Risk attribution and semi-heavy tailed distributions

Tesi di Dottorato di
EDIT RROJI
matricola 076099

Relatore : Prof. Marco Marchioro
Tutor : Prof. Emanuela Rosazza Gianin
A.A. 2013/2014
## Contents

### Introduction 4

1 Linear factor models 7
   1.1 Multifactor models in finance 8
   1.1.1 Kalman Filter 13
   1.2 Statistical Factors 19
      1.2.1 Principal Component Analysis 21
      1.2.2 Independent Component Analysis 33
      1.2.3 PCA vs ICA 38

2 Risk Measures: VaR, CVaR and Expectiles 39
   2.1 Risk measures: mathematical properties and conformity to Basel III 40
      2.1.1 Value At Risk and Expected Shortfall 40
   2.2 Elicitability and Expectiles 41
      2.2.1 Elicitability: forecast and comparison of estimation procedures 41
      2.2.2 Expectiles: coherent and elicitable 44
   2.3 Risk decomposition 50
      2.3.1 Homogeneity and Euler decomposition 50
      2.3.2 Risk measure decomposition formulas 51
      2.3.3 Empirical results 52

3 User defined risk factors 53
   3.1 Risk Contribution from user defined factors 54
      3.1.1 Risk contribution using PCA and ICA 57
   3.2 Custom Factor Attribution 58
   3.3 Custom and projected factors 59
      3.3.1 Decomposing the residual 59
      3.3.2 Application to the stock market 62
      3.3.3 Application to fixed income securities 63
4 Mixed Tempered Stable distribution

4.1 Semi-heavy tailed distributions ........................................ 67
4.2 Tempered Stable distribution ............................................ 68
4.3 Mixed Tempered Stable distribution ..................................... 72
   4.3.1 Definition and particular cases ..................................... 72
   4.3.2 Investigation using real market data ............................... 76
4.4 Risk Measures using the Mixed Tempered Stable distribution ......... 82
   4.4.1 Saddlepoint Approximation ........................................ 82
   4.4.2 Numerical Analysis .................................................. 84

5 Conclusions and Future Research ......................................... 89

Thank you.
Introduction

After the recent financial crisis there is much more emphasis on the sources of risk rather than just on the levels. The importance of attributing risk to each portfolio component can be deduced even from the fact that in the last few years a new principle arose in finance: the Risk Parity. It is an approach in portfolio management which focuses on allocation of risk rather than on the capital allocation (see Denis et al. [2011] for further details).

Value at risk (see RiskMetrics [1996]) and Expected Shortfall (see Tasche [2002]) have emerged as industry standards for measuring downside risk. Non-parametric approaches have gained the consensus of the practitioners since they are easier to implement and only the information in the return series is used. However, usually non-parametric methods imply more uncertain estimates for the risk measures considered as shown for example in Aussenegg and Miazhynskaia [2006]. Although risk decomposition in a parametric context is not straightforward, Boudt et al. [2009] show that the use of modified versions of VaR and ES based on asymptotic expansions simplifies the problem.

In this thesis we discuss the problem of risk attribution in a multifactor context using non-parametric approaches but we also introduce a new distribution for modeling returns. The risk measures considered are homogeneous since we exploit the Euler rule as in Tasche [1999]. Particular attention is given to the problem of attributing risk to user defined factors since the existing literature is limited when compared to other research arguments but of practical relevance. We point out the problems encountered during the analysis and present some methodologies that can be useful in practice. Each chapter combines both theoretical and practical issues.

The main original contributions can be found in Chapter 3 and Chapter 4 respectively in the introduction of a methodology for identifying idiosyncratic risk in presence of custom factors and the introduction of a new distribution named Mixed Tempered Stable.

The thesis is organized as follows. In Chapter 1 we review the existing literature on linear multifactor models. Starting from the usual Ordinary Least Squares regression we move on to a more sophisticated method for factor exposure estimation, the Kalman Filter (see Kalman [1960]). Principal Component Analysis (see Jolliffe [2002]) is applied for the identification of statistical factors in order to reduce dimensionality. At the end of the chapter a comparison with another mathe-
mational procedure, the Independent Component Analysis (see Hyvarinen [1999]), is given.

In Chapter 2 we review the main characteristics of two well known homogeneous risk measures: the Value At Risk (VaR) and the Expected Shortfall (ES). We discuss the problem of defining a unique risk measure from the regulator point of view and the mathematical properties required. The expectiles introduced in Newey and Powell [1987] as the solution of an Asymmetric Least Squares (ALS) regression have been considered only recently in the context of risk measurement. Coherence (see Artzner et al. [1999], Bellini et al. [2013]) and elicitability (see Gneiting [2011]) are two desirable properties that are both satisfied by the expectile based risk measure. The main difficulty in its use is the identification of the parameter that controls the tail weight in the ALS regression which in Kuan et al. [2009a] is interpreted as the level of prudentiality. From the quantile-expectile relation derived in Jones [1994], we find the implicit values for the weight parameter in datasets coming from equity and credit risk markets. The confidence levels used are those usually fixed by the regulators in the Value at Risk computation. The estimated parameter values are then used for comparing risk attribution results for the three considered risk measures.

Chapter 3 is devoted to the study of risk decomposition models for arbitrary factors chosen directly from the investors. We describe the approach proposed in Meucci [2007] and extend the analysis by considering independent factors obtained through the Independent Component Analysis Hyvarinen [1999] since the methodology can be applied to any linear transformation of the original risk factors that generates new uncorrelated variables. Starting from the work of Menchero and Poduri [2008] we derive a model for attributing risk to generic factors. The introduction of new factors, orthogonal to those considered by the investor, enables us to identify the idiosyncratic term and its contribution to risk. This result can be very useful in risk management because it is simple to implement and each component has a straightforward interpretation.

We propose a new distribution, named Mixed Tempered Stable (MixedTS), in Chapter 4. It is built as a Normal Variance Mean Mixture (NVMM) where the Normal assumption is substituted with the standardized Tempered Stable distribution. Assuming that the mixing random variable follows a Gamma distribution, the proposed model has the Variance Gamma (see Madan and Seneta [1990b]), the Tempered Stable (see Kim et al. [2008]) and the Geometric Stable distribution (see Kozubowski et al. [2011]) as special cases for particular parameters choice. The moment generating function exists and we give analytical formulas for the first four moments. The MixedTS, having as special cases some distributions widely applied in financial literature, is expected to be flexible in capturing different tail behaviors. By construction, no more than one independent component obtained through an ICA analysis can be Gaussian though it is interesting to see how the MixedTS fits the different components. Risk measures for the MixedTS are computed using numerical integration and the Saddlepoint approximation (Lugannani and Rice [1980]) for the tails.
Chapter 1

Linear factor models

In this chapter we review the literature on linear multifactor models for describing returns and discuss the problems arising in their use. The return decomposition is the starting point for the risk attribution process that will be discussed in the next chapters.

The Ordinary Least Squares regression is the simplest way of facing these kind of problems. It is well known that it considers the entire estimation period for computing factor exposures. By using the Kalman Filter (Kalman [1960]) we have a model where factor exposures can vary daily while maintaining the information obtained from the past. Kalman Filter is based on a state space model where the state variables are the factor exposures. The implementation is not easy and part of the data is needed for the algorithm stability. Both OLS and Kalman Filter can be applied when factor time series are given.

Principal Component Analysis Jolliffe [2002] generates statistical factors that are the projections of the original data on the eigenvectors associated to the dataset covariance matrix. In finance this technique is particularly useful for dimension reduction purposes. In a recent paper Longstaff et al. [2011] study the nature of sovereign credit risk and find that the majority of sovereign credit risk can be linked to global factors. They find that sovereign credit spreads are more related to the stock indexes than they to local economic measures. In the same spirit we apply the PCA analysis to credit spreads of European companies grouped by their rating. The dependence on different factors can suggest that the structure of each group is different.

Principal Component Analysis has been widely applied although the gaussianity hypothesis for the components is not in line with the stylized facts observed in financial markets. Independent Component Analysis (ICA) can be a valid alternative since it yields independent factors by maximizing the non-gaussianity of the components. Unfortunately, in the ICA analysis there is no single rule for ordering the components while the explained variance is a criterion for ordering the principal components. This fact justifies the difficulties of using ICA in dimension reduction problems in finance.
1.1 Multifactor models in finance

The simplest type of models are linear multifactor models since the relation between the variable of interest and the factors from which it is influenced is linear. Let us consider the portfolio return $\pi^t$ for the time interval $(t - 1; t]$ defined as:

$$\pi^t = \sum_{i=1}^{n} w^t_i r^t_i$$ (1.1)

where $w^t_i$ is the weight of the i-th asset in the portfolio. The portfolio is composed by n assets with the i-th asset log-return $r^t_i$, given its price $P^t_i$, for the period $(t - 1; t]$ is:

$$r^t_i = \ln \left( \frac{P^t_i}{P^{t-1}_i} \right)$$ (1.2)

We suppose that the investment decision is taken at time $t - 1$ and the holding period is the time interval $(t - 1; t]$. We assume a linear relation with a finite number of factors $F^t$:

$$r^t_i = \alpha^t_i + \sum_{j=1}^{n'} \beta^t_{ij} F^t_{ij} + \epsilon^t_i$$ (1.3)

with $\alpha^t_i$ being a constant and $\epsilon^t_i$ the residual term. The risk factors $F^2$ are financial variables like log returns on equities, bonds or indexes while from quantities like interest rates and spreads risk factors are built using absolute variations. The literature on this argument is rich and in continuous evolution. The main contributions in chronological order are:

- CAPM one factor model (see Sharpe [1964], Lintner [1965] and Black [1972]).
  It assumes a linear relationship between the expected return of a risky asset and its $\beta$:
  $$E[r^t_i] = r^t_f + \beta^t_i (E[r^t_{mkt}] - r^t_f)$$ (1.4)
  where $r^t_{mkt}$ is the return of the well diversified market portfolio and $r^t_f$ the risk free rate. The model states that $\beta^t_i$ defined as:
  $$\beta^t_i = \frac{Cov(r^t_i, r^t_{mkt})}{Var(r^t_{mkt})}$$ (1.5)
  drives the expected returns of asset i since it measures the part of the asset’s statistical variance that cannot be removed by the diversification provided in the market portfolio.

- Fama and French three factor model can be defined in the form (see Fama and French [1992] and

---

1. We use log returns instead of simple returns since they are additive, i.e. longer period returns are obtained by summation of short period returns.
2. From now on we will drop the time index and we will consider only one period returns.
1.1 Multifactor models in finance

Fama and French [1995]):

\[ E[r_i] = \beta_i^{mkt} E[r_{mkt}] + \beta_i^{SMB} E[r_{SMB}] + \beta_i^{HML} E[r_{HML}] \] (1.6)

They assume that all market returns can be explained by three factors: the exposure to the portfolio market, the exposure to value stocks (HML: High [book-to-market ratio] Minus Low), and the exposure to small stocks (SMB: Small [market capitalization]) where \( r_{SMB} \) measures the excess returns of small caps over big caps and \( r_{HML} \) the excess returns of value stocks over growth stocks.

Carhart [1997] four factor model considers the momentum in addition to the three factor model:

\[ E[r_i] = \beta_i^{mkt} E[r_{mkt}] + \beta_i^{SMB} E[r_{SMB}] + \beta_i^{HML} E[r_{HML}] + \beta_i^{Mom} E[r_{Mom}] \] (1.7)

The recent literature considers a higher number of factors. For example the Barra Integrated Model (BIM) considers more than 200 factors (see Shepard [2011]). It identifies first local risk factors and develops local models for different asset classes. These are then combined into a single risk model to capture cross-market correlations. The resulting model provides a structure for detailed risk decomposition of multi-asset class strategies.

Factors can have an economic interpretation, or can be statistical. In particular, in Straumann and Garidi [2011] they distinguish between six type of factors: macroeconomic, market, sector/industry, fundamental, technical and statistical.

Macroeconomic factors can be for example change in the growth rate of GDP or CPI, differences between returns on high yield and government bond indexes etc. Since most of the macroeconomic variables are released quarterly and risk management is performed daily, their use presents some problems. Returns on indexes like for example S&P500 or FTSEMib are market factors. Morgan Stanley Capital International (MSCI) releases daily estimates on GICS (Global Industry Classification Standard) indexes and the corresponding returns are the so called sector/industry factors.

In 2004, MSCI acquired Barra Inc. which uses these factors especially in the Barra Integrated Equity Model.

Firm specific quantities like price to book ratios and market capitalization are considered as fundamental factors while typical examples for technical factors are measures of liquidity. Statistical factors are different since they are not directly observable in the market but extracted via mathematical procedures.

These models have been used for return attribution in Brinson [1985] and risk contribution in Meucci [2007], Menchero and Poduri [2008] and Marchioro and Borrello [2013]. After the decision of the type of factors to consider, it remains to fix the number of factors to include in the model. Except for the cases when one is fully confident about the factors to include, the decision should be taken after considering some statistical tests. One simple way is to proceed step by step, in the sense that when we add an additional factor we compare the likelihood of the model with n factors to that of the n+1 factors model. Under the i.i.d assumption, for large time-series samples, the
ratio of the log likelihoods has an approximate chi-squared distribution. In particular, Connor and Korajczyk (1993) derive a test for the number of factors that is robust and is based on the decline in average idiosyncratic variance as additional factors are added.

In matrix notations, the linear factor model is:

$$ r = \alpha + \beta F + \epsilon $$

(1.8)

where $r$, $\alpha$ and $\epsilon$ are $1 \times T$ vectors. If we consider $K$ factors $\beta$ is a $1 \times K$ vector, while $F$ is a $K \times T$ matrix. Meucci [2010] puts the conditions for a model to be in the systematic-plus-idiosyncratic form: the residuals must be uncorrelated with each other and with the factors. In this paper he shows that the multifactor models used in finance are never of this form.

In models considering statistical risk factors both factors $F$ and exposures $\beta$ must be estimated but no constraints on the residual uncorrelation are posed. For risk factors directly observed in the market, matrix $F$ is fully known and what remains to estimate is only the factor exposures $\beta$ and the vector of constants $\alpha$. Models of this type are coined as time-series models.

The Ordinary Least Squares approach is obtained by considering time-series models for which factor exposures are the solution of an unconstrained minimization problem. In mathematical notations, the factor exposure estimates $\hat{\beta}$ resulting from the OLS approach are the solution of the problem:

$$ \hat{\beta} = \arg\min_{\beta} \| r - \beta F \| $$

(1.9)

where $\| . \|$ is the $L^2$-norm. Geometrically, it projects the returns in $r$ on the linear space spanned by the factors $F$. The minimization problem give the estimate of the factor exposures in matrix notations:

$$ \hat{\beta} = (F F')^{-1} F r $$

(1.10)

Particular attention deserves the decision of the time frame to consider for factor exposure estimation. In fig 1.2 we show the results of the OLS regression for four fund indexes vs 17 factors. The first fund considered is the Vanguard Asset allocation Fund which tracks the performance of its benchmark index, the $S&P 500$ index. The second fund is the Vanguard Balanced Fund which tracks the performance of the overall US stock market index using 60% of its capital and with the remaining tracks the performance of a market-weighted bond index. Allianz RCM Global Water and First Eagle Gold funds invest mainly in assets related to commodities. Indeed, the first invests on equity securities related to companies that have direct or indirect exposures to water related activities while the last fund invests 80% of its assets in gold or related securities. The factors used for explaining the returns for each fund are based on different asset classes. Two commodity indexes: Oil and Gold; three interest rate based risk factors: 3m Libor, 10y and 30 swap rates; the exchange rate EurUsd, the implied volatility index VIX and the 10 GICS sector indexes.
1.1 Multifactor models in finance

\[ r = r - \hat{\beta}_1 F_1 - \hat{\beta}_2 F_2 \]

\[ \hat{\beta}_1 F_1 + \hat{\beta}_2 F_2 \]

Figure 1.1: Geometric interpretation of the OLS regression: projection of returns in \( r \) on the linear space spanned by the regressors \( F_1 \) and \( F_2 \). The factor exposures estimates \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) minimize the residual \( \hat{e} \).

The evaluation date is 12/10/2012. The dataset is composed by daily log returns while the time frame estimation period changes. We consider respectively 1.5, 2, 2.5 and 3 years as estimation periods and compare the results. In general, factor exposures tend to remain stable but this is not always the case. Consider for example the Vanguard Balanced Fund exposure to the 3m Libor: choosing to use the last 2 years data instead of 2.5 years doubles the factor exposure.

The sign of the exposures remains the same since, depending on the target of each fund, the relation between each fund and the risk factor has an economic interpretation that is translated in a numeric value. For example, Allianz RCM Global Water fund seems to have a greater exposure to the Industrial GICS index. This is coherent with the idea that water influences the industrial sector. As we expected, the First Eagle Gold fund returns are highly influenced from the Gold returns. The OLS regression is very useful in giving an intuition of the factors that we must consider for a correct return/risk attribution model but we have to move on to more advanced instruments for computing factor exposures that are updated at least daily.
Vanguard Asset Allocation Inv
Tracks the performance of the SP500 index

|          | WTI Oil | Gold | US003M | SWAP10 | SWAP30 | EURUSD | VIX | INFT | FINL | CONS | HLTH | ENRS | COND | INDU | MATR | UTIL | TELS |
|----------|---------|------|--------|--------|--------|--------|-----|------|------|------|------|------|------|------|------|------|------|------|
| 1.5      | 2.9E-04 | 0.004| -0.002 | 0.010  | -0.008 | -0.009 | 1.5E-04 | 0.186 | 0.137 | 0.134 | 0.096 | 0.118 | 0.114 | 0.111 | 0.043 | 0.042 | 0.024 |
| 2        | 3.8E-04 | 0.004| 0.001  | 0.006  | -0.005 | -0.007 | -1.4E-04 | 0.185 | 0.143 | 0.127 | 0.098 | 0.120 | 0.112 | 0.038 | 0.034 | 0.024 |
| 2.5      | -9.5E-05| 0.005| 0.001  | 0.007  | -0.006 | -0.002 | 3.2E-04  | 0.182 | 0.147 | 0.119 | 0.103 | 0.119 | 0.111 | 0.111 | 0.037 | 0.041 | 0.022 |
| 3        | -1.3E-04| 0.004| 0.001  | 0.005  | -0.006 | -0.004 | 1.5E-04  | 0.185 | 0.148 | 0.120 | 0.106 | 0.111 | 0.111 | 0.111 | 0.036 | 0.039 | 0.023 |

With 60% of the capital tracks the performance of the overall US stock market and with 40% the performance of a market-weighted bond index.

<table>
<thead>
<tr>
<th></th>
<th>WTI Oil</th>
<th>Gold</th>
<th>US003M</th>
<th>SWAP10</th>
<th>SWAP30</th>
<th>EURUSD</th>
<th>VIX</th>
<th>INFT</th>
<th>FINL</th>
<th>CONS</th>
<th>HLTH</th>
<th>ENRS</th>
<th>COND</th>
<th>INDU</th>
<th>MATR</th>
<th>UTIL</th>
<th>TELS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>0.007</td>
<td>0.030</td>
<td>0.118</td>
<td>-0.050</td>
<td>0.025</td>
<td>0.108</td>
<td>0.018</td>
<td>0.158</td>
<td>0.065</td>
<td>0.136</td>
<td>0.079</td>
<td>0.036</td>
<td>0.136</td>
<td>0.032</td>
<td>0.027</td>
<td>0.027</td>
<td>0.047</td>
</tr>
<tr>
<td>2</td>
<td>0.006</td>
<td>0.031</td>
<td>0.130</td>
<td>-0.045</td>
<td>0.017</td>
<td>0.090</td>
<td>0.018</td>
<td>0.159</td>
<td>0.074</td>
<td>0.133</td>
<td>0.076</td>
<td>0.042</td>
<td>0.140</td>
<td>0.032</td>
<td>0.027</td>
<td>0.030</td>
<td>0.037</td>
</tr>
<tr>
<td>2.5</td>
<td>0.010</td>
<td>0.056</td>
<td>0.049</td>
<td>-0.033</td>
<td>0.003</td>
<td>0.098</td>
<td>0.022</td>
<td>0.157</td>
<td>0.069</td>
<td>0.149</td>
<td>0.095</td>
<td>0.031</td>
<td>0.144</td>
<td>0.040</td>
<td>0.018</td>
<td>0.051</td>
<td>0.037</td>
</tr>
<tr>
<td>3</td>
<td>0.009</td>
<td>0.049</td>
<td>0.056</td>
<td>-0.037</td>
<td>0.008</td>
<td>0.097</td>
<td>0.022</td>
<td>0.159</td>
<td>0.076</td>
<td>0.157</td>
<td>0.104</td>
<td>0.032</td>
<td>0.145</td>
<td>0.048</td>
<td>0.007</td>
<td>0.042</td>
<td>0.028</td>
</tr>
</tbody>
</table>

Allianz RCM Global Water
Seek the long-term capital appreciation. Invest in equity securities of companies which have direct or indirect exposure to water related activities.

<table>
<thead>
<tr>
<th></th>
<th>WTI Oil</th>
<th>Gold</th>
<th>US003M</th>
<th>SWAP10</th>
<th>SWAP30</th>
<th>EURUSD</th>
<th>VIX</th>
<th>INFT</th>
<th>FINL</th>
<th>CONS</th>
<th>HLTH</th>
<th>ENRS</th>
<th>COND</th>
<th>INDU</th>
<th>MATR</th>
<th>UTIL</th>
<th>TELS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>0.025</td>
<td>0.046</td>
<td>-0.103</td>
<td>0.006</td>
<td>-0.004</td>
<td>0.213</td>
<td>-0.022</td>
<td>0.085</td>
<td>0.081</td>
<td>-0.003</td>
<td>0.083</td>
<td>-0.013</td>
<td>0.010</td>
<td>0.123</td>
<td>0.086</td>
<td>0.170</td>
<td>0.014</td>
</tr>
<tr>
<td>2</td>
<td>0.020</td>
<td>0.051</td>
<td>-0.110</td>
<td>0.019</td>
<td>0.021</td>
<td>0.210</td>
<td>-0.018</td>
<td>0.071</td>
<td>0.065</td>
<td>-0.005</td>
<td>0.090</td>
<td>0.016</td>
<td>0.023</td>
<td>0.122</td>
<td>0.099</td>
<td>0.134</td>
<td>0.036</td>
</tr>
<tr>
<td>2.5</td>
<td>0.011</td>
<td>0.037</td>
<td>-0.066</td>
<td>0.006</td>
<td>-0.007</td>
<td>0.221</td>
<td>-0.018</td>
<td>0.072</td>
<td>0.059</td>
<td>0.003</td>
<td>0.084</td>
<td>0.015</td>
<td>0.022</td>
<td>0.300</td>
<td>0.079</td>
<td>0.067</td>
<td>0.047</td>
</tr>
<tr>
<td>3</td>
<td>0.022</td>
<td>0.024</td>
<td>-0.025</td>
<td>0.001</td>
<td>0.003</td>
<td>0.219</td>
<td>-0.017</td>
<td>0.085</td>
<td>0.039</td>
<td>0.017</td>
<td>0.084</td>
<td>0.017</td>
<td>0.016</td>
<td>0.356</td>
<td>0.074</td>
<td>0.054</td>
<td>0.052</td>
</tr>
</tbody>
</table>

First Eagle Gold
It invests 80% of its assets in gold or related securities.

<table>
<thead>
<tr>
<th></th>
<th>WTI Oil</th>
<th>Gold</th>
<th>US003M</th>
<th>SWAP10</th>
<th>SWAP30</th>
<th>EURUSD</th>
<th>VIX</th>
<th>INFT</th>
<th>FINL</th>
<th>CONS</th>
<th>HLTH</th>
<th>ENRS</th>
<th>COND</th>
<th>INDU</th>
<th>MATR</th>
<th>UTIL</th>
<th>TELS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>0.004</td>
<td>0.998</td>
<td>-0.071</td>
<td>0.012</td>
<td>0.030</td>
<td>0.012</td>
<td>-0.012</td>
<td>0.158</td>
<td>0.073</td>
<td>0.265</td>
<td>0.109</td>
<td>0.090</td>
<td>0.005</td>
<td>0.238</td>
<td>0.308</td>
<td>0.059</td>
<td>-0.035</td>
</tr>
<tr>
<td>2</td>
<td>0.013</td>
<td>0.945</td>
<td>-0.122</td>
<td>0.083</td>
<td>0.042</td>
<td>0.048</td>
<td>-0.013</td>
<td>0.141</td>
<td>0.047</td>
<td>0.286</td>
<td>0.133</td>
<td>0.085</td>
<td>0.084</td>
<td>0.295</td>
<td>0.473</td>
<td>0.060</td>
<td>-0.012</td>
</tr>
<tr>
<td>2.5</td>
<td>0.046</td>
<td>0.941</td>
<td>-0.072</td>
<td>0.075</td>
<td>-0.049</td>
<td>-0.003</td>
<td>-0.017</td>
<td>0.119</td>
<td>0.054</td>
<td>0.225</td>
<td>0.101</td>
<td>0.080</td>
<td>0.047</td>
<td>0.197</td>
<td>0.415</td>
<td>0.017</td>
<td>-0.033</td>
</tr>
<tr>
<td>3</td>
<td>0.033</td>
<td>0.954</td>
<td>-0.075</td>
<td>0.077</td>
<td>-0.054</td>
<td>0.002</td>
<td>-0.018</td>
<td>0.069</td>
<td>0.071</td>
<td>0.186</td>
<td>0.110</td>
<td>0.085</td>
<td>0.032</td>
<td>0.129</td>
<td>0.445</td>
<td>0.011</td>
<td>-0.043</td>
</tr>
</tbody>
</table>

Figure 1.2: We consider different funds and through a simple linear regression obtain the exposures to the selected risk factors. To emphasize the importance of estimation window selection we report results for four estimation windows: 1.5, 2, 2.5 and 3 years.
1.1 Multifactor models in finance

1.1.1 Kalman Filter

An alternative method for estimating factor exposures that does not require any decision on the time frame to consider is the Kalman Filtering (KF) introduced in Kalman [1960]. It is based on a state space model composed from two equations namely state and observation equation. The Kalman filter is a recursive algorithm that computes estimates for the unobserved variables at time \( t \), based on the available information up to that date. The filter structure and the update after receiving a new information give more reliable results than a simple Ordinary Least Squares regression. Here we outline how the Kalman Filter can be used for estimation.

