the GARCH(1,1) model. This model with a mean is given by

\[
R_t = \mu + \sigma_t Z_t \\
\sigma_t^2 = \omega + \alpha R_{t-1}^2 + \beta \sigma_{t-1}^2
\]

where \((Z_t)\) are independent identically distributed innovations with a standard normal distribution, see e.g. [17]. On day \(t_j\) the model parameters \(\mu, \omega, \alpha\) and \(\beta\) are estimated by quasi maximum likelihood from the observations of the last two years, i.e. \(R_{j-504}, \ldots, R_j\) assuming 252 trading days per year.

After the model parameters have been estimated on day \(t_j\) from historical data the process of logarithmic returns \((R_t)\) is simulated half a year ahead, i.e. until time \(t_j + 0.5\). In such a simulation \(\mu, \omega, \alpha\) and \(\beta\) are given and the time

**Fig. 3.** Implied volatility surface on 24/03/00.

**Table 1.** Models and the time periods used for their estimation.

<table>
<thead>
<tr>
<th>model</th>
<th>time period</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH in mean</td>
<td>2.0y</td>
</tr>
<tr>
<td>discrete Heston</td>
<td>2.0y</td>
</tr>
<tr>
<td>observed returns</td>
<td>1.0y</td>
</tr>
</tbody>
</table>
Empirical pricing kernels and investor preferences

series \((\sigma_i)\) and \((R_i)\) are unknown. The values of the DAX corresponding to the
simulated returns are then given by inverting the definition of the log returns:

\[ S_{t_i} = S_{t_{i-1}} \exp(R_i) \]

where we start with the observed DAX value on day \(t_j\). Repeating the simulation
\(N\) times we obtain \(N\) samples of the distribution of \(S_{t_j+0.5}\). We use \(N = 2000\)
simulations because tests have shown that the results become robust around this
number of simulations.

From these samples we estimate the probability density function of \(S_{t_j+0.5}\)
(given \((S_{t_{j-126}}, \ldots, S_{t_{j}})\)) by kernel density estimation. We apply the Gaussian
kernel and choose the bandwidth by Silverman’s rule of thumb, see e.g. [29].
This rule provides a trade-off between oversmoothing – resulting in a high bias
– and undersmoothing – leading to big variations of the density. We have more-
over checked the robustness of the estimate relative to this bandwidth choice.
The estimation results of a historical density are presented in figure 4 for the
day 24/03/2000. This density that represents a bullish market is has most of its
weight in the profit region and its tail for the losses is relatively light.

![Historical density on 24/03/2000 half a year ahead.](image)

As we use the Heston model for the estimation of the risk neutral density we
consider in addition to the described GARCH model a GARCH model that is a
discrete version of the Heston model. [21] show that the discrete version of the
square-root process is given by

\[ V_i = \omega + \beta V_{i-1} + \alpha(Z_{i-1} - \gamma \sqrt{V_{i-1}}) \]
and the returns are modelled by

$$R_i = \mu - \frac{1}{2} V_i + \sqrt{V_i} Z_i$$

where \((Z_i)\) are independent identically distributed innovations with a standard normal distribution. Having estimated this model by maximum likelihood on day \(t\), we simulate it half a year ahead and then smooth the samples of \(S_{t_j + 0.5}\) in the same way as in the other GARCH model.

In addition to these parametric models, we consider directly the observed returns over half a year

$$\tilde{R}_i = \frac{S_{t_{i+1}}}{S_{t_i}}$$.

In this way, we interpret these half year returns as samples from the distribution of the returns for half a year ahead. Smoothing these historical samples of returns gives an estimate of the density of returns and in this way also an estimate of the historical density of \(S_{t_j + 0.5}\).

### 3.4 Empirical pricing kernels

In contrast to many other studies that concentrate on the S&P500 index we analyze the German economy by focusing on the DAX, the German stock index. This broad index serves as an approximation to the German economy. We use two data sets: A daily time series of the DAX for the estimation of the subjective density and prices of European options on the DAX for the estimation of the risk neutral density.

In figure 5, we present the DAX in the years 1998 to 2004. This figure shows that the index reached its peak in 2000 when all the internet firms were making huge profits. But in the same year this bubble burst and the index fell afterwards for a long time. The historical density is estimated from the returns of this time series. We analyze the market utility functions in March 2000, July 2002 and June 2004 in order to consider different market regimes. We interpret 2000 as a bullish, 2002 as a bearish and 2004 as an unsettled market. These interpretations are based on table 2 that describes the changes of the DAX over the preceding 1 or 2 years. (In June 2004 the market went up by 11% during the previous 10 months.)

A utility function derived from the market data is a market utility function. It is estimated as an aggregate for all investors as if the representative investor existed. A representative investor is however just a convenient construction because the existence of the market itself implies that the asset is bought and sold, i.e. at least two counterparties are required for each transaction.

In section 2 we identified the market utility function (up to linear transformations) as

$$U(R) = \int_{R_0}^{R} K(x)dx$$
Market regimes in 2000, 2002 and 2004 described by the return $S_0/S_{0-\Delta}$ for periods $\Delta = 1.0\text{y}, 2.0\text{y}$.

<table>
<thead>
<tr>
<th>month</th>
<th>1.0y</th>
<th>2.0y</th>
</tr>
</thead>
<tbody>
<tr>
<td>03/2000</td>
<td>1.63</td>
<td>1.57</td>
</tr>
<tr>
<td>07/2002</td>
<td>0.66</td>
<td>0.54</td>
</tr>
<tr>
<td>06/2004</td>
<td>1.11</td>
<td>0.98</td>
</tr>
</tbody>
</table>

where $K$ is the pricing kernel for returns. It is defined by

$$K(x) = q(x)/p(x)$$

in terms of the historical and risk neutral densities $p$ and $q$ of returns. Any utility function (both cardinal and ordinal) can be defined up to a linear transformation, therefore we have identified the utility functions sufficiently. In section 3.3 we proposed different models for estimating the historical density. In figure 6 we show the pricing kernels resulting from the different estimation approaches for the historical density. The figure shows that all three kernels are quite similar: They have the same form, the same characteristic features like e.g. the hump and differ in absolute terms only a little. This demonstrates the economic equivalence of the three estimation methods on this day and this equivalence holds also for the other days. In the following we work with historical densities that are estimated by the observed returns.
Fig. 6. Empirical pricing kernel on 24/03/2000 (bullish market).

Besides the pricing kernel and the utility function we consider also the risk attitudes in the markets. Such risk attitudes are often described in terms of relative risk aversion that is defined by

$$ RRA(R) = -R \frac{U''(R)}{U'(R)}. $$

Because of $U' = cK = cq/p$ for a constant $c$ the relative risk aversion is also given by

$$ RRA(R) = -R \frac{q'(R)p(R) - q(R)p'(R)}{p^2(R)} \cdot \frac{q(R)}{p(R)} = R \left( \frac{p'(R)}{p(R)} - \frac{q'(R)}{q(R)} \right). $$

Hence, we can estimate the relative risk aversion from the estimated historical and risk neutral densities.

In figure 7 we present the empirical pricing kernels in March 2000, July 2002 and June 2004. The dates represent a bullish, a bearish and an unsettled markets, see table 2. All pricing kernels have a proclaimed hump located at small profits. Hence, the market utility functions do not correspond to standard specification of utility functions. We present the pricing kernels only in regions around the initial DAX (corresponding to a return of 1) value because the kernels explode outside these regions. This explosive behaviour reflects the typical pricing kernel form for losses. The explosion of the kernel for large profits is due to numerical problems in the estimation of the very low densities in this region. But we can see that in the
unsettled market the kernel is concentrated on a small region while the bullish
and bearish markets have wider pricing kernels. The hump of the unsettled
market is also narrower than in the other two regimes. The bullish and bearish
regimes have kernels of similar width but the bearish kernel is shifted to the
loss region and the bullish kernel is located mainly in the profit area. Moreover,
the figures show that the kernel is steeper in the unsettled markets than in the
other markets. But this steepness cannot be interpreted clearly because pricing
kernels are only defined up to a multiplicative constant.

![Graph](image)

**Fig. 7.** Empirical pricing kernel on 24/03/2000 (bullish), 30/07/2002 (bearish) and
30/06/2004 (unsettled or sidewards market).

The pricing kernels are the link between the relative risk aversion and the
utility functions that are presented in figure 8. These utility functions are only
defined up to linear transformations, see section 2. All the utility functions are
increasing but only the utility function of the bullish market is concave. This
concavity can be seen from the monotonicity of the kernel, see figure 7. Actually,
this non convexity can be attributed to the quite special form of the historical
density which has two modes on this date, see figure 4. Hence, we presume that
also this utility function has in general a region of convexity. The other two utility
functions are convex in a region of small profits where the bullish utility is almost
convex. The derivatives of the utility functions cannot be compared directly
because utility functions are identified only up to multiplicative constants. But
we can compare the ratio of the derivatives in the loss and profit regions for
the three dates because the constants cancel in these ratios. We see that the
derivatives in the loss region are highest in the bullish and lowest in the bearish
market and vice versa in the profit region. Economically these observations can
be interpreted in such a way that in the bullish market a loss (of 1 unit) reduces
the utility stronger than in the bearish market. On the other hand, a gain (of 1
unit) increases the utility less than in the bearish market. The unsettled market
shows a behaviour between these extreme markets. Hence, investors fear in a
good market situation losses more than in a bad situation and they appreciate
profits in a good situation less than in a bad situation.

Fig. 8. Market utility functions on 24/03/2000 (bullish), 30/07/2002 (bearish) and
30/06/2004 (unsettled or sidewards market).

Finally, we consider the relative risk aversions in the three market regimes.
These risk aversions are presented in figure 9, they do not depend on any con-
stants but are completely identified. We see that the risk aversion is smallest
in all markets for a small profit that roughly corresponds to the initial value
plus a riskless interest on it. In the unsettled regime the market is risk seeking
in a small region around this minimal risk aversion. But then the risk aversion
increases quite fast. Hence, the representative agent in this market is willing
to take small risks but is sensitive to large losses or profits. In the bullish and
bearish regimes the representative agent is less sensitive to large losses or profits
than in the unsettled market. In the bearish situation the representative agent
is willing to take more risks than in the bullish regime. In the bearish regime the
investors are risk seeking in a wider region than in the unsettled regime. In this sense they are more risk seeking in the bearish market. In the bullish market – on the other hand – the investors are never risk seeking so that they are less risk seeking than in the unsettled market.

Fig. 9. Relative risk aversions on 24/03/2000 (bullish), 30/07/2002 (bearish) and 30/06/2004 (unsettled or sidewards market).

The estimated utility functions most closely follow the specification proposed by [18]. The utility function proposed by [24] consists of one concave and one convex segment and is less suitable for describing the observed behaviour, see figure 10. Both utility functions were proposed to account for two opposite types of behaviour with respect to risk attitudes: buying insurance and gambling. Any utility function that is strictly concave fails to describe both risk attitudes. Most notable examples are the quadratic utility function with the linear pricing kernel as in the CAPM model and the CRRA utility function. These functions are presented in figure 10. Comparing this theoretical figure with the empirical results in figure 7 we see clearly the shortcoming of the standard specifications of utility functions to capture the characteristic hump of the pricing kernels.

4 Individual investors and their utility functions

In this section, we introduce a type of utility function that has two regions of different risk aversion. Then we describe how individual investors can be
aggregated to a representative agent that has the market utility function. Finally, we solve the resulting estimation problem by discretization and estimate the distribution of individual investors.

### 4.1 Individual Utility Function

We learn from figures 10 and 7 that the market utility differs significantly from the standard specification of utility functions. Moreover, we can observe from the estimated utility functions 8 that the loss part and the profit part of the utility functions can be quite well approximated with hyperbolic absolute risk aversion (HARA) functions, $k = 1, 2$:

$$U^{(k)}(R) = a_k (R - c_k)^\gamma_k + b_k,$$

where the shift parameter is $c_k$. These power utility functions become infinitely negative for $R = c_k$ and can be extended by $U^{(k)}(R) = -\infty$ for $R \leq c_k$, i.e. investors will avoid by all means the situation when $R \leq c_k$. The CRRA utility function has $c_k = 0$.

We try to reconstruct the market utility of the representative investor by individual utility functions and hence assume that there are many investors on
the market. Investor $i$ will be attributed with a utility function that consists of two HARA functions:

$$U_i(R) = \begin{cases} \max\{U(R,\theta_1,c_1);U(R,\theta_2,c_2,i)\}, & \text{if } R > c_1 \\ -\infty, & \text{if } R \leq c_1 \end{cases}$$

where $U(R,\theta,c) = a(R-c)^\gamma + b$, $\theta = (a,b,\gamma)^T$, $c_{2,i} > c_1$. If $a_1 = a_2 = 1$, $b_1 = b_2 = 0$ and $c_1 = c_2 = 0$, we get the standard CRRA utility function.

The parameters $\theta_1$ and $\theta_2$ and $c_1$ are the same for all investors who differ only with the shift parameter $c_2$, $\theta_i$ and $c_i$ are estimated from the lower part of the utility market function, where all investors probably agree that the market is “bad”. $\theta_2$ is estimated from the upper part of the utility function where all investors agree that the state of the world is “good”. The distribution of $c_2$ uniquely defines the distribution of switching points and is computed in section 4.3. In this way a bear part $U_{\text{bear}}(R) = U(R,\theta_1,c_1)$ and a bull part $U_{\text{bull}}(R) = U(R,\theta_1,c_2)$ can be estimated by least squares.

The individual utility function can then be denoted conveniently as:

$$U_i(R) = \begin{cases} \max\{U_{\text{bear}}(R);U_{\text{bull}}(R,c_i)\}, & \text{if } R > c_1; \\ -\infty, & \text{if } R \leq c_1. \end{cases}$$

Switching between $U_{\text{bear}}$ and $U_{\text{bull}}$ happens at the switching point $z$, whereas $U_{\text{bear}}(z) = U_{\text{bull}}(z,c_i)$. The switching point is uniquely determined by $c_i \equiv c_{2,i}$. The notations bear and bull have been chosen because $U_{\text{bear}}$ is activated when returns are low and $U_{\text{bull}}$ when returns are high.

Each investor is characterised by a switching point $z$. The smoothness of the market utility function is the result of the aggregation of different attitudes. $U_{\text{bear}}$ characterizes more cautious attitudes when returns are low and $U_{\text{bull}}$ describes the attitudes when the market is booming. Both $U_{\text{bear}}$ and $U_{\text{bull}}$ are concave. However, due to switching the total utility function can be locally convex.

These utility functions are illustrated in figure 11 that shows the results for the unsettled market. We observe/estimate the market utility function that does not correspond to standard utility approaches because of the convex region. We propose to reconstruct this phenomenon by individual utility functions that consist of a bearish part and a bullish part. While the bearish part is fixed for all investors the bullish part starts at the switching point that characterizes an individual investor. By aggregating investors with different switching points we reconstruct the market utility function. We describe the aggregation in section 4.2 and estimate the distribution of switching points in section 4.3. In this way we explain the special form of the observed market utility functions.

### 4.2 Market Aggregation Mechanism

We consider the problem of aggregating individual utility functions to a representative market utility function. A simple approach to this problem is to identify the market utility function with an average of the individual utility functions.
To this end one needs to specify the observable states of the world in the future by returns $R$ and then find a weighted average of the utility functions for each state. If the importance of the investors is the same, then the weights are equal:

$$U(R) = \frac{1}{N} \sum_{i=1}^{N} U_i(R),$$

where $N$ is the number of investors. The problem that arises in this case is that utility functions of different investors can not be summed up since they are incomparable.

Therefore, we propose an alternative aggregation technique. First we specify the subjective states of the world given by utility levels $u$ and then aggregate the outlooks concerning the returns in the future $R$ for each perceived state. For a subjective state described with the utility level $U$, such that

$$u = U_1(R_1) = U_2(R_2) = \ldots = U_N(R_N)$$

the aggregate estimate of the resulting returns is

$$R_A(u) = \frac{1}{N} \sum_{i=1}^{N} U_i^{-1}(u)$$

if all investors have the same market power. The market utility function $U_M$ resulting from this aggregation is given by the inverse $R_A^{-1}$. 

\[ \text{Fig. 11. Market utility function (solid) with bearish (dashed) and bullish (dotted) part of an individual utility function 5 estimated in the unsettled market of 30/06/2004.} \]
In contrast to the naive approach described at the beginning of this section, this aggregation mechanism is consistent under transformations: if all individual utility functions are changed by the same transformation then the resulting market utility is also given by the transformation of the original aggregated utility. We consider the individual utility functions $U_i$ and the resulting aggregate $U_M$. In addition, we consider the transformed individual utility functions $U_i^\phi(x) = \phi(U_i(x))$ and the corresponding aggregate $U_M^\phi$ where $\phi$ is a transformation. Then the aggregation is consistent in the sense that $U_M^\phi = \phi(U_M)$. This property can be seen from

\[
(U_M^\phi)^{-1}(u) = \frac{1}{N} \sum_{i=1}^{N} (U_i^\phi)^{-1}(u) \\
= \frac{1}{N} \sum_{i=1}^{N} U_i^{-1}\{\phi^{-1}(u)\} \\
= U_M^{-1}\{\phi^{-1}(u)\}
\]

The naive aggregation is not consistent in the above sense as the following example shows: We consider the two individual utility functions $U_1(x) = \sqrt{x}$ and $U_2(x) = \sqrt{x}/2$ under the logarithmic transformation $\phi = \log$. Then the naively aggregated utility is given by $U_M(x) = 3\sqrt{x}/4$. Hence, the transformed aggregated utility is $\phi(U_M(x)) = \log(3/4) + \log(x)/2$. But the aggregate of the transformed individual utility functions is

\[
U_M^\phi(x) = \frac{1}{2} \{\log(\sqrt{x}) + \log(\sqrt{x}/2)\} \\
= \frac{1}{2} \log \left(\frac{1}{2}\right) + \log(x)/2.
\]

This implies that $U_M^\phi \neq \phi(U_M)$ in general.

This described aggregation approach can be generalized in two ways: If the individual investors have different market power then we use the corresponding weights $w_i$ in the aggregation (6) instead of the uniform weights. As the number of market participants is in general big and unknown it is better to use a continuous density $f$ instead of the discrete distributions given by the weights $w_i$. These generalizations lead to the following aggregation

\[
R_A(u) = \int U^{-1}(\cdot, z)(u)f(z)dz
\]

where $U(\cdot, z)$ is the utility function of investor $z$. We assume in the following that the investors have utility function of the form described in section 4.1. In the next section we estimate the distribution of the investors who are parametrized by $z$. 
4.3 The Estimation of the Distribution of Switching Points

Using the described aggregation procedure, we consider now the problem of replicating the market utility by aggregating individual utility functions. To this end, we choose the parametric utility functions \( U(\cdot, z) \) described in 4.1 and try to recover with them the market utility \( U_M \). We do not consider directly the utility functions but minimize instead the distance between the inverse functions:

\[
\min f \left\| \int U^{-1}(\cdot, z)f(z)dz - U_M^{-1} \right\|_{L^2(\tilde{P})} 
\]

(7)

where \( \tilde{P} \) is image measure of the historical measure \( P \) on the returns under the transformation \( U_M \). As the historical measure has the density \( p \) the transformation theorem for densities implies that \( \tilde{P} \) has the density \( \tilde{p}(u) = p\{U^{-1}_M(u)\}/U'_M\{U^{-1}_M(u)\} \).

With this density the functional to be minimized in problem (7) can be stated as

\[
\int \left( \int U^{-1}(u, z)f(z)dz - U_M^{-1}(u) \right)^2 \tilde{p}(u) du
\]

\[
= \int \left( \int U^{-1}(u, z)f(z)dz - U_M^{-1}(u) \right)^2 p\{U^{-1}_M(u)\}/U'_M\{U^{-1}_M(u)\} du
\]

\[
= \int \left( \int U^{-1}(u, z)f(z)dz - U_M^{-1}(u) \right)^2 p\{U^{-1}_M(u)\}(U^{-1}_M)'(u) du
\]

because the derivative of the inverse is given by \((g^{-1})'(y) = 1/g'(g^{-1}(y))\). Moreover, we can apply integration by substitution to simplify this expression further

\[
\int \left( \int U^{-1}(u, z)f(z)dz - U_M^{-1}(u) \right)^2 p\{U^{-1}_M(u)\}(U^{-1}_M)'(u) du
\]

\[
= \int \left( \int U^{-1}U_M(x, z)f(z)dz - x \right)^2 p(x) dx.
\]

For replicating the market utility by minimizing (7) we observe first that we have samples of the historical distribution with density \( p \). Hence, we can replace the outer integral by the empirical expectation and the minimization problem can be restated as

\[
\min f \frac{1}{n} \sum_{i=1}^{n} \left( \int g\{U_M(x_i), z\}f(z)dz - x_i \right)^2
\]

where \( x_1, \ldots, x_n \) are the samples from the historical distribution and \( g = U^{-1} \).
Replacing the density $f$ by a histogram $f(z) = \sum_{j=1}^{J} \theta_j I_{B_j}(z)$ with bins $B_j$, $h_j = |B_j|$, the problem is transformed into

$$\min_{\theta_j} \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{J} \tilde{g}(i, j) \theta_j - x_i \right\}^2$$

where $\tilde{g}(i, j) = \int_{B_j} g\{U_M(x_i), z\} dz$.

Hence, the distribution of switching points can be estimated by solving the quadratic optimization problem

$$\min_{\theta_j} \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{J} \tilde{g}(i, j) \theta_j - x_i \right\}^2,$$

s.t. $\theta_j \geq 0$,

$$\sum_{j=1}^{J} \theta_j h_j = 1.$$

Such quadratic optimization problems are well known and their solutions can be obtained using standard techniques, see e.g. [25] or [32].

We present in figures 12–14 the estimated distribution of switching points in the bullish (24/03/2000), bearish (30/07/2002) and unsettled (30/06/2004) markets. The distribution density $f$ was computed for 100 bins but we checked the broad range of binwidths. The width of the distribution varies greatly depending on the regularisation scheme, for example as represented by the number of bins. The location of the distribution maximum, however, remains constant and independent from the computational method.

The maximum and the median of the distribution, i.e. the returns at which half of investors have bearish and bullish attitudes, depend on the year. For example, in the bullish market (Figure 12) the peak of the switching point distribution is located in the area of high returns around $R = 1.07$ for half a year. On the contrary, in the bearish market (Figure 13) the peak of switching points is around $R = 0.93$. This means that when the market is booming, such as in year 1999–2000 prior to the dot-com crash, investors get used to high returns and switch to the bullish attitude only for comparatively high $R$’s. An overall high level of returns serves in this respect as a reference level and investors form their judgements about the market relative to it. Since different investors have different initial wealth, personal habits, attitudes and other factors that our model does not take into account, we have a distribution of switching points. In the bearish market the average level of returns is low and investors switch to bullish attitudes already at much lower $R$’s.
5 Conclusion

We have analyzed in this paper empirical pricing kernels in three market regimes using data on the German stock index and options on this index. In the bullish, bearish and unsettled market regime we estimate the pricing kernel and derive the corresponding utility functions and relative risk aversions.

In the unsettled market of June 2004, the market investor is risk seeking in a small region around the riskless return but risk aversion increases fast for high absolute returns. In the bullish market of March 2000, the investor is on the other hand never risk seeking while he becomes more risk seeking in the bearish market of July 2002. Before the stock market crash in 1987 European options did not show the smile and the Black-Scholes model captured the data quite well. Hence, utility functions could be estimated at that times by power utility functions with a constant positive risk aversion. Our analysis shows that this simple structure does not hold anymore and discusses different structures corresponding to different market regimes.

The empirical pricing kernels of all market regimes demonstrate that the corresponding utility functions do not correspond to standard specifications of utility functions including [24]. The observed utility functions are closest to the general utility functions of [18]. We propose a parametric specification of these functions, estimate it and explain the observed market utility function by aggregating individual utility functions. In this way, we can estimate a distribution of individual investors.

The proposed aggregation mechanism is based on homogeneous investors in the sense that they differ only with switching points. Future research can...
Empirical pricing kernels and investor preferences

[Image: Utility Functions and Distribution of the Switching Points]

**Fig. 13.** Left panel: the market and fitted utility functions (indistinguishable in the graph). Right panel: the distribution of the reference points. 30 July 2002, a bearish market.

reveal how nonlinear aggregation procedures could be applied to heterogeneous investors.

**References**

Fig. 14. Left panel: the market and fitted utility functions (indistinguishable in the graph). Right panel: the distribution of the reference points. 30 June 2004, an unsettled market.

Calibration of a multiscale stochastic volatility model using European option prices

Lorella Fatone¹, Francesca Mariani², Maria Cristina Recchioni³ and Francesco Zirilli⁴

¹ Dipartimento di Matematica e Informatica, Università di Camerino
Via Madonna delle Carceri 9, 62032 Camerino (MC), Italy
lorella.fatone@unicam.it

² CERI- Centro di Ricerca Previsione Prevenzione e Controllo dei Rischi Geologici
Università di Roma “La Sapienza”, Piazza Umberto Piolozzi 9, 00038 Valmontone (RM), Italy
fra.mariani@libero.it

³ Dipartimento di Scienze Sociali “D. Serrani”, Università Politecnica delle Marche
Piazza Martelli 8, 60121 Ancona, Italy
m.c.recchioni@univpm.it

⁴ Dipartimento di Matematica “G. Castelnuovo”, Università di Roma “La Sapienza”
Piazzale Aldo Moro 2, 00185 Roma, Italy
f.zirilli@caspur.it

Abstract. In this paper we consider an explicitly solvable multiscale stochastic volatility model that generalizes the Heston model, and propose a new model calibration based on a nonlinear optimization problem. The model considered was introduced previously by the authors in [8] to describe the dynamics of an asset price and of its two stochastic variances. The risk neutral measure associated with the model and the risk premium parameters are introduced and the corresponding formulae to price call and put European vanilla options are derived. These formulae are given as one dimensional integrals of explicitly known integrands. We use these formulae to calibrate the multiscale model using European option prices as data, that is, to determine the values of the model parameters, of the correlation coefficients of the Wiener processes appearing in the model and of the initial stochastic variances implied by the “observed” option prices. The results obtained by solving the calibration problem are used to forecast future option prices. The calibration problem is translated into a suitable constrained nonlinear least squares problem. The proposed formulation of the calibration problem is applied to S&P 500 index data on the prices of European vanilla options in November 2005. This analysis points out some interesting facts.

Keywords. Multiscale stochastic volatility models, calibration model, option pricing.