Let \( \mathbf{r}_t \) be a \( n \times 1 \) be the vector of \( n \) asset returns at time \( t \). Suppose that each asset return depends linearly on \( p \) factors that are the rows of the deterministic \( n \times p \) matrix \( \mathbf{H}_t \). The state variables are the factor exposures in the \( p \times 1 \) vector \( \mathbf{\beta}_t \). A general linear state space model is of the form:

\[
\mathbf{r}_t = \mathbf{H}_t \mathbf{\beta}_t + \mathbf{\epsilon}_t \tag{1.11}
\]

\[
\mathbf{\beta}_t = A_t \mathbf{\beta}_{t-1} + \mathbf{v}_t \tag{1.12}
\]

where \( \mathbf{\epsilon}_t \sim N(0, \mathbf{R}_t) \) and \( \mathbf{v}_t \sim N(0, \mathbf{Q}_t) \). We assume \( \mathbf{\epsilon}_t \) and \( \mathbf{v}_t \) to be mutually and serially uncorrelated, meaning that they don’t depend on each other and on their past values. We will refer to equation 1.11 as the observation (or measurement) equation, while equation 1.12 is the state equation which describes the dynamic of the unobserved state variable \( \mathbf{\beta}_t \). If \( A_t = I \), the distribution of each factor loading in \( \mathbf{\beta} \) is a random walk. In financial applications it is reasonable to suppose \( A_t \) different from the identity matrix in order to model autocorrelated factor exposures because we observe persistence in financial market returns. In the case of a time homogeneous system, matrices \( A_t, R_t \) and \( Q_t \) don’t depend on time. We assume stationarity for the shocks and for the autocorrelation structure in the state variables, i.e we consider an homogeneous model.

Once the model is put in the state space form, the Kalman Filter is used to compute optimal forecasts of the normally distributed state space vector \( \mathbf{\beta}_t \), at a given time \( t \), based on the information we have up to the previous time \( t-1 \). The purpose of the filter is to update the estimate for the state vector as soon as we observe \( \mathbf{r}_t \). Consider the information set \( I_t \) generated by the return observations and by the estimates of the state space variable up to time \( t \):

\[
I_{t-1} = \{ \mathbf{r}_{t-1}, \ldots, \mathbf{r}_0, \hat{\mathbf{\beta}}_{t-1}, \ldots, \hat{\mathbf{\beta}}_0 \} \tag{1.13}
\]

It contains the past return observations \( \mathbf{r}_{t-1}, \ldots, \mathbf{r}_0 \) and the past estimates \( \hat{\mathbf{\beta}}_{t-1}, \ldots, \hat{\mathbf{\beta}}_0 \). The a priori estimates of the state variables are the best estimates using the information contained in \( I_{t-1} \):

\[
\hat{\mathbf{\beta}}^*_t = E[\mathbf{\beta}_t | I_{t-1}] \tag{1.14}
\]
The a priori error estimates are:

$$
\epsilon_t^* = \beta_t - \hat{\beta}_t^*
$$

(1.15)

with covariance matrix:

$$
P_t^* = E \left[ \epsilon_t^* \epsilon_t^{*\prime} \right]
$$

(1.16)

The posterior information set $I_t^*$ in addition to $I_{t-1}$ contains the observation vector $r_t$ and the a priori estimate $\hat{\beta}_t^*$:

$$
I_t^* = \{ r_t, r_{t-1}, ..., r_0, \hat{\beta}_t^*, \hat{\beta}_{t-1}, ..., \hat{\beta}_0 \}
$$

(1.17)

The a posteriori estimate for the state vector is obtained considering the a priori estimate and the return observation:

$$
\hat{\beta}_t = E [ \beta_t \mid I_t^* ]
$$

(1.18)

The a posteriori estimates for the estimation error and the covariance matrix are:

$$
\epsilon_t = \beta_t - \hat{\beta}_t
$$

(1.19)

$$
P_t = E [ \epsilon_t \epsilon_t^{\prime} ]
$$

(1.20)

In the Kalman Filter system the a posteriori estimate $\hat{\beta}_t$ is supposed to be the a priori estimate.
\( \hat{\beta}_t^* \) plus a correction term which depends on \( K_t \), the Kalman gain matrix:

\[
\hat{\beta}_t = \hat{\beta}_t^* + K_t \left( r_t - H_t \hat{\beta}_t^* \right)
\] (1.21)

The Kalman gain matrix \( K_t \) is a measure of the modification from the a priori \( \hat{\beta}_t \) to the a posteriori estimates after the observation of market returns \( r_t \). If we are fully confident in our a priori estimates, that implies zero elements for \( P_t^* \), the new observation \( r_t \) will not modify the estimation of the state space variables. Conversely, if the covariance matrix \( R \) of the measurement equation has all zero elements, our confidence in the measurements is high so the Kalman gain is high.

Suppose now that we have already the a priori error and the corresponding covariance matrix. As soon as we get a new observation \( r_t \), we update the estimation. After some simple calculations it is possible to have an expression for \( P_t \) in terms of the a priori error covariance matrix and the Kalman gain matrix:

\[
P_t = E \left[ \left( \beta_t - \hat{\beta}_t \right) \left( \beta_t - \hat{\beta}_t \right)' \right]
\] (1.22)

\[
= E \left[ \left( \beta_t - \hat{\beta}_t - K_t \left( r_t - H_t \hat{\beta}_t^* \right) \right) \left( \beta_t - \hat{\beta}_t - K_t \left( r_t - H_t \hat{\beta}_t^* \right) \right)' \right]
\] (1.23)

\[
= P_t^* - K_t H_t P_t^* - P_t^* H_t' K_t' + K_t H_t P_t^* H_t' K_t' + K_t R_t K_t'
\] (1.24)

We used the fact that vectors \( e_t \) and \( \epsilon_t \) are orthogonal. The intertemporal independence assumption for the errors implies that the error covariance matrix \( P_t \) is diagonal. The trace of \( P_t \) coincides with the sum of squared a posteriori errors. Observe that:

\[
tr \left( P_t \right) = tr \left( P_t^* \right) - tr \left( K_t H_t P_t^* \right) + tr \left( K_t H_t P_t^* H_t' K_t' \right) + tr \left( K_t R_t K_t' \right)
\] (1.25)

Since \( P_t \) is the expectation of the squared residuals, we want to minimize it meaning that after the observation of return \( r_t \) the a posteriori error becomes small. The Kalman gain matrix introduced in equation 1.21 is chosen to be the solution of the following minimization problem:

\[
\begin{align*}
\min_{K_t} & \quad tr(P_t) \\
\text{s.t.} & \quad \hat{\beta}_t = \hat{\beta}_t^* + K_t \left( r_t - H_t \hat{\beta}_t^* \right)
\end{align*}
\] (1.26)

Recall the following properties:

\[
\frac{\partial tr(KX)}{\partial K} = X'
\] (1.27)

\[
\frac{\partial tr(KBK')}{\partial K} = 2KB
\] (1.28)

where \( B \) is a symmetric matrix.
The minimum is obtained when:

$$
\frac{\partial \text{tr}(P_t)}{\partial K} = 0 - (H_t P_t^*)' - P_t^* H_t' + 2K_t H_t P_t^* H_t' + 2K_t R_t
$$

(1.29)

$$
= 0
$$

(1.30)

The solution is the desired Kalman gain matrix:

$$
K_t = P_t^* H_t' (H_t P_t^* H_t' + R_t)^{-1}
$$

(1.31)

Using this matrix and the observation $r_t$, we are able to update the estimation:

$$
\hat{\beta}_t^* = E[\beta_t | I_{t-1}] = E[A\beta_{t-1} + v_{t-1} | I_{t-1}] = A\hat{\beta}_{t-1}
$$

(1.32)

It is now possible to compute the covariance error of the a priori estimator:

$$
P_t^* = E \left[ (\beta_t - \hat{\beta}_t) (\beta_t - \hat{\beta}_t)' | I_{t-1} \right]
$$

(1.33)

$$
= E \left[ (A(\beta_{t-1} - \hat{\beta}_{t-1}) + v_{t-1}) (A(\beta_{t-1} - \hat{\beta}_{t-1}) + v_{t-1})' | I_{t-1} \right]
$$

(1.34)

$$
= AP_{t-1} A' + Q
$$

(1.35)

It is possible to have an explicit expression for $P_t$:

$$
P_t = E \left[ (\beta_t - \hat{\beta}_t) (\beta_t - \hat{\beta}_t)' | I_t \right]
$$

(1.36)

$$
= (I - K_t H_t) P_t^*
$$

(1.37)

The Kalman filter algorithm gives an a priori estimate of the factor loading $\hat{\beta}_t^*$ and immediately after we observe the actual return it updates the estimate in order to improve its prediction in the next step. The two steps are called the forecast step where we get the a priori estimates for the state space variables and the corresponding error covariance matrix:

$$
\hat{\beta}_t^* = A\hat{\beta}_{t-1}
$$

(1.38)

$$
P_t^* = AP_{t-1} A' + Q
$$

(1.39)

and the update during which the a posteriori quantities are obtained:

$$
K_t = P_t^* H_t' (H_t P_t^* H_t' + R_t)^{-1}
$$

(1.40)

$$
\hat{\beta}_t = \hat{\beta}_t^* + K_t (v_t - H_t \hat{\beta}_t^*)
$$

(1.41)

$$
P_t = (I - K_t H_t) P_t^*
$$

(1.42)
From the description of the main passages needed for the algorithm construction it can be easily noticed that the initial conditions we put on two state space equations highly influence the solution and the difficulty in its attainment. In fact, the assumption of stability for the error terms ease the update of the covariance matrices for the a priori and the a posteriori estimate errors. However, the implementation of this algorithm is not always possible since, as we will see later, about 50 to 100 observations are needed in order to reach stability. For market risk factors this is not a problem while for macroeconomic factors where the data available are not so frequent we may be not able to implement the KF algorithm.

![Figure 1.4: In this figure we describe how the Kalman Filter works. At time t, we have an a priori estimate for the factor exposure \( \hat{\beta}_t \). With the available information we compute the Kalman Gain matrix \( K_t \). The new observation \( r_t \) allows to update the state space variable estimation and get the a posteriori estimate \( \hat{\beta}_t \). Now an a priori estimate for the state space variable at time \( t + 1 \) becomes the starting point for the process to continue.](image-url)
Multifactor models become dynamic by specifying the dynamic for each factor or by using time varying factor exposure $\beta$. The easiest method to obtain time-varying factor exposures is by using rolling Ordinary Least Squares regression. However, this approach is costly in terms of computations. It is preferable to use a more sophisticated instrument like Kalman Filter since it generates sequences for factor exposures that minimizes the variance of the prediction error as we discussed before.

We plot in figure 1.4 the OLS and KF factor loadings for the Vanguard Fund Utilities index Fund when three factors are considered: the Utilities Index, the exchange rate EURUSD and and the Industry Index. The OLS univariate regression gives a constant factor loading for the entire period, while the KF allows for time-varying exposures. In the first 100 iterations the algorithm is not so stable but after it gives reliable estimates. Observe that we supposed $A=1$ in order to show that even the simplest choice can give time varying factor exposures. $Q=8.5e-05$ and $R=7.9e-05$ are maximum likelihood estimates based on conditional probabilities.

![Figure 1.5: In the figure are shown results of univariate OLS regression and Kalman Filter analysis of the Vanguard Fund Utilities index vs three factors: Utility index, EURUSD and Industry Index.](image)

Even if we use rolling OLS regression, the limit of having a mean factor exposure estimate for the entire period remains. On the other side, since the Kalman Filter algorithm is not so easy to understand its use in finance is limited.
1.2 Statistical Factors

Factor identification and dimensionality reduction are among the primary goals of multivariate financial time series analysis. They are useful in order to extract features and to find latent relations of risk drivers from high dimensional and complex portfolios. Attainment of these goals can be challenging.

Principal Component Analysis (PCA) is a mathematical procedure introduced in Pearson [1901] that uses an orthogonal transformation to convert a set of correlated variables into a set of uncorrelated variables called principal components (see Jolliffe [2002]). It yields uncorrelated factors but together with the Gaussian distribution assumption they become independent. In so doing, the problem of defining a dependence structure among the components is avoided. Many papers test the performance of principal component analysis when applied to financial data. Different problems were faced with the ahead mentioned technique like hedging bond portfolios by Falkenstein and Hanweck [1997], interest rate forecasting by Reisman and Zohar [2004], immunization by Soto [2004] and value at risk estimation by Abad and Benito [2007]. PCA simultaneously solves factor identification and dimension reduction problems but the assumption of Gaussianity is not realistic for the description of financial data.

An alternative method for factor identification is Independent Component Analysis (ICA). ICA extracts Independent Components (ICs) but there is no projection on the maximum variance direction. This technique has been implemented in stock returns analysis by Back and Weigend [1997], in high frequency analysis by Moody and Wu [1998], in immunization problems by Bellini and Salinelli [2003], in an intertemporal GARCH context by Wu et al. [2006] and in risk management by Chen et al. [2010]. These articles show a nice performance of the ICA method, when applied to financial data that are not gaussianly distributed. However, the literature on ICA applied to finance is not so rich due to the fact that the factors do not have an economic interpretation.

The two methods, PCA and ICA, are different since they are based on quantities describing two dependence measures. Correlation is a measure of linear relation between variables while independence measures the existence of any relation between them. Note that independence, meaning that it doesn’t exist any type of relation between variables, implies uncorrelation which considers only the linear relation between the variables. The vice versa is not always true.
**Definition 1.** The covariance of two r.v’s $X$ and $Y$ is:

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] \quad (1.43)$$

They are said to be uncorrelated if $\text{Cov}(X, Y) = 0$. The quantity:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} \quad (1.44)$$

is the correlation coefficient for the two r.v’s $X, Y$. Notice that if the two r.v’s are uncorrelated then:

$$E[XY] = E[X]E[Y] \quad (1.45)$$

**Definition 2.** Two r.v’s $X, Y$ are said to be independent if their joint probability distribution is the product of their marginal probability distributions, i.e:

$$P_{X,Y}(x, y) = P_X(x)P_Y(y) \quad (1.46)$$

for any realization $x, y$ of the two r.v’s.

From the Bayes rule we have that the conditional probability distribution for $X$ is:

$$P_{X|Y}(x|y) = P_{X,Y}(x, y)/P_Y(y) \quad (1.47)$$

giving an equivalent definition for independence:

$$P_{X|Y}(x|y) = P_X(x) \quad (1.48)$$

Notice that since:

$$E[XY] = \int \int xyP_{X,Y}(x, y)dxdy = \int \int xyP_X(x)P_Y(y)dxdy \quad (1.49)$$

we have that independence implies uncorrelation:

$$E[XY] = \int xP_X(x)dx \int yP_Y(y)dy = E[X]E[Y] \quad (1.50)$$

The vice versa is not always true except for the case when $X, Y$ are the marginals of a multivariate normal distribution.
1.2 Statistical Factors

1.2.1 Principal Component Analysis

There are many books explaining what Principal Component Analysis is. In particular Jolliffe [2002] gives in two sentences a general overview of the procedure:

The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables.

PCA is used in multivariate analysis since it allows to obtain new uncorrelated variables. We describe the idea and some important characteristics of this technique. Consider a dataset given by a matrix $X_{p \times T}$ with $p$ denoting the number of variables and $T$ the length of time observation. In matrix notations each variable is given by a generic row:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,T} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i,1} & x_{i,2} & \cdots & x_{i,T} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p,1} & x_{p,2} & \cdots & x_{p,T} \end{bmatrix}$$

(1.51)

We want to generate a new set of uncorrelated variables by linearly transforming the original variables $X$. The new variables are the rows of the matrix $Y_{p \times T}$ obtained as:

$$Y = A'X$$

(1.52)

and are called Principal Components (PCs). Geometrically, they are the projections of the original variables on the rows of $A_{p \times p}$ that contain the eigenvectors associated to the covariance matrix $\Sigma_X$ of the original variables defined as:

$$\Sigma_X = XX'$$

(1.53)

The covariance matrix $\Sigma_X$ is symmetric and positive semidefinite\(^3\). Its eigenvalues are non negative and its eigenvectors are orthogonal. There exist infinitely many eigenvectors associated to each eigenvalue but by adding the unit length condition, the corresponding eigenvectors are orthonormal.

The $i$-th principal component is obtained by multiplying the eigenvector associated to the $i$-th

\(^3\)This is a property that satisfy all the covariance matrices
largest eigenvalue of the dataset covariance matrix with the original variables, i.e.:

\[ y_i = \begin{bmatrix} y_{i,1} & y_{i,2} & \ldots & y_{i,T} \end{bmatrix} = \begin{bmatrix} a_{1,i} & a_{2,i} & \ldots & a_{p,i} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_p \end{bmatrix} \tag{1.54} \]

Putting together all the PCs we have:

\[ Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{p} a_{i,1} x_i \\ \sum_{i=1}^{p} a_{i,2} x_i \\ \vdots \\ \sum_{i=1}^{p} a_{i,k} x_i \\ \vdots \\ \sum_{i=1}^{p} a_{i,p} x_i \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{2,1} & \ldots & a_{i,1} & \ldots & a_{p,1} \\ a_{1,2} & a_{2,2} & \ldots & a_{i,2} & \ldots & a_{p,2} \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ a_{1,k} & a_{2,k} & \ldots & a_{i,k} & \ldots & a_{p,k} \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ a_{1,p} & a_{2,p} & \ldots & a_{i,p} & \ldots & a_{p,p} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_p \end{bmatrix} \tag{1.55} \]

Uncorrelation for the transformed variables implies a diagonal form for the covariance matrix \(\Sigma_Y = YY'\). The new and the original covariance matrices are connected since:

\[ \Sigma_Y = A'X(A'X)' = A'XX'A = A'\Sigma X A \tag{1.56} \]

Observe that \(\Sigma_X\) and \(\Sigma_Y\) are the covariance matrices respectively for the r.v.'s \(X\) and \(Y\). The sample covariances are obtained by dividing the theoretical covariances with \(T - 1\).

The variance of the \(i\)-th PC is:

\[ Var(y_i) = a_i'\Sigma X a_i \tag{1.57} \]

and it coincides with \(\lambda_i\) that is the \(i\)-th largest eigenvalue of the matrix \(\Sigma_X\). The uncorrelation of the PCs allows to linearly decompose the total variance, i.e.:

\[ Var(Y) = Var(y_1) + Var(y_2) + \ldots + Var(y_p) = \lambda_1 + \lambda_2 + \ldots + \lambda_p. \tag{1.58} \]

The PCA is known as a method used to reduce the number of variables considered in order to make easier the identification of common features among the original variables. This is not done without an information loss but in some circumstances with only few PCs we can reproduce a large portion of the original variability. Considering only the first \(r\) PCs ordered according to their variance, with \(r < p\), the explained variance is:

\[ \frac{\lambda_1 + \lambda_2 + \ldots + \lambda_r}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \tag{1.59} \]
1.2 Statistical Factors

We can go on two directions: fix \( r \) (the number of PCs) and compute the variance explained by the first \( r \) PCs or choose first the level of the variance to be explained and then find the minimum number of PCs for which it is possible. In a financial context, we aim at explaining most of the variance present in the dataset so we follow the second direction. Eigenvectors are calculated iteratively. The two most applied procedures are the singular value decomposition (SVD) and Jacobi Rotation method. The last one is slower but more accurate. See Golub and Van Loan [1996] for more details on these procedures.

PCs are uncorrelated variables each orthogonal to the previous one. Let us see how the PCA works for a given dataset \( X \):

\[
X = (x_1, x_2, ..., x_n) \quad (1.60)
\]

This procedure can be represented as a constrained maximization problem. Each PC is a convex combination of the observations:

\[
y_k = \sum_{i=1}^{p} a_{ik}x_i = a_k'X \quad (1.61)
\]

where \( a_k \), for any \( k > l \geq 1 \) is the solution of the problem:

\[
\begin{align*}
\max_{a_k} & \quad \text{Var} (y_k) \\
\text{s.t.} & \quad C_{ov} (y_k, y_l) = 0 \\
& \quad a_k'^* a_k = 1
\end{align*} \quad (1.62)
\]

The consequence of zero covariance is the orthogonality between the PCs.

Let us see more details on how the PCs are computed. In particular, to find \( a_1 \) we must observe that:

\[
\text{Var} (y_1) = \text{Var} (x_1 a_{11} + x_2 a_{21} + ... + x_n a_{n1}) = a_1' C_{ov} (x_1, ..., x_n) a_1 = a_1' \Sigma a_1 \quad (1.63)
\]

so that the maximization problem becomes:

\[
\begin{align*}
\max_{a_1} & \quad \text{Var} (y_1) \\
\text{s.t.} & \quad a_1'^* a_1 = 1
\end{align*} \quad (1.64)
\]

By using the technique of Lagrange multipliers we can rewrite the problem in the form:

\[
\max_{a_1} \left( a_1'^* \Sigma a_1 - \lambda_1 \left( a_1'^* a_1 - 1 \right) \right) \leftrightarrow \Sigma a_1 - \lambda_1 a_1 = 0 \quad (1.65)
\]

We recognize that this problem can be viewed as an eigenvalue decomposition since we have:

\[
\Sigma a_1 = \lambda_1 a_1 \quad (1.66)
\]
Notice that $a_1$ is an eigenvector of $\Sigma$ associated to the eigenvalue $\lambda_1$. Furthermore, this eigenvalue coincides with the variance of the first component. Indeed, the variance associated to the vector $y_1$ is:

$$Var (y_1) = a_1' \Sigma a_1 = a_1' \lambda_1 a_1 = \lambda_1$$  

(1.67)

Eigenvector $a_2$ is the solution of the following constrained problem:

$$\begin{cases}
\max_{a_2} Var (y_2) \\
s.t. \\
Cov (y_2, y_1) = 0 \\
a_2' a_2 = 1
\end{cases}$$  

(1.68)

The uncorrelation between the first two PCs gives a condition about the product of the eigenvectors $a_1$ and $a_2$:

$$Cov (a_1' X, a_2' X) = a_1' \Sigma a_2 = a_2' \lambda_1 a_1 = a_1' a_2' a_1 = 0 \Leftrightarrow a_2' a_1 = 0$$  

(1.69)

By writing the problem in the Lagrangian form we have:

$$\max_{a_2} \left( a_2' \Sigma a_2 - \lambda_2 (a_2' a_2) - \phi (a_2' a_1) \right) \Leftrightarrow \Sigma a_2 - \lambda_2 a_2 - \phi a_1 = 0 \Leftrightarrow a_2' \Sigma a_2 - \lambda_2 a_2' a_2 - \phi a_1 = 0$$  

(1.70)

The first two quantities are zero, so it remains that:

$$-\phi a_1' a_1 = 0 \Leftrightarrow \phi = 0$$  

(1.71)

The solution of the Lagrangian problem can be found in the eigenvalue problem:

$$\Sigma a_2 - \lambda_2 a_2 = 0 \Leftrightarrow \Sigma a_2 = \lambda_2 a_2$$  

(1.72)

In the same way, it can be shown that the maximum variance for the i-th PC score is i-th largest eigenvalue.