* The numerical calculations and data analysis reported in this paper were performed using the computing resources of CASPUR (Roma, Italy) under the contract “Multiscale stochastic volatility models in finance and insurance” granted to the Università di Roma “La Sapienza”. The support and sponsorship of CASPUR are gratefully acknowledged.
1 Introduction

The use of stochastic volatility models to describe asset price dynamics originates from empirical evidence that the price dynamics of assets are driven by processes with nonconstant volatility. In fact, it is well known that difficulties arise when models with constant volatility such as the Black and Scholes model [2] are used to describe asset price dynamics. Examples of such difficulties are the so-called volatility “smile” that appears in the volatilities implied by the observed option prices, and the presence of skewness and kurtosis in the approximate asset price probability density function deduced from empirical data. Several alternative models have been proposed to overcome these shortcomings, including “mean reverting” stochastic volatility models such as the Heston model [12]. The Heston model provides a satisfactory description of the price dynamics of several relevant assets, as shown in [13] and [16]. Moreover, the Heston model is explicitly solvable in the sense that the joint probability density function associated with the asset price and its stochastic variance can be written as a one-dimensional integral of an explicitly known integrand. The Heston model is a one-factor stochastic volatility model since the stochastic variance of the asset price is modeled as a real stochastic process. However, several empirical studies of option price data have shown that the term structure of the implied volatility of the prices of several assets (e.g., market indices and commodities) seems to be driven by two factors, one fluctuating on a fast time scale and another fluctuating on a longer time scale (see, for example, Alizadeh, Brandt, and Diebold [1]). We can conclude that, in some circumstances, one-factor stochastic volatility models are unable to fully capture the volatility smile and volatility dynamics (see [9], [6], [3], [5]). To overcome this inadequacy, several models that go beyond one-factor stochastic volatility models have been proposed; we group these models into two classes: multiscale stochastic volatility models [8], [11], [17] and jump models [15], [4]. Here we will concentrate on the multiscale stochastic volatility models. Several authors have devoted attention to such models, for example, Fouque, Papanicolaou, Sircar and Solna [10] developed a multiscale stochastic volatility model starting from ideas introduced in [1]. In [10], the proposed model is calibrated on real data to capture the volatility smile, and an explicit series expansion of the formula for the price of a European vanilla call option in the model considered in [10] is given. More recently, Wong and Chan [17] proposed a different multiscale stochastic volatility model and used it to price a long-term financial product called dynamic fund protection. Moreover, they reported a series expansion for the formula for the price of lookback options. Finally, the authors proposed a new multiscale stochastic volatility model that generalizes the Heston model and describes the dynamics of an asset price and of its two
stochastic variances using a system of three Ito stochastic differential equations [8]. The two stochastic variances vary on two different time scales. Under some hypotheses, the proposed model provides “explicitly” solvable and “easy to use” formulae to price in the model call and put European vanilla options can be deduced.

In this paper, we consider the multiscale model proposed in [8] and we use the formulae to price call and put European vanilla options deduced in [8] to solve a calibration problem, and a forecasting problem. The calibration problem can be stated as follows: given the observed prices of call and put European vanilla options on a given asset traded on a given day, determine the model parameters, the correlation coefficients of the Wiener processes appearing in the model and the initial stochastic variances implied by the observed prices. We use these implied values to forecast the prices of the options in the days following the day whose option prices have been used as data in the calibration. We translate the calibration problem into a constrained least squares optimization problem that is a generalization of the optimization problem considered in [8]. In fact, in [8], the objective function is defined using only the observed prices of the European call options. Here the objective function is defined using the observed prices of both call and put options. The new optimization problem formulated using both call and put prices has some interesting consequences, discussed in Section 3, compared to the optimization problem considered in [8]. In the numerical experiment (see Section 3) we consider European vanilla options on the S&P 500 index and we use the results of the calibration to forecast option prices. The forecasted option prices are compared to the observed prices; the results of this comparison are very satisfactory. A more detailed discussion of the multiscale model and a more extended analysis of the 2005 data relative to the S&P 500 and to its options in the year 2005 can be found in [8]. The website: http://www.econ.univpm.it/recchioni/finance/w7 contains auxiliary material including animations that assists in the understanding of this paper. More general information on the work of the authors and of their coauthors in mathematical finance can be found on the website http://www.econ.univpm.it/recchioni/finance.

The remainder of the paper is organized as follows. In Section 2, we describe the multiscale stochastic volatility model considered and, under some hypotheses, we derive an integral representation formula for its transition probability density function and for the price of European vanilla call and put options under the risk-neutral measure. In Section 3, we formulate and solve a calibration problem and a forecasting problem. We use observed option prices on the S&P 500 index to test the solution method of the calibration problem and the forecasting procedure. In Section 4, we present our conclusions.

2 The multiscale stochastic volatility model

Let $\mathbb{R}$ and $\mathbb{R}^+$ be sets of real and positive real numbers, respectively, and let $t$ be a real variable that denotes time. We consider the (vector valued real) stochastic process $(x_t, v_{1,t}, v_{2,t})$, $t > 0$, solution of the following system of stochastic
differential equations:

\[ dx_t = (\dot{\mu} + a_1 v_1 + a_2 v_2) dt + b_1 \sqrt{v_1} dW_t^{0,1} + b_2 \sqrt{v_2} dW_t^{0,2}, \quad t > 0, \]
\[ dv_{1,t} = \chi_1 (\theta_t - v_{1,t}) dt + \varepsilon_1 \sqrt{v_1} dW_t^1, \quad t > 0, \]  
\[ dv_{2,t} = \chi_2 (\theta_t - v_{2,t}) dt + \varepsilon_2 \sqrt{v_2} dW_t^2, \quad t > 0, \]  

where the quantities \( \dot{\mu}, a_1, b_1, \chi_1, \varepsilon_1, \theta_t, i = 1, 2, \) are real constants. The quantity \( \dot{\mu} \) is known as the drift rate. Note that elementary considerations suggest that we must require \( \chi_i \geq 0, \varepsilon_i \geq 0, \theta_i \geq 0, i = 1, 2. \) Moreover we require \( \frac{2 \chi_i \varepsilon_i}{\varepsilon_i} > 1, i = 1, 2. \) The condition \( 2 \chi_i \varepsilon_i > 1 \) guarantees that when \( v_{i,t} \) is positive with probability one at time \( t = 0, v_{i,t} \), the solution of (2), or of (3), remains positive with probability one for \( t > 0, i = 1, 2. \) Finally \( W_t^{0,1}, W_t^{0,2}, W_t^1, W_t^2, t > 0 \) are standard Wiener processes such that \( W_t^{0,1} = W_0^{0,1} = W_0^1 = W_0^2 = 0, \) and \( dW_t^{0,1}, dW_t^{0,2}, dW_t^1, dW_t^2, t > 0, \) are their stochastic differentials, and we assume that:

\[ \mathbb{E}(dW_t^{0,1} dW_t^{0,2}) = \mathbb{E}(dW_t^1 dW_t^1) = \mathbb{E}(dW_t^2 dW_t^2) = 0, t > 0, \] \[ \mathbb{E}(dW_t^{0,1} dW_t^1) = \rho_{0,1} dt, \quad \mathbb{E}(dW_t^{0,2} dW_t^2) = \rho_{0,2} dt, \quad t > 0. \] 

where \( \mathbb{E}(\cdot) \) denotes the expected value of \( \cdot, \) and \( \rho_{0,1}, \rho_{0,2} \in [-1,1] \) are constants known as correlation coefficients. Note that the autocorrelation coefficients of the stochastic differentials are equal to one (see [8] for further details).

We interpret \( x_t, t > 0, \) as the log-return of the asset price and \( v_{1,t}, v_{2,t}, t > 0, \) as the stochastic variances of \( x_t, t > 0. \) The fact that \( v_{1,t}, v_{2,t}, t > 0, \) are stochastic variances on different time scales is translated in the condition \( \chi_1 << \chi_2. \) With the above interpretation of \( (x_t, v_{1,t}, v_{2,t}), t > 0, \) the assumptions (4)-(5) appear natural.

Equations (1), (2) and (3) must be equipped with an initial condition, that is:

\[ x_0 = \tilde{x}_0, \quad v_{1,0} = \tilde{v}_{1,0}, \quad v_{2,0} = \tilde{v}_{2,0}, \] 

where \( \tilde{x}_0, \tilde{v}_{i,0}, i = 1, 2, \) are random variables that we assume to be concentrated in a point with probability one. For simplicity, we identify the random variables \( \tilde{x}_0, \tilde{v}_{i,0}, i = 1, 2, \) with the points where they are concentrated. Without loss of generality, we can choose \( \tilde{x}_0 = 0. \) Moreover we assume \( \tilde{v}_{i,0} \in \mathbb{R}^+, i = 1, 2, \) the quantities \( \tilde{v}_{i,0}, i = 1, 2, \) cannot be observed in real markets. Moreover, selecting values of \( a_1 = -1/2, a_2 = 0, b_1 = 1, b_2 = 0 \) in equations (1), (2), (3) corresponds to the fact that equations (1) and (2) are decoupled from equation (3), and that these values give the Heston model. In this sense the model (1), (2), (3) is a generalized version of the Heston model.

Let \( p_f(x, v_1, v_2, t, t', x', v'_1, v'_2, t'), (x, v_1, v_2), (x', v'_1, v'_2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+, t, t' \geq 0, t - t' > 0, \) be the transition probability density function associated with the stochastic differential system (1), (2), (3), that is, the probability density function of having \( x_t = x, v_{1,t} = v_1, v_{2,t} = v_2 \) given the fact that \( x_{t'} = x', v_{1,t'} = v'_1, v_{2,t'} = v'_2, \) when \( t - t' > 0. \) The transition probability density function \( p_f(x, v_1, v_2, t, t', x', v'_1, v'_2, t'), (x, v_1, v_2), (x', v'_1, v'_2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+, t, t' \geq 0, t - t' > 0, \)
0, as a function of the variables \((x, v_1, v_2, t)\) is the solution of the *Fokker Planck* equation with suitable initial and boundary conditions (see [8] for further details) and, as function of the variables \((x', v'_1, v'_2, t')\), satisfies the following backward equation:

\[
- \frac{\partial p_f}{\partial t'} = \frac{1}{2} \left( \frac{\partial^2 p_f}{\partial x'^2} + \frac{1}{2} \varepsilon_1^2 v'_1 \frac{\partial^2 p_f}{\partial v'_1^2} + \frac{1}{2} \varepsilon_2^2 v'_2 \frac{\partial^2 p_f}{\partial v'_2^2} + \varepsilon_b \rho_{0,1} v'_1 \frac{\partial^2 p_f}{\partial x' \partial v'_1} \right) + \varepsilon_2 b \rho_{0,2} v'_2 \frac{\partial p_f}{\partial x'} + \chi_1 (\theta_1 - v'_1) \frac{\partial p_f}{\partial v'_1} + \chi_2 (\theta_2 - v'_2) \frac{\partial p_f}{\partial v'_2} + (\tilde{\mu} + a_1 v'_1 + a_2 v'_2) \frac{\partial p_f}{\partial x'}
\]

with the final condition:

\[
p_f(x, v_1, v_2, t, x', v'_1, v'_2, t') = \delta(x' - x) \delta(v'_1 - v_1) \delta(v'_2 - v_2),
\]

\((x, v_1, v_2), (x', v'_1, v'_2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+, t \geq 0, t > t', 0 \leq t' < t, \)

and the appropriate boundary conditions (see [8], Section 2).

Letting \(\tau = t - t'\), we assume that the following integral representation formula for \(p_f(x, v_1, v_2, t, x', v'_1, v'_2, t')\), \((x, v_1, v_2), (x', v'_1, v'_2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+, t' \geq 0, t > t',\) holds:

\[
p_f(x, v_1, v_2, t, x', v'_1, v'_2, t') = \frac{1}{(2\pi)^5} \frac{2}{\varepsilon_1^2 \varepsilon_2^2} \int_{\mathbb{R}} dk e^{ikx} \int_{\mathbb{R}} dl_1 e^{\frac{2}{\varepsilon_1} dl_1} \int_{\mathbb{R}} dl_2 e^{\frac{2}{\varepsilon_2} dl_2} f(\tau, k, l_1, l_2, x', v'_1, v'_2),
\]

\((x, v_1, v_2), (x', v'_1, v'_2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+, \tau = t - t' > 0, \)

where \(i\) is the imaginary unit, \(f\) is the Fourier transform of the function obtained extending with zero the function \(p_f\) defined above as a function of the variables \((x, v_1, v_2)\) when \(v_1 \notin \mathbb{R}^+\) and/or \(v_2 \notin \mathbb{R}^+\), and \(k, l_1,\) and \(l_2\) are the conjugate variables of \(x, v_1,\) and \(v_2\) respectively (see [8], Section 2, for further details).

Using the arguments presented in [14] (pages 602-605), in [8] it was shown that:

\[
f(\tau, k, l_1, l_2, x', v'_1, v'_2) = e^{-ikx} e^{iA(\tau, k, l_1, l_2)} e^{-\frac{2}{\varepsilon_1} v'_1 B_1(\tau, k, l_1)} e^{-\frac{2}{\varepsilon_2} v'_2 B_2(\tau, k, l_2)},
\]

\((x', v'_1, v'_2) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+, (k, l_1, l_2) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}, \tau > 0, \)

where the functions \(A\) and \(B_i, i = 1, 2\) are given by:

\[
A(\tau, k, l_1, l_2) = -ik\tilde{\mu}\tau - \sum_{i=1}^{2} \left[ \frac{2\chi_i \theta_i}{\varepsilon_i^2} \left( \nu_i + \zeta_i \right) \tau + \ln \left( \frac{\left( \nu_i + \zeta_i - i l_i \right) e^{-2\zeta_i \tau} + (i l_i - \nu_i + \zeta_i)}{2\zeta_i} \right) \right],
\]

\((k, l_1, l_2) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}, \tau > 0, \)

\[
B_i(\tau, k, l_i) = \frac{(\nu_i - \zeta_i)(\nu_i + \zeta_i - i l_i) e^{-2\zeta_i \tau} + (\nu_i + \zeta_i)(i l_i - \nu_i + \zeta_i)}{(\nu_i + \zeta_i - i l_i) e^{-2\zeta_i \tau} + (i l_i - \nu_i + \zeta_i)},
\]

\((k, l_i) \in \mathbb{R} \times \mathbb{R}, \tau > 0, i = 1, 2, \)
where

\begin{align}
\nu_i &= -\frac{1}{2} (\chi_i + i k b_i \varepsilon_{i,k} \rho_{0,i}), \quad k \in \mathbb{R}, \quad i = 1, 2, \quad (13) \\
\zeta_i &= \frac{1}{2} (4 \nu_i^2 + \varepsilon_i^2 (b_i^2 k^2 + 2 i k a_i))^{1/2}, \quad k \in \mathbb{R}, \quad i = 1, 2. \quad (14)
\end{align}

Substituting equations (11) and (12) into equation (9) and integrating with respect to the variables \( l_1 \) and \( l_2 \) we can derive a representation formula for \( p_f(x, v_1, v_2, t, x', v_1', v_2', t') \), that is, for the transition probability density function of the stochastic process solution of (1), (2), (3), as a one dimensional integral of an explicitly known integrand (see [8] for further details). The formulae for pricing European vanilla call and put options with strike price \( K > 0 \) and maturity time \( T \) are derived from (9) using the no arbitrage pricing theory, that is, computing the option prices as expected values of a discounted payoff with respect to an equivalent martingale measure, also known as a risk-neutral measure (see for example [7], [15]). However to keep the exposition simple and since we use as data in the formulation of the calibration problem only option prices we can use to compute in the model the option prices the statistical measure associated with the model (1), (2), (3) whose density is given by (9), that is, we can incorporate the risk premium parameters into the parameters \( \chi_i \) and \( \theta_i \), \( i = 1, 2 \). In fact, in order to consider the risk neutral measure associated with (1), (2), (3), we should simply replace the parameters \( \chi_i \), \( \theta_i \), \( i = 1, 2 \) appearing in (1), (2), (3) with the parameters \( \tilde{\chi}_i \equiv \chi_i + \lambda_i, \tilde{\theta}_i \equiv \chi_i \theta_i / (\chi_i + \lambda_i), \quad i = 1, 2 \), where \( \lambda_i \in \mathbb{R}, \quad i = 1, 2 \) are the risk premium parameters (see [8] for more details) and we should impose the constraints \( \tilde{\chi}_i \geq 0, \tilde{\theta}_i \geq 0, \quad i = 1, 2 \). Now, writing the transition probability density function \( p_f \) as follows \( p_f(x, v_1, v_2, t, x', v_1', v_2', t') = \tilde{p}_f(x, v_1, v_2, t, x', v_1', v_2', t') e^{-2(\tilde{\theta} - \tilde{\chi}^2)}, (x, v_1, v_2), (x', v_1', v_2') \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+, t > 0, t' \geq 0, t - t' > 0 \), deriving a representation formula for \( \tilde{p}_f(x, v_1, v_2, t, x', v_1', v_2', t') \), and using the backward equation (7), we obtain the following formula for the price at time \( t = 0 \) of a European vanilla call option with time to maturity \( \tau = T - t > 0 \) (remember that \( t = 0 \) and strike price \( K \), when at time \( t = 0 \) the price of the underlying asset is given by \( S_0 \) and the stochastic variances are given by \( \tilde{v}_{1,0}, \tilde{v}_{2,0} \):

\begin{align}
C(\tau, K, S_0, \tilde{v}_{1,0}, \tilde{v}_{2,0}) &= \frac{S_0}{2 \pi} e^{-r \tau} e^{2 \tilde{\theta} \tau} \int_{-\infty}^{+\infty} -k^2 - 3k + 2 \cdot \frac{d\tilde{k}}{2\tau} \cdot \left( e^{-ix_0 \theta_i (\nu_i^2 + \zeta_i^2 + \log(s_{i,b}^2/(2\nu_i^2)))} \right) e^{-2i\tilde{\varepsilon}_{i,b} (\zeta_i^2 - \nu_i^2 s_{i,g}^2/(2\nu_i^2 s_{i,b}^2))}, \\
\quad \tau > 0, S_0 > 0, \tilde{v}_{1,0}, \tilde{v}_{2,0} > 0, \quad (15)
\end{align}

where \( r \) is the discount rate and the quantities \( \nu_i^c, \zeta_i^c, s_{i,b}^c, s_{i,g}^c \), \( i = 1, 2 \) are given by:

\begin{align}
\nu_i^c &= -\frac{1}{2} (\chi_i + i k b_i \varepsilon_{i,k} \rho_{0,i} - 2b_i \rho_{0,i} \varepsilon_{i}), \quad k \in \mathbb{R}, \quad i = 1, 2, \quad (16)
\end{align}
Calibration of a multiscale stochastic volatility model

\[ \zeta_i = \frac{1}{2} \left( 4(\nu_i^c)^2 + \varepsilon_i^2 (b_i^2)^2 + 2k a_i - 2k (a_i + b_i^2) \right)^{1/2}, \quad k \in \mathbb{R}, \quad i = 1, 2, \]

\[ s_{i,g}^c = 1 - e^{-2\zeta_i^c \tau}, \quad s_{i,b}^c = \zeta_i^c - \nu_i^c + (\zeta_i^c + \nu_i^c) e^{-2\zeta_i^c \tau}, \quad \tau > 0, \quad i = 1, 2. \]

Note that because we work with the risk neutral measure, the discount rate \( r \) must be chosen equal to \( \hat{\mu} \) (see formula (14) in [12]). We note that the relation between the log-returns \( x_t, t \geq 0 \) and the price \( S_t, t \geq 0 \), of the underlying asset is \( x_t = \log S_t / S_t, t > 0 \).

The formula for the price at time \( t = 0 \) of a European vanilla put option with time to maturity \( \tau = T-t > 0 \), (remember that \( t = 0 \)) and strike price \( K \), when at time \( t = 0 \) the price of the underlying asset is \( S_0 \) and the stochastic variances are \( \tilde{v}_{1,0}, \tilde{v}_{2,0} \) is given by:

\[
P(\tau, K, S_0, \tilde{v}_{1,0}, \tilde{v}_{2,0}) = \frac{K}{2\pi} e^{-\frac{\tau}{\hat{\mu}}} \int_{-\infty}^{+\infty} \frac{e^{-2\tau \theta} e^{i \theta \tau} \int_{-\infty}^{+\infty} \frac{e^{-ik(\log(S_0/K) + \hat{\mu} \tau) - \log(S_0/K)}}{-k^2 + 2i k + 2}} {\epsilon_{\sigma_i, \sigma_i}(2\zeta_i^p)} \tau/\varepsilon_i^2 e^{-2\varepsilon_i, \theta((\zeta_i^p)^2 - (\nu_i^p)^2) s_{i,b}^p/(\zeta_i^p s_{i,a}^p)}, \quad \tau > 0, \quad S_0 > 0, \quad \tilde{v}_{1,0}, \tilde{v}_{2,0} > 0,
\]

where the quantities \( \nu_i^p, \zeta_i^p, s_{i,g}^p \), and \( s_{i,b}^p \), are given by:

\[
\nu_i^p = -\frac{1}{2} (\chi_i + i k b_i \varepsilon_i \rho_{0,i} + b_i \rho_{0,i} \varepsilon_i), \quad k \in \mathbb{R}, \quad i = 1, 2,
\]

\[
\zeta_i^p = \frac{1}{2} \left( 4(\nu_i^p)^2 + \varepsilon_i^2 (b_i^2)^2 + 2k a_i - 2k (a_i + b_i^2) \right)^{1/2}, \quad k \in \mathbb{R}, \quad i = 1, 2,
\]

\[
s_{i,g}^p = 1 - e^{-2\zeta_i^p \tau}, \quad s_{i,b}^p = \zeta_i^p - \nu_i^p + (\zeta_i^p + \nu_i^p) e^{-2\zeta_i^p \tau}, \quad \tau > 0, \quad i = 1, 2.
\]

Note that the price of European call and put options at time \( t, 0 < t < T \) can be deduced from (15) and (19) with some obvious changes.

3 Calibration and forecasting problems: some experiments with real data

In this section, we conduct experiments on real data, using values of \( a_1 = a_2 = -\frac{1}{2} \) and \( b_1 = b_2 = 1 \) in equations (1), (2), (3). Let \( \Theta = (\varepsilon_1, \theta_{1,1}, \chi_1, \varepsilon_{0,1}, \hat{\mu}, \varepsilon_2, \theta_2, \rho_{0,2}, \chi_2, \varepsilon_{0,2}) \) be a vector comprised of the parameters of the multiscale model (note that the risk premium parameters can be included in the parameters \( \chi_i, \theta_i, i = 1, 2 \), of the correlation coefficients, and of the initial stochastic variances. Let \( m_c, m_p \) be two positive integers; we denote the data of the calibration problem with \( C^t(\bar{S}_t, T_i, K_i), i = 1, 2, \ldots, m_c \), and, \( P^t(\bar{S}_t, T_i, K_i), i = 1, 2, \ldots, m_p \), that is, the observed prices at time \( t \) of European vanilla call and put options, respectively, having maturity time \( T_i \) and strike price \( K_i, i = 1, 2, \ldots, \max(m_c, m_p), \) when the price of the underlying asset at time \( t \) is \( S_t \). Moreover, we denote the
prices of European vanilla call and put options obtained using (15) and (19) as $C_{t,\Theta}(\tilde{S}_t, T_i, K_i), i = 1, 2, \ldots, m_c$ and $P_{t,\Theta}(\tilde{S}_t, T_i, K_i), i = 1, 2, \ldots, m_p$, respectively, and choose the maturity time $\tau = \tau_i = T_i - t, i = 1, 2, \ldots, \max(m_c, m_p)$ and the asset price $S_0 = \tilde{S}_t$. Note that usually when a call option is traded for a couple $T_i, K_i$ the corresponding put option is also traded and vice versa; this is implicit in our assumption that it is possible to denote with $T_i, K_i, i = 1, 2, \ldots, \max(m_c, m_p)$ the couple maturity time, strike price of the option prices used as data. If necessary, our notation can be easily generalized to handle the data that are actually available.

Fig. 1. November 7, 2005: European vanilla call and put option prices (V) on the S&P 500 index forecasted using the multiscale model and prices observed in the market versus moneyness $K/S_0$

Let $\mathbb{R}^{11}$ be the 11-dimensional real Euclidean vector space and let $\mathcal{M}$ be the set of admissible vectors $\Theta$, that is:

$$\mathcal{M} = \{ \Theta = (\epsilon_1, \theta_1, \rho_{0,1}, \chi_1, \tilde{v}_{0,1}, \bar{\mu}, \epsilon_2, \theta_2, \rho_{0,2}, \chi_2, \tilde{v}_{0,2}) \in \mathbb{R}^{11} | \epsilon_i \chi_i, \theta_i \geq 0, i = 1, 2, \frac{2\chi_i \theta_i}{\epsilon_i} \geq 1, -1 \leq \rho_{0,i} \leq 1, \tilde{v}_{0,i} \geq 0, i = 1, 2 \}.$$

\[ (23) \]
at time $t$, $t \geq 0$. We calibrate the model (1), (3), (2) by solving the following constrained nonlinear least squares problem:

$$
\min_{\Theta \in \mathcal{M}} L_t(\Theta), \ t \geq 0, \tag{24}
$$

where the objective function $L_t(\Theta)$, $t \geq 0$, is defined as follows:

$$
L_t(\Theta) = \sum_{i=1}^{m_c} \left[ C^t(\tilde{S}_t, T_i, K_i) - C^t(\tilde{S}_t, T_i, K_i) \right]^2 + \sum_{i=1}^{m_p} \left[ P^t(\tilde{S}_t, T_i, K_i) - P^t(\tilde{S}_t, T_i, K_i) \right]^2, \ t \geq 0. \tag{25}
$$

The optimization problem (24) is a translation of the calibration problem for the model (1), (2), (3) stated in Section 1.

We solve the optimization problem (24) using a projected steepest descent method. This method is an iterative scheme that, starting from an initial vector $\Theta^0 \in \mathcal{M}$, generates a sequence $\{\Theta^n\}$, $n = 0, 1, \ldots$, of vectors $\Theta^n \in \mathcal{M}$, $n =$
0.1, . . . , moving along a descent direction obtained via a suitable projection on the constraints of the vector given by the negative of the gradient with respect to $\Theta$ of $L_t$. The procedure stops when the vector $\Theta^n$ generated satisfies for the first time the following condition:

$$L_t(\Theta^n) \leq e_{tol}, \text{ or } n > n_{max},$$

(26)

where $e_{tol}$, $n_{max}$ are positive constants that will be chosen as outlined below.

Experiments using synthetic and real data that show the adequacy of a formulation of the calibration problem similar to (23), (24), (25) and its ability to capture satisfactorily the “smile” effect can be found in [8].

Let us consider the values of the model parameters, the correlation coefficients, and the initial stochastic variances implied by the observed prices of the European vanilla call and put options on the S&P 500 index and by the value of the S&P 500 index in November 2005. The S&P 500 index is one of the leading indices of the New York Stock Exchange. We solve the calibration problem (24) using the call and put option prices available to us relative to the prices on November 3, 2005, and we have $m_c = 303$ and $m_p = 284$. The implied values obtained by solving the calibration problem using the data from November 3, 2005 are used to forecast the option prices of November 7 ($m_c = 303$, $m_p = 290$), November 14 ($m_c = 305$, $m_p = 295$), and November 28 ($m_c = 292$, $m_p = 265$), 2005. Note that since we use all available prices, the calibration procedure works simultaneously on out of money, at the money and in the money call and put option prices. In the stopping criterion (26), we use $e_{tol} = 0.07$ and $n_{max} = 10000$.

In the forecasting of option prices, we assume that the underlying asset price is known on the forecasting day and we forecast the values of the stochastic variances $v_{1,t}$, $v_{2,t}$, $t > t_0 = 0$ on the forecasting day. In particular, starting from the stochastic variances $\tilde{v}_{1,0}$, $\tilde{v}_{2,0}$ at time $t = t_0 = 0$ obtained from the calibration procedure, we forecast the stochastic variances using the mean values $\hat{v}_{1,t|\Theta}$, $\hat{v}_{2,t|\Theta}$, $t > t_0 = 0$, of the random variables $v_{1,t}$, $v_{2,t}$, $t > 0$, that is, we use the formulae (see [8] for further details):

$$\hat{v}_{1,t|\Theta} = E(v_{1,t|\Theta}) = \theta_1(1 - e^{-\chi_1(t-t_0)}) + e^{-\chi_1(t-t_0)}\tilde{v}_{1,0}, \ t > t_0 = 0,$$

(27)

$$\hat{v}_{2,t|\Theta} = E(v_{2,t|\Theta}) = \theta_2(1 - e^{-\chi_2(t-t_0)}) + e^{-\chi_2(t-t_0)}\tilde{v}_{2,0}, \ t > t_0 = 0.$$

(28)

Note that in our experiment, $t = t_0 = 0$ corresponds to November 3, 2005, and that we have assumed that the underlying asset price is known on the forecasting days (November 7, 14, and 28, 2005). The dates November 3, 7, 14, and 28 were chosen to show that the model parameters calibrated using data from the beginning of the month (November 3) can be used to accurately forecast the option prices approximately a week (November 7), two weeks (November 14) and a month (November 28) into the future.

Figures 1, 2, and 3 show the forecast prices of European vanilla call and put options on November 7, 14, and 28, 2005, respectively, compared with the observed prices. The agreement between the observed and forecast prices is very satisfactory for all dates, indicating that the future prices of the options could be
Fig. 3. November 28, 2005: European vanilla call and put option prices (V) on the S&P 500 index forecasted using the multiscale model and prices observed in the market versus moneyness K/S_0

well forecast using the the implied values obtained from the prices of November 3, 2005 in conjunction with formulae (15), (19), (27), (28) and the price of the underlying asset on the forecasting day. Comparison of the calibration procedure employed here, which uses both call and put option prices, and the calibration procedure of [8], which uses only call option prices, discloses that the procedure proposed here provides substantially better forecasts of at the money and out of the money option prices. The superiority of the proposed method becomes more evident as the time to maturity increases.

4 Conclusions

Previously, the authors showed [8] that the two factor model (1), (2), (3) captures the volatility smile better than the one factor model [12] and provides high-quality forecasts of European vanilla option prices when the time to maturity is large. Here we have presented a calibration procedure that yields improved forecasts of at the money and out of the money option prices; this improvement can be attributed to the proposed procedure giving more weight to these prices than the procedure used in [8]. In fact, these prices are small compared with those of in the money options, and hence play a minor role in the minimization
of the objective function (25) especially in the first steps of the minimization procedure. When only the call option prices are used in the objective function (as in [8]), the number of in the money and out of the money options traded in a given day can vary substantially depending on the asset price. When we add the put option prices as data, however, we balance the portfolio of options used as data. In fact, the call and put options traded in a given day usually have the same strike price and maturity time, so that when a put option is in the money the corresponding call option is out of the money and vice versa. Moreover, by adding the put option prices, we increase the number of data on at the money options, that is, we increase the weight of the at the money options data in the calibration. The above considerations account for the observed improvement in the quality of the forecasted prices of the at the money and out of the money options when we substitute the calibration procedure of [8] with the calibration procedure suggested here. These findings suggest that use of a weighted least squares procedure may further improve the model calibration. Finally we point out that the calibration problem studied here can be generalized by considering the parameters $a_i$, $b_i$, $i = 1, 2$, as unknowns to be determined in the calibration. The use of these four extra parameters could potentially improve the ability of the model to describe the financial data considered.

References

A reserve risk model
for a non-life insurance company

Salvatore Forte¹, Matteo Ialenti¹, and Marco Pirra²

¹ Department of Actuarial Sciences, Sapienza University of Roma,
Viale Regina Elena 295, 00185 Roma, Italy
{Salvatore.Forte, Matteo.Ialenti}@uniroma1.it

² Department of Management Sciences, University of Calabria,
Ponte Bucci Cubo 3 C, Rende (CS), 87030, Italy
Marco.Pirra@unical.it

Abstract. The aim of the study is the assessment of the reserve risk for
a non-life insurance company that writes two different lines of business;
the focus is on the determination of the current estimates incorporated
in the measurement of the liabilities (in some jurisdictions referred to as
technical provisions or actuarial reserves) of insurance contracts (with-
out risk margins) through a Bayesian stochastic methodology and on the
possible methods for the determination of risk margins above current es-
timates appropriate for the measurement of the liabilities for insurance
contracts for regulatory and general purpose financial reports. A new for-
mula for the calculation of the latter is proposed, analyzing its potential
advantages in comparison with the existing ones. In order to determine
the risk margin the reserve risk is calibrated on a one year time horizon
considering both the case of independence between lines of business and
the case of dependence.

Keywords. Current estimate, risk margin, Bayesian stochastic methods,
Markov Chain Monte Carlo methods.

M.S.C. classification. 62P05, 65C05, 91B30.

1 Introduction

The aim of the study is the assessment of the reserve risk for a non-life in-
surance company that writes two different lines of business (LoBs). First of all
the stochastic methodology for the assessment of the reserve for a single LoB
is described: among the existing solutions the choice is a Bayesian model which
represents an improvement of one of the most known deterministic models, the
Chain-Ladder method. The implementation of the Bayesian model is possible by
using Markov Chain Monte Carlo (MCMC) techniques, specifically the Gibbs al-
gorithm, as described in Gilks, [5], and Scollnik, [13]. Considering a multiLoB
insurance company the problem is the determination of a single distribution of the reserve referred to the LoBs jointly, i.e. the aggregate reserve. For this scope the methodology followed requires the use of copula functions. In the specific case the marginal distributions are the probability distribution functions of the reserve of each single LoB. The statistic used to determine the correlation between LoBs is the average cost of LoB. The average cost of LoB is one of the possible alternatives to determine the correlation between LoBs. The ideal approach should consider the two historical series of the amounts booked in the past for the reserves of the two LoBs, adjusting the data for the past inflation and other external effects, for example the settlement policy. Though this approach is not always applicable, especially when the Insurance Company has not sufficient historical data. In this work the choice is to estimate the dependence on the average costs: this choice is due to the fact that the claim reserves are amounts that the Insurance Companies book for the eventual payment of incurred claims. To this end the claim frequency is not so important since the number of claims is known (with the exception of IBNR), while the average costs have a relevant economic role. In order to determine the parameter that represents the dependence structure of each copula the canonical maximum likelihood method is followed. The point estimate is used to determine a preference on the goodness of the fit. In order to evaluate the distribution of the dependence parameter for each copula a simulation algorithm is used, specifically the Metropolis-Hastings algorithm in case of independent sampling. Having determined the distribution of dependence parameter the aggregate reserve is assessed varying the dependence structure and the corresponding dependence parameter. In the last part of the work, using the results obtained, the economic impact of the reserve risk is assessed. The reserve risk is part of the underwriting risk, that is intended as the risk of loss, or of adverse change, in the value of insurance liabilities due to inadequate pricing and provisioning. It should therefore capture the risk arising over the occurrence period (a period of one year over which an adverse event occurs) and their financial consequences over the whole run-off of the liabilities. Two different economic measures are used, the solvency capital requirement (SCR) and the risk margin. The solvency capital requirement corresponds to the economic capital a (re)insurance undertaking needs to hold in order to limit the probability of ruin to 0.5%; it is calculated using Value-at-Risk techniques: all potential losses, including adverse revaluation of assets and liabilities over the next 12 months are to be assessed. The SCR reflects the true risk profile of the undertaking, taking account of all quantifiable risks. The risk margin is the financial cost of uncertainty of liabilities over the whole run-off. The risk margin ensures that the overall value of technical provisions is equivalent to the amount a (re)insurance undertaking would expect to have to pay today if it transferred its contractual rights and obligations immediately to another undertaking. The cost of capital (CoC) method is used to assess the risk margin. The CoC relies on a projection of the Solvency Capital Required to face potential adverse events until the last payment of liabilities over the whole run-off of the reserves. A new proxy for the evaluation of the risk margin is presented that captures the impact.
of the correlations between LoBs. The paper is consistent with the most recent existing literature that treats these themes, both the assessment of the reserving risk with a one year horizon approach (Merz, [10], [11]) and the assessment of the parameter variability in dependency structures using MCMC techniques (Borowicz, [2]).

2 The current estimate in a Bayesian Framework

In the framework of the Solvency II Project, the European Commission requested the Committee of European Insurance and Occupational Pensions Supervisors to establish well defined solvency and supervisory standards in order to allow a convergent and harmonized application across EU of the general prudential principles in the determination of the insurance technical provisions and the required solvency capitals. In the Solvency II draft Directive framework, [3], the technical provisions have the following definition: “The value of technical provisions shall be equal to the sum of a best estimate and a risk margin; the best estimate shall be equal to the probability-weighted average of future cash flows, taking account of the time value of money (expected present value of future cash-flows), using the relevant risk-free interest rate term structure; the risk margin shall be such as to ensure that the value of the technical provisions is equivalent to the amount insurance and reinsurance undertakings would be expected to require in order to take over and meet the insurance and reinsurance obligations”. Current estimates have sometimes been referred to as “best estimates”, although the latter term has sometimes also been used to represent the estimate of the most likely possible (modal) outcome rather than the estimate of the probability-weighted expected (mean) value that will be discussed here and that most faithfully represents the current assessment of the relevant cash flows. Such estimates reflect unbiased expectations of the obligation at the report date and are determined on a prospective basis. A current estimate represents the expected present value of the relevant cash flows. In the case where the present value is based on a range of discount rates, it is appropriate to estimate the probability-weighted expected present value of these cash flows. The assumptions used to derive a current estimate reflect the current expectation based on all currently available information about the relevant cash flows associated with the financial item being measured. These expectations involve expected probabilities and conditions (scenarios) during the period in which the cash flows are expected to occur. An assessment of expected future conditions is made rather than blindly applying recent historical or current experience. Although historical or current experience is often the best source from which current expectations of future experience can be derived for a particular portfolio, current estimates of cash flows should not automatically consist of a reproduction of recent experience. In addition, although the observed experience might be relevant to the portfolio as it existed during the observation period, the current portfolio for which estimates are being made may differ in several respects – in many cases, it could be argued that the current portfolio is usually different than the observed portfolio.
Probabilities specify the degree of our belief in some proposition(s) under the assumption that some other propositions are true. The conditioning propositions have to include, at least implicitly, the information used to determine the probability of the conditioned proposition(s). Probability is a relation between conditioned hypothesis and conditioning information - it is meaningless to talk about the probability of a hypothesis without also giving the evidence on which that probability value is based. Bayes’ Theorem uses conditional probabilities to reflect a degree of learning. It is central to model empirical learning both because it simplifies the calculation of conditional probabilities and because it clarifies significant features of the subjectivist position. Learning is a process of belief revision in which a “prior” subjective probability P is replaced by a “posterior” probability Q that incorporates newly acquired information. This process proceeds in two stages: first, some of the subject’s probabilities are directly altered by experience, intuition, memory, or some other non-inferential learning process; second, the subject “updates” the rest of his/her opinions to bring them into line with his/her newly acquired knowledge. Let $Y[i, j]$ denote the claim amounts paid by the insurance company with a delay of $j - 1$ years for accidents reported in the year $i$, with $i, j = 1, ..., n$ (where $n$ represents the number of different generations). The value of $j$ is commonly known as the development period. Let $Z[i, j]$ denote the cumulative claim amount for accidents reported in the year $i$ with a delay of $j - 1$ years or less. For convenience, it is assumed that the observed data is in the traditional upper triangular form such that $Y[i, j]$ and $Z[i, j]$ are observed for $i = 1, ..., n$ and $j = 1, ..., n - i + 1$, and unobserved elsewhere. Define the single cell development factor $DF[i, j]$ as

$$DF[i, j] = \frac{Z[i, j + 1]}{Z[i, j]},$$

for $i = 1, ..., n$ and $j = 1, ..., n - 1$. At the end of reporting year $n$, these factors are only observed for $i = 1, ..., n - 1$, with $j = 1, ..., n - i$. Then these estimated development factors are used, in conjunction with (1), to develop estimates of the cumulative claim amounts in the lower triangle and, hence, of the missing incremental claim amounts and the loss reserve. There are many possible ways in which to construct estimates of the missing single cell development factors in each column. One of the most popular set of estimates, known as the volume weighted development factors, is given by:

$$WDF[j] = \frac{\sum_{i=1}^{n-j} Z[i, j + 1]}{\sum_{i=1}^{n-j} Z[i, j]}.$$

Observe that the volume weighted development factors are weighted averages of the single cell development factors, with the cumulative claim amounts appearing in the denominator of the latter used as the weights involved in the calculation of the former. The single cell development factors tend to be similarly valued, given the development year $j$. Thus, moving to a Bayesian framework means specifying stochastic models with equal means for single cell development factors sharing a common development year. Normal models are assumed, although others could
be entertained (e.g., gamma or lognormal). Thus,

\[ DF[i,j] \sim N\left( \theta_j, \frac{1}{\tau_{i,j}} \right) . \tag{3} \]

for \( i = 1, \ldots, n \) and \( j = 1, \ldots, n - 1 \). Observe that negative incremental claims are permitted by this model, since \( DF[i,j] \) may be less than 1. The second parameter appearing in this normal distribution is a precision, or inverse variance, parameter. The precision parameters \( \tau_{i,j} \) may be modelled in a variety of ways (for example they can be set equal to a common value for all values \( i \) and \( j \), i.e. \( \tau \)). Possibly, the parameters may be scaled by relevant weights, where the weights could be related to the written premium associated with the different years of origination, or perhaps to the number of claims associated with the different years of origination that settled in the latest development period (for example). Next, it is supposed that the underlying parameters \( \theta_{i,j} \) are drawn from a common normal distribution, i.e. :

\[ \theta_j \sim N\left( \mu_\theta, \frac{1}{\tau_\theta} \right) , \tag{4} \]

for \( j = 1, \ldots, n - 1 \). The remaining parameters \( \mu_\theta, \tau_\theta \), and \( \tau \) must be assigned prior density specifications in order to complete the definition of a full (i.e., fully specified) probability model. In particular:

\[ \tau \sim \Gamma(a, b) , \tag{5} \]

\[ \mu_\theta \sim N(c, d) , \tag{6} \]

\[ \tau_\theta \sim \Gamma(e, f) . \tag{7} \]

The parameters \( a, b, c, d, e, f \) are estimated on the lob historic experience. The Bayesian analysis of this model yields the posterior distribution (i.e., not just point estimates) for all unknown model parameters. This includes the posterior predictive distribution of the unobserved claim and cumulative claim amounts (i.e., the reserves). The implementation of the Bayesian model is possible by using Markov Chain Monte Carlo (MCMC) techniques, specifically the Gibbs algorithm. (Gilks, [5], Scollnik, [13]).

3 The aggregate current estimate

Considering a multiLoB insurance company the problem is the determination of a single distribution of the reserve referred to the LoBs jointly, i.e. the aggregate reserve. For this scope the methodology followed requires the use of copula functions. Copulas have become a popular multivariate modeling tool in many fields where the multivariate dependence is of great interest and the usual multivariate normality is in question. A copula is a multivariate distribution whose
marginals are all uniform over \((0, 1)\). For a \(p\)-dimensional vector \(U\) on the unit cube, a copula \(C\) is

\[
C (u_1, ..., u_p) = \Pr [U_1 \leq u_1, ..., U_p \leq u_p], \tag{8}
\]

Combined with the fact that any continuous random variable can be transformed to be uniform over \((0, 1)\) by its probability integral transformation, copulas can be used to provide multivariate dependence structure separately from the marginal distributions. Copulas first appeared in the probability metrics literature. Let \(F\) be a \(p\)-dimensional distribution function with margins \(F_1, ..., F_p\).

Sklar, \cite{14}, first showed that there exists a \(p\)-dimensional copula \(C\) such that for all \(x\) in the domain of \(F\),

\[
F (x_1, ..., x_p) = C \{F_1 (x_1), ..., F_p (x_p)\}. \tag{9}
\]

The last two decades witnessed the spread of copulas in statistical modelling. Joe, \cite{9}, and Nelsen, \cite{12}, are the two comprehensive treatments on this topic. A frequently cited and widely accessible reference is Genest and MacKay, \cite{4}, titled “The Joy of Copulas”, which gives properties of an important family of copulas, Archimedean copulas. Given a dataset, choosing a copula to fit the data is an important but difficult problem. The true data generation mechanism is unknown, for a given amount of data, it is possible that several candidate copulas fit the data reasonably well or that none of the candidate fits the data well. When maximum likelihood method is used, the general practice is to fit the data with all the candidate copulas and choose the ones with the highest likelihood.

Suppose that we observe \(n\) independent realizations from a multivariate distribution, \(\{(X_{i1}, ..., X_{ik})^T : i = 1, ..., n\}\). Suppose that the multivariate distribution is specified by \(k\) margins with cdf \(F_i\) and PDF \(f_i\), \(i = 1, ..., k\) and a copula with density \(c\). Let \(\lambda\) be the vector of marginal parameters and \(\alpha\) be the vector of copula parameters. The parameter vector to be estimated is \(\theta = (\lambda^T, \alpha^T)^T\). The loglikelihood function is:

\[
l (\theta) = \sum_{i=1}^{n} \log c \{F_1 (X_{i1}; \lambda), ..., F_{ik} (X_{ik}; \alpha)\} + \sum_{i=1}^{n} \sum_{j=1}^{k} \log f_i (X_{ij}; \lambda). \tag{10}
\]

The ML estimator of \(\theta\) is \(\hat{\theta}_{MLE} = \arg \max_{\theta \in \Theta} l (\theta)\), where \(\Theta\) represents the parameter space. Actuaries face a difficult task when estimating the parameter values for their chosen copula. In insurance, historical data is often limited. The parameter uncertainty can be significant for small data sets. Using fixed parameters in the copula for claims could lead to misestimation of risk, as there may not be enough historical data for the effect of dependence to become apparent. Common methods of assessing parameter uncertainty include “classical” statistical methods, such as the asymptotic normality of maximum likelihood estimates, and empirical approaches, such as non-parametric bootstrapping. However, these methods have drawbacks when applied to small samples, as is typical.
in insurance. An alternative general framework for the analysis of parameter uncertainty which does not have such problems is the use of Bayesian methods, in conjunction with Markov Chain Monte Carlo (MCMC) techniques, [5]. This is the followed approach.

4 The risk margin

Insurance obligations are, by their very nature, uncertain. The insurance industry exists to purchase uncertainty from policyholders by transferring at least part of this uncertainty for a price. Measurement of liabilities for insurance contracts is currently under discussion by the International Accounting Standards Board (IASB). The current likely measurement direction of the IASB is based on an exit value, i.e. the amount an insurer would expect to pay or receive at the current date if it transferred its outstanding rights and obligations under a contract to another entity. When deep liquid observable markets exist for financial instruments (such as for many financial assets), the observed exit price already provides an investor with an expected return sufficient for compensation for the risks in that investment relative to alternative investments. In this paper the market price includes both a current estimate of expected cash flows and a risk margin in excess of that amount. If there were a deep liquid market for insurance obligations, the observed market price for an insurance obligation would constitute the exit price. However, as no deep liquid market currently exists for insurance obligations, a model must be constructed that can produce exit values.

In putting this methodology into practice, it is assumed that a rational transferee would require something above the current estimate (even if transferor and transferee were to agree perfectly on the level of the current estimate). Otherwise, the transferee would expect to receive nothing for taking on the risk if everything does not work out as expected. The amount, the margin over current estimate, can therefore be regarded as an additional amount “for uncertainty”. It therefore can also be regarded as a compensation for the transferee for the risk of taking on an obligation to pay uncertain cash flows. In addition to serving as an element of the exit price, a risk margin makes it possible to absorb reasonable volatility in experience. If experience is more favorable than that assumed in the current estimate, without risk margins, the release of the excess risk margin creates a “profit” that serves as a reward for the investor that has taken the risk; if experience is worse than expected, the risk margin covers some part of the expected losses, also considering that is also a chance of achieving profits. Normally, a purchaser will not be willing to assume a risky obligation unless its expected reward for doing so not only covers the expected costs, but a margin for risk has been provided as well. Three basic approaches (sometimes referred to as methods), or more appropriately, families of approaches, (see IAA, [6], [7]) of determining risk margins have been used in the past:

1. Explicit assumption approaches. These risk margin methods use “appropriate” margins for adverse deviation on top of realistic “current estimate” assumptions.
2. Quantile methods. These risk margin methods express uncertainty in terms of the excess of a percentile (quantile) for a given confidence level above the expected value for a given period, such as the lifetime of the coverage.

3. Cost of capital methods. These risk margin methods are determined based on the cost of holding the capital needed to support the obligation. The cost of capital method is based on the explicit assumption that, at each point in time, the risk margin must be sufficient to finance the (solvency) capital otherwise a transferee will be unwilling to pay less than an amount that would fund future capital requirements. Reflection of the estimated current and future economic capital needs of a potential transferee ensures that the amount paid for the transferee for risk provides for the entire risk that will affect the purchaser. In contrast, the quantile and explicit assumption methods do not explicitly reflect current or future required capital. Having fixed the cost of holding the capital (e.g., \( CoC = 6\% \)), the risk margin at the valuation date \((t = 0)\) is determined as follows:

\[
RM_0 = \sum_{t=1}^{n-1} \frac{CoC \cdot SCR_t - 1}{(1 + i(0, t))^{t}},
\]

where \( SCR \) is the solvency capital requirement, \( i(0, t) \) is the interest rate.

5 The reserve risk

In the Solvency II draft Directive framework the Solvency Capital Requirement has the following definition: “The SCR corresponds to the economic capital a (re)insurance undertaking needs to hold in order to limit the probability of ruin to 0.5%, i.e. ruin would occur once every 200 years. The SCR is calculated using Value-at-Risk techniques, either in accordance with the standard formula, or using an internal model: all potential losses, including adverse revaluation of assets and liabilities over the next 12 months are to be assessed. The SCR reflects the true risk profile of the undertaking, taking account of all quantifiable risks, as well as the net impact of risk mitigation techniques.” Within this framework, the reserve risk is defined as a part of the underwriting risk, as follows: “Underwriting risk means the risk of loss, or of adverse change in the value of insurance liabilities, due to inadequate pricing and provisioning”. If we apply this framework to the reserve risk (see IAIS, [8]), the concept of time horizon should distinguish between a period of one year over which an adverse event occurs, i.e. “shock period”, and a period over which the adverse event will impact the liabilities, i.e. the “effect period”. In any case the reserve risk should capture the risks arising over the occurrence period and their financial consequences over the whole run-off of liabilities (for example, a court judgement or judicial opinion in one year – the shock period – may have permanent consequences for the value of claims and hence will change the projected cash flows to be considered over the full run-off of liabilities – the effect period). To illustrate the concept of a one year horizon year, let’s consider the following example. The goal is to assess the
reserve risk at 31.12\(N\) over a one year horizon, from the triangulation of losses over 12 underwriting years \([Uw(N-11);Uw(N)]\). Figure 1 is divided into 4 areas (\(A, B, C, D\)):

**Fig. 1.** Graphical representation of a non-life technical liability

**A:** This area contains the available data/information at 31.12\(N\) to assess the reserves at 31.12\(N\) (Noted \(R_n\)).

**B:** This area (soft grey) corresponds to a one year period beyond 31.12\(N\). This area represents the “shock period”. At the end of the shock period (i.e. at 31.12\(N+1\)), it will be possible to revise \(R_n\) a posteriori considering:
- the real payments of losses (noted \(P_{n+1}\)) over the period \([01.01.N+1;31.12.N+1]\),
- the valuation of reserves at 31.12\(N+1\) (noted \(R_{n+1}\)) regarding the available information until 31.12\(N+1\) for the underwriting years \([Uw(N-11);Uw(N)]\). The reserve risk at 31.12\(N\) measures the uncertainty of the valuation of reserves calculated at 31.12\(N\) regarding the additional information over the period \([01.01.N+1;31.12.N+1]\) that could change this valuation at 31.12\(N+1\) (The reserves at 31.12\(N+1\) do not include the liabilities related to the underwriting year \(N+1\). Indeed the risk associated with this underwriting year is captured in the premium risk). The reserve risk captures the difference between \([P_{n+1}+R_{n+1}]\) and \(R_n\).

**C:** Under the Solvency II framework and to calculate the reserve risk, this area represents the effect period beyond the shock period. This area contains additional information that could lead to revision of the reserves beyond 31.12\(N+1\). This additional information should not be taken into account. The use of the area C should be limited to the assessment of the financial consequences of the adverse events arising during the shock period.

**D:** This area contains the ultimate costs. These costs are used to assess the risk capital with a VaR methodology. The most usual actuarial methodologies are not consistent with the Solvency II framework since they capture all the adverse events arising beyond the one year horizon. Within the Solvency II
framework, it should not be a surprise that some long tail business where adverse movements in claims provisions emerge slowly over many years require less solvency capital than some short tail business exposed to catastrophe risks (for instance).

The uncertainty measurement of reserves in the balance sheet (called risk margin in the Solvency II framework) and the reserve risk do not have the same time horizon. It seems important to underline this point because it may be a source of confusion when the calibration is discussed. The risk margin captures uncertainty over the whole run-off of liabilities. The Solvency II draft Directive framework provides a definition of the risk margin: “The risk margin ensures that the overall value of the technical provisions is equivalent to the amount (re)insurance undertakings would expect to have to pay today if it transferred its contractual rights and obligations immediately to another undertakings; or alternatively, the additional cost, above the best estimate of providing capital to support the (re)insurance obligations over the lifetime of the portfolio” For non-life liabilities (which are non-hedgeable in general) the risk margin is the financial cost of uncertainty of liabilities over the whole run-off giving that this uncertainty is calibrated through the solvency filter: “Where insurance and reinsurance undertakings value the best estimate and the risk margin separately, the risk margin shall be calculated by determining the cost of providing an amount of eligible own funds equal to the Solvency Capital Requirement necessary to support the insurance and reinsurance obligations over the lifetime thereof.” Suppose the risk margin is assessed with the cost of capital (CoC) methodology. The level of the CoC relies essentially on the reserve risk calibration. If the reserve risk is over calibrated (i.e. for instance a calibration over the whole run-off of the reserves), the CoC methodology multiplies the level of prudence.

![Cost of capital calculation if the reserve risk is assessed over a one year horizon](image)

**Fig. 2.** Cost of capital calculation if the reserve risk is assessed over a one year horizon

For each year horizon, the CoC captures the cost of providing own funds equal to the Solvency Capital Requirement necessary to support the insurance
and reinsurance obligations over the run-off. If the duration of the run-off is \( N \) years, the CoC embeds \( N \) SCR valuations.