In real world applications we face the problem of deciding whether the Principal Component Analysis should be applied on the covariance or on the correlation matrix of the given dataset. Results are affected by the volatility of each variable when you do the PCA on the covariance matrix, whereas they are independent of volatility when PCA is based on the correlation matrix. Applying PCA to covariance or correlation is an empirical question. Note that PCA is not scale invariant and using the correlation matrix for PCA means to rescale the data such that they all have unit variance. This feature suggests that results of PCA on covariance and correlation matrices could differ significantly. If assets in the portfolio are of the
same type and with similar variance, the choice is not so relevant. But if we consider a portfolio composed of bonds and equities with bond yields being much less volatile than equity returns, PCA on covariance matrix would give higher weights to equities in the first component than if we use the correlation matrix. Significant differences in variable variances implies a greater influence of the variables with the largest variance in the first eigenvectors associated to the covariance matrix. However, sometimes using PCA on correlation matrix may be convenient. For example, consider a portfolio composed by a bond and a stock where bond return is expressed in basis points and stock return is expressed in units. Applying directly PCA on the covariance matrix would give misleading results since the two variables are not measured using the same unit. In some sense choosing to work with correlation matrices protects from these kind of problems but has the drawback that not all the information present in the data is used.

We will follow a hybrid approach composed by two steps. In the first step we standardize the variables by using the volatility of a previous fixed period that can be for example the volatility of the last quarter. Depending on the past, the relative variances are near one but they can be higher or lower since the market conditions change frequently. As a second step we perform PCA on the new covariance matrix whose diagonal elements will have similar values. In doing so, we maintain some of the information about greater volatility of some assets and in the same time have no big differences in the data units.

In particular we consider the five GICS indexes: Information Technology, Financial, Consumption Staple, Health and Energy. The dataset is composed by daily log returns going from 12/09/2010 to 12/09/2012. We first perform a PCA analysis on the covariance matrix computed using data from 01/01/2011 to 12/09/2012. For the same time frame we perform the PCA on the correlation matrix. It is possible to observe that now the variance explained from the first PC is lower than in the case where the covariance matrix. We compare the results with the hybrid approach. The volatility computed on the period 12/09/2010 - 31/12/2012 is used to standardize the data in the remaining period. The outcome is coherent with our expectations, in the sense that the first PC explains more than in the correlation case and less than in the covariance context.
<table>
<thead>
<tr>
<th></th>
<th>Cov</th>
<th>Inf</th>
<th>Fin</th>
<th>Cons</th>
<th>Hlth</th>
<th>Enr</th>
<th>Eigenvalues</th>
<th>Perc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inf</td>
<td>1.56E-04</td>
<td>1.74E-04</td>
<td>7.16E-05</td>
<td>9.79E-05</td>
<td>1.53E-04</td>
<td>7.05E-04</td>
<td></td>
<td>86%</td>
</tr>
<tr>
<td>Fin</td>
<td>1.74E-04</td>
<td>2.91E-04</td>
<td>9.72E-05</td>
<td>1.32E-04</td>
<td>2.11E-04</td>
<td>4.69E-05</td>
<td></td>
<td>92%</td>
</tr>
<tr>
<td>Cons</td>
<td>7.16E-05</td>
<td>9.72E-05</td>
<td>5.62E-05</td>
<td>6.14E-05</td>
<td>8.66E-05</td>
<td>3.44E-05</td>
<td></td>
<td>96%</td>
</tr>
<tr>
<td>Hlth</td>
<td>9.79E-05</td>
<td>1.32E-04</td>
<td>6.14E-05</td>
<td>9.21E-05</td>
<td>1.17E-04</td>
<td>2.50E-05</td>
<td></td>
<td>99%</td>
</tr>
<tr>
<td>Enr</td>
<td>1.53E-04</td>
<td>2.11E-04</td>
<td>8.66E-05</td>
<td>1.17E-04</td>
<td>2.27E-04</td>
<td>1.01E-05</td>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1.1: Covariance matrix using five sector indexes going from January 2011 to September 2012. The Consumer and Health sectors show a lower variance meaning that the first eigenvalues will be driven by the remaining sectors are Information Technology, Financial and Energy.

<table>
<thead>
<tr>
<th></th>
<th>Corr</th>
<th>Inf</th>
<th>Fin</th>
<th>Cons</th>
<th>Hlth</th>
<th>Enr</th>
<th>Eigenvalues</th>
<th>Perc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inf</td>
<td>1.00</td>
<td>0.82</td>
<td>0.77</td>
<td>0.82</td>
<td>0.82</td>
<td>4.22</td>
<td></td>
<td>84%</td>
</tr>
<tr>
<td>Fin</td>
<td>0.82</td>
<td>1.00</td>
<td>0.76</td>
<td>0.81</td>
<td>0.82</td>
<td>0.28</td>
<td></td>
<td>90%</td>
</tr>
<tr>
<td>Cons</td>
<td>0.77</td>
<td>0.76</td>
<td>1.00</td>
<td>0.85</td>
<td>0.77</td>
<td>0.19</td>
<td></td>
<td>94%</td>
</tr>
<tr>
<td>Hlth</td>
<td>0.82</td>
<td>0.81</td>
<td>0.85</td>
<td>1.00</td>
<td>0.81</td>
<td>0.18</td>
<td></td>
<td>97%</td>
</tr>
<tr>
<td>Enr</td>
<td>0.82</td>
<td>0.82</td>
<td>0.77</td>
<td>0.81</td>
<td>1.00</td>
<td>0.14</td>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1.2: Correlation matrix using the data going from January 2011 to September 2012. The explained variance by the first PC is lower than in the case when covariance matrix was used.

<table>
<thead>
<tr>
<th></th>
<th>Hybrid</th>
<th>Inf</th>
<th>Fin</th>
<th>Cons</th>
<th>Hlth</th>
<th>Enr</th>
<th>Eigenvalues</th>
<th>Perc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inf</td>
<td>1.02</td>
<td>0.82</td>
<td>0.75</td>
<td>0.88</td>
<td>0.88</td>
<td>4.42</td>
<td></td>
<td>85%</td>
</tr>
<tr>
<td>Fin</td>
<td>0.82</td>
<td>1.00</td>
<td>0.75</td>
<td>0.87</td>
<td>0.88</td>
<td>0.28</td>
<td></td>
<td>90%</td>
</tr>
<tr>
<td>Cons</td>
<td>0.75</td>
<td>0.75</td>
<td>0.94</td>
<td>0.89</td>
<td>0.79</td>
<td>0.20</td>
<td></td>
<td>94%</td>
</tr>
<tr>
<td>Hlth</td>
<td>0.88</td>
<td>0.87</td>
<td>0.89</td>
<td>1.10</td>
<td>0.91</td>
<td>0.19</td>
<td></td>
<td>98%</td>
</tr>
<tr>
<td>Enr</td>
<td>0.88</td>
<td>0.88</td>
<td>0.79</td>
<td>0.91</td>
<td>1.14</td>
<td>0.11</td>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1.3: Covariance matrix using hybrid data: the last quarter of 2010 was used to compute the basis standard deviation. The data going from January 2011 to September 2012 were then "standardized" and used for the covariance matrix computation. The explained variance by the first PC is lower than when we use covariance matrix for PCA and higher than in the correlation matrix case.
Now we apply the PCA analysis on Euro market Z-spread data grouped by sectors and subsectors in order to identify possible leading factors. The notion of spread is encountered frequently in finance, especially in fixed income. We call nominal spread the difference between the yield to maturity of a risky bond and the yield to maturity of a risk-free one with the same maturity.

The zero volatility spread (z-spread) is the fixed amount that should be added to the Libor curve in order to obtain the market price for a given bond. In the last decade credit risk has become very important. In fact for portfolios that contain bonds, the z-spreads are crucial risk factors. If we are in the market and have in the portfolio positions on a large number of bonds, identifying a smaller number of factors that can explain the movements of the z-spreads for different ratings reduces the number of risk factors considered in the portfolio management. In addition, the information we get can be used to identify the sectors or subsectors that have a higher influence for a given rating.

In the considered market we have that most of the investment grade issuers are AA, A or BBB rated. As uncertainty grows, in the market there is much more reluctance in assigning an AAA rating to a given issuer. Even for the A rated is not easy to have a large dataset since a lot of companies originally having this rating have been either uprated or more probably downrated. As we will show for the A group the most representative sector is the Public Sovereign since it is perceived as safer than other sectors whose issuers are private companies. However we were able to perform a complete analysis only for the time span going from 01/08/2011 to 29/02/2012. The considered period is not too long but it is interesting to observe the most important factors driving the z-spread of the AA rated companies.

<table>
<thead>
<tr>
<th>Util_Gas</th>
<th>Util_Transp</th>
<th>Fin_Bank</th>
<th>Util_Water</th>
<th>Fin_Finance</th>
<th>Basic_Industries</th>
<th>Indus_Healthcare</th>
<th>Indus_Manufacturing</th>
<th>Indus_Retail</th>
<th>Indus_Technology</th>
<th>Indus_Telecom</th>
<th>Pub_Sovereign</th>
<th>Util_Electric</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.410</td>
<td>1.209</td>
<td>1.037</td>
<td>0.684</td>
<td>1.152</td>
<td>1.008</td>
<td>0.623</td>
<td>0.831</td>
<td>0.544</td>
<td>0.572</td>
<td>0.531</td>
<td>1.159</td>
<td>1.251</td>
</tr>
<tr>
<td>1.209</td>
<td>1.037</td>
<td>0.684</td>
<td>1.152</td>
<td>1.008</td>
<td>0.544</td>
<td>0.572</td>
<td>0.831</td>
<td>0.544</td>
<td>0.572</td>
<td>0.531</td>
<td>1.159</td>
<td>1.251</td>
</tr>
<tr>
<td>1.037</td>
<td>0.684</td>
<td>1.152</td>
<td>1.008</td>
<td>0.544</td>
<td>0.572</td>
<td>0.831</td>
<td>0.544</td>
<td>0.544</td>
<td>0.572</td>
<td>0.531</td>
<td>1.159</td>
<td>1.251</td>
</tr>
<tr>
<td>0.684</td>
<td>1.152</td>
<td>1.008</td>
<td>0.544</td>
<td>0.572</td>
<td>0.831</td>
<td>0.544</td>
<td>0.544</td>
<td>0.544</td>
<td>0.572</td>
<td>0.531</td>
<td>1.159</td>
<td>1.251</td>
</tr>
</tbody>
</table>

Table 1.4: Rated A subsectors: Covariance matrix using the 2010 historical standard deviation as basis for standardizing the data.

The corresponding eigenvalues and explained variance are given in the table 1.10:

By observing the eigenvectors we can find the sectors that drive the market.

If we choose to build the covariance matrix based on the hybrid data with base the entire year we observe still big differences in terms of variability among the subsectors. A more accurate
Table 1.5: Eigenvalues associated to the A rated covariance matrix using 2010 as basis year

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>Explained Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.293</td>
<td>53%</td>
</tr>
<tr>
<td>2.955</td>
<td>68%</td>
</tr>
<tr>
<td>2.717</td>
<td>82%</td>
</tr>
<tr>
<td>0.979</td>
<td>87%</td>
</tr>
<tr>
<td>0.668</td>
<td>91%</td>
</tr>
<tr>
<td>0.430</td>
<td>93%</td>
</tr>
<tr>
<td>0.423</td>
<td>95%</td>
</tr>
<tr>
<td>0.321</td>
<td>97%</td>
</tr>
<tr>
<td>0.257</td>
<td>98%</td>
</tr>
<tr>
<td>0.214</td>
<td>99%</td>
</tr>
<tr>
<td>0.183</td>
<td>100%</td>
</tr>
<tr>
<td>0.009</td>
<td>100%</td>
</tr>
<tr>
<td>0.000</td>
<td>100%</td>
</tr>
<tr>
<td>0.000</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1.6: Eigenvectors of the A rated sectors

<table>
<thead>
<tr>
<th>Sector</th>
<th>Eigvec1</th>
<th>Eigvec2</th>
<th>Eigvec3</th>
<th>Eigvec4</th>
<th>Eigvec5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Util_Gas</td>
<td>0.353</td>
<td>0.028</td>
<td>0.101</td>
<td>0.163</td>
<td>0.317</td>
</tr>
<tr>
<td>Util_Transp</td>
<td>0.302</td>
<td>0.024</td>
<td>0.087</td>
<td>0.140</td>
<td>0.271</td>
</tr>
<tr>
<td>Fin_Bank</td>
<td>0.263</td>
<td>-0.325</td>
<td>-0.369</td>
<td>-0.066</td>
<td>-0.075</td>
</tr>
<tr>
<td>Util_Water</td>
<td>0.336</td>
<td>0.031</td>
<td>0.096</td>
<td>0.159</td>
<td>0.304</td>
</tr>
<tr>
<td>Fin_Finance</td>
<td>0.302</td>
<td>-0.386</td>
<td>-0.572</td>
<td>-0.144</td>
<td>-0.174</td>
</tr>
<tr>
<td>Fin_Insurance</td>
<td>0.191</td>
<td>-0.079</td>
<td>-0.190</td>
<td>-0.143</td>
<td>0.236</td>
</tr>
<tr>
<td>Basic_Industries</td>
<td>0.230</td>
<td>0.094</td>
<td>0.174</td>
<td>-0.077</td>
<td>-0.248</td>
</tr>
<tr>
<td>Indus_Healthcare</td>
<td>0.259</td>
<td>0.083</td>
<td>0.180</td>
<td>-0.001</td>
<td>-0.208</td>
</tr>
<tr>
<td>Indus_Manufacturing</td>
<td>0.239</td>
<td>0.055</td>
<td>0.168</td>
<td>-0.036</td>
<td>-0.261</td>
</tr>
<tr>
<td>Indus_Retail</td>
<td>0.275</td>
<td>0.081</td>
<td>0.193</td>
<td>0.048</td>
<td>-0.520</td>
</tr>
<tr>
<td>Indus_Technology</td>
<td>0.160</td>
<td>0.104</td>
<td>0.218</td>
<td>-0.908</td>
<td>0.153</td>
</tr>
<tr>
<td>Indus_Telecom</td>
<td>0.265</td>
<td>-0.013</td>
<td>0.079</td>
<td>0.160</td>
<td>-0.287</td>
</tr>
<tr>
<td>Pub_Sovereign</td>
<td>0.106</td>
<td>0.837</td>
<td>-0.531</td>
<td>-0.005</td>
<td>-0.034</td>
</tr>
<tr>
<td>Util_Electric</td>
<td>0.338</td>
<td>0.027</td>
<td>0.097</td>
<td>0.156</td>
<td>0.303</td>
</tr>
</tbody>
</table>
procedure could be based on considering only the last quarter of the year 2010 since volatility rose to higher levels in all the sectors. The standard deviation computed on the entire year gives a flatter measure than using only the more volatile period. We decided to standardize the 2011-2012 dataset using the volatile period and compare the results with the A rated group:

Table 1.7: Rated A subsectors : Covariance matrix using the last quarter of 2010 standard deviation.

<table>
<thead>
<tr>
<th></th>
<th>Utl_Gas</th>
<th>Utl_Transp</th>
<th>Fin_Bank</th>
<th>Fin_Water</th>
<th>Fin_Finance</th>
<th>Fin_Insurance</th>
<th>Basic_Industries</th>
<th>Indus_Healthcare</th>
<th>Indus_Manufacturing</th>
<th>Indus_Retail</th>
<th>Utl_Electric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utl_Gas</td>
<td>1.096</td>
<td>0.984</td>
<td>0.685</td>
<td>1.121</td>
<td>0.783</td>
<td>0.582</td>
<td>0.603</td>
<td>0.814</td>
<td>0.784</td>
<td>0.905</td>
<td>1.102</td>
</tr>
<tr>
<td>Utl_Transp</td>
<td>0.984</td>
<td>0.863</td>
<td>0.562</td>
<td>1.096</td>
<td>0.685</td>
<td>0.521</td>
<td>0.721</td>
<td>0.730</td>
<td>0.708</td>
<td>0.747</td>
<td>0.969</td>
</tr>
<tr>
<td>Fin_Bank</td>
<td>0.685</td>
<td>0.562</td>
<td>1.411</td>
<td>0.659</td>
<td>1.490</td>
<td>0.743</td>
<td>0.367</td>
<td>0.301</td>
<td>0.417</td>
<td>0.450</td>
<td>0.837</td>
</tr>
<tr>
<td>Fin_Water</td>
<td>1.121</td>
<td>1.066</td>
<td>0.659</td>
<td>1.156</td>
<td>0.779</td>
<td>0.594</td>
<td>0.824</td>
<td>0.834</td>
<td>0.806</td>
<td>0.853</td>
<td>1.130</td>
</tr>
<tr>
<td>Fin_Finance</td>
<td>0.763</td>
<td>0.685</td>
<td>1.490</td>
<td>0.779</td>
<td>2.468</td>
<td>0.940</td>
<td>0.419</td>
<td>0.464</td>
<td>0.480</td>
<td>0.497</td>
<td>0.909</td>
</tr>
<tr>
<td>Fin_Insurance</td>
<td>0.582</td>
<td>0.521</td>
<td>0.743</td>
<td>0.594</td>
<td>0.903</td>
<td>0.903</td>
<td>0.416</td>
<td>0.402</td>
<td>0.417</td>
<td>0.386</td>
<td>0.571</td>
</tr>
<tr>
<td>Basic_Industries</td>
<td>0.807</td>
<td>0.723</td>
<td>0.387</td>
<td>0.924</td>
<td>0.419</td>
<td>0.416</td>
<td>1.168</td>
<td>0.796</td>
<td>0.791</td>
<td>0.827</td>
<td>1.013</td>
</tr>
<tr>
<td>Indus_Healthcare</td>
<td>0.814</td>
<td>0.730</td>
<td>0.391</td>
<td>0.834</td>
<td>0.464</td>
<td>0.796</td>
<td>0.903</td>
<td>0.795</td>
<td>0.825</td>
<td>0.846</td>
<td>1.189</td>
</tr>
<tr>
<td>Indus_Manufacturing</td>
<td>0.789</td>
<td>0.709</td>
<td>0.417</td>
<td>0.806</td>
<td>0.480</td>
<td>0.417</td>
<td>0.791</td>
<td>0.785</td>
<td>0.855</td>
<td>0.847</td>
<td>1.055</td>
</tr>
<tr>
<td>Indus_Retail</td>
<td>0.832</td>
<td>0.747</td>
<td>0.450</td>
<td>0.851</td>
<td>0.497</td>
<td>0.366</td>
<td>0.837</td>
<td>0.855</td>
<td>0.847</td>
<td>1.013</td>
<td>2.073</td>
</tr>
<tr>
<td>Indus_Technology</td>
<td>0.405</td>
<td>0.364</td>
<td>0.141</td>
<td>0.414</td>
<td>0.132</td>
<td>0.235</td>
<td>0.487</td>
<td>0.426</td>
<td>0.438</td>
<td>0.441</td>
<td>0.401</td>
</tr>
<tr>
<td>Indus_Telecom</td>
<td>1.102</td>
<td>0.980</td>
<td>0.837</td>
<td>1.130</td>
<td>0.900</td>
<td>0.571</td>
<td>1.013</td>
<td>0.921</td>
<td>0.923</td>
<td>1.055</td>
<td>1.962</td>
</tr>
<tr>
<td>Pub_Sovereign</td>
<td>0.234</td>
<td>0.210</td>
<td>0.028</td>
<td>0.248</td>
<td>0.174</td>
<td>0.247</td>
<td>0.228</td>
<td>0.197</td>
<td>0.147</td>
<td>0.197</td>
<td>2.373</td>
</tr>
<tr>
<td>Utl_Electric</td>
<td>1.127</td>
<td>1.012</td>
<td>0.666</td>
<td>1.152</td>
<td>0.784</td>
<td>0.599</td>
<td>0.824</td>
<td>0.834</td>
<td>0.806</td>
<td>0.855</td>
<td>2.159</td>
</tr>
</tbody>
</table>

Table 1.8: Eigenvalues associated to the A rated covariance matrix using the last quarter of 2010 for the standard deviation used for the hybrid data

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>Explained Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.193</td>
<td>55%</td>
</tr>
<tr>
<td>2.700</td>
<td>16%</td>
</tr>
<tr>
<td>2.223</td>
<td>11%</td>
</tr>
<tr>
<td>0.807</td>
<td>4%</td>
</tr>
<tr>
<td>0.690</td>
<td>3%</td>
</tr>
<tr>
<td>0.513</td>
<td>2%</td>
</tr>
<tr>
<td>0.447</td>
<td>2%</td>
</tr>
<tr>
<td>0.340</td>
<td>2%</td>
</tr>
<tr>
<td>0.297</td>
<td>2%</td>
</tr>
<tr>
<td>0.226</td>
<td>1%</td>
</tr>
<tr>
<td>0.180</td>
<td>1%</td>
</tr>
<tr>
<td>0.008</td>
<td>1%</td>
</tr>
<tr>
<td>0.000</td>
<td>1%</td>
</tr>
<tr>
<td>0.000</td>
<td>1%</td>
</tr>
</tbody>
</table>

Now to explain at least 90 per cent of the total variability we need six factors, one more than the factors needed when we use the entire year 2010 for the estimation. The aim of PCA is to reduce the number of factors through which we can reproduce most of the variability originally observed in the market. At a first glance it seems that we are going on the wrong direction. If we concentrate only on the number of PCs yes, but if we look behind the second method gives more information. Indeed, by looking at the second and third eigenvectors we observe that now each PC, excluding the first one which gives almost equal weight, is dominated by one specific sector.
The second PC explains 14 per cent of the total variability and gives more weight to the finance sector. In particular it seems strongly influenced by the Financial Bank and Financial Finance subsectors. The third PC explains 12 per cent of the total variability and is highly dependent on the Public Sovereign subsector. The next two PCs explain each 4 per cent of the total variability but are driven by different sectors. The fourth PC influenced mostly by the Industrial sectors gives a quite big weight to the Industrial Technology subsector and negative weight to the Industrial Telecommunication. The fifth PC gives negative weight to the Industrial and Financial sectors while the dependence from the Utility seems strongly positive. Changing the time frame used for the estimation of the standard deviation we obtain different results.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Eigvec1</th>
<th>Eigvec2</th>
<th>Eigvec3</th>
<th>Eigvec4</th>
<th>Eigvec5</th>
<th>Eigvec6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Util_Gas</td>
<td>0.310</td>
<td>0.079</td>
<td>0.027</td>
<td>0.006</td>
<td>0.357</td>
<td>0.037</td>
</tr>
<tr>
<td>Util_Transp</td>
<td>0.278</td>
<td>0.071</td>
<td>0.024</td>
<td>0.005</td>
<td>0.320</td>
<td>0.033</td>
</tr>
<tr>
<td>Fin_Bank</td>
<td>0.244</td>
<td>-0.459</td>
<td>0.004</td>
<td>-0.034</td>
<td>-0.094</td>
<td>0.179</td>
</tr>
<tr>
<td>Util_Water</td>
<td>0.317</td>
<td>0.083</td>
<td>0.024</td>
<td>0.001</td>
<td>0.367</td>
<td>0.040</td>
</tr>
<tr>
<td>Fin_Finance</td>
<td>0.302</td>
<td>-0.706</td>
<td>-0.098</td>
<td>0.066</td>
<td>-0.197</td>
<td>-0.285</td>
</tr>
<tr>
<td>Fin_Insurance</td>
<td>0.202</td>
<td>-0.221</td>
<td>-0.094</td>
<td>0.320</td>
<td>0.172</td>
<td>0.187</td>
</tr>
<tr>
<td>Basic_Industries</td>
<td>0.262</td>
<td>0.224</td>
<td>0.042</td>
<td>0.093</td>
<td>-0.280</td>
<td>-0.057</td>
</tr>
<tr>
<td>Indus_Healthcare</td>
<td>0.255</td>
<td>0.187</td>
<td>0.046</td>
<td>0.079</td>
<td>-0.067</td>
<td>-0.304</td>
</tr>
<tr>
<td>Indus_Manufacturing</td>
<td>0.255</td>
<td>0.173</td>
<td>0.070</td>
<td>0.090</td>
<td>-0.163</td>
<td>-0.299</td>
</tr>
<tr>
<td>Indus_Retail</td>
<td>0.273</td>
<td>0.200</td>
<td>0.062</td>
<td>-0.063</td>
<td>-0.272</td>
<td>-0.450</td>
</tr>
<tr>
<td>Indus_Finance</td>
<td>0.132</td>
<td>0.173</td>
<td>0.041</td>
<td>0.708</td>
<td>-0.336</td>
<td>0.432</td>
</tr>
<tr>
<td>Indus_Telecom</td>
<td>0.368</td>
<td>0.085</td>
<td>0.098</td>
<td>-0.600</td>
<td>-0.354</td>
<td>0.521</td>
</tr>
<tr>
<td>Pub_Sovereign</td>
<td>0.085</td>
<td>0.158</td>
<td>-0.977</td>
<td>-0.058</td>
<td>-0.053</td>
<td>0.019</td>
</tr>
<tr>
<td>Util_Electric</td>
<td>0.319</td>
<td>0.081</td>
<td>0.028</td>
<td>0.006</td>
<td>0.367</td>
<td>0.038</td>
</tr>
</tbody>
</table>