![Cost of Capital calculation if the reserve risk is assessed over the whole run-off](image)

**Fig. 3.** Cost of Capital calculation if the reserve risk is assessed over the whole run-off

If the reserve risk is calibrated over the whole run-off, or, broadly speaking, if the reserve risk is over calibrated, the CoC creates undue layers of prudence with a leverage effect (see the \( N \) “clusters of risks” in Figure 2 versus \( N(N + 1)/2 \) “clusters of risks” in Figure 3).

### 6 A possible alternative

The Cost of Capital method for the assessment of the risk margin relies on a projection of the Solvency Capital Required to face potential adverse events until the last payment of liabilities, i.e. over the whole run-off of the reserves. Among the problems that can arise in the assessment two have to be considered inevitably: the projection of the capital requirement in future years and the double counting of the risk margin in the approach chosen. One of the possibilities for the calculation is based on the following formula:

\[
RM_0 = \sum_{t=1}^{n-1} CoC \cdot SCR_0 \cdot \frac{CE_{t-1}}{CE_0} \cdot \max\{1, \ln(1 + \gamma_{t-1})\} \cdot \frac{1}{(1 + i(0, t))^t},
\]

where \( RM \) represents the risk margin, \( CoC \) the cost of capital, \( SCR \) the solvency capital requirement, \( CE \) the current estimate, \( i(0, t) \) the interest rate, \( \gamma_t = \frac{CV(Res_t)}{CV(Res_0)} \) the ratio between coefficients of variation of the random variable \( Res \), which represents the outstanding claim reserve. The capital requirement is determined as follows:

\[
SCR_0 = VaR^{99.5\%}(Res_0) - RM_0 - CE_0,
\]
where $VaR_{99.5\%}(Res_0)$ represents the Value at Risk at the valuation date of the outstanding claim reserve at a 99.5% confidence level over a one-year time horizon. Substituting the (13) in (12) the risk margin becomes:

$$RM_0 = \frac{CoC \cdot (VaR_{99.5\%}(Res_0) - CE_0) \cdot ProFact}{1 + CoC \cdot ProFact}.$$  

(14)

where

$$ProFact = \sum_{t=1}^{n-1} \frac{CE_{t-1}}{CE_0} \cdot \max(1, \ln(1 + \gamma_{t-1})) \cdot \frac{1}{(1 + i(0,t))^t}.$$  

(15)

The assessment of the risk margin through (14) has some advantages:

- the solvency capital requirement follows the underlying driver, i.e. the current estimate;
- the formula considers that the variance increases as the time passes and consequently the SCR should increase as well, as the variance is a risk measure;
- the future variance of the current estimate is over estimated at the valuation date: this is due to the lack of information on the development factors for the extreme development years. The increase is mitigated through the use of the function;
- the double counting of risk margin both in the fair value and in the capital requirement is eliminated;
- the formula considers the real variance and the real Value-at-Risk of the current estimate instead of approximations and simplifications. When evaluating the aggregate current estimate, the coefficient of variation is determined considering the correlation between LoBs.

7 Numerical results

This paragraphs shows the results obtained through the methods described in the paper. The initial data set is represented by the run-off triangles of incremental payments of two distinct lines of business: Motor, other classes (LoB 3, Table 1) and Motor, third party liability (LoB 10, Table 2). The risk free interest rates adopted in the discounting are reported in Table 3. Table 4 and Table 5 report the values of the current estimate, the risk margin, the risk capital (the risk covered is the reserve risk, the Solvency Capital Requirement covers also other risks), comparing different possibilities described in literature. The different columns of the tables represent different way of calculations:

- I (method) : the risk margin is given by the 75% percentile;
- II (method) : the risk margin is obtained as indicated by the IAA (see [7]) ;
- III (method) : the risk margin is obtained as the difference between undiscounted reserve and the discounted one;
- IV (method) : the risk margin is obtained through the formula (14).
Table 1. Incremental Payments Triangle LoB 3: Motor, other classes, Values in Euro thousands

<table>
<thead>
<tr>
<th>$Y_{[i,j]}$</th>
<th>Development year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underwriting year</td>
<td>1</td>
</tr>
<tr>
<td>1999</td>
<td>15,762</td>
</tr>
<tr>
<td>2000</td>
<td>16,648</td>
</tr>
<tr>
<td>2001</td>
<td>18,333</td>
</tr>
<tr>
<td>2002</td>
<td>21,999</td>
</tr>
<tr>
<td>2003</td>
<td>25,271</td>
</tr>
<tr>
<td>2004</td>
<td>23,574</td>
</tr>
<tr>
<td>2005</td>
<td>26,833</td>
</tr>
<tr>
<td>2006</td>
<td>28,613</td>
</tr>
</tbody>
</table>

Table 2. Incremental Payments Triangle LoB 10: Motor, third party liability, Values in Euro thousands

<table>
<thead>
<tr>
<th>$Y_{[i,j]}$</th>
<th>Development year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underwriting year</td>
<td>1</td>
</tr>
<tr>
<td>1999</td>
<td>22,144</td>
</tr>
<tr>
<td>2000</td>
<td>25,875</td>
</tr>
<tr>
<td>2001</td>
<td>29,655</td>
</tr>
<tr>
<td>2002</td>
<td>31,031</td>
</tr>
<tr>
<td>2003</td>
<td>42,197</td>
</tr>
<tr>
<td>2004</td>
<td>44,481</td>
</tr>
<tr>
<td>2005</td>
<td>49,964</td>
</tr>
<tr>
<td>2006</td>
<td>49,848</td>
</tr>
</tbody>
</table>

Table 3. Risk free interest rates

<table>
<thead>
<tr>
<th>$t(\text{years})$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i(0,t)$</td>
<td>4.07%</td>
<td>4.12%</td>
<td>4.12%</td>
<td>4.12%</td>
<td>4.11%</td>
<td>4.12%</td>
<td>4.13%</td>
</tr>
</tbody>
</table>

Table 4. LoB 3 Values (Euro thousands)

<table>
<thead>
<tr>
<th>Values/Methods</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Estimate</td>
<td>19,274</td>
<td>19,274</td>
<td>19,274</td>
<td>19,274</td>
</tr>
<tr>
<td>Risk Margin</td>
<td>3,986</td>
<td>35</td>
<td>2,267</td>
<td>362</td>
</tr>
<tr>
<td>Fair Value</td>
<td>23,260</td>
<td>19,309</td>
<td>21,541</td>
<td>19,636</td>
</tr>
<tr>
<td>Risk Capital</td>
<td>13,382</td>
<td>17,333</td>
<td>15,101</td>
<td>17,006</td>
</tr>
</tbody>
</table>
Table 5. LoB 10 Values (Euro thousands)

<table>
<thead>
<tr>
<th>Values/Methods</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Estimate</td>
<td>210,860</td>
<td>210,860</td>
<td>210,860</td>
<td>210,860</td>
</tr>
<tr>
<td>Risk Margin</td>
<td>32,730</td>
<td>5,029</td>
<td>36,249</td>
<td>9,561</td>
</tr>
<tr>
<td>Fair Value</td>
<td>243,590</td>
<td>215,889</td>
<td>247,109</td>
<td>220,421</td>
</tr>
<tr>
<td>Risk Capital</td>
<td>114,310</td>
<td>142,011</td>
<td>110,791</td>
<td>137,479</td>
</tr>
</tbody>
</table>

The results obtained show both for LoB 3 and LoB 10 that the assessment of the risk margin is sensitive to the approach followed. In particular with the 75% percentile approach (I) the risk margin is close to the undiscounted approach (III), which is in use in the Italian market, and the values are much higher than the ones obtained with the cost of capital (II and IV). Comparing the values obtained with the approaches II and IV it can be noted that the consideration of the future variability (through the formula (14)) considerably increases the value of the risk margin (i.e. for LoB 10 the RM value passes from 5 mil to 10 mil nearly). Table 6 (aggregation without considering dependence LoBs) and Table 7 (aggregation considering dependence between LoBs) show the values of the current estimate, the risk margin, the risk capital when the two lines of business are considered jointly. For the dependence case the values are referred to the Gumbel Copula (which is the copula with the highest likelihood among the ones considered). The value of the dependence parameter is for prudence chosen as the 97.5% percentile of its distribution.

Table 6. Aggregate Values - LoB3 + LoB10, Independence (Euro thousands)

<table>
<thead>
<tr>
<th>Values/Methods</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Estimate</td>
<td>230,134</td>
<td>230,134</td>
<td>230,134</td>
<td>230,134</td>
</tr>
<tr>
<td>Risk Margin</td>
<td>34,785</td>
<td>5,005</td>
<td>38,515</td>
<td>8,041</td>
</tr>
<tr>
<td>Fair Value</td>
<td>264,919</td>
<td>235,139</td>
<td>268,649</td>
<td>238,175</td>
</tr>
<tr>
<td>Risk Capital</td>
<td>98,058</td>
<td>127,837</td>
<td>94,328</td>
<td>124,802</td>
</tr>
</tbody>
</table>

The results show that considering the aggregation between LoBs leads to a gain in comparison to the actual approach (that could be summarized in taking the values of each LoB calculated separately). The diversification impact is much higher if the risk capital is considered: if the approach IV is observed, the risk capital to be held is equal to Euro 135,1 mil. that is 12.5% lower compared to sum the two values calculated separately (Euro 154, 6 mil.). It is to be noted

\footnote{The risk free term structure is the one reported in the QIS3 Solvency II Technical Specifications, derived using swap rates rather than government bonds (see [1]).}
Table 7. Aggregate Values - LoB3 + LoB10, Copula Gumbel (Euro thousands)

<table>
<thead>
<tr>
<th>Values/Methods</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Estimate</td>
<td>230,134</td>
<td>230,134</td>
<td>230,134</td>
<td>230,134</td>
</tr>
<tr>
<td>Risk Margin</td>
<td>36,598</td>
<td>5,031</td>
<td>38,515</td>
<td>8,707</td>
</tr>
<tr>
<td>Fair Value</td>
<td>266,732</td>
<td>235,165</td>
<td>268,649</td>
<td>238,841</td>
</tr>
<tr>
<td>Risk Capital</td>
<td>107,249</td>
<td>138,816</td>
<td>105,332</td>
<td>135,140</td>
</tr>
</tbody>
</table>

also that the value of the risk capital could be underestimated in case the risks are considered independent (approach IV a risk capital 8% lower).

8 Conclusions

This work describes a possible solution for the calculation of the risk margin and compares the new formula proposed with the approaches already proposed in the literature. The methodology adopted follows the indications outlined in the Solvency II framework that encourages the development of stochastic actuarial models for the assessment of the technical provisions, though the analysis is limited to two lines of business. The numerical results of the case study proposed outline that the Cost of Capital approach (methods II and IV) is less prudent than the method that is actually adopted on the Italian market since the results are significantly lower than the ones obtained determining the Risk Margin through the ultimate cost method. The dependence between different LoBs has a significant impact on the estimate of the Reserve Risk Capital: therefore in the Solvency II framework a key issue will be the definition of the statistic to be used to determine this dependency and the estimation of the parameter, taking into account its variability, as shown in the applications presented in this work. If more LoBs were considered the effects of the dependence could be higher and the Capital requirement could be reduced even further. The intention is to continue the study of the possible solutions extending the analysis to the insurance company as a whole.

References

Parameter estimation for differential equations using fractal-based methods and applications to economics and finance

Herb Kunze¹, Davide La Torre¹,², Cinzia Colapinto²,
Jelena Loncar¹, Matteo Fini²

¹Department of Economics, Business and Statistics,
University of Milan, Italy
{davide.latorre, cinzia.colapinto, matteo.fini}@unimi.it
²Department of Mathematics and Statistics, University of Guelph,
Guelph, Ontario, Canada
{hkunze, jloncar}@uoguelph.ca

Abstract. Many problems from the area of economics and finance can be described using dynamical models. If time is the only independent variable and for which we work in a continuous framework, these models take the form of differential equations (DEs). These models can be studied through the direct problem and the inverse problem. The inverse problem consists of estimating the unknown parameters of the model starting from a set of observational data. We use fractal-based methods to get them. The method will be illustrated through several numerical examples and applications to economical and financial situations.

Keywords. Differential equations, collage methods, inverse problems, parameter estimation, stochastic differential equations, technological change, boat-fishery model.

M.S.C. classification. 65R32, 45Q05, 45R05.
J.E.L. classification. C02, C13, C63.

1 Introduction

Many problems from the area of economics and finance can be described using dynamical models. For them, in which time is the only independent variable and for which we work in a continuous framework, these models take the form of deterministic differential equations (DEs). We may study these models in two fundamental ways: the direct problem and the inverse problem. The direct problem is stated as follows: given all of the parameters in a system of DEs, find a solution or determine its properties either analytically or numerically. The inverse problem reads: given a system of DEs with unknown parameters and some observational data, determine the values of the parameters such that
the system admits the data as an approximate solution. The inverse problem is crucial for the calibration of the model; starting from a series of data we wish to describe them using differential equations in which the parameters have to be estimated from data samples. The solutions to the inverse problems are the estimations of the unknown parameters and we use fractal-based methods to get them.

The paper is organized as follows: in sections 1 and 2 we present the basic results concerning the solution of inverse problems for fixed point equations through the so called “collage theorem.” We then present some numerical examples: in section 3 and 4 we analyze inverse problems for two economic models arising in the contexts of technological change and resource management, and in section 5 we show how one can use the “collage method” for solving inverse problems for a class of stochastic differential equations.

## 2 Fixed point equations and inverse problems through the “collage theorem”

For the benefit of the reader, we now mention some important mathematical results which provide the basis for fractal-based approximation methods. Let us consider the fixed point equation $x = Tx$, where $(X, d)$ is a complete metric space and $T$ a contractive operator on $X$. The direct problem for a fixed point equation can be solved through the classical Banach theorem.

**Theorem 1.** (Banach) Let $(X, d)$ be a complete metric space. Also let $T : X \rightarrow X$ be a contraction mapping with contraction factor $c \in [0, 1)$, i.e., for all $x, y \in X$, $d(Tx, Ty) \leq cd(x, y)$. Then there exists a unique $\bar{x} \in X$ such that $\bar{x} = T\bar{x}$. Moreover, for any $x \in X$, $d(T^n x, \bar{x}) \rightarrow 0$ as $n \rightarrow \infty$.

A simple triangle inequality along with Banach’s theorem yields the following fundamental result.

**Theorem 2.** (“Collage Theorem” [2,1]) Let $(X, d)$ be a complete metric space and $T : X \rightarrow X$ a contraction mapping with contraction factor $c \in [0, 1)$. Then for any $x \in X$,

$$d(x, \bar{x}) \leq \frac{1}{1 - c} d(x, Tx),$$

(1)

where $\bar{x}$ is the fixed point of $T$.

The inverse problem is: given a target element $y$, can we find an operator $T$ with fixed point $\bar{x}$ so that $d(y, \bar{x})$ is sufficiently small. Thanks to the “Collage Theorem”, most practical methods of solving the inverse problem for fixed point equations seek to find an operator $T$ for which the collage distance $d(y, Ty)$ is as small as possible.

We now consider the case of random fixed point equations. Let $(\Omega, F, P)$ be a probability space. A mapping $T : \Omega \times X \rightarrow X$ is called a random operator if for any $x \in X$ the function $T(\cdot, x)$ is measurable. The random operator $T$ is said
Parameter estimation for differential equations...

...to be continuous/Lipschitz/contractive if, for a.e. \( \omega \in \Omega \), we have that \( T(\omega, \cdot) \) is continuous/Lipschitz/contractive ([11]). A measurable mapping \( x : \Omega \to X \) is called a random fixed point of the random operator \( T \) if \( x \) is a solution of the equation

\[
T(\omega, x(\omega)) = x(\omega), \quad \text{a.e.} \omega \in \Omega. \tag{2}
\]

We are concerned about the existence of solutions to such equations. Consider the space \( Y \) of all measurable functions \( x : \Omega \to X \). If we define the operator \( \tilde{T} : Y \to Y \) as \( (\tilde{T}x)(\omega) = T(\omega, x(\omega)) \) the solutions of this fixed point equation on \( Y \) are the solutions of the random fixed point equation \( T(\omega, x(\omega)) = x(\omega) \). If the metric \( d \) is bounded then the space \((Y, d_Y)\) is a complete metric space (see [4]) where

\[
d_Y(x_1, x_2) = \int_{\Omega} d_X(x_1(\omega), x_2(\omega))dP(\omega). \tag{3}
\]

The following result follows from the completeness of \((Y, d_Y)\) and Banach’s fixed point theorem. It states sufficient conditions for the existence of solutions.

**Theorem 3.** Suppose that

(i) for all \( x \in Y \) the function \( \xi(\omega) := T(\omega, x(\omega)) \) belongs to \( Y \),

(ii) \( d_Y(\tilde{T}x_1, \tilde{T}x_2) \leq cd_Y(x_1, x_2) \) with \( c < 1 \).

Then there exists a unique solution of \( \tilde{T}\bar{x} = \bar{x} \), that is, \( T(\omega, \bar{x}(\omega)) = \bar{x}(\omega) \) for a.e. \( \omega \in \Omega \).

The inverse problem can be formulated as: given a function \( \bar{x} : \Omega \to X \) and a family of operators \( \tilde{T}_a : Y \to Y \) find \( a \) such that \( \bar{x} \) is the solution of random fixed point equation

\[
\tilde{T}_a\bar{x} = \bar{x}, \tag{4}
\]

that is,

\[
T_a(\omega, \bar{x}(\omega)) = \bar{x}(\omega). \tag{5}
\]

The collage theorem can also be reformulated for this setting, using the same hypotheses as in Theorem 3. In both theorems, hypothesis (i) can be avoided if \( X \) is a Polish space.

3 An inverse problem for a technological competition model

A classical technological competition model can be formulated ([10]) as

\[
\frac{dx_1}{dt}(t) = f_1(x_1, x_2) = \frac{a_1}{K_1} x_1 (K_1 - x_1 - \alpha_2 x_2)
\]

\[
\frac{dx_2}{dt}(t) = f_2(x_1, x_2) = \frac{a_2}{K_2} x_2 (K_2 - x_2 - \alpha_1 x_1),
\]

where all of the parameters are positive and \( a_1, a_2, \alpha_1 \) and \( \alpha_2 \) are less than one. Observe that the nonnegative quadrant is invariant. This means that if we start
with \((x_1(0), x_2(0)) \geq (0, 0)\) then we have \((x_1(t), x_2(t)) \geq (0, 0)\) for all time. That is, in the applied meaningful cases, \(x_1\) and \(x_2\), as determined by our model, are always nonnegative. The linearization of the vector field \((f_1, f_2)\) is

\[
Df(x_1, x_2) = \begin{pmatrix}
\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\
\frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2}
\end{pmatrix} = \begin{pmatrix}
a_1 - \frac{a_1}{K_1} (2x_1 + \alpha_2 x_2) & \frac{a_1 \alpha_2}{K_1} x_2 \\
-a_2 \alpha_1 & -a_2 \alpha_2
\end{pmatrix} = \begin{pmatrix}
a_1 (2x_1 + \alpha_2 x_2) - \frac{a_1 \alpha_2}{K_1} x_2 \\
-a_2 \alpha_1 & -a_2 \alpha_2
\end{pmatrix}.
\]

Solving for equilibria, we obtain

\[
(0, 0), (0, K_2), (K_1, 0), \text{ and } (x_1^*, x_2^*) = \left( \frac{K_1 - \alpha_2 K_2}{1 - \alpha_1 \alpha_2}, \frac{K_2 - \alpha_1 K_1}{1 - \alpha_1 \alpha_2} \right).
\]

And so, evaluating the linearization at each of the equilibrium points, we calculate that

\[
Df(0, 0) = \begin{pmatrix}
a_1 & 0 \\
0 & a_2
\end{pmatrix}.
\]

Since \(Df(0, 0)\) is positive definite, we know that \((0, 0)\) is an unstable equilibrium (a source). On the other hand, at the next two points we obtain

\[
Df(0, K_2) = \begin{pmatrix}
\frac{a_1}{K_1} (K_1 - \alpha_2 K_2) & 0 \\
-a_2 \alpha_1 & -a_2 \alpha_2
\end{pmatrix}
\]

and

\[
Df(K_1, 0) = \begin{pmatrix}
-a_1 & -a_1 \alpha_2 \\
0 & \frac{a_2}{K_2} (K_2 - \alpha_1 K_1)
\end{pmatrix}.
\]

Each matrix has one negative eigenvalue, with the sign of other one determined by a relationship between \(K_1, K_2\), and one of the \(\alpha_i\)s. All three of these equilibrium points have at least one component equal to zero, corresponding to one of the competing technologies begin eliminated from the market. The origin is a special equilibrium point, in that we only arrive at it if we start at it: if neither of the two technologies is present at the start, both of them will never appear. The equilibrium \((0, K_2)\) corresponds to technology \(x_2\) triumphing over technology \(x_1\). The single negative eigenvalue corresponds to the case that the market only has technology \(x_2\) at the start. \(x_1\) never appears, so we arrive at an equilibrium state where technology \(x_2\) is the only one in the market. In the case that \(K_1 - \alpha_2 K_2 < 0\), it is possible for a market with both technologies present to approach a state where \(x_2\) has eliminated \(x_1\). If \(K_1 - \alpha_2 K_2 > 0\), then we can never reach \((0, K_2)\) if we start with both technologies present. Similar remarks can be made about \((K_1, 0)\). However, notice that if we try to make both of these boundary equilibria stable, we require both \(K_1 - \alpha_2 K_2 < 0\) and \(K_2 - \alpha_1 K_1 < 0\), which means that

\[
K_1 < \alpha_2 K_2 < \alpha_2 (\alpha_1 K_1) \Rightarrow 1 < \alpha_1 \alpha_2.
\]

But this is a contradiction if both \(\alpha_i\)s are less than one. As a result, we can make at most one of the nontrivial boundary equilibria stable. The final and
most interesting equilibrium point can correspond to coexistence of the two technologies in the case that both $x_1^*$ and $x_2^*$ are positive. This situation occurs when
\[ K_1 - \alpha_2 K_2 > 0 \text{ and } K_2 - \alpha_1 K_1 > 0. \]  
These conditions are familiar, corresponding to the case when both boundary equilibria cannot be reached by interior solutions. In this case, we calculate that

\[
Df(x_1^*, x_2^*) = \begin{pmatrix}
\frac{a_1}{K_1} - \alpha_2 & \frac{a_1}{K_1} - \frac{\alpha_2}{K_2}
\frac{a_2}{K_2} - \alpha_1 & \frac{a_2}{K_2} - \frac{\alpha_1}{K_1}
\end{pmatrix},
\]

with determinant

\[
a_1 a_2 \frac{(K_1 - \alpha_2 K_2)(K_2 - \alpha_1 K_1)}{K_2 K_1 (1 - \alpha_1 \alpha_2)} > 0.
\]

Since the determinant is positive and $(Df(x_1^*, x_2^*))_{11} < 0$, we conclude that if we are in the case where our system exhibits a positive equilibrium then it is asymptotically stable—in fact with basis of attraction the positive quadrant! Notice that if either inequality in (6) is replaced by the corresponding equation then our equilibrium point coalesces with one of the boundary equilibria. If either inequality is in fact negative, then the equilibrium point we are discussing is not physically realizable.

Figure 1 presents a solution trajectory in the case that $K_1 = 320, K_2 = 100, \alpha_1 = 0.2, \alpha_2 = 0.5, a_1 = 0.3, a_2 = 0.6$. Note that the inequalities in (6) are satisfied.

Fig. 1. All solution trajectories starting in the positive quadrant approach the positive equilibrium because (6) holds.
Figure 2 presents a solution trajectory in the case that \( K_1 = 125, K_2 = 320, \alpha_1 = 0.2, \alpha_2 = 0.5, \alpha_1 = 0.3, \) and \( a_2 = 0.6. \) In this case, we have no positive equilibrium, but the equilibrium point \((0, 320)\) is asymptotically stable.

![Graph showing solution trajectory](image)

**Fig. 2.** All solution trajectories starting in the positive quadrant approach \((0, K_2)\) in this case.

The inverse problem of interest to us is: given observed values \( x_1(t_i) \) and \( x_2(t_i) \) for \( 1 \leq i \leq N \), say, approximate the values of the parameters \( a_1, a_2, \alpha_1, \alpha_2, K_1, \) and \( K_2. \)

In [9], [3], [7], [8], [6], the collage theorem presented in Section 2 is used to solve such an inverse problem. Starting from the differential equation,

\[
\dot{x} = f(t, x), \quad x(0) = x_0, \tag{7}
\]

we consider the Picard integral operator associated with it,

\[
(Tx)(t) = x_0 + \int_0^t f(s, x(s)) \, ds. \tag{8}
\]

If \( f \) is Lipschitz in the variable \( x \), that is, \( |f(s, x_1) - f(s, x_2)| \leq K|x_1 - x_2| \), then \( T \) is Lipschitz on the space \( C([-\delta, \delta] \times [-M, M]) \) with Lipschitz constant \( c = \delta K \) [9]. Thus, for \( \delta \) sufficiently small, \( T \) is contractive with respect to the \( L^2 \) metric. Now let \( \delta' > 0 \) be such that \( \delta' K < 1. \) Let \( \{\phi_i\} \) be a basis of functions in \( L^2([-\delta', \delta'] \times [-M, M]) \), then

\[
f_a(s, x) = \sum_{i=1}^{+\infty} a_i \phi_i(s, x). \tag{9}
\]
Each sequence of coefficients $a = \{a_i\}_{i=1}^{\infty}$, then defines a Picard operator $T_a$. Suppose further that each function $\phi_i(s, x)$ is Lipschitz in $x$ with constants $K_i$.

Theorem 4. [9] Let $\|K\|_2 = \left(\sum_{i=1}^{\infty} K_i^2\right)^{\frac{1}{2}}$ and $\|a\|_2 = \left(\sum_{i=1}^{\infty} a_i^2\right)^{\frac{1}{2}}$. Then

$$|f_a(s, x_1) - f_a(s, x_2)| \leq \|a\|_2 \|K\|_2 |x_1 - x_2|$$

for all $s \in [-\delta', \delta']$ and $x_1, x_2 \in [-M, M]$.

Given a target solution $x(t)$, we now seek to minimize the collage distance $\|x - T_a x\|_2$. The square of the collage distance becomes

$$\Delta(a)^2 = \|x - T_a x\|_2^2 = \int_{t-\delta}^t \left| x(t) - \int_0^t \sum_{i=1}^{\infty} a_i \phi_i(s, x(s)) ds \right|^2 dt$$

and the inverse problem can be formulated as

$$\min_{a \in A} \Delta(a),$$

where $A = \{a \in \mathbb{R}^{+\infty} : \|K\|_2 \|a\|_2 < 1\}$. The minimization may be performed by means of classical minimization methods on a subspace of finite dimension. Of course, the approximation error goes to zero when the dimension goes to infinity.

We apply this approach to our technological competition model inverse problem. We use collage coding, finding the system of the form

$$\frac{dx_1}{dt}(t) = b_1 x_1 + c_1 x_1^2 + d_1 x_1 x_2$$

$$\frac{dx_2}{dt}(t) = b_2 x_1 + c_2 x_2^2 + d_2 x_1 x_2$$

for which the corresponding $L^2$ collage distance is minimized. Having found the coefficients $b_i, c_i, d_i$, $i=1,2$, we obtain the approximation of the physical parameters via

$$a_i = b_i, \quad K_i = -\frac{b_i}{c_i}, \quad \text{and} \quad a_2 = \frac{d_1}{c_1}, \quad \alpha_1 = \frac{d_2}{c_2}.$$  

**Example 1.** We set $K_1 = 125$, $K_2 = 320$, $a_1 = 0.2$, $a_2 = 0.5$, $a_1 = 0.3$, and $a_2 = 0.6$, and solve numerically the system of differential equations. We gather observed data by adding low amplitude Gaussian noise to sampled values of the numerical solution. For $x_1(t)$, we gather 100 sample values at the times $t = \frac{i}{100}$, $0 \leq i \leq 99$; we add normally distributed noise with distribution noise1. We fit a piecewise tenth-degree polynomial to each consecutive set of ten data points to produce our target function for $x_1(t)$. We follow the same procedure to produce a target function for $x_2(t)$, this time with noise distribution noise2. Finally, we minimize the collage distance to recover values of $b_i, c_i, d_i,$ and $x_{i0}$, from which we recover the approximations of $a_i, K_i$, and $\alpha_i$, $i=1,2$. The results (to five decimal places) obtained for different noise distributions are presented in Table 1. The values in the table are quite close to the true values, with the accuracy decreasing as the noise is increased.
4 An inverse problem for an economic resource model

We consider a common access fishery model,

\[ \dot{a}(t) = \gamma(pHb(t) - \bar{c})a(t) \]
\[ \dot{b}(t) = B(\bar{b} - b(t))b(t) - Ha(t)b(t), \]

where the first equation models fishing effort by the fishermen, quantified by the number of boats on the water, and the second equation models the fish population. (With some tweaking, this model is the equivalent to the self-regulating predator-prey model found in biomathematics. Here, the fish are analogous to the prey and fishermen (or boats) to the predators.)

This model is referred to as a “common access” model because there are no barriers to entry. That is, fishermen are free to enter and exit the industry as they wish without penalty, cost, legal restriction, or any other stipulations which make entry difficult. In practice, we have entry so long as profits are positive. In the case of zero profits, we will neither have entry nor exit until other factors influence the dynamics of interaction between our players, the fish and fishermen. For instance, if there is suddenly a large number of fish, more fishermen will enter the industry in hopes of realizing potential gains from profit. The meaning of each term in our model is given in the following list:

- \( a(t) \) = number of boats at time \( t \)
- \( b(t) \) = number of fish at time \( t \)
- \( \bar{b} \) = sustainable fish population, \( \bar{b} > 0 \)
- \( B \) = scaling term = \( \frac{\text{growth rate of fish}}{\bar{b}} \), \( 0 < B < 1 \)
- \( H \) = technological constant, converts effort into catch, \( 0 < H < 1 \)
- \( \bar{c} \) = marginal constant cost per boat, \( \bar{c} > 0 \)
- \( \bar{p} \) = market price per fish, \( \bar{p} > 0 \)
- \( R = pHb(t)a(t) \) = total industry revenue at time \( t \), \( R > 0 \)
- \( E = R - ca(t) \) = industry profits at time \( t \), \( E > 0 \)
- \( \gamma \) = scaling term, \( \gamma > 0 \)

---

Table 1. Collage Coding Results for the Technological Competition Model

<table>
<thead>
<tr>
<th>noise1</th>
<th>noise2</th>
<th>( a_1 )</th>
<th>( K_1 )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( K_2 )</th>
<th>( a_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N(0, 0.02) )</td>
<td>( N(0, 0.04) )</td>
<td>0.28235</td>
<td>123.72647</td>
<td>0.17625</td>
<td>0.58718</td>
<td>319.86466</td>
<td>0.50565</td>
</tr>
<tr>
<td>( N(0, 0.10) )</td>
<td>( N(0, 0.15) )</td>
<td>0.26212</td>
<td>122.09029</td>
<td>0.15334</td>
<td>0.57498</td>
<td>319.74032</td>
<td>0.51273</td>
</tr>
<tr>
<td>( N(0, 0.20) )</td>
<td>( N(0, 0.20) )</td>
<td>0.23672</td>
<td>119.72296</td>
<td>0.11802</td>
<td>0.57147</td>
<td>319.69409</td>
<td>0.52283</td>
</tr>
<tr>
<td>( N(0, 0.30) )</td>
<td>( N(0, 0.45) )</td>
<td>0.22304</td>
<td>118.25901</td>
<td>0.11801</td>
<td>0.55675</td>
<td>319.55805</td>
<td>0.52880</td>
</tr>
<tr>
<td>( N(0, 0.50) )</td>
<td>( N(0, 0.75) )</td>
<td>0.20793</td>
<td>116.46787</td>
<td>0.09335</td>
<td>0.54418</td>
<td>319.44395</td>
<td>0.53596</td>
</tr>
<tr>
<td>( N(0, 0.80) )</td>
<td>( N(0, 1.00) )</td>
<td>0.17142</td>
<td>111.14858</td>
<td>0.02406</td>
<td>0.51072</td>
<td>319.15910</td>
<td>0.55635</td>
</tr>
</tbody>
</table>
Parameter estimation for differential equations...

\[ \gamma_c = \text{rate at which fishermen leave the water, } c > 0. \]

In the first equation, \( ca(t) \) is the total cost of the boats per unit time, at time \( t \). The product \( pHb(t)a(t) \) is the total revenue per unit time. The difference \( pHb(t)a(t) - ca(t) \) is the profit at time \( t \). The first equation says that the rate of change of the fishing effort is proportional to the profit. The first term of the left hand side of the second equation, \( B(\bar{b} - b(t))b(t) \), is in the usual logistic form, describing the natural dynamics of the fish population. The bracketed term \( (\bar{b} - b(t)) \) is the element which makes this model self-regulating. The second term, \( Ha(t)b(t) \), represents the total harvest.

We find that the model has three equilibria

\[
(0, 0), \ (0, \bar{b}), \ \text{and} \ (a^*, b^*) = \left( \frac{B}{H} \left[ \bar{b} - \frac{\bar{c}}{H\bar{p}} \right], \ \frac{\bar{c}}{H\bar{p}} \right).
\]

The final equilibrium corresponds to coexistence of the fish and fisherman populations in the case that \( \bar{b} - \frac{\bar{c}}{H\bar{p}} \) is positive. The linearization of the vector field is

\[
Df(a, b) = \left( \begin{array}{cc} \gamma \bar{c} & 0 \\ -Hb & Bb - 2Bb - Ha \end{array} \right)
\]

Evaluating at the origin, we have

\[
Df(0, 0) = \left( \begin{array}{cc} 0 & 0 \\ 0 & Bb \end{array} \right).
\]

We conclude that \( (0, 0) \) is unstable. At the equilibrium point \( (0, \bar{b}) \), we find

\[
Df(0, \bar{b}) = \left( \begin{array}{cc} \gamma \bar{c} & 0 \\ -H\bar{b} & -Bb \end{array} \right),
\]

with eigenvalues of each sign. The equilibrium point is an unstable saddle point. The stable ray of contraction on the \( b \)-axis corresponds to the fact that in the absence of fishermen the fish population approaches the sustainable fish population value, \( \bar{b} \). Finally, in the case \( \bar{b} - \frac{\bar{c}}{H\bar{p}} > 0 \), at the positive equilibrium point \( (a^*, b^*) \), we determine that

\[
Df(a^*, b^*) = \left( \begin{array}{cc} 0 & \gamma \bar{c}B \left( \bar{b} - \frac{\bar{c}}{H\bar{p}} \right) \\ -\frac{\bar{c}}{\bar{p}} & -\frac{Bc}{H\bar{p}} \end{array} \right).
\]

We calculate that

\[
\det(Df(a^*, b^*)) = \gamma \bar{c}B \left( \bar{b} - \frac{\bar{c}}{H\bar{p}} \right) > 0
\]

\[
\text{trace}(Df(a^*, b^*)) = -\frac{B\bar{c}}{H\bar{p}} < 0.
\]

We conclude that the coexistence equilibrium is a stable sink.
To illustrate the results, we set $\bar{b} = 1000000$, $B = \frac{7}{480000}$, $H = 0.5$, $\bar{c} = 100000$, $\bar{p} = 2$, $\gamma = \frac{1}{20000}$, and use the initial values $b_0 = 100000$ and $a_0 = 22$ to generate the phase portrait in Figure 3.

![Fig. 3](image)

**Fig. 3.** Left to right: graphs of $a(t)$ versus $t$, $b(t)$ versus $t$, and the phase portrait $b(t)$ versus $a(t)$

We are now interested in solving the inverse problems: given data values $b(t_i)$, $i = 1, \ldots, M$ and $a(t_j)$, $j = 1, \ldots, N$, find values of the physical variables $b$, $B$, $H$, $c$, $p$, and $\gamma$ so that the solution to the system agrees approximately with the data.

**Example 2.** To generate solution data, we set $\bar{b} = 1000000$, $B = \frac{7}{480000}$, $H = 0.5$, $\bar{c} = 100000$, $\bar{p} = 2$, and $\gamma = \frac{1}{20000}$, solve numerically for $b(t)$ and $a(t)$, and sample the solutions at uniformly-spaced times in $[0, 1]$, adding low-amplitude Gaussian noise with amplitude $\varepsilon_b$ and $\varepsilon_a$, respectively. We fit piecewise polynomial target functions to these noisy data values and minimize the $L^2$ collage distance corresponding to the differential equations

\[
\begin{align*}
\dot{b}(t) &= c_1 b(t) + c_2 b^2(t) + c_3 a(t)b(t) \\
\dot{a}(t) &= c_4 a(t)b(t) + c_5 a(t).
\end{align*}
\]  

The results for different noise amplitudes are summarized in Table 2.

We observe that $c_1 = B\bar{b}$, $c_2 = -B$, $c_3 = -H$, $c_4 = \gamma \bar{p} H$, and $c_5 = -\gamma \bar{c}$. If we assume that $\bar{p} = 2$ is known, since it is the price determined by the market, we can calculate the remaining parameters from the minimal collage distance coefficient values. We obtain the results in Table 3. The values in the table lie quite close to the true values.

**5 An inverse problem for a class of stochastic differential equations**

Let us consider the following system of stochastic differential equations:

\[
\begin{align*}
\frac{d}{dt}X_t &= AX_t dt + B_t, \\
x(0) &= x_0.
\end{align*}
\]  

\[
\text{(17)}
\]
Parameter estimation for differential equations... 89

Table 2. Minimal Collage Distance Coefficients for the Resource Model Inverse Problem

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$\kappa$</th>
<th>$b_0$</th>
<th>$a_0$</th>
<th>$c_1$</th>
<th>$c_2$</th>
<th>$c_3$</th>
<th>$c_4$</th>
<th>$c_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>100000</td>
<td>22.0000</td>
<td>14.5833</td>
<td>-0.00001</td>
<td>-0.5000</td>
<td>0.00005</td>
<td>-5.0000</td>
</tr>
<tr>
<td>0.05</td>
<td>0.02</td>
<td>100680</td>
<td>21.9916</td>
<td>14.5735</td>
<td>-0.00001</td>
<td>-0.4998</td>
<td>0.00005</td>
<td>-4.9535</td>
</tr>
<tr>
<td>0.08</td>
<td>0.04</td>
<td>101134</td>
<td>22.0046</td>
<td>14.5343</td>
<td>-0.00001</td>
<td>-0.4998</td>
<td>0.00005</td>
<td>-4.9128</td>
</tr>
<tr>
<td>0.10</td>
<td>0.06</td>
<td>101476</td>
<td>22.0279</td>
<td>14.4795</td>
<td>-0.00001</td>
<td>-0.4998</td>
<td>0.00005</td>
<td>-4.8755</td>
</tr>
<tr>
<td>0.15</td>
<td>0.05</td>
<td>101964</td>
<td>21.9608</td>
<td>14.5714</td>
<td>-0.00001</td>
<td>-0.4990</td>
<td>0.00005</td>
<td>-4.8661</td>
</tr>
<tr>
<td>0.20</td>
<td>0.10</td>
<td>102815</td>
<td>22.0237</td>
<td>14.4100</td>
<td>-0.00001</td>
<td>-0.4983</td>
<td>0.00005</td>
<td>-4.7674</td>
</tr>
</tbody>
</table>

Table 3. Minimal Collage Distance Parameter Values for the Resource Model Inverse Problem

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$\kappa$</th>
<th>$b_0$</th>
<th>$a_0$</th>
<th>$b$</th>
<th>$B$</th>
<th>$H$</th>
<th>$\bar{c}$</th>
<th>$p$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>100000</td>
<td>22.0000</td>
<td>0.0000100</td>
<td>0.500</td>
<td>100000</td>
<td>2</td>
<td>0.0000500</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>0.02</td>
<td>100680</td>
<td>21.99</td>
<td>0.0000146</td>
<td>0.500</td>
<td>995241</td>
<td>2</td>
<td>0.0000496</td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>0.04</td>
<td>101134</td>
<td>22.01</td>
<td>0.0000143</td>
<td>0.500</td>
<td>99848</td>
<td>2</td>
<td>0.0000492</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>0.06</td>
<td>101476</td>
<td>22.03</td>
<td>0.0000138</td>
<td>0.500</td>
<td>99852</td>
<td>2</td>
<td>0.0000488</td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>0.05</td>
<td>101964</td>
<td>21.96</td>
<td>0.0000150</td>
<td>0.499</td>
<td>99543</td>
<td>2</td>
<td>0.0000489</td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td>0.10</td>
<td>102815</td>
<td>22.02</td>
<td>0.0000136</td>
<td>0.498</td>
<td>99455</td>
<td>2</td>
<td>0.0000479</td>
<td></td>
</tr>
</tbody>
</table>

where $X: \mathbb{R} \times \Omega \to \mathbb{R}^n$, $A$ is a (deterministic) matrix of coefficients and $B_t$ is a classical vector Brownian motion. An inverse problem for this kind of equation can be formulated as: given an i.d. sample of observations of $X(t, \omega)$, say $(X(t, \omega_1), \ldots, X(t, \omega_n))$, get an estimation of the matrix $A$. For this purpose, let us take the integral over $\Omega$ of both sides of the previous equation and suppose that $X(t, \omega)$ is sufficiently regular; recalling that $B_t \sim \mathcal{N}(0, t)$, we have

$$
\int_{\Omega} \frac{dx}{dt} dP(\omega) = \frac{d}{dt} \mathbb{E}(X(t, \cdot)) = AE(X(t, \cdot))
$$

This is a deterministic differential equation in $\mathbb{E}(X(t, \cdot))$. From the sample of observations of $X(t, \omega)$ we can then get an estimation of $\mathbb{E}(X(t, \cdot))$ and then use of approach developed for deterministic differential equations to solve the inverse problem for $A$. The essential idea from [5] is that each realization $x(\omega_j, s)$, $j = 1, \ldots, N$, of the random variable $x(\omega, s)$ is the solution of a fixed point equation

$$
x(\omega_j, s) = \int_0^s \phi(\omega_j, t, x(\omega_j, t)) dt + x_0(\omega_j)
$$

Thus, for each target function $x(\omega_j, s)$, we can find the constant values $x_0(\omega_j)$ and $a_i(\omega_j)$ via collage coding. Upon treating each realization, we will have de-
Herb Kunze et al.

determined $x_0(\omega_j)$ and $a_i(\omega_j)$, $i = 1, \ldots, M$, $j = 1, \ldots, N$. We then construct the approximations

$$
\mu \approx \mu_N = \frac{1}{N} \sum_{j=1}^{N} x_0(\omega_j) \text{ and } \nu_i \approx (\nu_i)_N = \frac{1}{N} \sum_{j=1}^{N} a_i(\omega_j),
$$

(19)

where we note that results obtained from collage coding each realization are independent. Using our approximations of the means, we can also calculate that

$$
\sigma^2 \approx \sigma_N^2 = \frac{1}{N-1} \sum_{j=1}^{N} (x_0(\omega_j) - \mu_N)^2, \quad \sigma^2_i \approx (\sigma_i)_N^2 = \frac{1}{N-1} \sum_{j=1}^{N} (a_i(\omega_j) - (\nu_i)_N)^2.
$$

As a numerical example, we consider the first-order system

$$
\frac{d}{dt} x_t = a_1 x_t + a_2 y_t + b_t \\
\frac{d}{dt} y_t = b_1 x_t + b_2 y_t + c_t
$$

Setting $a_1 = 0.5$, $a_2 = -0.4$, $b_1 = -0.3$, $b_2 = 1$, $x_0 = 0.9$, and $y_0 = 1$, we construct observational data values for $x_t$ and $y_t$ for $t_i = \frac{i}{N}$, $1 \leq i \leq N$, for various values of $N$. For each of $M$ data sets, different pairs of Brownian motion are simulated for $b_t$ and $c_t$. Figure 4 presents several plots of $b_t$ and $c_t$ for $N = 100$.

**Fig. 4.** Example plots of $b_t$ and $c_t$ for $N = 100

In Figure 5, we present some plots of our generated $x_t$ and $y_t$, as well as phase portraits for $x_t$ versus $y_t$. For each sample time, we construct the mean of the observed data values, $x^*_i$ and $x^*_i$, $1 \leq i \leq N$. We minimize the squared collage distances

$$
\Delta^2_x = \frac{1}{N} \sum_{i=1}^{N} \left( x^*_i - x_0 - \frac{1}{N} \sum_{j=1}^{i} (a_1 x^*_j + a_2 y^*_j) \right)^2
$$
Parameter estimation for differential equations...

Fig. 5. Example plots of $x_t$, $y_t$, and $x_t$ versus $y_t$ for $N = 100$

and

$$
\Delta_y^2 = \frac{1}{N} \sum_{i=1}^{N} \left( y_{t_i}^* - y_0 - \frac{1}{N} \sum_{j=1}^{i} \left( b_1 x_{t_j}^* + b_2 y_{t_j}^* \right) \right)^2
$$

to determine the minimal collage parameters $a_1$, $a_2$, $b_1$, and $b_2$. The results of the process are summarized in Table 4.

Table 4. Minimal collage distance parameters for different $N$ and $M$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$M$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>100</td>
<td>0.2613</td>
<td>-0.2482</td>
<td>-0.2145</td>
<td>0.9490</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>0.3473</td>
<td>-0.3496</td>
<td>-0.2447</td>
<td>0.9709</td>
</tr>
<tr>
<td>100</td>
<td>300</td>
<td>0.3674</td>
<td>-0.3523</td>
<td>-0.2494</td>
<td>0.9462</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0.3775</td>
<td>-0.3015</td>
<td>-0.1989</td>
<td>0.9252</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0.3337</td>
<td>-0.3075</td>
<td>-0.2614</td>
<td>0.9791</td>
</tr>
<tr>
<td>200</td>
<td>300</td>
<td>0.4459</td>
<td>-0.3858</td>
<td>-0.2822</td>
<td>0.9718</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>0.4234</td>
<td>-0.3246</td>
<td>-0.2894</td>
<td>0.9838</td>
</tr>
<tr>
<td>300</td>
<td>200</td>
<td>0.3834</td>
<td>-0.3263</td>
<td>-0.3111</td>
<td>1.0099</td>
</tr>
<tr>
<td>300</td>
<td>300</td>
<td>0.5094</td>
<td>-0.4260</td>
<td>-0.3157</td>
<td>0.9965</td>
</tr>
</tbody>
</table>

6 Concluding Remarks

In this paper, we have considered three inverse problems drawn from applications in economics and finance. The fundamental approach for solving the problems is rooted in fractal-based analysis. The results in the paper demonstrate the usefulness of the collage method. It is worth mentioning that the method does not require significant computational power or time.
References

Jump telegraph processes and a volatility smile

Nikita Ratanov

Facultad de Economía, Universidad del Rosario,
C.14, No.4-69, Bogotá, Colombia
nratanov@urosario.edu.co

Abstract. We continue to study financial market models based on generalized telegraph processes with alternating velocities. The model is supplied with jumps occurring at the times of velocity switchings. This model is arbitrage-free and complete if the directions of jumps in stock prices are in a certain correspondence with their velocity and with the behaviour of the interest rates. A risk-neutral measure and arbitrage-free formulae for a standard call option are constructed. A new version of convergence under suitable scaling to the Black-Scholes model is proved, and the explicit limit is obtained. Next, we examine numerically the explicit formulae for call prices to obtain the behaviour of implied volatilities. Moreover, this model has some features of models with memory. The historical volatility of jump telegraph model is similar to historical volatility of the moving average type model.

Keywords. telegraph process, option pricing, volatility smile.

M.S.C. classification. 91B28, 60J75.

1 Introduction

The famous Black-Scholes formula has well known shortages and it is rarely used to price options. It is commonly accepted that Black-Scholes pricing formula underprices deep-in-the-money and out-of-the-money options and it overprices at-the-money options, see [15] and e.g. [16]. This observation provokes a growing interest in the construction of more and more complicated stochastic volatility models based on stochastic dynamics of the Black-Scholes implied volatility (see a review of these activities e.g. in [4]). The volatility implied by the Black-Scholes formula is used in these models as common language to explain how the option should be priced. Usually the implied volatility as a function of moneyness $K/S_0$ forms a so called “volatility smile”. On the other hand this “smile-consistent” approach proposes the quantity sophistication instead of fundamental explanation of difficulties. Moreover, models of this type increase Markov dimension of the market.

To explain market’s movements we propose a rather new model based on telegraph-like processes. This paper continues our previous research [20] of such
a model. Suppose that the log-returns are driven by a telegraph process, i.e. they move with pair of constant velocities alternating one to another at Poisson times. To make the model more adequate and to avoid arbitrage opportunities the log-return movement should be supplied with jumps occurring at times of the tendency switchings.

As a basis for building the model in Section 2, we take a counting Poisson process $N = N(t), t \geq 0$ with alternating transition intensities $\lambda_{\pm} > 0$. The process $\sigma(t) = (-1)^{N(t)}$ (or $\sigma(t) = -(\sigma(t))$ with values $\pm 1$ displays a current market state. Using $\sigma(t), t \geq 0$, we define processes $c_{\sigma(t)} = c_{\pm}, h_{\sigma(t)} = h_{\pm}, h_{\pm} > -1, r_{\sigma(t)} = r_{\pm}, r_{\pm} \geq 0$. Processes $T_s$ and $J_s$ are defined as $T_s(t) = \int_{0}^{t} c_{\sigma(\tau)} d\tau$ and $J_s(t) = \int_{0}^{t} h_{\sigma(\tau)} dN(\tau)$. The evolution of the risky asset $S(t)$ is determined by a stochastic exponential of the sum $T_s + J_s$. The risk-free asset is given by the usual exponential of the process $T_s = T_s(t) = \int_{0}^{t} r_{\sigma(\tau)} d\tau, t \geq 0$. Here and below the subscript $s$ indicates the starting value $s = \sigma(0) = \pm 1$ of the market’s state $\sigma(t)$.

In view of such trajectories, the market is set up as a continuous process that evolves with velocity $c_{\pm}$ or $c_{-}$, changes the direction of movement from $c_{\pm}$ to $c_{\mp}$ and exhibits jumps of size $h_{\pm}$ whenever velocity changes. The interest rate in the market is stochastic with values $r_{\pm}$.

The processes $T_s(t) + J_s(t), t \geq 0$ are given by the pair of states $(c_{\pm}, \lambda_{\pm}, h_{\pm})$. They are called jump telegraph processes with states $(c_{\pm}, \lambda_{\pm}, h_{\pm})$. This model is regarded as jump telegraph market model.

In Section 2 we describe the model in detail. This section contains also the explicit expressions for means and variances which are exploited to describe historical volatility in Section 4. For the beginning all parameters are supposed to be deterministic, which leads to completeness of the market. The case of random jump values and random velocities creates incomplete market model and it will be reported anywhere later.

Such a model looks attractive because of finite propagation velocity and the intuitively clear comportment. Under respective scaling it converges to Black-Scholes model. Section 3 is concerned with this convergence and the definition of volatility in jump telegraph model. It contains a new version of scaling theorem (cf. [20] and [21]), and a new fundamental and natural explanation of volatility. It permits us to define the volatility of the jump telegraph model depending on the velocities $c_{\pm}$, the jumps values $h_{\pm}$ and the switching intensities $\lambda_{\pm}$. Further, (Section 4), we consider a historical volatility as $\text{HV}(t) = \sqrt{\text{Var}S(t)/t}$, and then an implied volatility as $\text{IV}(t) = \sqrt{\text{Var}S(t)/t}$, where $\text{Var}S(t)$ is implied variances of jump telegraph model with respect to the Black-Scholes dynamics. The implied volatility $\text{IV}(t)$ with various values of log-moneyness $\mu$ forms the so called volatility smile. Volatility smiles of various shapes are presented in Section 5.

Telegraph processes have been studied before in different probabilistic aspects (see, for instance, Goldstein [9] (1951), Kac [11], [12] (1974) and Zacks [22] (2004)). These processes have been exploited for stochastic volatility modelling (Di Masi et al. [7] (1994)) as well as for obtaining a “telegraph analog” of the
Black-Scholes model (Di Crescenzo and Pellicer [6] (2002)). In contrast with the paper by Di Crescenzo and Pellecer, we use more complicated and delicate construction of such a model to avoid arbitrage and to develop an adequate option pricing theory in this framework. Recently telegraph processes was applied to actuarial problems [14].

Parameters of telegraph market model was calibrated in the working paper of De Gregorio and Lacus [5]. This calculations are based on weekly closings of the Dow-Jones industrial average July 1971 - Aug 1974 and returns of IBM stock closings. In Section 5 we use these calibrated data to estimate the implied volatility (see Table 3 and Figure 4).

2 Jump telegraph processes with alternating intensities

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\lambda_{\pm}$ be positive numbers. First, we consider two counting Poisson processes $N_+ = N_+(t)$, $N_- = N_-(t)$, $t \geq 0$ with values $\{0, 1, 2, \ldots\}$ and alternating intensities $\lambda_{\pm} := \lambda_{\pm}$, i.e. for $n = 0, 1, 2, \ldots$ as $\Delta t \to 0$

$$
\mathbb{P}\{N_-(t + \Delta t) = 2n + 2 \mid N_-(t) = 2n + 1\} = \lambda_+ \Delta t + o(\Delta t),
$$

$$
\mathbb{P}\{N_+(t + \Delta t) = 2n + 2 \mid N_+(t) = 2n + 1\} = \lambda_- \Delta t + o(\Delta t),
$$

$$
\mathbb{P}\{N_-(t + \Delta t) = 2n + 1 \mid N_-(t) = 2n\} = \lambda_- \Delta t + o(\Delta t),
$$

$$
\mathbb{P}\{N_+(t + \Delta t) = 2n + 1 \mid N_+(t) = 2n\} = \lambda_+ \Delta t + o(\Delta t).
$$

Further we will consider all stochastic processes subscribed by + or - to be adopted to filtration $\mathcal{F} = (\mathcal{F}_t^\pm)_{t \geq 0}$ or $\mathcal{F} = (\mathcal{F}_t^-)_{t \geq 0}$ generated by $N_+ = N_+(t)$ and $N_- = N_-(t)$ respectively.