Table 1.9: Eigenvectors of the A rated sectors obtained with the hybrid data whose base standard deviation is computed using the last quarter of 2010

We perform the same analysis for the BBB rated issuers. We use the hybrid approach based on the standard deviation estimated during the last quarter of 2010. By looking at the eigenvectors, we observe that now we cannot claim that the PCs give a higher weight to a specific sector i.e each new factor is now a weighted average of the different sectors. We now need eight factors to explain 90 per cent of the total variability. The variance explained by each PC is not very large. Unfortunately there is not a unique recipe that tells us how long we must go back in the estimation period and how long these estimates can be used. In turbulent periods, volatility changes rapidly so the covariances of the hybrid data will probably be quite different from one. On the other side if the market is quiet volatility moves slowly so the covariances obtained from the hybrid data will be near one. If we observe a severe change of the market conditions probably this is due to one or more factors in particular so the update of the basis standard deviation becomes compulsory. However, even if market conditions remain stable, the factors driving the market may change. We
1.2 Statistical Factors

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>Explained Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.721</td>
<td>48%</td>
</tr>
<tr>
<td>3.908</td>
<td>59%</td>
</tr>
<tr>
<td>3.215</td>
<td>68%</td>
</tr>
<tr>
<td>2.507</td>
<td>75%</td>
</tr>
<tr>
<td>1.696</td>
<td>80%</td>
</tr>
<tr>
<td>1.521</td>
<td>84%</td>
</tr>
<tr>
<td>1.259</td>
<td>88%</td>
</tr>
<tr>
<td>0.955</td>
<td>90%</td>
</tr>
<tr>
<td>0.850</td>
<td>93%</td>
</tr>
<tr>
<td>0.736</td>
<td>95%</td>
</tr>
<tr>
<td>0.594</td>
<td>97%</td>
</tr>
<tr>
<td>0.464</td>
<td>98%</td>
</tr>
<tr>
<td>0.417</td>
<td>99%</td>
</tr>
<tr>
<td>0.285</td>
<td>100%</td>
</tr>
<tr>
<td>0.022</td>
<td>100%</td>
</tr>
<tr>
<td>0.000</td>
<td>100%</td>
</tr>
<tr>
<td>0.000</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1.10: Eigenvalues associated to the BBB rated covariance matrix using the last quarter of 2010 for the standard deviation estimation

<table>
<thead>
<tr>
<th></th>
<th>Eigvec1</th>
<th>Eigvec2</th>
<th>Eigvec3</th>
<th>Eigvec4</th>
<th>Eigvec5</th>
<th>Eigvec6</th>
<th>Eigvec7</th>
<th>Eigvec8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility Gas_Pipelines</td>
<td>0.310</td>
<td>0.233</td>
<td>0.124</td>
<td>0.141</td>
<td>0.028</td>
<td>0.016</td>
<td>0.037</td>
<td>0.031</td>
</tr>
<tr>
<td>Utility Reg_Transp</td>
<td>0.445</td>
<td>0.334</td>
<td>0.178</td>
<td>0.203</td>
<td>0.041</td>
<td>0.023</td>
<td>0.054</td>
<td>0.045</td>
</tr>
<tr>
<td>Utility Water</td>
<td>0.298</td>
<td>0.225</td>
<td>0.118</td>
<td>0.130</td>
<td>0.031</td>
<td>0.001</td>
<td>0.040</td>
<td>0.030</td>
</tr>
<tr>
<td>Utility Electric</td>
<td>0.300</td>
<td>0.225</td>
<td>0.120</td>
<td>0.136</td>
<td>0.027</td>
<td>0.015</td>
<td>0.036</td>
<td>0.030</td>
</tr>
<tr>
<td>Financial Bank</td>
<td>0.182</td>
<td>0.004</td>
<td>0.126</td>
<td>-0.193</td>
<td>-0.163</td>
<td>0.093</td>
<td>0.087</td>
<td>0.095</td>
</tr>
<tr>
<td>Financial Finance</td>
<td>0.252</td>
<td>-0.175</td>
<td>0.177</td>
<td>-0.364</td>
<td>-0.664</td>
<td>0.367</td>
<td>-0.045</td>
<td>-0.091</td>
</tr>
<tr>
<td>Financial Insurance</td>
<td>0.340</td>
<td>-0.272</td>
<td>0.159</td>
<td>-0.620</td>
<td>0.354</td>
<td>-0.470</td>
<td>0.137</td>
<td>-0.012</td>
</tr>
<tr>
<td>Industrial Basic_Industries</td>
<td>0.168</td>
<td>-0.167</td>
<td>0.004</td>
<td>-0.004</td>
<td>-0.037</td>
<td>0.255</td>
<td>-0.014</td>
<td>-0.054</td>
</tr>
<tr>
<td>Industrial Consumer_Products</td>
<td>0.147</td>
<td>-0.244</td>
<td>-0.159</td>
<td>0.166</td>
<td>0.133</td>
<td>0.107</td>
<td>-0.052</td>
<td>-0.032</td>
</tr>
<tr>
<td>Industrial Energy</td>
<td>0.232</td>
<td>0.075</td>
<td>0.030</td>
<td>0.116</td>
<td>0.063</td>
<td>-0.189</td>
<td>-0.317</td>
<td>-0.415</td>
</tr>
<tr>
<td>Industrial Manufacturing</td>
<td>0.248</td>
<td>-0.235</td>
<td>-0.063</td>
<td>0.042</td>
<td>0.053</td>
<td>-0.010</td>
<td>-0.233</td>
<td>0.251</td>
</tr>
<tr>
<td>Industrial Media</td>
<td>0.125</td>
<td>-0.250</td>
<td>-0.138</td>
<td>0.138</td>
<td>0.131</td>
<td>0.061</td>
<td>-0.004</td>
<td>-0.290</td>
</tr>
<tr>
<td>Industrial Property</td>
<td>0.142</td>
<td>-0.268</td>
<td>-0.205</td>
<td>0.368</td>
<td>-0.531</td>
<td>-0.595</td>
<td>0.268</td>
<td>-0.015</td>
</tr>
<tr>
<td>Industrial Retail</td>
<td>0.234</td>
<td>-0.379</td>
<td>-0.193</td>
<td>0.180</td>
<td>0.137</td>
<td>0.176</td>
<td>-0.335</td>
<td>0.460</td>
</tr>
<tr>
<td>Industrial Technology</td>
<td>0.109</td>
<td>-0.204</td>
<td>-0.108</td>
<td>0.076</td>
<td>0.137</td>
<td>0.236</td>
<td>0.069</td>
<td>-0.651</td>
</tr>
<tr>
<td>Industrial Telecom</td>
<td>0.194</td>
<td>0.387</td>
<td>-0.832</td>
<td>-0.330</td>
<td>-0.076</td>
<td>0.004</td>
<td>-0.057</td>
<td>-0.006</td>
</tr>
<tr>
<td>Industrial Transportation</td>
<td>0.097</td>
<td>-0.103</td>
<td>-0.141</td>
<td>0.067</td>
<td>0.183</td>
<td>0.280</td>
<td>0.784</td>
<td>0.140</td>
</tr>
</tbody>
</table>

Table 1.11: The first 8 eigenvectors associated to the covariance matrix of the BBB rated issuers
face the same problem as in curve construction or market timing problems: it is not a science but a matter of art and making the right decision is left to the sensitivity of the subject based on its past experience.
1.2.2 Independent Component Analysis

Independent component analysis (ICA) (see Hyvarinen [1999]) is a mathematical procedure for revealing hidden factors driving the realizations of the observed variables. As in the PCA case, we have the multivariate data matrix $X_{p \times T}$ and we want to find the independent components that are the rows of matrix $S_{p \times T}$. Again we suppose that the observed variables are a linear transformation of the original factors through a matrix $A_{p \times p}$, i.e.:

$$X = AS$$  \hspace{1cm} (1.73)

Except for the linearity all the other assumptions presented in PCA are relaxed. In this model the original sources $S$, called independent components (ICs), are mixed with matrix $A$ giving the observations in $X$.

$$S \rightarrow A \rightarrow X \rightarrow W \rightarrow Y$$

In fact, we do not find directly matrix $A$ and the unobservable variables in $S$ but we look for a demixing matrix $W_{p \times p}$ from which we obtain $Y_{p \times T}$ whose rows are the estimates for the unobservable sources.

$$Y = WX$$  \hspace{1cm} (1.74)

Matrix $W$ is obtained using any ICA algorithm which performs an optimization procedure whose objective function depends on a measure of independence of the ICs. Variance maximization is now substituted with independence for finding the hidden factors.

In literature we can find different approaches and related algorithms for the application of this model. ICA can be derived in various ways, for example by maximization of non gaussianity, maximum likelihood estimation or minimization of mutual information. Although there are some recent papers on parametric approaches (see for example Lu et al. [2009] for an example on parametric ICA for modeling financial time series), most of the literature is about non-parametric ICA.

In the general definition, each vector $Y_i$ represents an independent component that is actually an estimate of the i-th source $S_i$. There is an ambiguity in the ICA method in the sense that an estimated signal $Y_i$ determines up to a multiple the variance of the source $S_i$. That is, there exist infinite values for $\alpha_i$ such that this decomposition holds:

$$X_i = \tilde{A}_i \tilde{S}_i$$  \hspace{1cm} (1.75)

where $\tilde{A}_i$ is obtained by dividing with $\alpha_i$ the corresponding elements in matrix $A$ while we must multiply the elements in $S$ with $\alpha_i$ to obtain the elements in $\tilde{S}_i$. It is possible to choose $\alpha_i$ in that way that we have a unit variance signal but the sign ambiguity still remains.

Another problem is the fact that it is not possible to uniquely determine the order of the
independent components. The reason is that given a permutation matrix $P$ and its inverse the model can be decomposed in:

$$X = AP^{-1}PS$$  \tag{1.76}

Any algorithm chosen for factor identification can suppose that the original independent variables are $PS$ instead of $S$ with estimated mixing matrix being $AP^{-1}$.

It is reasonable to affirm that the observed variables are more gaussian than the distribution of some factors. Indeed, if we think at the observed variables as a linear transformation of more than two r.v’s, whose distribution is unknown for us, from the Central Limit Theorem we have that the univariate distributions can be highly non gaussian.

As we already mentioned one way of estimating the ICs we can maximize a measure of non-gaussianity for each component. FastICA is an algorithm (see Hyvarinen [1999]) based on a fixed-point iteration scheme maximizing non-Gaussianity. Since we will use it in many application let us recall the main passages in its construction.

The ICs are assumed to be non-gaussian. A measure of non-gaussianity is the negentropy $J(Y)$ defined as a difference of two quantities:

$$J(y) = H(\hat{y}) - H(y)$$  \tag{1.77}

where $\hat{y}$ is a random vector coming from a gaussian distribution while the random vector $y$ can be generated using any distribution. Observe that $H(y)$ is a measure of entropy for the random vector $y$:

$$H(y) = \int -p(y) \log p(y) dy$$  \tag{1.78}

The negentropy can be seen as a measure of the distance from the perfect disorder of the gaussian world. It is possible to use an approximation for the negentropy which facilitates computations:

$$J(y) \approx \left[E \{G(\hat{y})\} - E \{G(y)\}\right]^2$$  \tag{1.79}

for particular choices of function G. In particular, we consider choose:

$$G(Y) = e^{-y^2/2}$$  \tag{1.80}

Before applying the ICA algorithm the original dataset must be preprocessed, i.e centered and whitened. Consider a dataset composed by two r.v’s uniformly distributed on $[-30, 70]$. Each original variable is centered meaning that we subtract the vector mean to the dataset. The centered data are then whitened, i.e the transformed dataset $\hat{X}$ is such that:

$$E[\hat{X}\hat{X}'] = I$$  \tag{1.81}
1.2 Statistical Factors

Two variables generated uniformly from $-30$ to $70$

(a) True signals

(b) Observed variables

Figure 1.6: In the left figure we plot the two original signals $S_1$ and $S_2$ that are uniformly distributed on $[-30, 70]$. The observed quantities $X_1$ and $X_2$ are linear transformations of the signals $S_1$ and $S_2$. The plot in the right refers to $X_1 = 0.3 \ast S_1 - 0.6 \ast S_2$ and $X_2 = 0.72 \ast S_1 + 0.17 \ast S_2$.

A common way to transform the data is by using the matrix obtained from the eigenvalue and eigenvectors decomposition of $\text{Cov}(X) = E[XX']$. If we put the eigenvectors in matrix $E$ and $D$ is a diagonal matrix whose elements are the corresponding eigenvalues $D = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$ then

$$X = EDE'$$

(1.82)

The performed transformation:

$$\hat{X} = ED^{-\frac{1}{2}}E'X$$

(1.83)

yields the whitened data. Since:

$$\hat{X} = ED^{-\frac{1}{2}}E'AS = \hat{A}S$$

(1.84)

the new mixing matrix $\hat{A}$ is orthogonal because:

$$E[\hat{X}\hat{X}'] = \hat{A}E[SS']\hat{A}' = \hat{A}\hat{A}' = I$$

(1.85)

FastICA is an iterative algorithm and it finds the direction for the weight vector $w$ maximizing the non-Gaussianity of the projection $w'$ for the data $\hat{x}$. The function $g(\cdot)$ is the derivative of the function $G$ defined in equation 1.80. The main steps of the algorithm to obtain one IC are:

- 1 Start with a random weight vector $w$
- 2 Define $w^+ \leftarrow E\left\{\hat{x}g(w' \hat{x})\right\} - E\left\{g'(w' \hat{x})\right\}w$
36

Linear factor models

Figure 1.7: Preprocessing the dataset means to center the data and then whiten them. The last term means that the covariance matrix of the transformed dataset is diagonal.

• 3 Divide $w \leftarrow w^+ / \|w^+\|

• 4 Compare $w$ with the previous one. If the difference is bigger than a fixed value $\epsilon$, it means that there is no convergence so the update the value as in 2.

In multivariate analysis the identification of the main factors that generate the variables that we observe is an important step. Under the stationary hypothesis it gives the possibility to use the past information to predict the future. Here we apply the ICA analysis and try to identify the factors driving the returns on stock markets. We consider the ten GICS sectors as factors for the VFIAX index and apply the FastICA algorithm to find 10 independent components. In 1.2.2 differences between the components can be immediately noticed.

Madan [2006a] suggests to consider as independent factors only those that differ substantially from the normal distribution while the remaining components are grouped together to form the idiosyncratic component. The difficulty in the component interpretation explains the limited use of this instrument in finance although from a mathematical point of view it is very interesting.
1.2 Statistical Factors

Figure 1.8: We consider the ten GICS sectors as factors for the VFIAX index and apply the FastICA algorithm to find 10 ICs that generate our portfolio return. In the figure we show the time series of the ICs.
1.2.3 PCA vs ICA

PCA and ICA yield both linear transformations of the variables but they are based on quite different hypothesis and yield components that have different statistical properties. In fact we have that PCs are uncorrelated and gaussian while ICs are independent and supposed non gaussian. Only Second-order statistics are needed in the PCA analysis since derivation is based on variance maximization while higher-order statistics are considered for non-gaussianity identification in the ICA analysis. Although explained variance is a criterion for ordering the principal components, no single criterion exists for ordering the independent components. This is the reason why in literature alternative methods that at least produce a sparse mixing matrix are considered (see Zhang et al. [2009]). The exact amplitude and sign of the independent components cannot be determined.

Since in the ICA analysis we try to maximize the non-gaussianity of the components, the distribution of the components are fitted better using distributions with heavier tails. In figure 1.2.3 we fit both the Normal and the Variance Gamma (see Madan and Seneta [1990b]) distribution to the first three ICs and PCs of the ten GICS. In Chapter 4 we will discuss the ICA analysis in a parametric context in order to identify the differences among the components.

Figure 1.9: The pictures show the first three ICs and PCs of the ten GICS. The Normal and Variance distribution were fitted to each of them to evidence the differences. In particular, the first IC has a peaked distribution and the normal assumption doesn't hold.
Chapter 2

Risk Measures: VaR, CVaR and Expectiles

Risk analysis is most commonly considered in the context of a particular risk measure leading to concepts such as mean-variance optimization or diversification. The first risk measure considered was the standard deviation and one of the most well-known work based on this quantity is Markowitz [1952] which considers a risk-return optimization problem. In the last twenty years, the attention given to the correct definition of risk increased. The most used risk measure is the ValueAtRisk (VaR henceforth) RiskMetrics [1996].

Despite its popularity, VaR has two main disadvantages. The first is that it is not a coherent measure in the sense of Artzner et al. [1999] meaning that the diversification principle is not ensured. The second drawback is due to the fact that it does not fully take account for the tail risk. Indeed, being a point of the distribution, it does not consider the amount of loss below its value. The international regulators are thinking about moving to another risk measure for the capital requirements and, although the Expected Shortfall (ES henceforth) is a coherent risk measure, it does not seem to be the best alternative. Indeed its backtesting is more difficult to perform and, moreover, methodologies used for the estimation of the ES are not directly comparable.

In fact risk measures for which direct comparison is possible satisfy the elicitability property (see Gneiting [2011]). A statistical functional is elicitabile if it is a solution of a minimization of a suitable score function. In this context a particular attention was recently given to another statistical quantity: the expectile. Introduced in Newey and Powell [1987], it is the result of an Asymmetric Least Square regression. The expectile based risk measure seems to have nice properties.

In particular, in this chapter we review the definition of a risk measure and its mathematical properties. Then we focus on the VaR and the ES before introducing the elicitability property. An empirical investigation is performed on equity and credit risk market data for the extraction of implicit values for the weight parameter in the expectile definition.
2.1 Risk measures: mathematical properties and conformity to Basel III

Let $\chi$ be the family of the random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$ describing the profit and loss (or return) of a given portfolio (or asset). The risk measure associated to $X \in \chi$ is defined as follows:

**Definition 3.** A *risk measure* is a map $\rho : \chi \rightarrow \mathbb{R}$ meaning that $\rho(X) \in \mathbb{R}$.

From an economic point of view $\rho(X)$ is referred to capital requirements. Depending on the sign it can have two interpretations:

- **Positive**: $\rho(X)$ is the additional amount of money to give as a margin so that the new position becomes acceptable.
- **Negative**: $\rho(X)$ is the maximum amount of money that can be withdrawn in order to maintain the acceptability of the position.

Portfolio managers use risk measures to quantify the riskiness of their position for a fixed horizon of time $T$ and usually a confidence level $\alpha$.

From a theoretical point of view, a very appealing class of risk measures are the coherent ones. Following the axiomatic approach introduced in Artzner et al. [1999], $\rho(X)$ is a coherent risk measure if it satisfies the following properties:

- **Translation Invariance** For all $\lambda \in \mathbb{R}$ and for all $X \in \chi$ we have $\rho(X - \lambda) = \rho(X) - \lambda$.
- **Monotonicity** For all $X, Y \in \chi$ such that $X \leq Y$ we have $\rho(X) \leq \rho(Y)$.
- **Positive Homogeneity** For all $\lambda \geq 0$ and for all $X \in \chi$ we have $\rho(\lambda X) = \lambda \rho(X)$.
- **Subadditivity** For all $X, Y \in \chi$ we have $\rho(X + Y) \leq \rho(X) + \rho(Y)$

2.1.1 Value At Risk and Expected Shortfall

Let $X$ be a r.v defined on the probability space $(\Omega, \mathcal{F}, P)$. $q \in \mathbb{R}$ is an $\alpha$-quantile with $\alpha \in (0,1)$ if:

$$P(X < q) \leq \alpha \leq P(X \leq q) \quad (2.1)$$

The biggest $\alpha$-quantile of $X$ is:

$$q^+_\alpha(X) = \inf \{ x \in \mathbb{R}; F_X(x) > \alpha \} \quad (2.2)$$

while the lowest $\alpha$-quantile of $X$ is defined as:

$$q^-_\alpha(X) = \inf \{ x \in \mathbb{R}; F_X(x) \geq \alpha \} \quad (2.3)$$
2.2 Elicitability and Expectiles

For continuous rv’s we have that $q^+_\alpha(X) = q^-_\alpha(X) = q_\alpha(X)$ though we can use both for defining VaR. In particular, we define the ValueAtRisk with $\alpha$ level of confidence as:

$$VaR_\alpha(X) = -q_\alpha(X)$$  \hspace{1cm} (2.4)

The notion of Value at Risk was first introduced in a technical report of J.P. Morgan in 1996. Its great success is due to the fact that it entered in the requirements of the Basel committee although it does not satisfy the subadditivity property. Another crucial risk measure is the Expected Shortfall Acerbi [2002] defined as:

$$ES_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha VaR_\beta(X) \, d\beta$$  \hspace{1cm} (2.5)

In the particular case where the distribution of $X$ is continuous, Expected Shortfall coincides with the Conditional Value at Risk (CVaR) introduced in Rockafellar and Uryasev [2000]. For $\alpha \in (0, 1)$ the CVaR at confidence level $\alpha$ is defined by:

$$CVaR_\alpha(X) = \int_{-\infty}^{+\infty} \alpha F^{\alpha}_X(z)$$  \hspace{1cm} (2.6)

where

$$F^{\alpha}_X(z) = \begin{cases} 0 & z < VaR_{\alpha,t}(X) \\ \frac{F_X(z) - \alpha}{1-\alpha} & z \geq VaR_{\alpha,t}(X) \end{cases}$$  \hspace{1cm} (2.7)

In this thesis we will use the Expected Shortfall for continuous and empirical distributions. The interpretation will be that of a conditional mean based on Tail Conditional(TCE) definition:

$$TCE_\alpha(X) = E[-X | X \leq -VaR_\alpha(X)]$$  \hspace{1cm} (2.8)

That is, for us CVaR, ES and TCE will refer to the same quantity. As we have seen, the two considered risk measures have both Pros and Cons. In particular VaR is not subadditive, its estimators are stable and backtesting is straightforward. On the other side we have that ES is coherent, is sensitive to extreme values but its estimators are less stable than the VaR estimators as shown in Yamai and Yoshida [2005]. Backtesting for VaR it is simple and a lot of work has been done (see Christoffersen [2010] for a general overview) while for ES is not the same since it is not an elicitable functional as explained in the next section.

2.2 Elicitability and Expectiles

2.2.1 Elicitability: forecast and comparison of estimation procedures

Savage [1971] introduced the notion of elicitation for statistical functionals. In this section we follow the presentation given in Gneiting [2011].
Let $F$ be the distribution function of a r.v $Y$. Suppose we make a probabilistic forecast $F$ for the events concerning $Y$. The future observations $Y$ are the realizations of a random variable and we need a scoring method to compare the predictive performance of forecasters. There are two types of forecasts: probabilistic and point forecasts. The first considers the problem of finding any probability measure $F \in \mathcal{F}$ related to $Y$ while in the second case we seek for a forecast of the distribution function $F$. For example the forecast of the mean, the mode and the quantile associated to a distribution enter in the last type of problems. We will focus on point forecast results but before explaining how all this is related to the considered risk measures let us recall some important results known in literature. Consider a family $\mathcal{F}$ of probability measures on $(I, \mathcal{B}(I))$.

**Definition 4.** Statistical functional or functional
A functional is a map $T : \mathcal{F} \rightarrow P(I)$ where $P(I)$ is the power set of the interval $I$.

**Definition 5.** Consistent Scoring function
A scoring function $S$ is said to be consistent for a specified functional $T$ defined on the the class of probability measures $\mathcal{F}$ if are satisfied both conditions:

- $E_F S(x, Y)$ exists and it is finite $\forall x \in I$ and $\forall F \in \mathcal{F}$.
- $E_F S(t, Y) \leq E_F S(x, Y)$ for all $x \in I$, $\forall F \in \mathcal{F}$ and $\forall t \in T(F)$.

where the scoring function is usually an error measure.

**Definition 6.** Strictly Consistent Scoring function
A scoring function $S$ is said to be strictly consistent for a specified functional $T$ defined on the the class of probability measures $\mathcal{F}$ if:

- $E_F S(t, Y) = E_F S(x, Y)$ $\forall x \in I$, $\forall F \in \mathcal{F}$ and $\forall t \in T(F)$.

The equality of the expectations ensures that $t \in T(F)$ implies $x \in T(F)$.

**Definition 7.** Elicitable functional
A functional $T$ is elicitable if exists a scoring function $S$ that is strictly consistent for it.

Recently, Bellini and Bigozzi [2013] suggest a slightly more restrictive definition more suited for financial applications by adding the convexity requirement for the scoring function. The following result is useful in the case we want to show that a given functional is not elicitable: Osband [1985] shows that the level sets of an elicitable functional are convex. This property requires that if $F_0 \in \mathcal{F}$, $F_1 \in \mathcal{F}$ and for $p \in (0, 1)$ $F_p = (1 - p)F_0 + pF_1 \in \mathcal{F}$ the conditions $t \in T(F_0)$ and $t \in T(F_1)$ imply that $t \in (F_p)$. In fact, if $T$ is elicitable, meaning that it exists a strictly consistent scoring function $S$, the linearity property of expectations and the elicitability definition, for any $x \in I$ suggest that:

$$E_{F_p} S(t, Y) = pE_{F_0} S(t, Y) + (1 - p)E_{F_1} S(t, Y) \leq pE_{F_0} S(x, Y) + (1 - p)E_{F_1} S(x, Y) = E_{F_p} S(x, Y)$$

(2.9)
2.2 Elicitability and Expectiles

ensuring that \( t \in (F_p) \).

For a given functional \( T \), in order to get precise proofs, we need some smoothness conditions on the scoring function \( S \). A common choice is the class of Generalized Piecewise linear functions of order \( \alpha \):

**Definition 8.** Generalized Piecewise linear (GPL) of order \( \alpha \)

A function \( S_g : I \times I \to [0, +\infty) \) of the form:

\[
S_g(x, y) = (1 - \alpha)[g(x) - g(y)]1_{y \leq x} + \alpha[g(x) - g(y)]1_{y > x}
\]

with \( g : I \to \mathbb{R} \) being a non decreasing function.

In the particular case where \( g \) is a linear function, \( S_g \) coincides with the asymmetric piecewise loss function used in quantile regression. The function \( S_g \), for \( g \) linear, is a strictly consistent scoring function for the functional \( q_\alpha(Y) = F_\alpha^{-1}(\alpha) \). It is simple now to understand why VaR is an elicitable functional. This property ensures that exists competing alternatives that are directly comparable which to measure the error we make by assuming a particular distribution for the data considered. The sufficiency condition for elicitation doesn’t depend on the smoothness conditions for the GPL function. The CVaR functional is not elicitable since the level set of the functionals are not convex. In Gneiting [2011] a counterexample for the class \( \mathcal{F} \) containing measures with finite support. In particular suppose \( a, b, c, d \in I \) with \( a < b < c < 0.5(b + d) \) and define two probability measures:

\[
F_1 = \alpha \delta_a + 0.5(1 - \alpha)[\delta_b + \delta_d] F_2 = \alpha \delta_c + (1 - \alpha)\delta_{(b+d)/2}
\]

It is easy to observe that for \( \alpha \leq 1/3 \):

\[
CVaR_\alpha(F_1) = CVaR_\alpha(F_1) = 0.5(b + d)
\]

CVaR doesn’t have convex level sets since:

\[
CVaR_\alpha(0.5(F_1 + F_2)) = 0.25(b + c + 2d) > 0.5(b + d)
\]

The CVaR associated to the convex combination of the probability measures is higher than the weighted sum of the single CVaRs since \( b < d \).

The BIS Consultative Document BIS [2012] poses the question of what are the likely operational constraints with moving from VaR to ES, including any challenges in delivering robust backtesting, and how might these be best overcome. The absence of elicitation property probably will justify the embargo for the use of the ES for capital requirements.
2.2.2 Expectiles: coherent and elicitable

For $\tau \in (0, 1)$, the $\tau$-expectile of a r.v $X$ with finite mean is defined as:

$$e_\tau(X) = \arg\min_{m \in \mathbb{R}} E[\tau \max(m - X, 0) + (1 - \tau) \max(X - m, 0)]$$  \hspace{1cm} (2.14)

It can easily be shown that $e_\tau(X)$ satisfies the relation:

$$e_\tau(X) = \frac{\tau E[X I_{\{X \leq e_\tau(X)\}}] + (1 - \tau) E[X I_{\{X > e_\tau(X)\}}]}{\tau P[X \leq e_\tau(X)] + (1 - \tau) P[X > e_\tau(X)]}$$  \hspace{1cm} (2.15)

The risk measure associated to the expectile is:

$$\rho(X) = -e_\tau(X)$$  \hspace{1cm} (2.16)

The main characteristics that make it a candidate as a risk measure are the fact that it is homogenous and coherent for particular choice of $\tau$ (see Bellini et al. [2013]). Under the condition of finiteness for the second moment of $F$, the $\tau$-expectile is the optimal point forecast if the scoring function is:

$$S_\tau(x, y) = |1_{x \geq y} - \tau| (x - y)^2$$  \hspace{1cm} (2.17)

Bellini and Bigozzi [2013] showed that expectiles are the only coherent risk measures to be elicitable it exists a clear methodology to perform backtesting. The alternative procedures can be ranked by using the scoring function since if the scoring function $S(x, y)$ is consistent for the considered risk measure $\rho(X)$, then the averaged realized score:

$$\frac{1}{n} = \sum_{i=1}^{n} S(\rho_i, x_i)$$  \hspace{1cm} (2.18)

should be small.

The main difficulty in the use of the expectile based risk measure is the choice of the parameter $\tau$. The arbitrary choice for the parameter value can be faced by recalling that the expectile parameter $\tau$ and quantiles are connected since (see Jones [1994]):

$$\tau = \frac{\alpha q_\alpha - \int_{-\infty}^{q_\alpha} x dF(x)}{E[X] - 2 \int_{-\infty}^{q_\alpha} x dF(x) - (1 - 2\alpha) q_\alpha}$$  \hspace{1cm} (2.19)

It is though possible, once fixed $\alpha$, to obtain a value for $\tau$ for the given distribution of the data considered.

In order to have a better intuition of the parameters appearing in the quantile and expectile definitions, we perform a naive backtesting analysis. The S&P500 returns for the time-period
2.2 Elicitability and Expectiles

12/09/2009-12/09/2012 are modeled as:

\[ r_t = \sigma_t \epsilon_t \]  \hspace{1cm} (2.20)

and following Bollerslev [1986] we suppose \( \sigma_t \sim \text{GARCH}(1,1) \), that is:

\[ \sigma_t^2 = \omega + \beta_1 r_{t-1}^2 + \beta_2 \sigma_{t-1}^2 \]  \hspace{1cm} (2.21)

where \( \beta_1 \) and \( \beta_2 \) are the model parameters. From the GARCH(1,1) fitting, we get the sequence \( \hat{\sigma}_t \) \( t = 1, \ldots, n \). Consider:

\[ \hat{\epsilon}_t = \frac{r_t}{\hat{\sigma}_t} \]  \hspace{1cm} (2.22)

and compute the expectile of this sequence for fixed \( \tau \) as the solution of an Asymmetric Least Square Regression:

\[ e_\tau = \arg\min_m \frac{1}{n} \sum_{i=1}^{n} [\tau (\hat{\epsilon}_t - m)^2 I_{\hat{\epsilon}_t \geq m} + (1 - \tau)(m - \hat{\epsilon}_t)^2 I_{\hat{\epsilon}_t < m}] \]  \hspace{1cm} (2.23)

As a first attempt, we fix two values for \( \tau \) respectively 0.05 and 0.01. The expectiles obtained daily for these two parameter values are plotted together with the \( \alpha \)-quantile for \( \alpha = 0.05 \). The

Figure 2.1: The S&P500 returns for the time-period 12/09/2009-12/09/2012 is modeled using a GARCH(1,1). The daily forecast for the quantities \( e_{0.05}, e_{0.01} \) and \( q_{0.05} \) is then compared with the actual returns in order to compute the exceedance rate.
time frame considered for the estimation contains 200 observations. Daily we check whether the actual return observed is less than the expectile or quantile estimated in the previous step. The returns for which this happen form the sequence of exceedances or hittings. The exceedance rate is defined as the fraction of the exceedances over the total number of observations. We compute $e_{0.01}$ and obtain an exceedance rate of 4.36% while for $e_{0.05}$ the exceedance rate is 12.55%. The exceedance rate for $VaR_{0.05}$ is 5.45%. From this simple exercise we confirm the fact that while $\alpha$ can be interpreted as the probability of observing returns lower than the corresponding $VaR_\alpha$ value, for $\tau$ this is not possible. In fact, it has been interpreted as the prudential parameter in Kuan et al. [2009b] since it the weight given to the tails in the ALS regression problem. Now we perform an empirical analysis based on the expectile-quantile relation given in equation 2.19. Fixed $\alpha$, we use the empirical distribution to compute the value for $\tau$. First we apply our analysis to the S&P500 and to the ten GICS return distributions. Then we move on to the credit market and consider the daily z-spread variations for eleven European countries.
From the empirical distribution of the daily log returns for the S&P500 index, we get the $\tau$ for $\alpha$ values that are usually used in finance: 0.01, 0.05 and 0.10.

![Graph](image)

**Figure 2.2:** In this figure we plot the $\tau$ for different $\alpha$ values using the empirical distributions of the S&P500 index. The values were obtained from the formula that relates $\alpha$ to $\tau$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.01</th>
<th>0.05</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$</td>
<td>0.0011</td>
<td>0.0119</td>
<td>0.0459</td>
</tr>
<tr>
<td>$q_\alpha$</td>
<td>-0.0226</td>
<td>-0.0159</td>
<td>-0.0107</td>
</tr>
<tr>
<td>$e_\tau$</td>
<td>-0.0227</td>
<td>-0.0161</td>
<td>-0.0108</td>
</tr>
</tbody>
</table>

**Table 2.1:** In this table we give the $\tau$ values obtained using the corresponding empirical distribution for the S&P500 index when $\alpha$ is 0.01, 0.05 and 0.10. The corresponding empirical expectile and quantile values are given and we can see they differ quite slightly.
For the same $\alpha$ values, we repeat the analysis for the ten GICS sector indexes.

Figure 2.3: In this figure we plot $\tau$ values for different $\alpha$ using the empirical distributions of the ten GICS. Observe that for $\alpha = 0.05$ the ten sectors $\tau$-s are more concentrated than the $\tau$-s obtained for $\alpha = 0.10$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.01</th>
<th>0.05</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$</td>
<td>INFT</td>
<td>0.0003</td>
<td>0.0134</td>
</tr>
<tr>
<td></td>
<td>FINL</td>
<td>0.0010</td>
<td>0.0155</td>
</tr>
<tr>
<td></td>
<td>CONS</td>
<td>0.0008</td>
<td>0.0143</td>
</tr>
<tr>
<td></td>
<td>HLTH</td>
<td>0.0009</td>
<td>0.0118</td>
</tr>
<tr>
<td></td>
<td>ENRS</td>
<td>0.0011</td>
<td>0.0128</td>
</tr>
<tr>
<td></td>
<td>COND</td>
<td>0.0012</td>
<td>0.0158</td>
</tr>
<tr>
<td></td>
<td>INDU</td>
<td>0.0009</td>
<td>0.0143</td>
</tr>
<tr>
<td></td>
<td>MATR</td>
<td>0.0014</td>
<td>0.0118</td>
</tr>
<tr>
<td></td>
<td>UTIL</td>
<td>0.0014</td>
<td>0.0112</td>
</tr>
<tr>
<td></td>
<td>TELS</td>
<td>0.0011</td>
<td>0.0125</td>
</tr>
</tbody>
</table>

Table 2.2: In this table we give the $\tau$ value obtained using the corresponding empirical distribution for the ten GICS when $\alpha$ is 0.01, 0.05 or 0.10.

Observe that for higher $\alpha$ values, the set of implicit values for $\tau$ becomes larger. However, since the distributions of the ten GICS sector indexes are similar, the corresponding $\tau$-s are similar too.
Let us consider the dataset composed by the daily variations of the z-spreads for eleven sovereigns in the euro zone going from 20/10/2010 to 20/10/2012. With respect to the equity dataset, now there is bigger variability in the $\tau$ values for each fixed $\alpha$. Moreover, we systematically observe higher values for $\tau$ than those obtained for the same $\alpha$ values when the GICS dataset is considered.

Figure 2.4: In this figure we show $\tau$ for different $\alpha$ using the empirical distributions of eleven European sovereign z-spreads.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.01</th>
<th>0.05</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td>0.0039</td>
<td>0.0328</td>
<td>0.0831</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.0061</td>
<td>0.0306</td>
<td>0.0723</td>
</tr>
<tr>
<td>Finland</td>
<td>0.0043</td>
<td>0.0292</td>
<td>0.0697</td>
</tr>
<tr>
<td>France</td>
<td>0.0036</td>
<td>0.0307</td>
<td>0.0702</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.0024</td>
<td>0.0344</td>
<td>0.0915</td>
</tr>
<tr>
<td>Italy</td>
<td>0.0009</td>
<td>0.0322</td>
<td>0.0804</td>
</tr>
<tr>
<td>Netherl</td>
<td>0.0034</td>
<td>0.0230</td>
<td>0.0700</td>
</tr>
<tr>
<td>Portugal</td>
<td>0.0041</td>
<td>0.0179</td>
<td>0.0746</td>
</tr>
<tr>
<td>Slovakia</td>
<td>0.0034</td>
<td>0.0205</td>
<td>0.0610</td>
</tr>
<tr>
<td>Spain</td>
<td>0.0026</td>
<td>0.0295</td>
<td>0.0814</td>
</tr>
<tr>
<td>Germany</td>
<td>0.0017</td>
<td>0.0228</td>
<td>0.0570</td>
</tr>
</tbody>
</table>

Table 2.3: In this table we give the $\tau$ value obtained using the corresponding empirical distribution for the sovereign z-spread daily variations when $\alpha$ is 0.01, 0.05 or 0.10.
The difference in the results suggests that for fixed $\alpha$ values, the implicit $\tau$ depend on the shape of the dataset distribution. In fact, if we compute the kurtosis level for the $z$-spread dataset, the value is much higher than when we consider equity indexes. The results show that if we want to use an expectile based risk measure, the parameter $\tau$ should not be chosen arbitrarily but it should be dependent on some statistics of the distribution considered. From the two exercises performed, it seems that $\tau$ can be defined as an increasing function of the kurtosis level. Further investigations should be performed in order to model the relation between them.

<table>
<thead>
<tr>
<th>INFT</th>
<th>FINL</th>
<th>CONS</th>
<th>HLTH</th>
<th>ENRS</th>
<th>COND</th>
<th>INDU</th>
<th>MATR</th>
<th>UTIL</th>
<th>TELS</th>
</tr>
</thead>
</table>

Table 2.4: Kurtosis level for the empirical distribution of the daily log returns for the ten GICS indexes.

<table>
<thead>
<tr>
<th>Austria</th>
<th>Belgium</th>
<th>Finland</th>
<th>France</th>
<th>Ireland</th>
<th>Italy</th>
<th>Netherland</th>
<th>Portugal</th>
<th>Slovakia</th>
<th>Spain</th>
<th>Germany</th>
</tr>
</thead>
</table>

Table 2.5: Kurtosis level for the empirical distribution of the sovereign $z$-spread daily variation for eleven countries. The departure from the normal distribution is evident.

### 2.3 Risk decomposition

We cannot define universally the most appropriate financial risk measure. Users have to make choices of the properties of risk measures that are most important for their purposes. Let us see how the three considered risk measures can be used in risk decomposition. The homogeneity property results fundamental.

#### 2.3.1 Homogeneity and Euler decomposition

A function $f(x, y)$ is said to be homogeneous of order $n$ if:

$$f(\lambda x, \lambda y) = \lambda^n f(x, y)$$

(2.24)

For homogeneous functions holds the Euler theorem:

$$\frac{\partial f(x, y)}{\partial x} x + \frac{\partial f(x, y)}{\partial y} y = nf(x, y)$$

(2.25)

This theorem can be generalized for more variables and it is the basis for the risk contribution methodology introduced in Tasche [1999].

Consider a linear model relating portfolio returns $\pi$ with factors in $F$:

$$\pi = \beta F$$

(2.26)
Any homogeneous risk measure \( R(\pi) \) can be written as the sum:

\[
R(\pi) = \sum_{i=1}^{n} \beta_i \frac{\partial R(\pi)}{\partial \beta_i} = \sum_{i=1}^{n} RC_i
\]  

(2.27)

where the risk contribution of the \( i \)-th risk factor is:

\[
RC_i(\pi) = \beta_i \frac{\partial R(\pi)}{\partial \beta_i}
\]  

(2.28)

This result has been widely used in finance since it gives a simple rule for attributing risk (see for example Meucci [2007] and Marchioro and Borrello [2013]).

### 2.3.2 Risk measure decomposition formulas

Sharpe [2002] affirms that a mere mathematical decomposition of portfolio risk can not necessarily be seen as risk contribution. However, this mathematical decomposition of portfolio risk has a financial meaning. Consider the following risk measures:

**Value at Risk**

\[
VaR_\alpha(\pi) = -q^+_\alpha(\pi) = -\inf \{x \in \mathbb{R} | F_\pi(x) > \alpha \}
\]  

(2.29)

**Expected Shortfall (or CVaR)**

\[
ES_\alpha(\pi) = -E[\pi | \pi \leq -VaR_\alpha(\pi)]
\]  

(2.30)

**Expectile based risk measure**

\[
\rho_\tau(\pi) = -\frac{\tau E[\pi I_{\{\pi \leq e_\tau(\pi)\}}]}{\tau P[\pi \leq e_\tau(\pi)]} + \frac{(1 - \tau) E[\pi I_{\{\pi > e_\tau(\pi)\}}]}{(1 - \tau) P[\pi > e_\tau(\pi)]}
\]  

(2.31)

and the corresponding risk contributions based on the:

**Value-at-risk (see Gourieoux et al. [2000])**

\[
RC_i = -E[F_i | \pi = VaR_\alpha(\pi)] \beta_i
\]  

(2.32)

**Expected Shortfall (see Scaillet [2002])**

\[
RC_i = -E[F_i | \pi \leq -VaR_\alpha(\pi)] \beta_i
\]  

(2.33)

**Expectiles (see Tasche (2013))**

\[
RC_i = -\frac{\tau E[F_i I_{\{\pi \leq e_\tau(\pi)\}}]}{\tau P[\pi \leq e_\tau(\pi)]} + \frac{(1 - \tau) E[F_i I_{\{\pi > e_\tau(\pi)\}}]}{(1 - \tau) P[\pi > e_\tau(\pi)]} \beta_i
\]  

(2.34)

The formulas described are easy to implement. In fact, we will perform risk attribution considering the three described risk measures in the following sections.
2.3.3 Empirical results

Here, we consider the data composed by daily log returns from 11/02/2010 to 11/02/2013 on the SP500 and the ten GICS. The ES at level $\alpha = 0.05$ and the expectile for $\tau = 0.012$ were computed for the general index and for each sector. The portion of risk associated to each sector seems to remain stable going from one measure to the other.

<table>
<thead>
<tr>
<th></th>
<th>SPX</th>
<th>INFT</th>
<th>FINL</th>
<th>CONS</th>
<th>HLTH</th>
<th>ENRS</th>
<th>COND</th>
<th>INDU</th>
<th>MATR</th>
<th>UTIL</th>
<th>TELS</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES_{0.05}</td>
<td>0.0191</td>
<td>0.00037</td>
<td>0.00015</td>
<td>0.00015</td>
<td>0.00024</td>
<td>0.00022</td>
<td>0.00024</td>
<td>0.00008</td>
<td>0.00004</td>
<td>0.00003</td>
<td>2.72E-05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19.6%</td>
<td>19.3%</td>
<td>8.0%</td>
<td>8.1%</td>
<td>12.7%</td>
<td>11.7%</td>
<td>12.8%</td>
<td>4.3%</td>
<td>1.9%</td>
<td>1.5%</td>
<td>0.1%</td>
<td></td>
</tr>
<tr>
<td>e_{0.012}</td>
<td>0.0153</td>
<td>0.00028</td>
<td>0.00013</td>
<td>0.00010</td>
<td>0.00021</td>
<td>0.00017</td>
<td>0.00019</td>
<td>0.00007</td>
<td>0.00003</td>
<td>0.00002</td>
<td>2.70E-05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20.8%</td>
<td>18.6%</td>
<td>8.8%</td>
<td>6.6%</td>
<td>13.9%</td>
<td>11.2%</td>
<td>12.7%</td>
<td>4.4%</td>
<td>1.7%</td>
<td>1.0%</td>
<td>0.2%</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.6: Risk decomposition when the factors are the GICS
Chapter 3

User defined risk factors

In the previous chapters we discussed the problem of identifying the factors that can be useful in explaining the overall return/risk. As we already saw, the considered factors can have a financial interpretation or can be the result of a mathematical procedure though interpretation is not straightforward. In this chapter we face the problem of attributing risk to factors that are directly chosen by the investor and different from those that fully determine the total portfolio/asset returns.

Now we suppose that the explicative factors have already been identified. We want to quantify the influence of a financial variable that doesn’t enter in the set of risk factors considered. For example, consider a portfolio of fixed rate bonds whose risk factors are the changes in value of zero rates and credit spreads and suppose that an abrupt movement in the Oil index price is observed. Though not considered before, the exposure of the portfolio to the Oil index becomes relevant. Since this variable is not uncorrelated with the initial risk factors, its inclusion in the risk decomposition model requires some efforts in order to avoid the creation of possible collinear factors. This problem has become particularly relevant especially since the recent financial crisis has suggested that market variables are more related to each other than one could think. Menucci [2007] considers the problem in the context where the new defined factors are uncorrelated from the residuals and linear transformations of the original risk factors. We extend the analysis by attributing risk to the ICs since the conditions posed in the approach are satisfied in the ICA analysis.

However, in order to be more realistic, we study the problem in the case of non linear transformations. Menchero and Poduri [2008] propose a model that is justified by the need of attributing risk to the allocation decisions made by the portfolio manager. We consider the question from a different point of view. After the allocation decision has been made, we define some new factors uncorrelated with the user defined factors. They are the projections of the original factors on the space orthogonal to that generated from the user defined factors. In this way we are able to decompose the return vector in three uncorrelated quantities: the user defined factors, the projected...
factors and the idiosyncratic error term. The model proposed is easy to implement and each risk component has a financial interpretation.