Processes $\sigma_+(t) = (-1)^{N_+(t)}$ and $\sigma_-(t) = (-1)^{N_-}(t)$ indicate a current state: if $0 < \tau_1 < \tau_2 < \tau_3 < \ldots$ is a Poisson flow, then interarrival times $\tau_{n+1} - \tau_n$, $n = 0, 1, 2, \ldots$ are independent and exponentially distributed with parameter $\lambda_{\sigma_+(\tau_n)}$. Subscripts $\pm$ here respect to the initial state of the medium.

Let $c_-, c_+$ and $h_-, h_+$ be real numbers, $h_+ > -1$, $h_- < c_+$, $T_-, T_+ = T_+(t)$, $t \geq 0$ are defined as follows:

$$
T_\pm(t) = \int_0^t c_{\sigma_\pm(t')}dt'.
$$

We define also right continuous pure jump processes $J_- = J_-(t)$, $J_+ = J_+(t)$, $t \geq 0$, which are driven by the same Poisson processes:

$$
J_{\pm}(t) = \int_0^t h_{\sigma_\pm(t')}dN_{\pm}(t') = \sum_{j=1}^{N_+(t)} h_{\sigma_-(\tau_j)}.
$$
Processes $X_\tau = T_\tau + J_\tau$ and $X_+ = T_+ + J_+$ are referred to as jump telegraph processes with parameters $c_\pm, \lambda_\pm, h_\pm$.

The following theorem could be interpreted as a version of the Doob-Meyer decomposition for telegraph processes.

**Theorem 1.** Jump telegraph processes $X_\tau = X_\tau(t)$ and $X_+ = X_+(t)$, $t \geq 0$ with parameters $c_\pm, h_\pm, \lambda_\pm$ are martingales if and only if $c_+ = -\lambda_+, h_+$ and $c_- = -\lambda_-, h_-$.  

**Proof.** See [20], Theorem 1.

Next, we study the properties of telegraph processes under a change of measure. Let $T_\ast^\ast$ be the telegraph process with the states $(c_\pm, \lambda_\pm)$, and $J_\ast^\ast = \sum_{j=1}^{N_\ast(j)} c_\ast^{(\tau_j-)/\lambda_\ast(\tau_j-)}$ be the jump process with jump values $h_\pm = -c_\pm/\lambda_\pm > -1$. Consider a probability measure $\mathbb{P}_s^\ast$ with the following local density (with respect to $\mathbb{P}_s$):

$$Z_\ast(t) = \frac{d\mathbb{P}_s^\ast}{d\mathbb{P}_s}|_t = \mathcal{E}_t(T_\ast^\ast + J_\ast^\ast), \quad 0 \leq t \leq T, \ s = \pm. \quad (1)$$

Here $\mathcal{E}_t(\cdot)$ denotes stochastic exponential.

Using properties of stochastic exponentials, we obtain

$$Z_\ast(t) = e^{T_\ast(t) \kappa_\ast^\ast(t)}, \quad (2)$$

where $\kappa_\ast^\ast(t) = \prod_{\tau \leq t} (1 + \Delta J_\ast^\ast(\tau))$ with $\Delta J_\ast^\ast(\tau) = J_\ast^\ast(\tau) - J_\ast^\ast(\tau^{-})$.

The process $\kappa_\ast^\ast = \kappa_\ast^\ast(t), t \geq 0$ can be represented as $\kappa_\ast^\ast(t) = \kappa_\ast^{N_\ast(t),s}$. Here the sequence $\kappa_\ast^{n,s}$ is defined as follows:

$$\kappa_\ast^{n,s} = \kappa_\ast^{n-1,-s}(1 + h_\ast^s), n \geq 1, \quad \kappa_\ast^{0,s} \equiv 1. \quad (3)$$

It means that if $n = 2k$,

$$\kappa_\ast^{n,s} = (1 + h_\ast^s)^k(1 + h_-^s)^k;$$

and if $n = 2k + 1$,

$$\kappa_\ast^{n,s} = (1 + h_\ast^s)^{k+1}(1 + h_-^s)^k.$$

**Theorem 2 (Girsanov theorem).** Under the probability measure $\mathbb{P}_s^\ast$,

- process $N_\ast = N_\ast(t), t \geq 0$ is a Poisson process with intensities $\lambda_-^\ast = \lambda_--c_-^s = \lambda_-(1 + h_-^s)$ and $\lambda_+^\ast = \lambda_+ - c_+^s = \lambda_+(1 + h_+^s)$.
- process $T_\ast = T_\ast(t), t \geq 0$ is a telegraph process with states $(c_-, \lambda_-^\ast)$ and $(c_+, \lambda_+^\ast)$.

Probability measure $\mathbb{P}_s^\ast$ becomes the martingale measure for jump telegraph process $T_\ast + J_\ast$, if it is constructed using parameters $c_-^s = \lambda_- + \frac{c_-}{h_-}, c_+^s = \lambda_+ + \frac{c_+}{h_+}, h_-^s = -1 - \frac{c_-}{\lambda_-, h_-}$ and $h_+^s = -1 - \frac{c_+}{\lambda_+, h_+}$.
Theorem 3. Let $\Delta$ be a pair of jump telegraph processes with parameters $<c, h, \lambda >$. Their probability densities $p_\pm^{(n)}$ solve the system

$$\begin{cases}
\frac{\partial p_+^{(n)}}{\partial t} + c_+ \frac{\partial p_+^{(n)}}{\partial x} = -\lambda_+ [p_+^{(n)}(x, t) - p_+^{(n-1)}(x-h_+, t)], \\
\frac{\partial p_-^{(n)}}{\partial t} + c_- \frac{\partial p_-^{(n)}}{\partial x} = -\lambda_- [p_-^{(n)}(x, t) - p_+^{(n-1)}(x-h_-, t)]
\end{cases}$$

(5)

with zero initial conditions $p_\pm^{(n)}|_{t=0} = 0$, $n \geq 1$ and $p_\pm^{(0)}(x, t) = e^{-\lambda \pm t} \delta(x-c \pm t)$.

Proof. See [19], equation (2.12).

System (5) has the following solution (see e.g. [21]), $p_\pm^{(n)}(x, t) = g_\pm^{(n)}(x - j_\pm^{(n)}, t)$, where

$$j_\pm^{(n)} = \begin{cases}
k(h_+ + h_-), & n = 2k, \\
k(h_+ + h_-) + h_\pm, & n = 2k + 1 \forall k \in \mathbb{Z}.
\end{cases}$$

and

$$g_+^{(n)}(x, t) = e^{-\mu t - \nu x} \lambda_+^{\lfloor n/2 \rfloor} \lambda_-^{\lfloor n/2 \rfloor} (c_+ - c_-)^n (c_+ t - x)^{n - \lfloor n/2 \rfloor - 1} (x - c_- t)^{\lfloor n/2 \rfloor} 1_{\{c_- t < x < c_+ t\}},$$

$$g_-^{(n)}(x, t) = e^{-\mu t - \nu x} \lambda_+^{\lfloor n/2 \rfloor} \lambda_-^{\lfloor n/2 \rfloor} (c_+ - c_-)^n (c_+ t - x)^{\lfloor n/2 \rfloor - (n - \lfloor n/2 \rfloor - 1)} (x - c_- t)^{n - \lfloor n/2 \rfloor} 1_{\{c_- t < x < c_+ t\}},$$

$n \geq 1$. Here

$$\nu = \frac{\lambda_+ - \lambda_-}{c_+ - c_-}, \quad \mu = \lambda \pm - \nu c \pm = \frac{c_+ \lambda_- - c_- \lambda_+}{c_+ - c_-}.$$
Fig. 1. Probability densities of telegraph process \( T_\pm(t) \) (absolutely continuous part) with values \( t = 1, c_\pm = \pm 4, h_\pm = \mp 0.2 \) and with \( \lambda_\pm = 5 \) or \( \lambda_\pm = 20 \).

\[
p_\pm(x, t) = \sum_{n=0}^{\infty} p_\pm^{(n)}(x, t) = \sum_{n=0}^{\infty} g_\pm^{(n)}(x - j_\pm^{(n)}, t),
\]

(6)

and functions \( p_\pm \) satisfy the following system (see equation (2.9) in [21]):

\[
\begin{align*}
\frac{\partial p_+}{\partial t} + c_+ \frac{\partial p_+}{\partial x} &= -\lambda_+ [p_+(x, t) - p_-(x - h_+, t)], \\
\frac{\partial p_-}{\partial t} + c_- \frac{\partial p_-}{\partial x} &= -\lambda_- [p_-(x, t) - p_+(x - h_-, t)],
\end{align*}
\]

(7)

with the initial condition \( p_\pm(x, 0) = \delta(x) \).

The densities \( p_\pm \) with certain \( c_\pm, h_\pm \) and different \( \lambda_\pm \) are presented in Fig.1.

Representation (6) of the telegraph process densities is adapted to the following rule of measure change. If the intensities \( \lambda_\pm \) of the driving Poisson process are changed to \( \bar{\lambda}_\pm \), then the densities of telegraph process will take the form

\[
\bar{p}_\pm(x, t) = e^{-(\bar{\mu} - \mu)t - \bar{\nu} - \nu} x \sum_{n=0}^{\infty} p_\pm^{(n)} \times \kappa_\pm^{(n)},
\]

(8)

where \( \kappa_\pm^{(n)} = (\bar{\lambda}_+/\lambda_+)^{n-[n/2]} (\bar{\lambda}_-/\lambda_-)^{[n/2]} \), \( \kappa_-^{(n)} = (\bar{\lambda}_+/\lambda_+)^{[n/2]} (\bar{\lambda}_-/\lambda_-)^{n-[n/2]} \),

\[
\nu = \frac{\bar{\lambda}_+ - \bar{\lambda}_-}{c_+ - c_-} \quad \text{and} \quad \bar{\mu} = \frac{c_+ \bar{\lambda}_+ - c_- \bar{\lambda}_-}{c_+ - c_-}.
\]

Applying (7) one can easily obtain the following system for expectations (see [21], Corollary 2.6).
Lemma 1. Let \( f = f(x) \) and \( \mu_{\pm} = \mu_{\pm}(t), \ t \geq 0 \) be smooth deterministic functions, \( X_{\pm} \) be jump telegraph processes with parameters \( < c_{\pm}, h_{\pm}, \lambda_{\pm}> \). Then functions

\[
u_{\pm} = u_{\pm}(x, t) = \mathbb{E}f(x - \mu_{\pm}(t) + X_{\pm}(t))
\]

form a solution of the system

\[
\begin{align*}
\frac{\partial u_{+}}{\partial t} - (c_+ - \mu_+) \frac{\partial u_{+}}{\partial x} &= -\lambda_+ [u_+(x, t) - u_-(x + \beta_+(t), t)] \\
\frac{\partial u_{-}}{\partial t} - (c_- - \mu_-) \frac{\partial u_{-}}{\partial x} &= -\lambda_- [u_-(x, t) - u_+(x + \beta_-(t), t)]
\end{align*} 
\]  \tag{9}

with \( \beta_+(t) = h_+ - (\mu_+(t) - \mu_-)(t) \), \( \beta_-(t) = h_- - (\mu_-(t) - \mu_+)(t) \).

Here \( \mu_{\pm} = \frac{\partial u_{\pm}}{\partial x} \).

From (9) we deduce formulae for mean value and variance of a jump telegraph process

\[
m_{\pm}(t) = \mathbb{E}(X_{\pm}(t)), \quad s_{\pm}(t) = \text{Var}(X_{\pm}(t)).
\]

Indeed, with the choices \( f(x) = x, \mu_{\pm} = 0 \) and \( f(x) = x^2, \mu_{\pm} = m_{\pm}(t) \) we get respectively

\[
\frac{dm_{\pm}}{dt} = Am + v_1 \tag{10}
\]

and

\[
\frac{ds_{\pm}}{dt} = As + v_2. \tag{11}
\]

Here

\[
A = \begin{bmatrix}
-\lambda_+ & \lambda_+ \\
\lambda_- & -\lambda_-
\end{bmatrix}, \quad m = \begin{bmatrix}
m_+(t) \\
m_-(t)
\end{bmatrix}, \quad s = \begin{bmatrix}
s_+(t) \\
s_-(t)
\end{bmatrix},
\]

\[
v_1 = \begin{bmatrix}
c_+ + \lambda_+h_+ \\
c_- + \lambda_-h_-
\end{bmatrix} \quad \text{and} \quad v_2 = \begin{bmatrix}
\lambda_+(h_+ + m_+ - m_-)^2 \\
\lambda_-(h_- + m_+ - m_-)^2
\end{bmatrix}.
\]

System (10)-(11) can be explicitly resolved:

\[
m(t) = \frac{t}{2\lambda} \left( C_1 \begin{bmatrix}
1 \\
1
\end{bmatrix} + C_2\Phi_\lambda(t) \begin{bmatrix}
\lambda_+ \\
-\lambda_-
\end{bmatrix} \right), \quad \Phi_\lambda(t) = \frac{1 - e^{-2\lambda t}}{2\lambda}, \tag{12}
\]

where \( C_1 = \lambda_+(c_+ + \lambda_+h_+) + \lambda_+(c_- + \lambda_-h_-), \ C_2 = c_+ - c_- + \lambda_+h_+ - \lambda_-h_- \).

Then

\[
m_+ - m_- = \frac{C_2}{2\lambda}(1 - e^{-2\lambda t}) = C_2\Phi_\lambda(t)
\]

and

\[
s(t) = \int_0^t e^{(t-\tau)A}v_2(\tau)d\tau \tag{13}
\]

with

\[
v_2(\tau) = \begin{bmatrix}
\lambda_+(h_+ - C_2\tau\Phi_\lambda(\tau))^2 \\
\lambda_-(h_- + C_2\tau\Phi_\lambda(\tau))^2
\end{bmatrix}.
\]
With this in hand we can easy find the limits of \( s_\pm(t)/t \) as \( t \to 0 \) and as \( t \to \infty \):

\[
\lim_{t \to 0} s_\pm(t)/t = \lambda_\pm h_\pm^2,
\]
\[
\lim_{t \to \infty} s_\pm(t)/t = \frac{\lambda_+ \lambda_-}{(\lambda_+ + \lambda_-)^2} \left[ (\lambda_- (h_+ + h_-) + c_- - c_+)^2 + (\lambda_+ (h_+ + h_-) + c_+ - c_-)^2 \right].
\]

(14)

We shall use these limits in Section 4 to evaluate the comportment of historical volatilities.

In symmetric case \( \lambda_+ = \lambda_- := \lambda \) the formulae for solutions of (10)-(11) can be simplified as follows. Setting \( A = (c_+ + c_-)/2, a = (c_+ - c_-)/2, B = (h_+ + h_-)/2, b = (h_+ - h_-)/2, \gamma_+ = -2a(a/\lambda + h_+), \gamma_- = -2a(a/\lambda - h_-) \) we have

\[
m_\pm(t) = [A + \lambda B \pm (a + \lambda b)\Phi_\lambda(t)] t, \tag{15}
\]

\[
s_\pm(t) = [a^2/\lambda + \lambda B^2 + (a + \lambda b)^2\Phi_\lambda(t)/\lambda + \gamma_\pm \Phi_\lambda(t) \pm 2B(a + \lambda b)e^{-2\lambda t}] t. \tag{16}
\]

These formulae are presented in [21] (see Theorem 2.7, formulae (2.25)-(2.26)).

3 Jump telegraph market model and diffusion rescaling

We consider a market with one stock and a bond.

The stock price \( S(t) = S_\pm(t), t \geq 0 \) follows the equation

\[
dS(t) = S(t-)dX(t), \quad S(0) = S^0, \quad \sigma(0) = \pm 1, \tag{17}
\]

where \( X(t) = X_\pm(t) = T_\pm(t) + J_\pm(t), t \geq 0 \) is the jump telegraph process with parameters \( c_\pm, h_\pm, \lambda_\pm >, \sigma(0) = \pm 1 \) indicates initial market trend. Integrating we have

\[
S(t) = S^0 \xi_t(X) = S^0 \exp(T(t))\kappa(t), \quad \kappa(t) = \prod_{n=0}^{N(t)} (1 + h_{\sigma(\tau_n)^-}). \tag{18}
\]

The bond price is

\[
B(t) = \exp(\mathcal{T}(t)), \tag{19}
\]

where \( \mathcal{T} = \mathcal{T}_\pm(t), t \geq 0 \) be the telegraph process with velocities \( r_\pm \geq 0 \), which is driven by the same inhomogeneous Poisson process:

\[
\mathcal{T}(t) = \int_0^t r_{\sigma(\tau')}d\tau'.
\]
Model (18)-(19) is named jump telegraph market model.

The probability measure $P^*$ is the martingale measure for pricing process $\tilde{S}(t) \equiv B(t)^{-1}S(t)$. Process $\tilde{S}(t) = S_0 \exp(T(t) - T(t))\kappa(t)$, $t \geq 0$ is again the stochastic exponent of jump telegraph process with parameters $< c_\pm - r_\pm, h_\pm, \lambda_\pm >$. So with no loss of generality we assume $r_\pm = 0$. Thus, the stock price process $S(t)$ is a nonnegative $P^*$-martingale. By Theorem 1 under measure $P^*$ the driving Poisson process $N$ has intensities $\lambda_+^* = -c_- / h_-$ and $\lambda_-^* = -c_+ / h_+$, and by Theorem 2 change of measure is defined by $c_\pm^* = \lambda_\pm - \lambda_\pm^*$.

It is well known that under suitable scaling the telegraph process $\tilde{T}(t), t \geq 0$ converges to a Brownian motion: if $c_+ \to +\infty, c_- \to -\infty, \lambda_\pm \to \infty$ such that $c_+ / \sqrt{\lambda_+} \to \sigma$, $c_- / \sqrt{\lambda_-} \to -\sigma$, then the telegraph process $T(t), t \geq 0$ converges in distribution to $\sigma w(t), t \geq 0$, where $w$ denotes the standard Brownian motion. This convergence was first proved in [11]; see more details and some extensions in [18].

Thus, it is reasonable to obtain a similar rescaling result for jump telegraph model (18). To separate the drift from the diffusion component we consider the telegraph processes $\tilde{T}_\pm(t), t \geq 0$, driven by the same Poisson process as $T_\pm$ and with velocities $a_+$ and $-a_-$, where $a_+ = \frac{c_+ - c_-}{\sqrt{\lambda_+} + \sqrt{\lambda_-}} \sqrt{\lambda_+}$ and $a_- = \frac{c_+ - c_-}{\sqrt{\lambda_+} + \sqrt{\lambda_-}} \sqrt{\lambda_-}$.

Notice that $a_+ + a_- = c_+ - c_- \text{ and } \frac{a_+}{a_+ + a_-} = \frac{\sqrt{\lambda_+}}{\sqrt{\lambda_+} + \sqrt{\lambda_-}}$.

It is easy to see that

$$T_+(t) - \tilde{T}_+(t) = T_-(t) - \tilde{T}_-(t) = At,$$

where $A = \frac{c_+ \sqrt{\lambda_+} - c_- \sqrt{\lambda_-}}{\sqrt{\lambda_+} + \sqrt{\lambda_-}} = \frac{c_+ a_+ + c_- a_-}{c_+ - c_-}$.

Further we assume $\lambda_\pm \to +\infty, c_+ - c_- \to +\infty$ and

$$\frac{c_+ - c_-}{\sqrt{\lambda_+} + \sqrt{\lambda_-}} \to \sigma, \quad \sqrt{\frac{\lambda_+}{\lambda_-}} \to \gamma$$

for some $\sigma, \gamma \geq 0$.

To control jump and drift components we suppose $h_\pm \to 0$ such that for some $\alpha_\pm, \delta \in (-\infty, \infty)$

$$\sqrt{\lambda_\pm} h_\pm \to \alpha_\pm$$

and

$$\Delta := A + \frac{\sqrt{\lambda_+} \lambda_-}{\sqrt{\lambda_+} + \sqrt{\lambda_-}} (\sqrt{\lambda_+} h_+ + \sqrt{\lambda_-} h_-) \to \delta.$$  

Notice that $\Delta = \frac{\sqrt{\lambda_+}}{\sqrt{\lambda_+} + \sqrt{\lambda_-}} (c_+ + \lambda_- h_-) + \frac{\sqrt{\lambda_-}}{\sqrt{\lambda_+} + \sqrt{\lambda_-}} (c_+ + \lambda_+ h_+)$ and set $\beta^2 = \frac{\alpha_+^2 + \gamma \alpha_-^2}{\Delta^2}$.

The following theorem generalizes previous author’s results (see Theorem 3.3 [21] and Theorem 4 [20]). The scaling property (24) can be applied to interpret a volatility in jump telegraph model (18).
Theorem 4. Under scaling (21)-(23) jump telegraph model (18) converges in
distribution to the Black-Scholes model:

\[ S(t) \xrightarrow{D} S_0 \exp\{vw(t) + (\delta - \beta^2/2)t\}, \quad (24) \]

where \( v = \sqrt{\sigma + (\gamma \alpha_+ - \alpha_-)/(1 + \gamma)} \).

Remark 1. Under the martingale measure \( \mathbb{P}^* \) transition intensities take a form
\(-c_+/h_+\). Thus the drift vanishes, \( \Delta = \sqrt{\lambda_-/\lambda_+ + \lambda_-/\lambda_+}(c_- + \lambda_- h_- + \lambda_+ h_+) = 0. \) Moreover, in this case \( \sigma = 1 + \frac{c_- - c_+}{\sqrt{\lambda_+ + \lambda_-}} = \left( \frac{\alpha_+ - \alpha_-}{1 + \gamma} \right)^{\pm} \). The limiting volatility \( v \) in this case coincides with \( \beta; \ v = \beta = \sqrt{\frac{\alpha_+ + \gamma \alpha_-}{1 + \gamma}} \).

Proof. Let \( f_{\pm}(z, t) = \mathbb{E}X_t(zX_{\pm}) = \mathbb{E}e^{zT_t(z)}P_{\pm}(t)^2 \) be the moment-generating function. We prove here the convergence

\[ f_{\pm}(z, t) \to \exp\{(\delta - \beta^2/2) z t + v^2/z^2 t/2\}, \quad (25) \]

which is sufficient for the convergence of pointwise distributions in (24).

Using (20) and the rule (8) we note that

\[ f_{\pm}(z, t) = e^{Azt} \mathbb{E}e^{zT_t(z)}P_{\pm}(t)^2 = e^{Azt+(\bar{\mu}-\mu)t} \int_{-\infty}^{\infty} e^{(z+\nu)x} \bar{p}_{\pm}(x, t) \, dx. \]

Here \( \bar{p}_{\pm} \) are (generalized) probability densities of the telegraph processes \( T_{\pm} \) with velocities \( a_+ \) and \( -a_- \), which are controlled by the Poisson process with alternating intensities \( \lambda_{\pm} = \lambda_+ (1 + h_\pm)^2 \); furthermore \( \bar{\mu} = (a_- \lambda_+ + a_+ \lambda_-)/(a_+ + a_-) \), \( \mu = (a_- \lambda_+ + a_+ \lambda_-)/(a_+ + a_-) \) and \( \nu = (\lambda_+ - \lambda_-)/(a_+ + a_-) \), \( \nu = (\lambda_+ - \lambda_-)/(a_+ + a_-) \).

Since under the scaling (21) \( a_+/\sqrt{\lambda_+}, a_-/\sqrt{\lambda_-} \to \sigma \) and thus the processes \( T_+(t), t \geq 0 \) and \( T_-(t), t \geq 0 \) converge to \( \sigma w(t), t \geq 0 \), then

\[ \bar{p}_{\pm}(x, t) \to \frac{1}{\sigma \sqrt{2\pi t}} e^{-x^2/(2\sigma^2)}. \]

Further notice that

\[ p - \nu = \frac{\lambda_+ [(1 + h_+)^2 - 1] - \lambda_- [(1 + h_-)^2 - 1]}{a_+ + a_-} \sim z \frac{\lambda_+ h_+ - \lambda_- h_-}{a_+ + a_-} \to z \frac{\gamma \alpha_+ - \alpha_-}{\sigma (1 + \gamma)}. \quad (26) \]

Moreover
Jump telegraph processes and a volatility smile

\[ \bar{\mu} - \mu = \frac{a_+ (\bar{\lambda}_- - \lambda_-) + a_- (\bar{\lambda}_+ - \lambda_+)}{a_+ + a_-} \]

\[ = \frac{a_+ \lambda_-}{a_+ + a_-} [(1 + h_-)^2 - 1] + \frac{a_- \lambda_+}{a_+ + a_-} [(1 + h_+)^2 - 1] \]

\[ = \frac{z \sqrt{\lambda_+ \lambda_-}}{\sqrt{\lambda_+ + \sqrt{\lambda_-}} + \sqrt{\lambda_-}} \left[ \sqrt{\lambda_+ h_+} + \sqrt{\lambda_- h_-} \right] \]

\[ + \frac{z^2 - z}{2} \left[ \frac{\sqrt{\lambda_-}}{\sqrt{\lambda_+} + \sqrt{\lambda_-}} \lambda_+ h_+^2 + \frac{\sqrt{\lambda_+}}{\sqrt{\lambda_+} + \sqrt{\lambda_-}} \lambda_- h_-^2 \right]. \] (27)

Applying (21)-(23) and summarizing the above statements, we obtain the convergence (25). \( \Box \)

**Remark 2.** Condition (23) in this theorem means that the total drift \( \Delta \equiv A + \frac{\sqrt{\lambda_+ \lambda_-}}{\sqrt{\lambda_+ + \sqrt{\lambda_-}}} (\sqrt{\lambda_+ h_+} + \sqrt{\lambda_- h_-}) \) is asymptotically finite. Here \( A = \frac{a_+ - c_+ + a_+ - c_-}{c_+ - c_-} \) is generated by the velocities of the telegraph process, and the summand \( \frac{\sqrt{\lambda_+ \lambda_-}}{\sqrt{\lambda_+ + \sqrt{\lambda_-}}} (\sqrt{\lambda_+ h_+} + \sqrt{\lambda_- h_-}) \) represents the drift component (possibly with infinite asymptotics) that is motivated only by jumps. If the limits of \( \lambda_\pm h_\pm \) are finite, then \( A \to const \), and \( \alpha_+ = \alpha_- = 0 \). In this case the volatility of limit is \( v = \sigma = \lim a_\pm / \sqrt{\lambda_\pm} \).

Hence in model (18)-(19) value \( a_+/\sqrt{\lambda_+} = a_-/\sqrt{\lambda_-} = (c_+ - c_-)/(\sqrt{\lambda_+} + \sqrt{\lambda_-}) \) can be interpreted as “telegraph” component of volatility, and \( \sqrt{\lambda_\pm h_\pm} \) are volatility components engendered by jumps.

In general, the limiting volatility \( v = \sqrt{(\sigma + (\gamma \alpha_+ - \alpha_-)/(1 + \gamma))^2 + \beta^2} \) depends both on “telegraph” and jump components. So it is natural to define volatility in jump telegraph model as (see (26)-(27))

\[ vol = \sqrt{\left( \frac{c_+ - c_-}{\sqrt{\lambda_+} + \sqrt{\lambda_-}} \right)^2 \left( 1 + \frac{\lambda_+ h_+ - \lambda_- h_-}{c_+ - c_-} \right)^2 + \frac{\sqrt{\lambda_+ \lambda_- h_+^2} + \sqrt{\lambda_+ \lambda_- h_-^2}}{\sqrt{\lambda_+} + \sqrt{\lambda_-}}}. \] (28)

### 4 Historical and implied volatilities in the jump telegraph model

#### 4.1 Historical volatility

Historical volatility is defined as

\[ \text{HV}(t) = \sqrt{\frac{\text{Var}\{\log S(t + \tau)/S(\tau)\}}{\tau}}. \] (29)
For classical Black-Scholes model \( \log S(t+\tau)/S(\tau) \sim \mathcal{D} \equiv at + \sigma w(t) \) (where \( w = w(t), \ t \geq 0 \) is a standard Brownian motion), the historical volatility is constant: \( \text{HV}_{\text{BS}}(t) \equiv \sigma \).

In a moving-average type model, which is described by (see [2])

\[
\log S(t)/S(0) = at + \sigma w(t) - \sigma \int_{0}^{t} \int_{-\infty}^{\tau} pe^{-(q+p)(\tau-u)} du w(u),
\]

\((\sigma, q, q+p > 0)\) the historical volatility has a more tricky structure

\[
\text{HV} = \frac{\sigma}{2\lambda} \sqrt{q^2 + p(2q + p)} \Phi_{\lambda}(t) \quad (30)
\]

with \( 2\lambda = q + p \) and \( \Phi_{\lambda}(t) = \frac{1 - e^{-2\lambda t}}{2\lambda t} \). Recently this type of models have been applied to capture memory effects of the market [8], [10].

The historical volatility of jump telegraph model (18) takes the form

\[
\text{HV}(t) = \sqrt{\frac{1}{t} \int_{0}^{t} e^{(t-\tau)A} v(\tau) d\tau}, \quad (31)
\]

where \( v = \begin{bmatrix} v_+(\tau) \\ v_-(\tau) \end{bmatrix} \) is defined as in (13), but with \( \ln(1 + h_\pm) \) instead of \( h_\pm \):

\[
v_+(\tau) = \lambda_+ [\ln(1 + h_+) - C \tau \Phi_\lambda(\tau)]^2, \quad v_-(\tau) = \lambda_- [\ln(1 + h_-) + C \tau \Phi_\lambda(\tau)]^2.
\]

Here as usual, subscripts \( \pm \) denote the initial state of the market, \( C = c_+ - c_- + \lambda_+ \ln(1 + h_+) - \lambda_- \ln(1 + h_-) \) and \( \Phi_\lambda(\tau) = \frac{1 - e^{-(\lambda_+ + \lambda_-)\tau}}{2\lambda_+ + \lambda_-} \).

Historical volatility in jump telegraph model has the following very natural limiting behaviour (see (14)):

\[
\lim_{t \to 0} \text{HV}_\pm(t) = \sqrt{\lambda_\pm \ln(1 + h_\pm)}, \quad \lim_{t \to \infty} \text{HV}_\pm(t) = \sqrt{\frac{\lambda_+ \lambda_-}{2\lambda^3} [(\lambda_- B - a)^2 + (\lambda_+ B + a)^2]}
\]

\((B = \frac{1}{2} \ln(1 + h_+)(1 + h_-), \ a = (a_+ + a_-)/2; \) see (14)). These limits look reasonable: the limit at 0 is engendered by jumps only, the limit at \( \infty \) contains both “velocity” component and a long term influence of jumps.

Using (16) and (29), in the symmetric case \( \lambda_+ = \lambda_- = \lambda \) formula (31) takes the form similar to (30)

\[
\text{HV}_\pm(t) = \sqrt{\frac{a^2}{\lambda} + \lambda B^2 + (a + \lambda b)^2 \Phi_\lambda(t) / \lambda + \gamma\Phi_\lambda(t) \pm 2B(a + \lambda b)e^{-2\lambda t}}.
\]

The limits of historical volatility under a standard diffusion scaling (see Theorem 4) are more complicated. Nevertheless, in the symmetric case \( \lambda_+ = \lambda_- = \lambda \) the limit at 0 is engendered by jumps only, the limit at \( \infty \) contains both “velocity” component and a long term influence of jumps. The limits look reasonable: the limit at 0 is engendered by jumps only, the limit at \( \infty \) contains both “velocity” component and a long term influence of jumps.
\[ \lambda_- = \lambda, \] we have under the scaling conditions \( \lambda, a \to \infty, h_\pm \to 0, a^2/\lambda \to \sigma^2, \sqrt{\lambda h_\pm} \to \alpha_\pm \) that the historical volatility \( \mathcal{H}V_\pm(t) \) defined by (31) converges to \( \sqrt{\sigma^2 + (\alpha_+ + \alpha_-)^2/4} \).

Notice, that under the martingale measure \( P^* \), we have \( \lambda = -c_\pm/h_\pm, \sigma = (-\alpha_+ + \alpha_-)/2, \) and the diffusion limit of historical volatility equals to \( \nu = \sqrt{(\alpha_+^2 + \alpha_-^2)/2} \), which coincides with the volatility expression in Remark 1.

### 4.2 Implied volatility

Define the Black-Scholes call price function \( f(\mu, v), \mu = \log K \) by

\[
f(\mu, v) = \begin{cases} F\left(\frac{-\mu}{\sqrt{v}} + \frac{\sqrt{v}}{2}\right) - e^{\mu} F\left(\frac{-\mu}{\sqrt{v}} - \frac{\sqrt{v}}{2}\right), & \text{if } v > 0, \\ (1 - e^{\mu})^+, & \text{if } v = 0. \end{cases}
\]

The processes \( V_\pm(\mu, t), t \geq 0, \mu \in \mathbb{R} \) defined by the equation

\[
\mathbb{E}\left[\left(\frac{S(t + \tau)}{S(\tau)} - e^{\mu}\right)^+ | \mathcal{F}_\tau\right] = f(\mu, V_{\sigma(\tau)}(\mu, t))
\]

are referred to as implied variance processes.

The implied volatilities \( \mathcal{I}V_\pm(\mu, t) \) are

\[
\mathcal{I}V_\pm(\mu, t) = \sqrt{\frac{V_\pm(\mu, t)}{t}}.
\]

The LHS of (32) is defined exactly by following. In the framework of this model option pricing formulae and hedging strategies are completely constructed (see [20]).

\[
\mathbb{E}\left[\left(\frac{S(t + \tau)}{S(\tau)} - e^{\mu}\right)^+ | \mathcal{F}_\tau\right] = u_s(\mu, t; \lambda_\pm) - e^{\mu} u_s(\mu, t; \lambda_\pm^*), \quad s = \sigma(\tau),
\]

where \( \lambda_\pm = \lambda_\pm^*(1 + h_\pm), \lambda_\pm^* = -c_\pm/h_\pm > 0 \). Functions \( u_s \) can be expressed as

\[
u_s(\mu, t; \lambda_\pm) = \sum_{n=0}^{\infty} u_s^{(n)}(\mu - b_s^{(n)}, t), \quad s = \pm,
\]

where \( b_s^{(n)} = \ln \kappa_n = \sum_{j=0}^{n} \ln(1 + h_{s \pm}(\tau_j -)) \) are drift parameters engendered by jumps. Summands \( u_s^{(n)} \) of this sum has the following structure: for \( n \geq 1 \)

\[
u_s^{(n)}(\mu, t) = \begin{cases} 0, & y > c_+ t \\ u_s^{(n)}(p, q), & c_- t \leq y \leq c_+ t, \quad p = \frac{c_+ t - y}{c_+ - c_-, q = \frac{y - c_- t}{c_+ - c_-}} \end{cases}
\]
and \( u^{(0)}_+(y, t) = \begin{cases} 0, & \text{if } p < 0 \\
 e^{-\lambda_+ t}, & \text{if } p \geq 0 \end{cases} \)
 and \( u^{(0)}_-(y, t) = \begin{cases} e^{-\lambda_- t}, & \text{if } q < 0 \\
 0, & \text{if } q \geq 0 \end{cases} \).

Functions \( \rho^{(n)}_\pm(t) \) in (35) have a form
\[
\rho^{(n)}_\pm(t) = e^{-\lambda \pm \rho} A^{(n)}_\pm(t).
\]

Here \( A^{(n)}_+ = (\lambda_+)^{(n+1)/2} (n)_{(n+1)/2}, A^{(n)}_- = (\lambda_-)^{(n+1)/2} (n)_{(n-1)/2} \)
and \( P^{(n)}_\pm(t) = \frac{\rho^n}{n!} \cdot _1 F_1 (m_n^{(\pm)} + 1; n + 1; -\delta t), m^{(\pm)}_n = [n/2], m^{(-)}_n = [(n-1)/2], \)
\( \delta = \lambda_+ - \lambda_- \). Here we exploit a hypergeometric function \( _1 F_1 (\alpha; \beta; z) \)
which is defined as
\[
_1 F_1 (\alpha; \beta; z) = 1 + \sum_{k=1}^{\infty} \frac{\alpha (\alpha + 1) \ldots (\alpha + k - 1)}{k! \beta (\beta + 1) \ldots (\beta + k - 1)} z^k = 1 + \sum_{k=1}^{\infty} \frac{\alpha_k}{k!} z^k.
\]
(see Abramowitz and Stegun [1]). Notice that \( P^{(2n+1)}_+ = P^{(2n+1)}_- = P^{(2n+1)}_0, n = 0, 1, 2, \ldots \)
Moreover \( w^{(n)}_\pm = e^{-\lambda_+ q - \lambda_- p} A^{(n)}_\pm e^{(n)}_\pm, \ p, q > 0, \)
where \( e^{(1)}_\pm = P^{(1)}(p) = (1 - e^{-\delta p})/\delta, \)
and
\[
\begin{align*}
\nu^{(2n)}_+ &= \nu^{(2n)}_-(p, q) = P^{(2n)}_+(p) + q P^{(2n-1)}_-(p) + \sum_{k=2}^{n} \frac{q^k}{k!} \sum_{j=0}^{k-2} \delta^{k-j-2} \beta_{k-1,j} P^{(2n-j-2)}_-(p), \\
\nu^{(2n)}_- &= \nu^{(2n)}_-(p, q) = P^{(2n)}_-(p) + \sum_{k=2}^{n-1} \frac{q^k}{k!} \sum_{j=0}^{k-2} \delta^{k-j} \beta_{k+1,j} P^{(2n-j)}_-(p), \\
\nu^{(2n+1)}_+ &= \nu^{(2n+1)}_-(p, q) = P^{(2n+1)}_+(p) + \sum_{k=1}^{n} \frac{q^k}{k!} \sum_{j=0}^{k-1} \delta^{k-j-1} \beta_{k,j} P^{(2n-j)}_-(p).
\end{align*}
\]

Here the coefficients \( \beta_{k,j}, j < k \) are defined as follows: \( \beta_{k,0} = \beta_{k,1} = \beta_{k,k-2} = \beta_{k,k-1} = 1, \)
\[
\beta_{k,j} = \frac{(k-j)(k-j)!}{[j/2]!}.
\]

Remark 3. In the symmetric case \( \lambda_+ = \lambda_- = \lambda \) we have \( P^{(n)}_\pm(t) = \frac{t^n}{n!} \)
and functions \( u^{(n)}_\pm \) can be simplified as follows
\[
\begin{align*}
u^{(n)}_+(y, t) &= e^{-\lambda t} \frac{\lambda^n}{n!} \sum_{k=0}^{m^{(\pm)}_n} \binom{n}{k} p^{n-k} q^k \text{ if } p, q > 0 \\
&= t^n, \quad \text{if } q < 0.
\end{align*}
\]

The detailed proof of (34)-(37) see in [20].
5 Numerical results

We performed the numerical valuation of the jump telegraph volatility (28) and the historical volatility (31), which are compared with the implied volatilities (33) with respect to different moneyness and to the initial market states. The implied volatilities are calculated by the explicit formulae (33)-(37). First, we consider the symmetric case: $\lambda_{\pm} = 10, c_{\pm} = \pm 1$ and $h_{\pm} = \mp 0.1$. In Figure 2 we plot implied volatilities of this simple case. Table 1 lists call prices and implied volatilities of this volatility smile numerically. Notice that these frowned smiles of implied volatilities $\text{IV}_-$ and $\text{IV}_+$ intersect at $K/S_0 \approx 1.17$.

Table 1. Symmetric smile, $t = 1, S_0 = 100, \lambda_{\pm} = 10, h_{\pm} = \mp 0.1, c_{\pm} = \pm 1$

<table>
<thead>
<tr>
<th>$K$</th>
<th>40</th>
<th>70</th>
<th>100</th>
<th>117</th>
<th>130</th>
<th>160</th>
<th>190</th>
<th>220</th>
<th>250</th>
<th>280</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_-$</td>
<td>0.0013</td>
<td>31.6774</td>
<td>12.7370</td>
<td>6.9036</td>
<td>4.1565</td>
<td>1.1433</td>
<td>0.2632</td>
<td>0.0478</td>
<td>0.0058</td>
<td>0.0002</td>
</tr>
<tr>
<td>$c_+$</td>
<td>0.0026</td>
<td>31.7257</td>
<td>12.7680</td>
<td>6.9039</td>
<td>4.1382</td>
<td>1.1128</td>
<td>0.2430</td>
<td>0.0390</td>
<td>0.0032</td>
<td>0.00</td>
</tr>
<tr>
<td>$\text{IV}_-$</td>
<td>0.2670</td>
<td>0.3147</td>
<td>0.3206</td>
<td>0.3200</td>
<td>0.3186</td>
<td>0.3180</td>
<td>0.3109</td>
<td>0.2875</td>
<td>0.2545</td>
<td>0.00</td>
</tr>
<tr>
<td>$\text{IV}_+$</td>
<td>0.2811</td>
<td>0.3175</td>
<td>0.3214</td>
<td>0.3200</td>
<td>0.3180</td>
<td>0.3109</td>
<td>0.3010</td>
<td>0.2875</td>
<td>0.2671</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Fig. 2. Symmetric smile, $t = 1, S_0 = 100, \lambda_{\pm} = 10, h_{\pm} = \mp 0.1, c_{\pm} = \pm 1$, $\text{IV}_\pm = 0.3162$, jump telegraph volatility = 0.3162

Table 2 and Figure 3 show the implied volatility picture for skewed movement, when the market prices have a drift: both velocities are positive, and to avoid an arbitrage we suppose jump values to be negative. This figure has unstable oscillations for deep-out-of-the-money options. Moreover, only in this case
historical and jump telegraph volatilities are less than implied volatilities values for at-the-money options.

Table 2. Skewed smile, $t = 1$, $S_0 = 100$, $\lambda_\pm = 10$, $h_- = -0.03$, $h_+ = -0.19$, $c_- = 0.3$, $c_+ = 1.9$

<table>
<thead>
<tr>
<th>$K$</th>
<th>50</th>
<th>100</th>
<th>150</th>
<th>200</th>
<th>250</th>
<th>300</th>
<th>350</th>
<th>400</th>
<th>450</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_-$</td>
<td>50.8133</td>
<td>17.6956</td>
<td>5.5624</td>
<td>1.8243</td>
<td>0.6350</td>
<td>0.2325</td>
<td>0.0882</td>
<td>0.0347</td>
<td>0.0127</td>
<td>0.0053</td>
</tr>
<tr>
<td>$c_+$</td>
<td>50.9762</td>
<td>18.5944</td>
<td>6.3367</td>
<td>2.2640</td>
<td>0.8586</td>
<td>0.3454</td>
<td>0.1413</td>
<td>0.0621</td>
<td>0.0279</td>
<td>0.0099</td>
</tr>
<tr>
<td>$\text{IV}_-$</td>
<td>0.4475</td>
<td>0.4473</td>
<td>0.4539</td>
<td>0.4590</td>
<td>0.4620</td>
<td>0.4632</td>
<td>0.4630</td>
<td>0.4624</td>
<td>0.4577</td>
<td>0.4565</td>
</tr>
<tr>
<td>$\text{IV}_+$</td>
<td>0.4662</td>
<td>0.4704</td>
<td>0.4776</td>
<td>0.4827</td>
<td>0.4856</td>
<td>0.4875</td>
<td>0.4868</td>
<td>0.4873</td>
<td>0.4867</td>
<td>0.4766</td>
</tr>
</tbody>
</table>

Fig. 3. Skewed smile, $t = 1$, $S_0 = 100$, $\lambda_\pm = 10$, $h_- = -0.03$, $h_+ = -0.19$, $c_- = 0.3$, $c_+ = 1.9$, $\text{HV}_- = 0.4198$, $\text{HV}_+ = 0.4402$; jump telegraph volatility=0.4301

Finally, we calculate exactly the case which was considered in the work of A. De Gregorio and S.M. Iacus [5]. In this paper values of the parameters was statistically estimated. The numerical work are based on weekly closings of the Dow-Jones industrial average July 1971 - Aug 1974. We admit the values of alternating intensities $\lambda_\pm$ and alternating market trends $c_\pm$, proposed by [5]. Assuming these parameters have respect to martingale measure we calibrate jump values as $h_\pm = -c_\pm/\lambda_\pm$.

The model was taken asymmetric with $\lambda_- = 48.53$, $\lambda_+ = 34.61$, $h_- = -0.0126$, $h_+ = -0.0358$, $c_- = 0.61$, $c_+ = 1.24$. It respects to simulations of a preferably bullish market with small jump corrections. The main feature of
this market is in the redundancy of small jumps. The calibrated martingale distribution is strongly asymmetric.

The behaviour of implied volatility in the jump-telegraph model for these data surprisingly resembles the calibration results for stochastic volatility models of the Ornstein-Uhlenbeck type (see [17], fig. 5.1, where implied volatilities of OU-stochastic volatility model was depicted) and for jump-diffusion models (see Table 2 of [3] which contains the implied volatilities calibrated with respect to jump-diffusion model. All calculations there are prepared considering a data set of European call options on S&P 500 index).

Figure 5 depicts an implied volatility surface with respect to strike prices and maturity times.

Table 3. Dow-Jones smile, \( t = 1, S_0 = 100, \lambda_- = 48.53, \lambda_+ = 34.61, h_- = -0.0126, h_+ = -0.0358, c_- = 0.61, c_+ = 1.24 \)

<table>
<thead>
<tr>
<th>( K )</th>
<th>50</th>
<th>70</th>
<th>100</th>
<th>130</th>
<th>160</th>
<th>190</th>
<th>220</th>
<th>250</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c_- )</td>
<td>50.0002</td>
<td>30.1167</td>
<td>6.8313</td>
<td>0.4913</td>
<td>0.0146</td>
<td>0.0002</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>( c_+ )</td>
<td>50.0002</td>
<td>30.1215</td>
<td>6.8838</td>
<td>0.5117</td>
<td>0.0162</td>
<td>0.0003</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>( IV_- )</td>
<td>0.1809</td>
<td>0.1762</td>
<td>0.1714</td>
<td>0.1684</td>
<td>0.1663</td>
<td>0.1645</td>
<td>0.1628</td>
<td>0.1608</td>
</tr>
<tr>
<td>( IV_+ )</td>
<td>0.1819</td>
<td>0.1773</td>
<td>0.1728</td>
<td>0.1699</td>
<td>0.1679</td>
<td>0.1662</td>
<td>0.1646</td>
<td>0.1629</td>
</tr>
</tbody>
</table>

Fig. 4. Dow-Jones smile, \( t = 1, S_0 = 100, \lambda_- = 48.53, \lambda_+ = 34.61, h_- = -0.0126, h_+ = -0.0358, c_- = 0.61, c_+ = 1.24, HV_- = 0.1630, HV_+ = 0.1642; \) jump telegraph volatility=0.1661
Fig. 5. Skewed smile, $S_0 = 100, \lambda_\pm = 10, h_- = -0.03, h_+ = -0.19, c_- = 0.3, c_+ = 1.9$

References


Jump telegraph processes and a volatility smile

Estimating the true embedded risk management cost of total return strategies

Francesco Sandrini\(^1\), David Edelman\(^2\), and Dermot Ryan\(^1\)

\(^1\) Pioneer Investments Management, 
1 George’s Quay Dublin 2, Dublin 
{Francesco.Sandrini, Dermot.Ryan}@pioneerinvestments.com 
http://www.pioneerinvestments.com

\(^2\) Michael Smurfit Business School, 
UCD, Dublin 
davide@ucd.it 
http://www.ucd.ie

Abstract. We propose to measure the value added by periodic portfolio rebalancing in actively managed strategies. Using Monte-Carlo simulation and dynamic stochastic programming we simulate the pay-off of an actively managed strategy. We seek to replicate this pay-off using a static investment based on the same Monte-Carlo scenarios and the same investment timeframe, but including in the static portfolio some derivative strategies not available to the active manager. We contend that the allocation to the derivative strategies quantifies the value added by active management. We then test the sensitivity of the solution to various parameters of the problem.

Keywords. Stochastic optimization, option hedging, simulated annealing.

M.S.C.classification. 47N10, 91G20, 91G60.


1 Introduction

The goal of this paper is to simulate an actively managed total return portfolio as a pseudo ALM problem of managing a portfolio of assets against a simulated LIBOR rate and to try to quantify the value added by the active manager under different circumstances. We do this by attempting to replicate the managers final payoff using an optimal static portfolio including options.

1.1 Description of the approach

We model two simplified investment strategies. In the first strategy an active manager manages a portfolio consisting of a "safe" money market index and one or more "risky" assets (equities, fixed income or credit), against a benchmark of the money market index, re-balancing the strategy periodically. In the second a passive manager makes a static investment in the same instruments plus a number of derivative instruments whose underlying assets are the risky assets, benchmarked against the money market index. The exposure to option instruments in some sense quantifies the value added by the active management. Using such a framework we can measure the difficulty of the active manager's job under different hypotheses.

The modelling is performed within a Monte Carlo framework. We generate a large number of forward trajectories for each of the assets under consideration in order to generate an expected distribution of final wealth for the dynamic strategy, and then attempt to match that pay-off as closely as possible with a static portfolio based on the same scenarios.

This problem presents several numerical challenges. The dynamic problem requires solving a large linear optimization problem with a large number of variables and constraints (see [21]). While the static optimization problem involves a much smaller number of variables and constraints, the utility function we will consider is extremely badly behaved, with many local minima and discontinuities, making it completely unsuited to variance-based optimization methods and thus we must consider alternative optimization methods.

1.2 An actively managed portfolio

We mimic the active manager's decision process using multi stage optimization and stochastic programming. Consider for example a situation in which a manager chooses an initial allocation between a money market index and an equity index to be held for one period, at which stage the portfolio may be rebalanced to a new allocation to be held for a further period. The target of the manager is to beat the benchmark (the money market index) at the end of the second period.

The factors which influence the manager's decision include the allowed budget of risk (defined, for example, in terms of a maximum probability of shortfalling the target) and the expected returns on the assets but also the knowledge that he will be free to change his allocation in the future in response to the actual performance of the assets during the first period. For example, in a scenario in which the equity underperforms expectations in the first period, thereby increasing the probability of a shortfall at the end of period two, the manager may decide to increase his holding of the money market index for the second period in order to minimize the expected downside. Thus the manager can make more efficient use of the risk budget with his initial allocation, in the knowledge that he will be able to rebalance after one period.
This decision is essentially a two-stage linear stochastic program, where a decision maker takes some action at the first stage, after which a random event occurs, affecting the outcome of the first stage decision. A recourse decision can then be made in the second stage that compensates for any bad effects that might have been experienced as a result of the first stage decision. The optimal policy from such a model is a first stage policy and a collection of recourse decisions (a decision rule) defining which second stage action should be taken in response to each random outcome.

We model this procedure using a Monte Carlo tree which branches at the start of the first period and then again in each scenario at the start of the second period. Then we seek a strategy consisting of an allocation at the start of the first period and an allocation for each possible scenario at the start of period two, which is optimal in some sense. The exact nature of the solution will be determined by the choice of utility function used to evaluate different strategies, which must reflect both the investment style under consideration and the budget of risk available to the manager.

### 1.3 A static portfolio

Next we try to replicate the payoff achieved by an active manager using a static “buy and hold” strategy. The ability of the active manager to rebalance his portfolio in response to portfolio performance, moving 100% into the money market index if necessary in order to hedge against any downside, introduces a measure of optionality into the active portfolio’s expected payoff. Therefore, including a put option on the equity index with strike price equal to the expected final value of the benchmark in the static portfolio increases the ability of the passive manager to replicate the active payoff. We can also allow the active manager to buy a call on the equity with a much higher payoff, allowing us to mimic the ability of the active manager to increase risk in well performing scenarios.

We model the static portfolio by using the same scenarios for the underlying assets as above, and seeking the allocation between the assets and a number of optional strategies on the assets which produces a payoff that “best” (in a sense to be described below) matches the payoff of the dynamic strategy.

### 1.4 Overview of the paper

In the next section we will present a more thorough mathematical description of the two stages of the problem. We will then enumerate the various options in the way we tackle the problem. In Section 3 we will present the results of a control case and then we will examine the effect of varying each of the options listed in Section 2.
2 Problem description

2.1 Mathematical formulation of the optimization problem

The forward asset price scenarios are arranged in a tree structure as shown in Figure 1. Letting $T$ denote the time horizon of the simulation in years and $f$ denote the number of sampling times per year, then the set $T_S$ of sampling times is given by $T_S = \{0, \frac{1}{f}, \frac{2}{f}, \ldots, 1, \ldots, 2, \ldots, T - \frac{f}{f}, T\}$. Let $A$ denote the set of assets we are considering; for each scenario we will simulate the price of each asset $a \in A$ at each time $t \in T_S$. In Figure 1 the vertical dimension represents time and there is a node on each scenario at each sampling time.

Let $T_D \subset T_S$ denote the portfolio decision times at which rebalancing may occur ($T_D$ must include $t = 0$). As explained above, each scenario branches at each time $t \in T_D$, and this branching behaviour is regular in the sense that the same number of new scenarios is the same in each scenario at a given $t \in T_D$. It should be clear that the structure of the tree is completely specified by $T, f, T_D$ and by the number of new branches at each $t \in T_D$.

We also use the following notation; $N$ denotes the set of all nodes of the tree, $N_D$ the set of all decision nodes and $n_0$ the unique node at time $t = 0$. To each node $n$ we assign a probability, $P(n)$, which is the reciprocal of the number of scenarios at that time and for each node $n \in N - \{n_0\}$ we denote by $p(n) \in N$ the unique node preceding $n$ on the same scenario.

Let $a' \in A$ be the target asset (i.e. the money market rate). For each node we simulate an asset price for each asset and thus for each node $n \in N - n_0$ we can define $r_{n,a}$ to be the return on asset $a$ between nodes $p(n)$ and $n$.

![Scenario tree: the broken line represents a single scenario](image-url)
Let $x_{n,a}$ be the amount of asset $a$ held at node $n$ and let $x_{n,a}^+$ and $x_{n,a}^-$ denote respectively the amount of $a$ bought and sold at node $n \in N_D$. Let $W(n)$ denote the total portfolio wealth at node $n$ and let $B(n)$ denote the target at node $n$.