3.1 Risk Contribution from user defined factors

In this section we discuss the problem of attributing risk to user defined risk factors following the approach introduced in Meucci [2007]. We consider the return vector \( r_{1 \times t} = \beta F \) that can be attributed to \( n \) factors:

\[
F_{n \times t} = \begin{bmatrix}
F_1 \\
F_2 \\
\vdots \\
F_i \\
\vdots \\
F_n
\end{bmatrix} = \begin{bmatrix}
F_{1,1} & F_{1,2} & \cdots & F_{1,t} \\
F_{2,1} & F_{2,2} & \cdots & F_{2,t} \\
\vdots & \vdots & \ddots & \vdots \\
F_{i,1} & F_{i,2} & \cdots & F_{i,t} \\
\vdots & \vdots & \ddots & \vdots \\
F_{n,1} & F_{n,2} & \cdots & F_{n,t}
\end{bmatrix}
\] (3.1)

with exposures given by the vector:

\[
\beta_{1 \times n} = \begin{bmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_i \\
\vdots \\
\beta_n
\end{bmatrix}
\] (3.2)

The risk associated to the asset is the weighted sum of each factor marginal risk:

\[
R(\beta) = \sum_{i=1}^{n} \beta_i \frac{\partial R(\beta)}{\partial \beta_i}.
\] (3.3)

As we have seen in Chapter 2, if we choose to consider the expected shortfall as the risk measure:

\[
R(\beta) = E[-\beta F | r \leq -VaR_{\alpha}(\beta F)]
\] (3.4)

the marginal contribution to risk associated to the i-th factor is:

\[
\frac{\partial R(\beta)}{\partial \beta_i} = -E[F_i | r \leq -VaR_{\alpha}(\beta F)]
\] (3.5)

Consider now k new factors \( F' \), with \( k \leq n \), obtained as a linear transformation of the original factors through matrix \( P_{k \times n} \):

\[
F'_{k \times t} = P_{k \times n} F_{n \times t} = \begin{bmatrix}
P_{1,1} & P_{1,2} & \cdots & P_{1,n} \\
P_{2,1} & P_{2,2} & \cdots & P_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
P_{k,1} & P_{k,2} & \cdots & P_{k,n}
\end{bmatrix} \begin{bmatrix}
F_{1,1} & F_{1,2} & \cdots & F_{1,t} \\
F_{2,1} & F_{2,2} & \cdots & F_{2,t} \\
\vdots & \vdots & \ddots & \vdots \\
F_{i,1} & F_{i,2} & \cdots & F_{i,t} \\
\vdots & \vdots & \ddots & \vdots \\
F_{n,1} & F_{n,2} & \cdots & F_{n,t}
\end{bmatrix}
\] (3.6)
3.1 Risk Contribution from user defined factors

The new defined factors $\tilde{F}$ don’t span the entire market and an error term appears:

$$ r = \tilde{\beta}\tilde{F} + \epsilon $$  \hspace{1cm} (3.7)

The new exposures $\tilde{\beta}$ need to be defined following any criteria. In a linear regression problem the exposures are chosen such that the new variables (here $\tilde{F}$) are uncorrelated with the residuals:

$$ Cov(\tilde{F}, \epsilon') = 0 $$  \hspace{1cm} (3.8)

Substitute the return definition in terms of the original factors $F$:

$$ Cov(F, r' - \tilde{F}'\tilde{\beta}') = 0 $$  \hspace{1cm} (3.9)

and after performing some linear algebra passages we have:

$$ Cov(PF, F'\beta') - Cov(PF, F'P')\tilde{\beta}' = 0 $$  \hspace{1cm} (3.10)

The exposures to the new factors are:

$$ \tilde{\beta}' = \left[ P \Sigma F P \right]^{-1} P \Sigma F \beta' $$  \hspace{1cm} (3.11)

where $\Sigma F$ is the covariance matrix of the original set of risk factors. Given the rule that identifies the exposures $\tilde{\beta}$, we can compute the risk contribution for the new factors. The part of $r$ that is not attributed to the new factors $\tilde{F}_{k \times t} = P_{k \times n} F_{n \times t}$ is attributed to the remaining theoretical factors $\hat{F}$, i.e:

$$ r = \tilde{\beta}\tilde{F} + \bar{\beta}\hat{F} $$  \hspace{1cm} (3.12)

$$ \hat{F}_{(n-k) \times t} = \hat{P}_{(n-k) \times n} F_{n \times t} $$  \hspace{1cm} (3.13)

where $\hat{P}_{(n-k) \times n}$ is such that the matrix:

$$ \hat{P}_{n \times n} = \begin{bmatrix} P \\ \hat{P} \end{bmatrix} $$  \hspace{1cm} (3.14)

is invertible. The portfolio return vector can be written in the form:

$$ r = \beta F = \beta \tilde{P}^{-1}\tilde{P}F = \begin{bmatrix} \tilde{\beta} & \bar{\beta} \end{bmatrix} \begin{bmatrix} \tilde{F} \\ \hat{F} \end{bmatrix} $$  \hspace{1cm} (3.15)
The total risk using the Euler rule is:

\[
R = \tilde{\beta} \frac{\partial R}{\partial \tilde{\beta}} + \hat{\beta} \frac{\partial R}{\partial \hat{\beta}} = \begin{bmatrix} \beta_1 & \cdots & \beta_k \end{bmatrix} \begin{bmatrix} \frac{\partial R}{\partial \beta_1} \\ \vdots \\ \frac{\partial R}{\partial \beta_k} \end{bmatrix} + \begin{bmatrix} \hat{\beta}_1 & \cdots & \hat{\beta}_{(n-k)} \end{bmatrix} \begin{bmatrix} \frac{\partial R}{\partial \hat{\beta}_1} \\ \vdots \\ \frac{\partial R}{\partial \hat{\beta}_{(n-k)}} \end{bmatrix}
\] (3.16)

The complete vector of exposures is:

\[
\begin{bmatrix} \tilde{\beta} \\ \hat{\beta} \end{bmatrix} = \beta \bar{P}^{-1}
\] (3.17)

from where we have:

\[
\begin{bmatrix} \tilde{\beta} \\ \hat{\beta} \end{bmatrix} \begin{bmatrix} P \\ \hat{P} \end{bmatrix} = \beta
\] (3.18)

In order to get the marginal exposures we need to express \( \beta \) as a linear transformation of \( \tilde{\beta} \) making clear the relation between them:

\[
\beta = \tilde{\beta}P + \hat{\beta}\hat{P}
\] (3.19)

Recall the differentiation rule

\[
\frac{\partial X A}{\partial X} = A
\] (3.20)

and the chain rule applied to matrices and get that:

\[
\frac{\partial R}{\partial \beta} = \begin{bmatrix} \frac{\partial R}{\partial \beta_{1,1}} & \frac{\partial R}{\partial \beta_{1,2}} & \cdots & \frac{\partial R}{\partial \beta_{1,k}} \\ \frac{\partial R}{\partial \beta_{2,1}} & \frac{\partial R}{\partial \beta_{2,2}} & \cdots & \frac{\partial R}{\partial \beta_{2,k}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial R}{\partial \beta_{k,1}} & \frac{\partial R}{\partial \beta_{k,2}} & \cdots & \frac{\partial R}{\partial \beta_{k,k}} \end{bmatrix} \begin{bmatrix} \frac{\partial \beta_{1,1}}{\partial \beta_{1,1}} & \frac{\partial \beta_{1,1}}{\partial \beta_{1,2}} & \cdots & \frac{\partial \beta_{1,1}}{\partial \beta_{1,k}} \\ \frac{\partial \beta_{1,2}}{\partial \beta_{1,1}} & \frac{\partial \beta_{1,2}}{\partial \beta_{1,2}} & \cdots & \frac{\partial \beta_{1,2}}{\partial \beta_{1,k}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \beta_{k,1}}{\partial \beta_{k,1}} & \frac{\partial \beta_{k,1}}{\partial \beta_{k,2}} & \cdots & \frac{\partial \beta_{k,1}}{\partial \beta_{k,k}} \end{bmatrix}
\] (3.21)

The contribution to risk of each new factor becomes the product of the initial contribution to risk with the transposed matrix used for the factor transformation:

\[
\frac{\partial R}{\partial \beta} = \frac{\partial R}{\partial \tilde{\beta}} P'
\] (3.22)

This approach requires only the knowledge of the transformation matrix but it can be applied in the cases when the new considered factors are uncorrelated and linear transformations of the original risk factors. In his paper Meucci [2007] performs an empirical analysis using the PCs as new uncorrelated factors. We go on and use this approach for decomposing risk when the factors considered are the ICs. The independence of the ICs ensures their uncorrelation, though the conditions required are satisfied.
3.1 Risk Contribution from User Defined Factors

3.1.1 Risk contribution using PCA and ICA

Let us consider the same dataset as in Section 2.5, i.e. we consider the S&P500 index as a portfolio whose returns depend on the initial risk factors, the GICS sector indexes. First, we investigate the risk attributed to the 7 PCs that explain at least 98% of the portfolio total variability. The risk measures considered are the Expected Shortfall for $\alpha = 0.05$ and the expectile for $\tau = 0.012$. As we expected, the first PC contribution to the portfolio risk is substantial for both the risk measures considered.

<table>
<thead>
<tr>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
<th>PC7</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES_0.05</td>
<td>0.0192</td>
<td>0.0004</td>
<td>-0.0003</td>
<td>-4.97E-05</td>
<td>-3.69E-06</td>
<td>-1.64E-05</td>
<td>-5.93E-05</td>
</tr>
<tr>
<td>100.6%</td>
<td>1.9%</td>
<td>-1.6%</td>
<td>-0.3%</td>
<td>0.0%</td>
<td>-0.1%</td>
<td>-0.3%</td>
<td>-0.2%</td>
</tr>
<tr>
<td>e_0.012</td>
<td>0.0153</td>
<td>0.0004</td>
<td>-0.0002</td>
<td>1.92E-05</td>
<td>-8.27E-06</td>
<td>-1.96E-05</td>
<td>0.0001</td>
</tr>
<tr>
<td>98.5%</td>
<td>2.4%</td>
<td>-1.6%</td>
<td>0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>0.7%</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

Table 3.1: Risk decomposition using PCs as user defined factors. The analysis is performed considering on the S&P500 and considering as initial risk factors the GICS sector indexes.

When the factors considered are the ICs, it is not possible to order them according to the variance explained. Being in front of an open problem, we decide to consider the same number of factors as suggested in the PCA analysis. That is, we consider the first 7 ICs as the sources of randomness in the market while the remaining 3 are grouped together in the residual term. Observe that, as we supposed, the contribution to risk of the ICs does not decrease systematically as we go on with the components. The risk attribution to the residual term is higher than that of the seventh IC. This suggests that further investigation should be performed on the residual term in order to extract any other IC that influences directly the portfolio risk.

Since the distribution of the statistical factors can be quite different from those of the initial risk factors, the sign of the contribution to the total risk can be negative. In the next section we will attribute risk to user defined risk factors that are based on observed financial quantities and we expect that the sign of the contribution to risk for each factor to be coherent with the sign of the correlation coefficient of the portfolio and the factor considered.

<table>
<thead>
<tr>
<th>IC1</th>
<th>IC2</th>
<th>IC3</th>
<th>IC4</th>
<th>IC5</th>
<th>IC6</th>
<th>IC7</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES_0.05</td>
<td>0.0189</td>
<td>-0.0048</td>
<td>-0.0004</td>
<td>0.0005</td>
<td>0.0005</td>
<td>0.0036</td>
<td>0.0002</td>
</tr>
<tr>
<td>98.9%</td>
<td>-25.0%</td>
<td>-2.4%</td>
<td>2.6%</td>
<td>2.9%</td>
<td>18.8%</td>
<td>1.2%</td>
<td>2.9%</td>
</tr>
<tr>
<td>e_0.012</td>
<td>0.0153</td>
<td>-0.0039</td>
<td>-0.0004</td>
<td>0.0003</td>
<td>0.0005</td>
<td>0.0031</td>
<td>0.0001</td>
</tr>
<tr>
<td>99.6%</td>
<td>-25.7%</td>
<td>-2.6%</td>
<td>2.3%</td>
<td>3.4%</td>
<td>20.1%</td>
<td>1.0%</td>
<td>1.9%</td>
</tr>
</tbody>
</table>

Table 3.2: Risk decomposition using ICs as user defined risk factors. The analysis is performed considering on the SPX and considering as initial risk factors the GICS sector indexes.
3.2 Custom Factor Attribution

An intuitive attribution of risk requires that the sources of risk to be aligned with the sources of return. Improper use of return sources may lead to a number of problems like mismatch between risk and return sources, non-intuitive marginal contributions to risk or wrongly naming aggressive positions as risk reducing. Risk forecasting is an ex-ante analysis. Generally a portfolio manager knows the factor exposures $\beta$ but factor return distributions $F$ must be predicted. During the performance attribution process, that is an ex-post analysis, $\beta$ and $F$ are known at the end. The discussion in this section is about risk forecasting, i.e an ex-ante analysis and based on the methodology introduced in Menchero and Poduri [2008].

The starting point is the return attribution model. Suppose that each asset return depends on $K$ factors and on an idiosyncratic component:

$$ r = \beta F + \epsilon_{\text{orig}} $$

(3.23)

where the idiosyncratic returns $\epsilon_{\text{orig}}$ are mutually uncorrelated and uncorrelated with factor returns. In particular, we have that $r$ is the $1 \times T$ vector of asset returns, $\beta$ is the $1 \times K$ vector containing the weights generated from the allocation decisions, $F$ is the $K \times T$ matrix of factor returns and $\epsilon$ is the $1 \times T$ vector of residuals. Factors $F$ can be used for describing returns but if they don’t match the manager investment process may be wrong for performance attribution. Let $F^Y$ be the custom factors such that they reflect the investment process:

$$ r = \beta^Y F^Y + \epsilon^Y $$

(3.24)

with the residuals $\epsilon^Y$ being neither mutually uncorrelated nor uncorrelated with factors $F^Y$. Most of the time factors $F^Y$ are not able to explain all the risk but there is some interest in understanding the part of return/risk attributed to a specific decision on investing in a particular asset. In terms factors $F$ and $F^Y$ the portfolio return becomes:

$$ r = \beta^Y F^Y + \beta F + \epsilon_{\text{Y,orig}} $$

(3.25)

The possible linear dependence among $\beta^Y$ and $\beta$ makes difficult the estimation used by regression. To avoid collinear $\beta$ and $\beta^Y$ create the exposures to residual factors $\tilde{\beta}$ such that:

$$ \beta^Y \tilde{\beta} = 0 $$

(3.26)

We want to decompose total risk in contributions coming from custom and residual factors. The projection operator which preserves only the components perpendicular to $\beta^Y$ is:

$$ P_{\beta^Y \perp} = I - \beta^Y (\beta^Y \beta^Y)^{-1} \beta^Y $$

(3.27)
and is used to obtain the exposures to the residual factors:

\[
\tilde{\beta} = P_{\beta Y} \beta
\]

(3.28)

The projection operator which preserves only the components within the space of \(\beta^Y\) is:

\[
P_{\beta^Y} = \beta^Y (\beta^Y \beta^Y)^{-1} \beta^Y
\]

(3.29)

Observe that:

\[
P_{\beta^Y} + P_{\beta^Y \perp} = I
\]

(3.30)

and obtain:

\[
r = \beta^Y F^Y + \tilde{\beta} \tilde{F} + \tilde{\epsilon}
\]

(3.31)

where \(F^Y\) is the \(L \times T\) vector of custom factors, \(\tilde{F}\) is the \(K \times T\) matrix of residual factors, \(\tilde{\epsilon}\) is the \(1 \times T\) vector of idiosyncratic returns, \(\beta^Y\) is the \(1 \times L\) vector of custom factor exposures and \(\tilde{\beta}\) is the \(1 \times K\) vector of residual factor exposures. Through OLS regression factor returns estimation are:

\[
F_{\text{ols}}^Y = (\beta^Y \beta^Y)^{-1} \beta^Y r
\]

(3.32)

\[
\tilde{F}_{\text{ols}} = (\tilde{\beta} \tilde{\beta})^{-1} \tilde{\beta} r
\]

(3.33)

If we think at \(r\) as the matrix of the returns of \(n\) assets\(^1\) and consider the weight \(n \times 1\) vector \(w\), the portfolio return \(\pi\), defined as:

\[
\pi = w' r
\]

(3.34)

becomes the sum of returns attributable to custom, residual factors and idiosyncratic factors:

\[
\pi = \pi_Y + \pi_{\tilde{F}} + \pi_{\tilde{\epsilon}}
\]

(3.35)

where \(\pi_Y = w' P_Y r\), \(\pi_{\tilde{F}} = w' P_{\tilde{F}} r\) and \(\pi_{\tilde{\epsilon}} = w' (I_N - P_Y - P_{\tilde{F}}) r\).

### 3.3 Custom and projected factors

#### 3.3.1 Decomposing the residual

In this section we consider ex-post analysis, in the sense that once the factor returns are observed we can perform risk attribution to the factors. The approach is similar to that discussed in Menchero and Poduri [2008] but with some differences. As usual, we start with the hypothesis that the return can linearly be decomposed using \(n\) factors that we call \(F^X\). The problem of attributing risk to a new set of factors in \(F^Y\) and contemporary to the old factors \(F^X\) is faced at first by considering a two step-regression. Here we describe the main steps. Given the new factors \(F^Y\) we determine

\(^1r\) is an \(n \times t\) matrix
the exposures to them by:

$$\beta^Y = (F^Y F^Y)^{-1} F^Y r$$  \hfill (3.36)

Denoting the new residuals with $\delta$ we have the decomposition of the returns in terms of the new factors and the residual term:

$$r = \beta^Y F^Y + \delta$$  \hfill (3.37)

Since $F^Y$ is an arbitrary chosen factor, its contribution to the portfolio return/risk can be quite limited. In order to extract more information from the $\delta$ term, we perform another regression analysis this time using as regressors the original factors $F^X$. The relation of $\delta$ with factors $F^X$ is:

$$\delta = \beta^{Proj} F^X + \tilde{\epsilon}$$  \hfill (3.38)

where:

$$\beta^{Proj} = (F^X F^X)^{-1} F^X \delta$$  \hfill (3.39)

The new decomposition is:

$$r = \beta^Y F^Y + \beta^{Proj} F^X + \tilde{\epsilon}$$  \hfill (3.40)

but remains the problem that $F^Y$ and $F^X$ could be linearly dependent.

We want to attribute risk to factors $F^Y$ but in the same time the contribution of the idiosyncratic term must be extracted from the first regression step. That is, we aim at having a model of the form:

$$r = \beta^Y F^Y + \beta^{\tilde{X}} F^{\tilde{X}} + \epsilon^i$$  \hfill (3.41)

where $\epsilon^i$ is the idiosyncratic residual uncorrelated with the custom factors $F^Y$ and $F^{\tilde{X}}$ contains the factors uncorrelated both with the idiosyncratic residual and the custom factors.

We start by identifying what remains unexplained from the old factors $F^X$ when we consider only the factors in $F^Y$. This is done through a simple linear regression model:

$$F^X = \bar{\beta}^Y F^Y + \bar{u}$$  \hfill (3.42)

The residuals $\bar{u}$, that by construction are uncorrelated with $F^Y$, are defined as new factors $F^{\tilde{X}}$. These are the projections of the old factors $F^X$ on the space orthogonal to $F^Y$:

$$F^{\tilde{X}} = F^X - \bar{\beta}^Y F^Y = F^X [I - F^Y (F^Y F^Y)^{-1} F^Y]$$  \hfill (3.43)

In this way we are able to decompose the return vector $r$ in three uncorrelated quantities:

$$r = \begin{bmatrix} \beta^Y & \beta^{\tilde{X}} \end{bmatrix} \begin{bmatrix} F^Y \\ F^{\tilde{X}} \end{bmatrix} + \epsilon^i$$  \hfill (3.44)
3.3 Custom and projected factors

\[ \tilde{u} = F_X - \beta_1 F_Y^1 - \beta_2 F_Y^2 \]

Figure 3.1: Geometric representation of the residuals $\tilde{u}$ orthogonal to the space generated by the custom factors $F_Y$

The estimated exposures are the result of an OLS regression of two uncorrelated groups of variables:

\[
\begin{bmatrix}
\beta_Y \\
\beta_{\tilde{X}}
\end{bmatrix} = r
\begin{bmatrix}
F_Y & F_{\tilde{X}}
\end{bmatrix}'
\left[
\begin{bmatrix}
F_Y & F_{\tilde{X}}
\end{bmatrix}
\begin{bmatrix}
F_Y & F_{\tilde{X}}
\end{bmatrix}'
\right]^{-1}
\]

(3.45)

The new introduced methodology is composed by three simple steps:

1) Given the custom factors $F_Y$ and using equation (3.43) we obtain the new residual factors $F_{\tilde{X}}$.

2) The new exposures are:

\[
\begin{bmatrix}
\beta_Y \\
\beta_{\tilde{X}}
\end{bmatrix} = \left[r F_Y (F_Y F_Y')^{-1} r F_{\tilde{X}} (F_{\tilde{X}} F_{\tilde{X}}')^{-1}\right]
\]

(3.46)

3) As a final step, we compute the idiosyncratic residuals: $e^i = r - (\beta_Y F_Y + \beta_{\tilde{X}} F_{\tilde{X}})$. The procedure is simple and doesn’t require a lot of storage space. Once we have the return decomposition model, for the discussed risk measures the attribution to each factor is straightforward. Any portfolio manager can offer to his clients a set of custom factors and daily update the residual factors. In this way the client receives the information required and in the same time gets an idea of where does the residual risk come from. The new factors $F_{\tilde{X}}$ have a simple interpretation since they contain the returns that are not idiosyncratic and not explained from $F_Y$. 
3.3.2 Application to the stock market

Here we apply the procedure described in the previous section to the equity market and in particular we consider the returns of the VFIAX fund index which replicates the performance of the SPX index. The evaluation date is 16/10/2012 and we use two years of historical data in computing the risk measures. As a first step we compute the exposures to the factors used for explaining the returns, i.e daily log returns on GICS indexes.

<table>
<thead>
<tr>
<th>Factor</th>
<th>COND</th>
<th>CONS</th>
<th>ENRS</th>
<th>FINL</th>
<th>HLTH</th>
<th>INDU</th>
<th>INFT</th>
<th>MATR</th>
<th>TELS</th>
<th>UTIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )</td>
<td>0.110</td>
<td>0.130</td>
<td>0.122</td>
<td>0.141</td>
<td>0.098</td>
<td>0.112</td>
<td>0.187</td>
<td>0.040</td>
<td>0.024</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.009)</td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.006)</td>
</tr>
</tbody>
</table>

Table 3.3: Exposures of the VFIAX fund index to the 10 GICS sectors considered as factors computed considering the period 16/10/2010 till 16/10/2012.

The Global index return vector ( quoted in the U.S market and downloaded from the Bloomberg terminal) is considered as a custom factor. We compute the projections of the GICS returns on the space orthogonal to that generated from the custom factor and applying the procedure introduced before, we get the idiosyncratic return vector. In figure 3.3.2 we plot the fund returns and the sequence of the idiosyncratic residuals. To confirm the fact, that to the residual term is attributed a small part of the fund risk we show the results of the decomposition of the Expected Shortfall in 3.3.2 for the three groups of factors: the custom, the projected and the idiosyncratic factors.

![Figure 3.2](image)

Figure 3.2: In this figure we show the time series of the VFX returns and the idiosyncratic residual vector when the Global index is considered as custom factor.
3.3 Custom and projected factors

<table>
<thead>
<tr>
<th></th>
<th>VFIAX</th>
<th>FY</th>
<th>FX</th>
<th>e^i</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES0.05</td>
<td>0.0298</td>
<td>0.0249</td>
<td>0.0048</td>
<td>4.79E-05</td>
</tr>
</tbody>
</table>

|          | 83.57% | 16.27% | 0.16% |

Table 3.4: Decomposition of the VFIAX $ES_{0.05}$ considering as new factor $F_Y$ the return on the Global index to which is most of the risk.

3.3.3 Application to fixed income securities

Here we give two examples of the methodology introduced in the previous section. In particular, we consider the returns of two bonds both with time to maturity 5 years paying coupons annually. The issuer of the first bond is assumed to have a z-spread similar to the Industrial Sector B rated companies while the second bond z-spread is that of the Utility sector A rated companies. Coherently, we assume a coupon rate of 6.875% for the first bond and a coupon rate of 2% for the second bond. The evaluation date is 16/11/2012 while the maturity date is 16/11/2017.

Given function $g$ such that a bond price is:

$$P = g(\beta X_i, F_X)$$ (3.47)

we can define $\beta X_i = -MD X_i$, i.e minus the modified duration of the i-th zero rate or i-th z-spread.

The log-returns can be approximated using the modified duration associated to each risk factor:

$$r_t \approx -\sum_{i=1}^{5} MD z_{iy} \Delta z_{iy} - MD spread \Delta z_{spread}$$ (3.48)

The formula used to compute the modified duration is:

$$MD z_{iy} = \frac{P(z_{iy} + 0.0001) - P(z_{iy})}{2P(z_{iy})0.0001}$$ (3.49)

We consider the daily realizations for the risk factors and use them to generate bond return scenarios. As custom factor we consider the returns on the Oil index whose prices are downloaded from the Bloomberg terminal.

We consider the Expected Shortfall as a risk measure. We expect the bond in the Industrial sector to have a higher exposure to the Oil index than the second bond. The two step-regression and the new projection factor method are used to determine the residuals. In the second approach, being the idiosyncratic residuals the part of return that can not be attributed neither to the custom factor nor to the projected factors, the risk contribution of the residual term should be lower.
At the evaluation date the price of the first bond is $P_1 = 96.92$ with z-spread = 645.39 basis points. The notional amount is $N = 100$ and the coupon rate = 6.875%. The original risk factors are the change in value of the zero rates corresponding to the payment maturities and of the z-spread. The new factor with respect to which we want to perform the return and risk decomposition is $F_Y$ that is the log return of the Oil Index quoted in the market. We apply the two procedures discussed in the previous section, i.e. the two-step regression and the orthogonal factor projection procedure. The residuals obtained with both methods are showed in figure 3.3.

<table>
<thead>
<tr>
<th>Factor Exposure</th>
<th>$\Delta z_{1y}$</th>
<th>$\Delta z_{2y}$</th>
<th>$\Delta z_{3y}$</th>
<th>$\Delta z_{4y}$</th>
<th>$\Delta z_{5y}$</th>
<th>$\Delta z_{spread}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.067</td>
<td>0.125</td>
<td>0.174</td>
<td>0.216</td>
<td>3.864</td>
<td>4.447</td>
</tr>
</tbody>
</table>

Table 3.5: Here we report the factor exposures for the first bond that in this case are minus the modified duration for the considered zero rate or z-spread.

At the evaluation date the price of the second bond is $P_1 = 95.85$ with z-spread = 208.5 basis points. The notional amount is $N = 100$ and coupon rate = 2%. The original risk factors are the change in value of the zero rates corresponding to the payment maturities and of the z-spread. The new factor with respect to which we want to perform the return and risk decomposition is $F_Y$ that is the log return of the Oil Index quoted in the market. We apply the two procedures discussed in the previous section, i.e. the two-step regression and the orthogonal factor projection procedure. The residuals obtained with both methods are showed in fig 3.4.
Table 3.6: Decomposition of the $ES_{0.05}$ of the bond in the Industrial sector considering as new factor $F^Y$ the return on Oil index. The greater part of the risk is attributed to the projected factors $F^X$.

<table>
<thead>
<tr>
<th>BondIndu</th>
<th>$F^Y$</th>
<th>$F^X$</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ES_{0.05} (r)$</td>
<td>0.0128</td>
<td>1.70E-04</td>
<td>0.0127 -1.03E-04</td>
</tr>
<tr>
<td></td>
<td>1.3%</td>
<td>99.5%</td>
<td>-0.8%</td>
</tr>
</tbody>
</table>

Table 3.7: Here we report the factor exposures for the second bond that in this case are minus the modified duration for the considered zero rate or z-spread.

Table 3.8: Decomposition of the second bond $ES_{0.05}$ considering as new factor $F^Y$ the return on Oil index. The greater part of the risk is attributed to the projected factors $F^X$.

<table>
<thead>
<tr>
<th>BondIndu</th>
<th>$F^Y$</th>
<th>$F^X$</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ES_{0.05} (r)$</td>
<td>8.27E-03</td>
<td>1.80E-06</td>
<td>0.0083 -3.39E-05</td>
</tr>
<tr>
<td></td>
<td>0.02%</td>
<td>100.39%</td>
<td>-0.41%</td>
</tr>
</tbody>
</table>
4.1 Semi-heavy tailed distributions

The Stable distribution has gained a great popularity in modeling economic and financial time series, starting from the seminal work of Mandelbrot [1963]. However, empirical evidence is not coherent with the Stable model assumption since the return distribution shows tails heavier than Normal but thinner than the Stable ones.

A drawback of the Stable distribution is that only fractional moments of order \( p \leq \alpha \) with \( \alpha \in (0, 2) \) exist and, consequently, the standard hypothesis for applying the Central Limit Theorem do not hold. For this reason several researchers have considered the Tempered Stable distribution as an valid alternative in modeling financial returns.

The Tempered Stable distribution can be obtained by multiplying the Lévy density of an \( \alpha \)-Stable with a decreasing tempering function (see Cont and Tankov [2003]). Performing this operation, the tail behavior of the new distribution changes from heavy to semi-heavy characterized by exponential instead of power decay. The existence of the conventional moments is ensured and the Tempered Stable satisfies the conditions of the classical Central Limits Theorem. This is an advantage in modeling asset returns with respect to the Stable distribution. Indeed the return distributions tend to be closed to normal when we consider monthly or annual data (see Cont [2001] for survey of the Stylized Facts).

In this chapter, we propose a new distribution, namely the Mixed Tempered Stable (MixedTS henceforth). The idea is to build the new distribution in the similar way of the Normal Variance Mean Mixture (NVMM) where the Normal assumption is substituted by the standardized Tempered Stable. In this way, we are able to overcome some limits of the NVMM. In particular, the asymmetry and the kurtosis do not depend only on the mixing random variable but also on the
Tempered Stable.
Assuming that the mixing random variable follows a Gamma distribution, the proposed model has the Variance Gamma (see Madan and Seneta [1990b] and Loregian et al. [2012]), the Tempered Stable (see Cont and Tankov [2003]) and the Geo-Stable distributions (see Kozubowski et al. [1997]) as special cases. The new distribution is applied to real data. In particular, we conduct two experiments. In the first, we build a Garch(1,1) with MixedTS innovations and estimate it using the financial time series. In the second, we consider a multifactor model to describe the log returns of the Vanguard Fund Index which tries to replicate the performance of the S&P 500 index. As factors, we consider the Global Industry Classification Standard (GICS) indexes developed by MSCI-Barra provider. We capture the GICS dependence structure using the Independent Component Analysis introduced by Comon [1994] and developed by Hyvarinen et al. [2001].

Once fixed the number of the independent components, we use the Mixed Tempered Stable to model each of them. Since the observed fund returns are modeled as a linear sum of the independent ones, the single factor return density could be non-gaussian and/or semi-heavy tailed. This is confirmed by our results and the different values obtained for the MixedTS parameters suggest that the component distributions can be of quite different nature. The Independent Component Analysis has been already used in finance to model the interest rates term structure (see Bellini and Salinelli [2003]). Recently Madan [2006b] used a non-Gaussian factor model and modelled the components, coming from an ICA analysis, using the Variance Gamma distribution. Our multifactor model can be seen as a generalization of this latter since, as observed above, the Variance Gamma is a particular case of the Mixed Tempered Stable distribution. This portfolio return decomposition can be used in risk management. Marginal contribution of each linear transformed factor can be easily computed for given homogeneous risk measures as described in Chapter 3.

The outline of the chapter is as follows. In Section 2 we describe the main characteristics of the Tempered Stable distribution and make some observations useful for the derivation of the new distribution. In Section 3 we introduce the Mixed Tempered Stable distribution, its main features and illustrate the fitting of the MixedTS distribution to the Independent Components. Section 5 gives some results for the computation of the risk measures considered in this thesis considering the new distribution.

### 4.2 Tempered Stable distribution

In this section we review the main features of the Tempered Stable distribution. A random variable $X$ follows a Tempered Stable distribution if its Lévy measure is given by:

$$
\nu(dx) = \left( C_+ e^{-\lambda_+ x} x^{1+\alpha_+} 1_{x>0} + C_- e^{-\lambda_- |x|} \frac{|x|^{1+\alpha_-}}{|x|^{1+\alpha_-}} 1_{x<0} \right) dx
$$

(4.1)
4.2 Tempered Stable distribution

with \( \alpha_+ , \alpha_- \in (0, 2) \) and \( C_+, C_- , \lambda_+, \lambda_- \in (0, +\infty) \).

The characteristic function is obtained by solving the integral [see Cont and Tankov, 2003]:

\[
E [e^{iuX}] = \exp \left[ iu\gamma + \int_{\mathbb{R}} \left( e^{iux} - 1 - iux \right) \nu (dx) \right] \\
= \exp \{ iu\gamma + C_+ \Gamma (-\alpha_+) \left( \lambda_+ - iu \right)^{\alpha_+} - \lambda^{\alpha_+} \} \\
+ \ C_- \Gamma (-\alpha_-) \left( \lambda_- + iu \right)^{\alpha_-} - \lambda^{\alpha_-} \}
\]

(4.2)

where \( \gamma \in \mathbb{R} \). As observed in Küchler and Tappe [2013], for \( \alpha_+ , \alpha_- \in (0, 1) \), the Tempered Stable is obtained as a difference of two independent one sided tempered stable distributions introduced in Tweedie [1984]. The corresponding process shows finite variation with infinite activity.

The interest of researchers for the Tempered Stable distribution is confirmed by the fact that the following cases have been investigated in literature:

- For \( C_+ = C_- = C \) and \( \alpha_+ = \alpha_- = \alpha \), the CGMY distribution is obtained [see Carr et al., 2002].
- If \( \alpha_+ = \alpha_- \) and \( \lambda_+ = \lambda_- \) we get the truncated Lévy flight introduced in Koponen [1995].
- Computing the limit for \( \alpha_+ = \alpha_- \to 0^+ \) we get the Bilateral Gamma distribution [see Küchler and Tappe, 2008a,b, 2009].
- Taking the limit for \( \alpha_+ = \alpha_- \to 0^+ , C_+ = C_- \) and \( \lambda_+ = \lambda_- \) the Variance Gamma is obtained [see Madan and Seneta, 1990a; Loregian et al., 2012, for estimation].

In this chapter, we consider the same restrictions as in [see Kim et al., 2008], i.e.: \( \alpha_+ = \alpha_- = \alpha \) and \( \gamma = \mu - \Gamma (1 - \alpha) \left( C_+ \lambda_+^{\alpha-1} - C_- \lambda_-^{\alpha-1} \right) \). In this case the distribution of \( X \sim CTS (\alpha, \lambda_+, \lambda_-, C_+, C_-, \mu) \) is called classical tempered stable and \( E (X) = \mu \). Its characteristic function is given by:

\[
E [e^{iuX}] = \Phi (u; \alpha, \lambda_+, \lambda_-, C_+, C_- , \mu) \\
= \exp \{ iu\mu - iu\Gamma (1 - \alpha) \left( C_+ \lambda_+^{\alpha-1} - C_- \lambda_-^{\alpha-1} \right) \} \\
+ \ C_+ \Gamma (-\alpha) \left( \lambda_+ - iu \right)^{\alpha} - \lambda^{\alpha} \} + C_- \Gamma (-\alpha) \left( \lambda_- + iu \right)^{\alpha} - \lambda^{\alpha} \} \\
\]

The cumulants of the r.v \( X \) can though be obtained taking the derivatives of the characteristic function :

\[
c_n (X) := \frac{1}{n} \frac{\partial^n}{\partial u^n} \ln \left( E [e^{iuX}] \right) \bigg|_{u=0} 
\]

(4.3)

Given the distribution parameters, for \( n = 1 \) we have :

\[
c_1 (X) = \mu 
\]

(4.4)
and for \( n \geq 2 \):

\[
c_n(X) = \Gamma(n - \alpha)(C_+ \lambda_+^{\alpha-n} + (-1)^n C_- \lambda_-^{\alpha-n})
\]  

(4.5)

Therefore using (4.5) the following quantities are computed:

\[
\begin{align*}
E(X) &= c_1(X) = \mu \\
Var(X) &= c_2(X) = \Gamma(2 - \alpha) \left[ C_+ \lambda_+^{\alpha-n} + (-1)^n C_- \lambda_-^{\alpha-n} \right] \\
\gamma_1 &= \frac{c_3(X)}{c_2^2(X)} = \frac{\Gamma(3 - \alpha)[C_+ \lambda_+^{\alpha-3} - C_- \lambda_-^{\alpha-3}]}{c_2^3(X)} \\
\gamma_2 &= 3 + \frac{c_4(X)}{c_2^2(X)} = 3 + \frac{\Gamma(4 - \alpha)[C_+ \lambda_+^{\alpha-4} - C_- \lambda_-^{\alpha-4}]}{c_2^3(X)}
\end{align*}
\]

(4.6)

From the skewness formula it can be noticed that the difference between \( C_+ \lambda_+^{\alpha-3} \) and \( C_- \lambda_-^{\alpha-3} \) drives the asymmetry while for \( C_+ = C_- \) we must look only at the two tempering parameters \( \lambda_+ \) and \( \lambda_- \).

**Remark 9.** Fixing \( \lambda_+ = \lambda_- = \lambda \), \( \mu = 0 \) and \( C_+ = C_- = C \), the characteristic function of a symmetric Tempered Stable distribution is given by:

\[
E[\exp(iuX)] = \exp[iu\mu + C \Gamma(1 - \alpha) [(\lambda - iu)^\alpha + (\lambda + iu)^\alpha - 2\lambda^\alpha]]
\]

(4.7)

If \( r = \sqrt{u^2 + \lambda^2} \) and \( \theta = \arctg \left( \frac{u}{\lambda} \right) \), the characteristic function in (4.7) becomes:

\[
E[\exp(iuX)] = \exp[C \Gamma(1 - \alpha) [r^\alpha e^{-i\theta} + r^\alpha e^{i\theta} - 2\lambda^\alpha]]
\]

The last equality holds due to the Euler relation \( e^{i\theta} + e^{-i\theta} = 2 \cos(\alpha \theta) \). Taking the limit for \( \lambda \to 0^+ \) we obtain the symmetric \( \alpha \)-stable distribution since:

\[
\lim_{\lambda \to 0^+} \exp \left[ C \Gamma(1 - \alpha) \left( 2 \left( u^2 + \lambda^2 \right) \frac{\alpha}{2} \cos(\alpha \theta) - 2\lambda^\alpha \right) \right] = \exp \left[ C \Gamma(1 - \alpha) \left[ 2 |u|^\alpha \cos \left( \frac{\alpha \pi}{2} \right) \right] \right]
\]

The r.v \( X \) has zero mean and unit variance for \( \mu = 0 \) and:

\[
C = C_+ = C_- = \frac{1}{\Gamma(2 - \alpha)(\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2})}
\]

(4.8)

The corresponding r.v is standardized, i.e \( X \sim \text{stdCTS}(\alpha, \lambda_+, \lambda_-) \). It useful to observe that in the standardized CTS distribution \( C \) is fully determined once given the values for \( \alpha, \lambda_+ \) and \( \lambda_- \). Its characteristic exponent, defined as \( L_{\text{stdCTS}}(u; \alpha, \lambda_+, \lambda_-) = \log E[e^{iuX}] \), is given by:

\[
L_{\text{stdCTS}}(u; \alpha, \lambda_+, \lambda_-) = \frac{(\lambda_+ - iu)^\alpha - \lambda_+^\alpha + (\lambda_- + iu)^\alpha - \lambda_-^\alpha}{\alpha(\alpha - 1)(\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2})} + \frac{iu(\lambda_+^{\alpha-1} - \lambda_-^{\alpha-1})}{(\alpha - 1)(\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2})}
\]

(4.9)

For \( \alpha \to 2 \) the limiting distribution coincides with the normal distribution:

\[
\lim_{\alpha \to 2} L_{\text{stdCTS}}(u; \alpha, \lambda_+, \lambda_-) = -\frac{u^2}{2}
\]
Remark 10. Condition (4.8) implies that the convergence to $\alpha$-stable distribution is not possible since the characteristic exponent would converge to zero. 

The standardized Classical Tempered Stable distribution has the following property that makes it appealing for mixtures.

Proposition 11. Let $\tilde{X} \sim \text{stdCTS} \left( u : \alpha, \lambda_+, \lambda_- \sqrt{h} \right)$ and $h \in (0, +\infty)$ then the random variable $Y \overset{d}{=} \sqrt{h} \tilde{X}$ has the following characteristic exponent:

$$\ln E \left[ e^{iuY} \right] = h \left[ \frac{(\lambda_+ - iu)^\alpha - \lambda_+^\alpha + (\lambda_- + iu)^\alpha - \lambda_-^\alpha}{\alpha (\alpha - 1) (\lambda_+^\alpha - \lambda_-^\alpha)} + \frac{iu (\lambda_+^{\alpha-1} - \lambda_-^{\alpha-1})}{(\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2})}\right]$$ (4.10)

Moreover if $h \in \mathbb{N}$ we have:

$$Y \overset{d}{=} \sum_{j=1}^{h} X_j$$ (4.11)

where $X_j$ are iid stdCTS $(\alpha, \lambda_+, \lambda_-)$

Proof. To derive the (4.10) exponent we evaluate (4.9) in $\sqrt{hu}$ and obtain:

$$\ln E \left[ e^{iuY} \right] = \frac{(\lambda_+ \sqrt{h} - iu \sqrt{h})^\alpha - h \frac{\lambda_+^\alpha}{\alpha (\alpha - 1)} + (\lambda_- \sqrt{h} + iu \sqrt{h})^\alpha - h \frac{\lambda_-^\alpha}{\alpha (\alpha - 1)}}{(\lambda_+^\alpha - \lambda_-^\alpha)} + \frac{iu \sqrt{hu} (h \frac{\lambda_+^{\alpha-1} - \lambda_-^{\alpha-1}}{\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2}})}{\alpha (\alpha - 1) (\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2})}$$ (4.12)

factorizing $h$ we get:

$$\ln E \left[ e^{iuY} \right] = \left[ \frac{h \frac{\lambda_+^{\alpha-1} - \lambda_-^{\alpha-1}}{\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2}} + \frac{iu \sqrt{hu} (h \frac{\lambda_+^{\alpha-1} - \lambda_-^{\alpha-1}}{\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2}})}{\alpha (\alpha - 1) (\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2})}}{h \frac{\lambda_+^{\alpha-1} - \lambda_-^{\alpha-1}}{\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2}}} \right]$$ (4.13)

simplifying we obtain the result (4.10).

To prove (4.11) we use the iid assumption for $X_j$ and the characteristic exponent of the random variable $\sum_{j=1}^{h} X_j$ becomes:

$$\ln E \left[ \exp \left( \sum_{j=1}^{h} X_j \right) \right] = h \ln E \left[ \exp \left( X_1 \right) \right]$$ (4.14)

where $X_1 \sim \text{stdCTS} \left( \alpha, \lambda_+, \lambda_- \right)$. Using (4.9) we obtain the characteristic exponent of $\sqrt{h} \tilde{X}$ that implies the condition (4.11). \qed
4.3 Mixed Tempered Stable distribution

In this section, using the prop. 11, we build a new distribution that is shown to have some nice mathematical and statistical characteristics.

4.3.1 Definition and particular cases

Definition 12. We say that a continuous random variable $Y$ follows a Mixed Tempered Stable distribution if:

$$Y \overset{d}{=} \sqrt{V} \tilde{X}$$

(4.15)

where $\tilde{X} | V \sim \text{stdTS}(\alpha, \lambda_+, \sqrt{V}, \lambda_-, \sqrt{V})$. $V$ is a Lévy distribution defined on positive axis and its m.g.f always exists.

The logarithm of the m.g.f. is:

$$\Phi_V(u) = \ln \left[ E \left[ \exp \left( u \sqrt{V} \tilde{X} \right) \right] \right]$$

(4.16)

We compute the characteristic exponent for the new distribution and apply the law of iterated expectation:

$$E \left[ e^{iu\sqrt{V} \tilde{X}} \right] = E \left\{ E \left[ e^{iu\sqrt{V} \tilde{X}} | V \right] \right\} = \exp \left[ \Phi_V \left( L_{\text{stdCTS}}(u; \alpha, \lambda_+, \lambda_-) \right) \right]$$

(4.17)

The characteristic function identifies the distribution at time one of a time changed Lévy process. From [see Sato, 1998; Carr and Wu, 2004] the distribution is infinitely divisible.

Proposition 13. The first four moments for the MixedTS are:

$$\begin{align*}
E \left[ \sqrt{V} \tilde{X} \right] &= 0 \\
\text{Var} \left[ \sqrt{V} \tilde{X} \right] &= E \left[ V \right] \\
\gamma_1 &= \left( 2 - \alpha \right) \frac{\lambda_+^{3-\alpha} - \lambda_-^{3-\alpha}}{(\lambda_+^{3-\alpha} + \lambda_-^{3-\alpha})} E^{-1/2} \left[ V \right] \\
\gamma_2 &= \left[ 3 + (3 - \alpha) \left( 2 - \alpha \right) \frac{\lambda_+^{3-\alpha} + \lambda_-^{3-\alpha}}{(\lambda_+^{3-\alpha} + \lambda_-^{3-\alpha})} \right] \frac{E[V^2]}{E[V]} 
\end{align*}$$

Figure 4.1 shows the behaviour of the skewness for different combinations of $\lambda_+$ and $\lambda_-$ and fixed $\alpha$.

In figure 4.2 we have the behaviour of the kurtosis for different values of $\lambda_+$, $\lambda_-$ and $\alpha$. 
Figure 4.1: Consider the case when $V \sim \Gamma(1, 1)$ and fix some value for $\alpha$. We plot the skewness curve levels for different combinations of $\lambda_+$ and $\lambda_-$ to have an idea of the possible skewness values. In the particular case when they coincide, the skewness is zero. The effect of an higher $\alpha$ is the reduction of the skewness level (kept fixed values of the other parameters). The distribution of the MixedTS becomes symmetric for $\alpha = 2$.

Figure 4.2: Consider the case when $V \sim \Gamma(1, 1)$ and fix some value for $\alpha$. We plot the kurtosis values for different combinations of $\lambda_+$ and $\lambda_-$ to have an idea of the possible kurtosis values. The effect of an higher $\alpha$ is the reduction of the kurtosis level. If the fixed value for $\alpha$ is 1.9 the curve level for kurtosis tend to be close to 3 and the limiting case of kurtosis=3 is obtained for $\alpha = 2$. 
If we assume that $V \sim \Gamma(a, \sigma^2)$ the characteristic exponent in (4.17) becomes:

$$E \left[ \exp \left( iu \sqrt{V} X \right) \right] = \exp \left[ -a \ln \left( 1 - \sigma^2 \frac{(\lambda_+ - iu)^\alpha - (\lambda_+)^\alpha + (\lambda_+ + iu)^\alpha - (\lambda_-)^\alpha}{\alpha (\alpha - 1)(\lambda_+)^{\alpha - 2} + (\lambda_-)^{\alpha - 2}} \right) - \sigma^2 \frac{iu(\lambda_+^\alpha - 1 - \lambda_-^\alpha - 1)}{(\alpha - 1)(\lambda_+^{\alpha - 2} + \lambda_-^{\alpha - 2})} \right]$$

(4.18)

For $\sigma = \frac{1}{\sqrt{a}}$ we compute the limit for $a$ going to infinity and obtain the stdCTS as a special case (see Figure 4.3(b)). The symmetric Variance Gamma distribution is obtained by choosing $\alpha = 2$ as shown in Figure 4.3(a).

**Proposition 14.** Choosing

$$\begin{align*}
\lambda_+ &= \lambda_- = \lambda \\
a &= 1 \\
\sigma &= \lambda^{\frac{\alpha - 2}{2}} \gamma \frac{\alpha (\alpha - 1)}{\cos(\alpha \frac{\pi}{2})}
\end{align*}$$

(4.19)

and computing the limit for $\lambda \to 0^+$ we obtain the Geometric Stable distribution.

Indeed substituting the conditions (4.20) in the characteristic exponent of the new distribution we get:

$$E \left[ \exp \left( iu \sqrt{V} X \right) \right] = \exp \left[ - \ln \left( 1 - \frac{(\lambda_+ - iu)^\alpha - (\lambda_+)^\alpha + (\lambda_+ + iu)^\alpha - (\lambda_-)^\alpha}{2\alpha (\alpha - 1)} \right) \right]$$

Applying the limit and following the same arguments for the convergence of the Tempered Stable to the symmetric distribution we get:

$$E \left[ \exp \left( iu \sqrt{V} X \right) \right] \rightarrow \left( 1 - |u|^\alpha \frac{\cos(\alpha \frac{\pi}{2})}{\alpha (\alpha - 1)} \right)^{-1}$$

for $\alpha \neq 1$. The convergence to the Geometric Stable distribution is shown in Figure 4.3(c).
4.3 Mixed Tempered Stable distribution

(a) The symmetric VG distribution is a particular case of the MixedTS and it is obtained for $\alpha = 2$. We fix $\mu_0 = 0$, $\mu = 0$, $\sigma = 1.2$, $\alpha = 1.7$, $\lambda_+ = 1.2$ and $\lambda_- = 8$. In the figure we plot the MixedTS for different $\alpha$ values. The distribution associated is asymmetric but the limiting case not.

(b) The symmetric CTS distribution is a particular case of the MixedTS and it is obtained for $\alpha = 2$. We fix $\mu_0 = 0$, $\mu = 0$, $\lambda_+ = 1.2$ and $\lambda_- = 8$ and $\alpha = 1.4$. In the figure we plot the MixedTS for different values of $\alpha$.

(c) To show the convergence of the MixedTS distribution to the Geometric Stable one we plot it for different $\lambda$ values given $\mu_0 = 0$, $\mu = 0$, $\alpha = 1$ and $\gamma = 1$. Recall that the GeoStable is obtained as a limiting case when $\sigma = \lambda^{\frac{\alpha-2}{2}} \gamma^{\frac{1}{2}} \sqrt{\frac{\alpha(\alpha-1)}{\cos(\alpha \pi)}}$ and $\lambda \to 0$. Observe that as $\lambda$ gets smaller the tails get heavier. We cut the plot since in zero the GeoStable distribution has a peak going to $+\infty$.

Figure 4.3: Special cases of the GeoStable distribution
4.3.2 Investigation using real market data

This section is devoted to the empirical investigation of the MixedTS distribution in modeling asset returns. We consider daily log returns of the Vanguard Fund Index which tries to replicate the Standard and Poor 500 performance. It seems quite natural to consider as portfolio risk factors the daily log-returns of the 10 GICS sector indexes since each of the S&P500 members belongs to one of them. The data are daily log returns ranging from 14-June-2010 to 20-September-2012 obtained from the Bloomberg data provider. We denote by \( X^{(t)} \) the vector of risk factors at time \( t \) with the \( i \)-th component \( X^{(t)}_i = \log S^{(t)}_i / S^{(t-1)}_i \) being the \( i \)-th GICS daily log-return. We can now define a linear model for the portfolio returns:

\[
r_P = \beta X + \epsilon. \tag{4.21}
\]

where \( \beta \) is the \( 1 \times n \) vector of factor exposures, \( X \) is the \( n \times t \) matrix of factor returns and \( \epsilon \) is the \( 1 \times t \) vector of residuals.

Note that we drop the \( t \) index since we report in matrix notations the whole time-series for our quantities. In the simplest case \( \beta \) and \( \epsilon \) are obtained through an OLS regression. Here the number of risk factors is \( n = 10 \) and the residuals are centered due to the particular choice of our portfolio.

The \( R^2 \) is about 99 per cent meaning that the explained variance is quite high. In Table 4.1 we report factor exposures and observe that they are coherent with the market capitalization associated to each sector. We apply the FastICA Algorithm to the GICS returns and find the components that will be the ones that maximize the non-gaussianity condition present in the optimization algorithm.

If we think the ICs to be the columns of a matrix \( S \), the risk factors time series can be seen as linear transformations of the independent signal sources. In the mixing matrix \( A \in \mathbb{R}^{n \times n} \) is contained the information about the weight of the single original source in the market sector, i.e \( X = AS \).

The portfolio return distribution can be obtained as a linear combination of \( n \) independent distributions. The Central Limit theorem suggests that the distribution of the single IC can be heavy tailed and/or asymmetric since what we observe is the realization vector of a sum of distributions.