The following constraints completely define the optimization problem:

\begin{align*}
  x_{n,a} & \geq 0 \text{ for all } a \in A, n \in N, \quad (1) \\
  x_{n,a}^+ & \geq 0, x_{n,a}^- \geq 0 \text{ for all } a \in A, n \in N_D, \quad (2) \\
  \sum_{a \in A} (x_{n,a}^+ - x_{n,a}^-) & = 0 \text{ for all } n \in N_D, \quad (3) \\
  x_{n,a} & = x_{p(n),a}(1 + r_{n,a}) \text{ for all } n \in N - N_D, \quad (4) \\
  x_{n,a} & = x_{p(n),a}(1 + r_{n,a}) + x_{n,a}^+ - x_{n,a}^- \text{ for all } n \in N_D - \{n_0\}, \quad (5) \\
  W(n) & = \sum_{a \in A} x_{n,a} \text{ for all } n \in N, \quad (6) \\
  B(n) & = B(p(n))(1 + r_{n,a'}) \text{ for all } n \in N - n_0, \quad (7)
\end{align*}

and

\begin{align*}
  W(n_0) & = B(n_0) = 100. \quad (8)
\end{align*}

For the static problem we also have the extra constraint

\begin{align*}
  x_{n,a}^+ = 0, x_{n,a}^- = 0 \text{ for all } a \in A, n \in N_D, \quad (9)
\end{align*}

i.e. no rebalancing is permitted.

### 2.2 Dynamic Setup

The choice of objective function will determine the shape of the distribution of final wealth resulting from the optimal allocation strategy - thus we need to find an objective function which reflects the investment style of total return strategies as well as the available budget of risk.

Typically the simulated pay-off distribution from a total return strategy assumes a quite skewed shape. The manager trades off a little decrease in expected absolute returns for active downside protection (perhaps specified as a pre-determined probability of a shortfall relative to the benchmark). In the examples we will use the *Semivariance Utility function*, defined as follows;

\begin{align*}
  U(\{x_{n,a}\}) = \sum_{n \in N_D} P_n (\beta W^-(n) - (1 - \beta)W(n)) \quad (10)
\end{align*}

where

\begin{align*}
  W^-(n) = \begin{cases} 
    B(n) - W(n) & \text{if } W(n) < B(n) \\
    0 & \text{otherwise}
  \end{cases}
\end{align*}
The relative risk aversion of utility as a function of wealth is defined as

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

(see for example [7]). Thus for \( W > B \) this utility displays constant relative risk aversion (\( R(W) = 0 \)) meaning that the relative allocation to risky assets does not change as wealth increases on the upside. For \( W < B \) we find that

\[ R(W) = \frac{-1}{1 - \frac{1 - \beta}{2W}} \]

which is negative for \( \beta > \frac{1}{2W+1} \), meaning that the utility displays decreasing relative risk aversion on the downside. Thus the allocation to risky assets will decrease as the wealth decreases leading to a truncated left tail in the expected final wealth distribution.

The degree of downside risk aversion is determined by the value of \( \beta \). In practice we choose \( \beta \in [0, 1) \) to reflect the risk budget of the portfolio. For values of \( \beta \) close to 1, the objective function penalizes a shortfall more heavily than it rewards excess return, while decreasing \( \beta \) allows the manager more latitude for tolerating a shortfall in some scenarios if it is rewarded by a higher excess return in other scenarios.

By only applying the objective function at the final nodes we also allow the manager more freedom to choose a risky initial allocation than if the objective were applied at all times, or even just the rebalancing times. Fig. 2 illustrates how the desired right skewed payoff emerges over time, in this case for a three stage optimization problem.

![Fig. 2. Distribution of dynamic wealth: Results after three re-balancing stages.](image-url)
2.3 Static Setup

For the static portfolio we broaden the investment universe to include a number of option strategies based on the risky asset(s) and seek an allocation that matches as closely as possible the payoff produced by the active management.

Thus in this optimization problem the benchmark is the distribution of final wealth of the active strategy and the objective is to find the static allocation strategy for which the distribution of final wealth is as close as possible to the dynamic distribution.

We consider strategies consisting of long and short positions in put options and call options. In some cases we also allow the optimizer to determine the optimal strike price of each option so as to be able to achieve the closest possible fit.

Let $F_A$ denote the cumulative density function of the active strategy. Let $S$ represent any choice of static asset allocation strategy (including the choice of strike prices) and let $F_S$ denote the cumulative density function of the associated final NAV. Then the utility function is defined as

$$U(S) = \max_{x \in (-\infty, \infty)}(|F_A(x) - F_S(x)|),$$

that is the test statistic of the two-sample Kolmogorov-Smirnov test.

This problem sets a numerically challenging problem. The objective function has a large number of discontinuities and local minima and is completely unsuited to traditional variance based optimization methods. In addition the fact that the strike prices may be variables of the problem makes it quite different to traditional asset allocation optimizations. Thus in this case we use the Adaptive Simulated Annealing technique [3], a guided random search engine suitable for problems with many local minima.

2.4 Scenario generation

We present two different processes for generating future price scenarios for the assets; Geometric Brownian Motion and multivariate GARCH.

**Geometric Brownian Motion** The process is defined as:

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW(t)$$

$$dW(t) \overset{D}{=} \sqrt{dt} Z$$

where $Z$ is a standard Gaussian random variable. The drift $\mu$ and volatility $\sigma$ is set equal to the historical risk premium and volatility for the sake of simplicity. The correlations are based upon historical correlations and are modelled via Cholesky decomposition.
GARCH DCC Time varying correlations are often estimated with Multivariate Garch models that are linear in squares and cross products of the data. A new class of multivariate models called dynamic conditional correlation (DCC) models was proposed by Engle [8] 1999,2002). This family of models is a very useful way to describe the evolution over time of the correlation matrix of large systems. The DCC-GARCH model overcomes the computational constraints of the multivariate GARCH models by adopting a two-step procedure. In the first step, a set of univariate GARCH models is estimated for each asset return. In the second stage, a simple specification is used to model the time-varying correlation matrix, which is obtained using the standardised residuals from the first stage. A particularly appealing feature of the model is that it preserves the simple interpretation of univariate GARCH models, while providing a consistent estimate of the correlation matrix. The inefficiency inherent in the two stage estimation process is coped with by modifying the asymptotic covariance of the correlation estimation parameters. Correlations are critical inputs for many of the common tasks of financial management, including risk management. Hedges require estimates of the correlation between the returns of the assets in the hedge. If the correlations and volatilities are changing, then the hedge ratio should be adjusted to account for the most recent information. By using this specification we aim to account for this feature and the variability induced by the dynamic re-balancing activity. Similarly, structured products such as rainbow options that are designed with more than one underlying asset, have prices that are sensitive to the correlation between the underlying returns. A forecast of future correlations and volatilities is the basis of any pricing formula.

We’ll use a standard GARCH DCC framework to generate Monte Carlo scenarios in replacement of the GBM described in the general set up. We follow the general approach as described by Engle (2002) [8].

\[
\begin{align*}
    r_t & \sim N(0, D_t R_t D_t) \\
    D_t^2 &= \text{diag}(\omega_i) + \text{diag}(\kappa_i) \circ r_{t-1} r_{t-1}' + \text{diag}(\lambda_i) D_{t-1}^2 \\
    \epsilon_t &= D_t^{-1} r_t \\
    Q_t &= S \circ (u - A - B) + A \circ \epsilon_{t-1} \epsilon_{t-1}' + B \circ Q_{t-1} \\
    R_t &= \text{diag}(Q_t)^{-1} Q_t \text{diag}(Q_t^{-1})
\end{align*}
\]

Where \( R \) is a correlation matrix containing the conditional correlations. It’s not the goal of the paper to provide a detailed description of the model, please refer to Engle (2002). The assumption of normality in the first equation gives rise to a likelihood function. Without this assumption, the estimator will still have the QML interpretation. The second equation simply expresses the assumption that each of the assets follows a univariate GARCH process.

2.5 Objective and risk neutral measures

Consider put or call options on a given underlying asset with different strikes but the same expiration. If we obtain market prices for those options, we can apply the Black-Scholes (1973) model to back-out implied volatilities. Intuitively, we
might expect the implied volatilities to be identical. In practice, it is likely that they will not be.

Most derivatives markets exhibit persistent patterns of volatilities varying by strike. In some markets, those patterns form a smile. In others, such as equity index options markets, it is more of a skewed curve. This has motivated the name volatility skew. In practice, either the term "volatility smile" or "volatility skew" (or simply skew) may be used to refer to the general phenomena of volatilities varying by strike.

There are various explanations for why volatilities exhibit skew. Different explanations may apply in different markets. In most cases, multiple explanations may play a role. Some explanations relate to the idealized assumptions of the Black-Scholes approach to valuing options. Almost every one of those assumptions - log normally distributed returns, return homoskedasticity, etc. - could play a role. For example, in most markets, returns appear more leptokurtic than is assumed by a log normal distribution. Market leptokurtosis would make way out-of-the-money or way in-the-money options more expensive than would be assumed by the Black-Scholes formulation. By increasing prices for such options, the volatility smile could be the markets' indirect way of achieving such higher prices within the imperfect framework of the Black-Scholes model. Other explanations relate to relative supply and demand for options. In equity markets, the volatility skew could reflect investors' fear of market crashes which would cause them to bid up the prices of options at strikes below current market levels.

It is generally considered best practice to use implied volatility when pricing equity options (see for example [18]). We will compare the effects of using historical volatility versus implied volatility. In fact, as we will see, the outcome of our experiment is relatively insensitive to the method used to price the options, and in most cases we have considered historical volatility for the sake of objectivity and expediency.

3 Results

Below we will present the results of the experiments. We performed a basic version of the job as a control case and then observed the sensitivity of the results to changing different options. The variables of the problem which we vary are:

- the risk budget of the active manager,
- the asset return generating process (i.e. GBM or GARCH DCC),
- the investment universe,
- the dimensions of the tree, i.e. the number of rebalancings and the number of scenarios,
- the use of risk neutral or objective measures in pricing the options.

3.1 Case 1: The control case

In the control version of the problem we use the following parameters. The simulation horizon is set at 3 year, \( T = 3 \), with rebalancing at 0, 1 and 2 years.
Francesco Sandrini et al.

\( T_D = \{0, 1, 2\} \) with branching factors of respectively 50, 50 and 40, giving 100,000 terminal scenarios. The sampling frequency is weekly \( (f = 52) \) and the scenarios are generated using Geometric Brownian Motion.

We consider an investment universe consisting of two asset classes, money market and equities. To simulate these assets we have used JP Morgan 3M Cash Index and the Dow Jones EuroStoxx index respectively (both denominated in Euro), based on a 5 year sample of weekly data (2001-2006). The money market index also acts as the benchmark.

Of course, it is not possible to directly invest in an index such as the EuroStoxx (although one can replicate such an investment using index futures or exchange traded funds). However the intention is that the return and volatility of the index could be considered as a proxy for the return and vitality one could expect of an investment in the relevant asset class.

We consider an investment universe consisting of two asset classes, money market and equities. To simulate these assets we have used JP Morgan 3M Cash Index and the Dow Jones EuroStoxx index respectively (both denominated in Euro), based on a 5 year sample of weekly data (2001-2006). The money market index also acts as the benchmark.

We have used the SemiVariance loss objective function described in 2.2, applied at terminal nodes and we have set \( \beta = 0.3 \). For the static portfolio we have included two derivatives.

In this case we do not allow the optimizer to vary the strike prices, but instead we have chosen two option strategies which we believe should allow the static manager to better replicate the active strategy.

The first is a put option on the equity index with strike price 108.2, which is the expected value at \( t = 3 \) of a portfolio invested 100% in the money market index. This is intended to mimic the active manager’s ability to create downside protection by increasing the allocation to the safe asset if the risk of a shortfall becomes too high.

The second is a call option on the equity with strike price equal to 140.0, which gives us access to more upside in the scenarios in which the equity performs particularly strongly. This is analogous to the ability of the active manager to increase the equity allocation when the risk of a shortfall is low.

For simplicity we have used flat historical volatility in pricing the options. We will see the effect of risk-neutral pricing in a later section.

The results of the dynamic and static portfolio allocation decisions are shown in Table 1. Fig 3 displays the distribution of final wealth across 100,000 scenarios of the two strategies.

### 3.2 Effect of varying the risk parameter

We consider the effect on the outcome of our problem of changing the value of the risk parameter \( \beta \). We consider two different versions of the control case with tighter risk control - a medium risk case with \( \beta = 0.7 \) and a low risk case with \( \beta = 0.9 \).

**Case 2(a): \( \beta = 0.7 \).** Table 2 summarizes the results of the medium risk problem. In comparison to the control case, two differences are notable. Firstly the lower allocation to derivatives reflects the lower allocation to equities in the dynamic portfolio and the lower level of dynamic hedging performed by the
Estimating the true embedded risk management...

![Graph](image)

**Fig. 3.** Case 1: Control Case

**Table 1.** Case 1: Control Case

<table>
<thead>
<tr>
<th></th>
<th>Dynamic</th>
<th>Static</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Asset Allocation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Money Market Index</td>
<td>80.7%</td>
<td>18.21%</td>
</tr>
<tr>
<td>Equity Index</td>
<td>19.3%</td>
<td>80.25%</td>
</tr>
<tr>
<td>Put Option (Equity, strike 108.3)</td>
<td>1.12%</td>
<td>4.26%</td>
</tr>
<tr>
<td>Call Option (Equity, strike 140.0)</td>
<td>0.42%</td>
<td></td>
</tr>
<tr>
<td><strong>Final result</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benchmark wealth</td>
<td>108.3</td>
<td>108.3</td>
</tr>
<tr>
<td>NAV</td>
<td>114</td>
<td>112.5</td>
</tr>
<tr>
<td>Standard deviation NAV</td>
<td>13.3</td>
<td>9.3</td>
</tr>
<tr>
<td>Probability NAV &gt; benchmark</td>
<td>0.56</td>
<td>0.58</td>
</tr>
<tr>
<td>ks-statistic</td>
<td></td>
<td>4.26%</td>
</tr>
</tbody>
</table>
active manager. Secondly the lower value of the ks-statistic indicated that a more accurate replication job is possible in comparison to the higher risk case.

**Table 2.** Case 2(a): Medium Risk

<table>
<thead>
<tr>
<th>Asset Allocation</th>
<th>Dynamic</th>
<th>Static</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t = 0$</td>
<td>$t = 1$</td>
</tr>
<tr>
<td>Money Market Index</td>
<td>96.3%</td>
<td>95.1%</td>
</tr>
<tr>
<td>Equity Index</td>
<td>3.7%</td>
<td>4.9%</td>
</tr>
<tr>
<td>Put Option (Equity, strike 108.3)</td>
<td>0.29%</td>
<td></td>
</tr>
<tr>
<td>Call Option (Equity, strike 140.0)</td>
<td>0.03%</td>
<td></td>
</tr>
<tr>
<td>Final result</td>
<td>Dynamic</td>
<td>Static</td>
</tr>
<tr>
<td>Benchmark wealth</td>
<td>108.28</td>
<td>108.28</td>
</tr>
<tr>
<td>NAV</td>
<td>109.12</td>
<td>109.38</td>
</tr>
<tr>
<td>Standard deviation NAV</td>
<td>2.62</td>
<td>1.78</td>
</tr>
<tr>
<td>Probability NAV $&gt;$ benchmark</td>
<td>0.56</td>
<td>0.58</td>
</tr>
<tr>
<td>ks-statistic</td>
<td></td>
<td>4.03%</td>
</tr>
</tbody>
</table>

**Case 2(b): $\beta = 0.9$.** Table 3 summarizes the results of the low risk case. Note again that, in comparison to the $\beta = 0.3$

**Table 3.** Case 2(b): Low Risk

<table>
<thead>
<tr>
<th>Asset Allocation</th>
<th>Dynamic</th>
<th>Static</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t = 0$</td>
<td>$t = 1$</td>
</tr>
<tr>
<td>Money Market Index</td>
<td>98.6%</td>
<td>98.2%</td>
</tr>
<tr>
<td>Equity Index</td>
<td>1.4%</td>
<td>1.8%</td>
</tr>
<tr>
<td>Put Option (Equity, strike 108.3)</td>
<td>0.098%</td>
<td></td>
</tr>
<tr>
<td>Call Option (Equity, strike 140.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final result</td>
<td>Dynamic</td>
<td>Static</td>
</tr>
<tr>
<td>Benchmark wealth</td>
<td>108.3</td>
<td>108.3</td>
</tr>
<tr>
<td>NAV</td>
<td>108.7</td>
<td>108.6</td>
</tr>
<tr>
<td>Standard deviation NAV</td>
<td>0.74</td>
<td>0.89</td>
</tr>
<tr>
<td>Probability NAV $&gt;$ benchmark</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>ks-statistic</td>
<td></td>
<td>2.93%</td>
</tr>
</tbody>
</table>
3.3 Changing the data generating process

Next we examine the sensitivity of the solution to the data generating solution. We compare the results of the control case with the solution to an identical problem with only the scenario generating process changed. In Table 4 we note the larger allocation to options compared to the control case. This reflects the extra value added by active management in the context of time varying correlations and volatilities. Also the ks-statistic is slightly higher than in the control case.

**Table 4.** Case 3: Garch scenarios

<table>
<thead>
<tr>
<th>Asset Allocation</th>
<th>Dynamic</th>
<th>Static</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t = 0</td>
<td>t = 1</td>
</tr>
<tr>
<td>Money Market Index</td>
<td>66.1%</td>
<td>57.4%</td>
</tr>
<tr>
<td>Equity Index</td>
<td>33.9%</td>
<td>42.6%</td>
</tr>
<tr>
<td>Put Option (Equity, strike 108.2)</td>
<td></td>
<td>1.92%</td>
</tr>
<tr>
<td>Call Option (Equity, strike 140.0)</td>
<td></td>
<td>0.12%</td>
</tr>
<tr>
<td>Final result</td>
<td>Dynamic</td>
<td>Static</td>
</tr>
<tr>
<td>Benchmark wealth</td>
<td>108.3</td>
<td>108.3</td>
</tr>
<tr>
<td>NAV</td>
<td>117.5</td>
<td>115.7</td>
</tr>
<tr>
<td>Standard deviation NAV</td>
<td>17.5</td>
<td>13.4</td>
</tr>
<tr>
<td>Probability NAV &gt; benchmark</td>
<td>0.63</td>
<td>0.65</td>
</tr>
</tbody>
</table>

3.4 Changing the investment universe

Next, we increase the size of the investment universe to 4 assets by adding a government bond index and a corporate bond index, calibrated to the JP Morgan EMU Government Bond Index and the MSCI EMU Credit (Investment Grade) Index (2001-2006) (see Fig 4).

A total return manager would typically utilize a much broader universe of asset classes in order to exploit the benefit of diversification. Unfortunately in this case we are constrained by the numerical complexity of the optimization problems. We consider the four asset case for the purposes of comparison with the two asset case.

Table 5 summarizes the results. In comparison with the control case we see a much lower allocation to options. This indicates the greater ease with which a manager can beat his benchmark in this context (the dynamic portfolio displays a far higher Sharpe Ratio). The ks-statistic is slightly lower than the control case as the extra asset classes make it easier to replicate a given strategy. One could extrapolate from these results that the effect of adding further assets would allow us to better replicate the dynamic strategy.
Fig. 4. The historical performance of the indices used to calibrate the 4 assets classes

<table>
<thead>
<tr>
<th>Asset Allocation</th>
<th>Dynamic</th>
<th>Static</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( t = 0 )</td>
<td>( t = 1 )</td>
</tr>
<tr>
<td>Money Market Index</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Equity Index</td>
<td>41.8%</td>
<td>28%</td>
</tr>
<tr>
<td>Bond Index</td>
<td>0%</td>
<td>1.2%</td>
</tr>
<tr>
<td>Corp Bond Index</td>
<td>58.2%</td>
<td>70.8%</td>
</tr>
<tr>
<td>Put Option (Equity, strike 108.3)</td>
<td>0.41%</td>
<td></td>
</tr>
<tr>
<td>Call Option (Equity, strike 140.0)</td>
<td>0.17%</td>
<td></td>
</tr>
<tr>
<td>Final result</td>
<td>Dynamic</td>
<td>Static</td>
</tr>
<tr>
<td>Benchmark wealth</td>
<td>108.3</td>
<td>108.3</td>
</tr>
<tr>
<td>NAV</td>
<td>123.6</td>
<td>122</td>
</tr>
<tr>
<td>Standard deviation NAV</td>
<td>18.8</td>
<td>15.3</td>
</tr>
<tr>
<td>Probability NAV &gt; benchmark</td>
<td>0.84</td>
<td>0.82</td>
</tr>
<tr>
<td>ks-statistic</td>
<td></td>
<td>3.62%</td>
</tr>
</tbody>
</table>
3.5 Changing the rebalancing frequency

In this section we change the frequency of rebalancing, comparing the results of the control case to the results obtained using a tree with branching at times 0, \( \frac{1}{2} \), 1, 1 \( \frac{1}{2} \) and 2 years and using 300,000 \((20 \times 15 \times 10 \times 10 \times 10)\) scenarios.

The extra rebalancing times produces a notable effect of the dynamic payoff, with an expected NAV of 121.5 after three years. This presents a much harder challenge for the static portfolio to achieve. To facilitate the task we include two extra options on the equity index, so that the static portfolio now includes two puts and two calls. Furthermore, we allow the optimizer to chose the strike prices of all but one of the options (the first put option, as before, has a strike of 108.3, being the expected value of a 100% money market portfolio).

However, even with these extra degrees of freedom, the optimal static portfolio is a much poorer fit to the dynamic portfolio compared to any of the cases considered so far, as evidenced by the ks-statistic of 7.58%. This, in spite of the much higher overall exposure to derivatives (Table 6).

Unfortunately the numerical difficulty of the optimization problem increases exponentially with the addition of extra rebalancing times, but certainly the evidence of the extra value added by more active management is clear.

<table>
<thead>
<tr>
<th>Asset Allocation</th>
<th>Dynamic</th>
<th>Static</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t = 0</td>
<td>t = 0.5</td>
</tr>
<tr>
<td>Money Market Index</td>
<td>81.8%</td>
<td>60.4%</td>
</tr>
<tr>
<td>Equity Index</td>
<td>18.2%</td>
<td>39.6%</td>
</tr>
<tr>
<td>Equity Put 1 (strike 108.3)</td>
<td>-0.98%</td>
<td></td>
</tr>
<tr>
<td>Equity Put 2 (strike 84.01)</td>
<td>-1.0%</td>
<td></td>
</tr>
<tr>
<td>Equity Call 1 (strike 122.2)</td>
<td>1.49%</td>
<td></td>
</tr>
<tr>
<td>Equity Call 2 (strike 141.7)</td>
<td>1.49%</td>
<td></td>
</tr>
</tbody>
</table>

Final result

<table>
<thead>
<tr>
<th></th>
<th>Dynamic</th>
<th>Static</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benchmark wealth</td>
<td>108.3</td>
<td>108.3</td>
</tr>
<tr>
<td>NAV</td>
<td>121.5</td>
<td>117.5</td>
</tr>
<tr>
<td>Standard deviation NAV</td>
<td>23.6</td>
<td>19.7</td>
</tr>
<tr>
<td>Prob NAV &gt; benchmark</td>
<td>0.69</td>
<td>0.64</td>
</tr>
<tr>
<td>ks-statistic</td>
<td></td>
<td>7.58%</td>
</tr>
</tbody>
</table>

3.6 Changing from objective to risk neutral pricing

In this section we test the effects of changing from flat historical volatilities to market implied volatilities in pricing the options in the static portfolio. Fig. 5 displays the data sample we used of to calibrate our model.
Table 7 displays the results. The change from the control case is slight. The effect of using the volatility smile is to increase the price of the options, consequently we can afford less exposure. The effect is that while the option exposure is only slightly reduced (1.41% compared to 1.54% in the control case), the fit is significantly poorer (5.55% compared to 4.26%).

**Table 7. Case 6: Volatility smile**

<table>
<thead>
<tr>
<th>Asset Allocation</th>
<th>Dynamic</th>
<th>Static</th>
</tr>
</thead>
<tbody>
<tr>
<td>Money Market Index</td>
<td>80.7%</td>
<td>76.36%</td>
</tr>
<tr>
<td>Equity Index</td>
<td>19.3%</td>
<td>22.23%</td>
</tr>
<tr>
<td>Put Option (Equity, strike 108.2)</td>
<td></td>
<td>1.33%</td>
</tr>
<tr>
<td>Call Option (Equity, strike 140.0)</td>
<td></td>
<td>0.08%</td>
</tr>
<tr>
<td>Final result</td>
<td>Dynamic</td>
<td>Static</td>
</tr>
<tr>
<td>Benchmark wealth</td>
<td>108.3</td>
<td>108.3</td>
</tr>
<tr>
<td>NAV</td>
<td>114</td>
<td>112.3</td>
</tr>
<tr>
<td>Standard deviation NAV</td>
<td>13.3</td>
<td>9</td>
</tr>
<tr>
<td>Probability NAV &gt; benchmark</td>
<td>0.56</td>
<td>0.6</td>
</tr>
<tr>
<td>ks-statistic</td>
<td></td>
<td>5.55%</td>
</tr>
</tbody>
</table>
4 Conclusion

We have attempted to quantify the value of active portfolio management in a total return context and to examine the effect on the quantification of various choices considered in our setup. Table 8 summarizes our findings.

<table>
<thead>
<tr>
<th>Case</th>
<th>Total option exposure p.a.</th>
<th>Goodness of fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>0.51%</td>
<td>4.26%</td>
</tr>
<tr>
<td>Medium Risk</td>
<td>0.11%</td>
<td>4.03%</td>
</tr>
<tr>
<td>Low Risk</td>
<td>0.04%</td>
<td>2.93%</td>
</tr>
<tr>
<td>Garch Scenarios</td>
<td>0.68%</td>
<td>4.61%</td>
</tr>
<tr>
<td>4 Assets</td>
<td>0.20%</td>
<td>3.62%</td>
</tr>
<tr>
<td>More Rebalancing</td>
<td>1.25%</td>
<td>7.58%</td>
</tr>
<tr>
<td>Volatility Smile</td>
<td>0.47%</td>
<td>5.55%</td>
</tr>
</tbody>
</table>

We interpret the first column as the value added by the active management compared to a static strategy. Thus the results indicate:

- The value added by active management increases with increasing risk budget. For the medium risk strategy which delivers approximately 27 bps p.a. over the benchmark the value added is of the order of 11 bps. In comparison, the control case delivers approximately 180 bps over the benchmark with a value added of 51 bps.
- The value added by the manager increases when a more realistic asset return generating process (GARCH) is used.
- The value added by the manager decreases with a broadening of the investment universe. Increasing the investment universe allows the manager to exploit the benefits of diversification.
- The value added by the manager (and the difficulty of replicating the results statically) increase dramatically with increased rebalancing.
- The results are relatively insensitive to a change from objective to risk neutral option pricing.

References