We use the Jarque-Bera test to check for non-normality in the univariate case. Simultaneously...
sample skewness and kurtosis are compared with the corresponding normal values. In Table 4.2 we show the sample skewness, kurtosis and p-values for the Jarque-Bera test of the 10 ICs.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std</th>
<th>Skew</th>
<th>Ex-Kurt</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>VFIAX</td>
<td>0.0005</td>
<td>0.0121</td>
<td>-0.4693</td>
<td>3.9683</td>
<td>-0.0688</td>
<td>0.0463</td>
</tr>
<tr>
<td>COND</td>
<td>0.0007</td>
<td>0.0129</td>
<td>-0.5674</td>
<td>3.0566</td>
<td>-0.0690</td>
<td>0.0472</td>
</tr>
<tr>
<td>CONS</td>
<td>0.0006</td>
<td>0.0078</td>
<td>-0.3871</td>
<td>3.2413</td>
<td>-0.0390</td>
<td>0.0332</td>
</tr>
<tr>
<td>ENRS</td>
<td>0.0006</td>
<td>0.0158</td>
<td>-0.3878</td>
<td>3.3909</td>
<td>-0.0864</td>
<td>0.0687</td>
</tr>
<tr>
<td>FINL</td>
<td>0.0002</td>
<td>0.0175</td>
<td>-0.3415</td>
<td>4.1493</td>
<td>-0.1052</td>
<td>0.0789</td>
</tr>
<tr>
<td>HLTH</td>
<td>0.0006</td>
<td>0.0101</td>
<td>-0.4205</td>
<td>3.9130</td>
<td>-0.0540</td>
<td>0.0456</td>
</tr>
<tr>
<td>INDU</td>
<td>0.0004</td>
<td>0.0142</td>
<td>-0.4390</td>
<td>2.7993</td>
<td>-0.0711</td>
<td>0.0495</td>
</tr>
<tr>
<td>INFT</td>
<td>0.0007</td>
<td>0.0130</td>
<td>-0.2948</td>
<td>2.0849</td>
<td>-0.0596</td>
<td>0.0445</td>
</tr>
<tr>
<td>MATR</td>
<td>0.0005</td>
<td>0.0159</td>
<td>-0.3575</td>
<td>2.5824</td>
<td>-0.0756</td>
<td>0.0593</td>
</tr>
<tr>
<td>TELS</td>
<td>0.0007</td>
<td>0.0097</td>
<td>-0.2897</td>
<td>2.9581</td>
<td>-0.0550</td>
<td>0.0426</td>
</tr>
<tr>
<td>UTIL</td>
<td>0.0004</td>
<td>0.0089</td>
<td>-0.1124</td>
<td>4.7863</td>
<td>-0.0563</td>
<td>0.0414</td>
</tr>
</tbody>
</table>

Table 4.2: The reported statistics for the fund VFIAX and for the GICS show that the corresponding distributions are negatively skewed and with tails heavier than the one generated from a normal distribution.

The portfolio return can be decomposed in the form:

$$r_p = \beta^F F + \beta^N N + \epsilon.$$  \hspace{1cm} (4.22)

with $F \in \mathbb{R}^{l \times t}$ being the matrix containing the $l$ rows of the $S$ matrix containing the components we decided to be meaningful in the market and with $N \in \mathbb{R}^{(n-l) \times t}$ the remaining considered as noise. The new exposures $\beta^F$ and $\beta^N$ are obtained by multiplying of the initial exposures $\beta$ with the associated rows in the mixing matrix $A$. The linear decomposition of returns in terms of uncorrelated factors eases the computation of the characteristic function:

$$E[e^{iur_p}] = E[e^{i\sum_{t=1}^{t=1} \beta^F_{t}F_{t}}] = \prod_{i=1}^{l} E[e^{i\beta^F_{t}F_{t}}]E[e^{i\epsilon}]$$  \hspace{1cm} (4.23)

The linear relation in the portfolio return model can be used to compute the marginal contribution to return/risk of each of the chosen IC or in a portfolio optimization problem as in Madan [2006a]. We emphasize the fact that our main focus is not introducing a new method on how to use ICA in finance but to stress how the flexibility of the Mixed Tempered Stable distribution can allow to capture contemporaneously the different shapes of each IC.
As measures of fit we consider the Mortara index $A_1$ and the quadratic K. Pearson index $A_2$:

$$A_1 = \frac{1}{n} \sum_{j=1}^{S} |n_j - \hat{n}_j|$$

$$A_2 = \sqrt{\frac{1}{n} \sum_{j=1}^{S} \left(\frac{n_j - \hat{n}_j}{n_j}\right)^2}$$

where $S$ is the number of classes, $n_j$ are the observed frequencies while $\hat{n}_j$ are the theoretical ones. The $n$ is the summation of the theoretical frequencies.

Table 4.3 reports the estimated mixing matrix obtained using the FastICA algorithm.

<table>
<thead>
<tr>
<th>Mixing Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>-0.0035</td>
</tr>
<tr>
<td>-0.0008</td>
</tr>
<tr>
<td>0.0030</td>
</tr>
<tr>
<td>-0.0022</td>
</tr>
<tr>
<td>-0.0021</td>
</tr>
<tr>
<td>-0.0036</td>
</tr>
<tr>
<td>-0.0028</td>
</tr>
<tr>
<td>-0.0019</td>
</tr>
<tr>
<td>-0.0029</td>
</tr>
<tr>
<td>-0.0012</td>
</tr>
</tbody>
</table>

Table 4.3: Estimated Mixing Matrix obtained applying the FastICA algorithm to the dataset composed by the GICS. Only 8 out of the 10 Independent Components are chosen to use as factors. The remaining two are considered as noise.

The empirical densities of the independent components are shown in figure 4.4. In table 4.5 we give the MixedTS fitted parameters for each component and some measures of fit.

<table>
<thead>
<tr>
<th>Statistics Ics</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>JB-Statistic</td>
</tr>
</tbody>
</table>

Table 4.4: We report the skewness, kurtosis and JB-test for each component.

The four ICs with the highest JB statistic are considered as factors while the remaining ones as noise. The VFIAX return density was reconstructed using the MixedTS distribution for the factors.
4.3 Mixed Tempered Stable distribution

Figure 4.4: The MixedTS was fitted to each IC empirical density. The fitted parameters are reported in 4.5. It is easy to observe that the new distribution contemporaneously captures density shapes which are asymmetric and/or fat tailed.

and assuming normality for the noise. For comparison we plot the normal distribution fitted to the fund return density in figure 4.5.

Table 4.5: We fitted the MixedTS distribution to the ICs considered as factors in our model. In the table we show the parameters and the fitting measures.
Figure 4.5: The four ICs with the highest JB statistic are considered as factors while the remaining ones as noise. The VPIAX return density was reconstructed using the MixedTS distribution for the factors and assuming normality for the noise. For comparison we plot the normal distribution fitted to the fund return density.
Table 4.6: We fitted the MixedTS and VG distribution to the empirical density of each sector and obtained the corresponding parameters for both models.
4.4 Risk Measures using the Mixed Tempered Stable distribution

4.4.1 Saddlepoint Approximation

In some cases we need to approximate the tail of the return density. Edgeworth expansions are frequently used to approximate distributions when the higher order moments are available. This methodology seems to work well in the center of distribution but not in the tails. In particular, it often produces negative values for densities in the tails. Saddlepoint expansion can be understood as an refinement of Edgeworth expansion on the tails. Wong [2008] proposes the saddle point technique to derive the small sample asymptotic distribution for the sample expected shortfall under a standard normal null hypothesis. In Cizek et al. [2011] is given a general introduction about how to approximate the density and the cumulative distribution function (cdf) of a continuous r.v. $X$ whose moment generating function (mgf) $M_X(t)$ exists in an open set around zero. Here, we describe the main passages. We define the cumulant generating function (cgf)

$$K_X(t) = \ln [M_X(t)]$$

and the solution $\hat{s} = s(x)$ of the equation $x = K_X'(\hat{s})$ is the Saddle-Point at $x$ for every $x \in \text{support}(X)$. The first order Saddle-Point approximation for the density $f_X(x)$ is:

$$\hat{f}_X(x) = \frac{1}{\sqrt{2\pi K_X''(\hat{s})}} \exp \{K_X(\hat{s}) - x\hat{s}\}$$

while the approximation for the cdf $F_X(x)$ obtained by Lugannani and Rice [1980] is:

$$\hat{F}_X(x) = \Phi (\hat{\omega}) + \phi (\hat{\omega}) \left\{ \frac{1}{\hat{\omega}} - \frac{1}{\hat{u}} \right\}, \quad x \neq E[X]$$

for $\hat{\omega} = \text{sgn}(\hat{s})\sqrt{2\hat{s}x - 2K_X(\hat{s})}$, $\hat{u} = \hat{s}\sqrt{K_X' (\hat{s})}$. $\Phi (z)$ and $\phi (z)$ are the cdf and the density of a standard normal random variable. The condition $K_X'(0) = E[X]$ implies that $\hat{s} = 0$ is the saddlepoint for $E[X]$. Since $K_X(0) = 0$ and $\hat{\omega} = 0$ we can not use directly the expression for $x = E[X]$ but linear interpolation around the mean ensures continuity of the $\hat{F}_X(x)$. Second order approximation formula requires higher order derivatives of the cumulant generating function. Defining:

$$\hat{\kappa}_i = \frac{K_X'(\hat{s})}{(K_X'(\hat{s}))^{\frac{1}{2}}}$$

and

$$a_i = \frac{\hat{\kappa}_4}{8} - \frac{5\hat{\kappa}_4^2}{24}$$

(4.27)
the second order saddle-point approximation formulas for the density and the cdf are respectively:

\[
\begin{align*}
\tilde{f}(x) &= \hat{f}(x) (1 + a_1) \\
\tilde{F}(x) &= \hat{F}(x) - \Phi(\hat{\omega}) \left\{ \hat{u}^{-1} \left( \frac{\hat{\kappa}_4}{8} - \frac{5}{24} \hat{\kappa}_3^2 \right) - \hat{u}^{-3} - \frac{\hat{\kappa}_3^2}{2\hat{u}^2} + \hat{\omega}^{-3} \right\}, \quad x \neq E[X]
\end{align*}
\]

(4.29)

The saddlepoint approximation can be used for computing some risk measures, in particular Martin [2006] gives an approximation for the integral appearing in the Expected Shortfall (ES) formula:

\[
ES_\alpha(X) = \frac{1}{\alpha} \int_{-\infty}^{q} x f_X(x) dx \approx \frac{1}{\alpha} \left[ \mu_X F_X(q) - f_X(q) \frac{q - \mu_X}{\hat{\omega}} \right]
\]

(4.30)

Replacing the true density in (4.30) with the saddle-point first order approximation formula (4.25), the ES can be evaluated using the following first-order approximation formula:

\[
\hat{ES}_\alpha(X) = \Phi(\hat{\omega}_q) \mu_X + \sqrt{\frac{1}{2\pi}} \exp \left\{ -\frac{\hat{\omega}_q^2}{2} \right\} \left[ \frac{\mu_X}{\hat{\omega}_q} - \frac{q}{\hat{u}_q} \right]
\]

(4.31)

where \(\hat{\omega}_q, \hat{u}_q\) are evaluated in \(x = q\).

Using the second order saddle-point approximation (4.29) the corresponding Expected Shortfall approximated formula is:

\[
\hat{\hat{ES}}_\alpha = \Phi(\hat{\omega}_q) \mu_X + \sqrt{\frac{1}{2\pi}} \exp \left\{ -\frac{\hat{\omega}_q^2}{2} \right\} \left[ \frac{\mu_X}{\hat{\omega}_q} - \frac{q}{\hat{u}_q} + \left( \frac{q}{\hat{u}_q} - \frac{\mu_z}{\hat{u}_q^3} - \frac{\mu_X}{\hat{u}_q^3} + \frac{q\hat{\kappa}_3}{2\hat{u}_q^2} - \frac{q\hat{a}_1}{\hat{u}_q} - \frac{1}{\hat{a}_q^2} \right) \right]
\]

(4.32)

These formulas can be easily handled in a parametric context and the obtained estimates are more stable than the historical quantities.
4.4.2 Numerical Analysis

In the previous sections we tested the ability of the new distribution to reproduce the features of the returns observed in the market. Now, we focus on some methodologies that allow the computation of risk measures in a parametric context in order to exploit the results obtained for the Mixed tempered Stable distribution. As a first exercise, we fit the marginal MixedTS directly to each GICS sector index empirical distribution and in table 4.7 we report the estimated parameters for each index.

<table>
<thead>
<tr>
<th></th>
<th>VFIAX</th>
<th>COND</th>
<th>CONS</th>
<th>ENRS</th>
<th>FINL</th>
<th>HLTH</th>
<th>INDU</th>
<th>INFT</th>
<th>MATR</th>
<th>TELS</th>
<th>UTIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_0$</td>
<td>-0.0011</td>
<td>0.0035</td>
<td>-0.0023</td>
<td>0.0023</td>
<td>0.0041</td>
<td>-0.0078</td>
<td>0.0073</td>
<td>0.0030</td>
<td>0.0046</td>
<td>0.0116</td>
<td>0.0064</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.0021</td>
<td>-0.0016</td>
<td>0.0031</td>
<td>0.0000</td>
<td>-0.0019</td>
<td>0.0092</td>
<td>-0.0039</td>
<td>-0.0014</td>
<td>-0.0030</td>
<td>-0.0149</td>
<td>-0.0040</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.0045</td>
<td>0.0028</td>
<td>0.0062</td>
<td>0.0033</td>
<td>0.0047</td>
<td>0.0029</td>
<td>0.0069</td>
<td>0.0085</td>
<td>0.0048</td>
<td>0.0026</td>
<td>0.0040</td>
</tr>
<tr>
<td>$\lambda_+$</td>
<td>1.6337</td>
<td>1.9358</td>
<td>5.0716</td>
<td>1.8305</td>
<td>1.4665</td>
<td>6.8170</td>
<td>0.6007</td>
<td>1.8138</td>
<td>1.7252</td>
<td>2.1552</td>
<td>0.8187</td>
</tr>
<tr>
<td>$\lambda_-$</td>
<td>0.8290</td>
<td>2.0906</td>
<td>0.7012</td>
<td>2.0792</td>
<td>1.9103</td>
<td>0.7407</td>
<td>0.7000</td>
<td>0.7000</td>
<td>0.7000</td>
<td>0.7000</td>
<td>2.7837</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.7000</td>
<td>1.9900</td>
<td>0.7000</td>
<td>1.9900</td>
<td>1.9900</td>
<td>1.9900</td>
<td>1.9900</td>
<td>1.9900</td>
<td>1.9900</td>
<td>1.9900</td>
<td>1.9900</td>
</tr>
</tbody>
</table>

Table 4.7: Estimated Parameters for each Index in the dataset

We evaluate the cumulative distribution function using the formula that through the characteristic function $\phi_X(t)$ allows us to evaluate the cdf $F_X(x)$:

$$F_X(x) = \frac{1}{2} - \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-itx} \phi_X(t)}{it} dt$$

(4.33)

We consider the VaR computed using the Inverse Fourier Transform and compare it with that obtained with historical simulation approach. The comparison for 5% and 10% levels of confidence is reported in Table 4.8

<table>
<thead>
<tr>
<th></th>
<th>VaR$^{MixedTS}_{0.05}$</th>
<th>VaR$^{Emp}_{0.05}$</th>
<th>VaR$^{MixedTS}_{0.1}$</th>
<th>VaR$^{Emp}_{0.1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VFIAX</td>
<td>0.0131</td>
<td>0.0130</td>
<td>0.0085</td>
<td>0.0092</td>
</tr>
<tr>
<td>COND</td>
<td>0.0118</td>
<td>0.0121</td>
<td>0.0086</td>
<td>0.0097</td>
</tr>
<tr>
<td>CONS</td>
<td>0.0118</td>
<td>0.0107</td>
<td>0.0072</td>
<td>0.0079</td>
</tr>
<tr>
<td>ENRS</td>
<td>0.0169</td>
<td>0.0166</td>
<td>0.0101</td>
<td>0.0112</td>
</tr>
<tr>
<td>FINL</td>
<td>0.0155</td>
<td>0.0154</td>
<td>0.0098</td>
<td>0.0105</td>
</tr>
<tr>
<td>HLTH</td>
<td>0.0128</td>
<td>0.0123</td>
<td>0.0082</td>
<td>0.0095</td>
</tr>
<tr>
<td>INDU</td>
<td>0.0118</td>
<td>0.0120</td>
<td>0.0108</td>
<td>0.0097</td>
</tr>
<tr>
<td>INFT</td>
<td>0.0169</td>
<td>0.0145</td>
<td>0.0119</td>
<td>0.0111</td>
</tr>
<tr>
<td>MATR</td>
<td>0.0154</td>
<td>0.0168</td>
<td>0.0110</td>
<td>0.0140</td>
</tr>
<tr>
<td>TELS</td>
<td>0.0142</td>
<td>0.0129</td>
<td>0.0104</td>
<td>0.0104</td>
</tr>
<tr>
<td>UTIL</td>
<td>0.0118</td>
<td>0.0115</td>
<td>0.0085</td>
<td>0.0089</td>
</tr>
</tbody>
</table>

Table 4.8: In this table we compare the empirical VaR with the theoretical one obtained for the ten GICS and the VFIAX fund index.
For each index, we compare the empirical ES estimates with the estimates using four different methodologies.

1. Numerical integration of the expected value, i.e., the distribution is obtained through the Inverse Fourier Transform. We recall that, under the assumption of the existence for the $E(X)$ \(^1\), the Expected Shortfall can be written as (see Cizek et al. [2011]):

$$ES_\alpha(X) := E[X | X \leq x_\alpha] = x_\alpha - \frac{1}{\alpha} \int_{-\infty}^{x_\alpha} F(u) \, du$$

(4.34)

2. The second approach is based on Monte Carlo simulation. The random number generator is built using the Inverse Transform Sampling method (see Glasserman [2003]) based on the following two steps:

   a. Generate a random number $\hat{u}$ from a uniform distribution $U([0, 1])$.

   b. Find $x$ such that the equation $F_X(x) = \hat{u}$.

3. First order Saddle Point Approximation formula.

4. Second order Saddle Point Approximation formula.

Tables 4.9 and 4.10 report the ES computed with the four methodologies and in the last column we have the empirical ES obtained using the historical simulation. In this case the windows size is the last year of observations. In table 4.9, we have computed the ES at 5% level of confidence, while in table 4.10 the ES is at 10% level of confidence.

---

\(^1\)For the Mixed Tempered Stable distribution, this condition is ensured by the existence of moment generating function. The condition of existence of the first moment is less obvious when we obtain the Geo Stable as a special case.
### Mixed Tempered Stable distribution

<table>
<thead>
<tr>
<th></th>
<th>$ES_{0.05}^{MixedTS}$</th>
<th>$ES_{0.05}^{Fourier}$</th>
<th>$ES_{0.05}^{SPA1}$</th>
<th>$ES_{0.05}^{SPA2}$</th>
<th>$ES_{0.05}^{Emp}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VFIAX</td>
<td>0.0173</td>
<td>0.0179</td>
<td>0.0205</td>
<td>0.0204</td>
<td>0.0170</td>
</tr>
<tr>
<td>COND</td>
<td>0.0157</td>
<td>0.0158</td>
<td>0.0155</td>
<td>0.0164</td>
<td>0.0170</td>
</tr>
<tr>
<td>CONS</td>
<td>0.0176</td>
<td>0.0177</td>
<td>0.0200</td>
<td>0.0196</td>
<td>0.0163</td>
</tr>
<tr>
<td>ENRS</td>
<td>0.0184</td>
<td>0.0185</td>
<td>0.0182</td>
<td>0.0192</td>
<td>0.0230</td>
</tr>
<tr>
<td>FINL</td>
<td>0.0192</td>
<td>0.0194</td>
<td>0.0194</td>
<td>0.0202</td>
<td>0.0206</td>
</tr>
<tr>
<td>HLTH</td>
<td>0.0177</td>
<td>0.0188</td>
<td>0.0209</td>
<td>0.0204</td>
<td>0.0171</td>
</tr>
<tr>
<td>INDU</td>
<td>0.0210</td>
<td>0.0210</td>
<td>0.0235</td>
<td>0.0233</td>
<td>0.0197</td>
</tr>
<tr>
<td>INFT</td>
<td>0.0209</td>
<td>0.0206</td>
<td>0.0231</td>
<td>0.0223</td>
<td>0.0191</td>
</tr>
<tr>
<td>MATR</td>
<td>0.0215</td>
<td>0.0211</td>
<td>0.0212</td>
<td>0.0218</td>
<td>0.0233</td>
</tr>
<tr>
<td>TELS</td>
<td>0.0186</td>
<td>0.0188</td>
<td>0.0189</td>
<td>0.0192</td>
<td>0.0216</td>
</tr>
<tr>
<td>UTIL</td>
<td>0.0161</td>
<td>0.0161</td>
<td>0.0165</td>
<td>0.0165</td>
<td>0.0170</td>
</tr>
</tbody>
</table>

Table 4.9: In this table we give the results of computing the ES for $\alpha = 0.05$

<table>
<thead>
<tr>
<th></th>
<th>$ES_{0.1}^{MixedTS}$</th>
<th>$ES_{0.1}^{Fourier}$</th>
<th>$ES_{0.1}^{SPA1}$</th>
<th>$ES_{0.1}^{SPA2}$</th>
<th>$ES_{0.1}^{Emp}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VFIAX</td>
<td>0.0142</td>
<td>0.0147</td>
<td>0.0161</td>
<td>0.0166</td>
<td>0.0140</td>
</tr>
<tr>
<td>COND</td>
<td>0.0128</td>
<td>0.0129</td>
<td>0.0123</td>
<td>0.0141</td>
<td>0.0139</td>
</tr>
<tr>
<td>CONS</td>
<td>0.0132</td>
<td>0.0139</td>
<td>0.0158</td>
<td>0.0159</td>
<td>0.0126</td>
</tr>
<tr>
<td>ENRS</td>
<td>0.0153</td>
<td>0.0152</td>
<td>0.0145</td>
<td>0.0165</td>
<td>0.0184</td>
</tr>
<tr>
<td>FINL</td>
<td>0.0154</td>
<td>0.0155</td>
<td>0.0151</td>
<td>0.0173</td>
<td>0.0166</td>
</tr>
<tr>
<td>HLTH</td>
<td>0.0133</td>
<td>0.0145</td>
<td>0.0154</td>
<td>0.0156</td>
<td>0.0139</td>
</tr>
<tr>
<td>INDU</td>
<td>0.0165</td>
<td>0.0175</td>
<td>0.0183</td>
<td>0.0195</td>
<td>0.0153</td>
</tr>
<tr>
<td>INFT</td>
<td>0.0168</td>
<td>0.0183</td>
<td>0.0199</td>
<td>0.0198</td>
<td>0.0160</td>
</tr>
<tr>
<td>MATR</td>
<td>0.0172</td>
<td>0.0170</td>
<td>0.0168</td>
<td>0.0184</td>
<td>0.0191</td>
</tr>
<tr>
<td>TELS</td>
<td>0.0152</td>
<td>0.0155</td>
<td>0.0153</td>
<td>0.0160</td>
<td>0.0163</td>
</tr>
<tr>
<td>UTIL</td>
<td>0.0129</td>
<td>0.0131</td>
<td>0.0132</td>
<td>0.0140</td>
<td>0.0133</td>
</tr>
</tbody>
</table>

Table 4.10: In this table we give the results of computing the ES for $\alpha = 0.10$
The last comparison based on marginal distribution regards the expectile based risk measure. As observed in Chapter 2 a plausible value of the weight parameter for the considered dataset is 0.012. The parametric expectile is obtained by using the second approach described for the ES computation, i.e by using Monte Carlo technique.

The results are reported in Table 4.11.

<table>
<thead>
<tr>
<th></th>
<th>$\tau_{0.012}$ MixedTS</th>
<th>$\tau_{0.012}$ Emp</th>
</tr>
</thead>
<tbody>
<tr>
<td>VFIAX</td>
<td>0.0135</td>
<td>0.0136</td>
</tr>
<tr>
<td>COND</td>
<td>0.0127</td>
<td>0.0134</td>
</tr>
<tr>
<td>CONS</td>
<td>0.0130</td>
<td>0.0132</td>
</tr>
<tr>
<td>ENRS</td>
<td>0.0141</td>
<td>0.0184</td>
</tr>
<tr>
<td>FINL</td>
<td>0.0156</td>
<td>0.0162</td>
</tr>
<tr>
<td>HLTH</td>
<td>0.0152</td>
<td>0.0135</td>
</tr>
<tr>
<td>INDU</td>
<td>0.0156</td>
<td>0.0157</td>
</tr>
<tr>
<td>INFT</td>
<td>0.0183</td>
<td>0.0152</td>
</tr>
<tr>
<td>MATR</td>
<td>0.0173</td>
<td>0.0188</td>
</tr>
<tr>
<td>TELS</td>
<td>0.0156</td>
<td>0.0175</td>
</tr>
<tr>
<td>UTIL</td>
<td>0.0137</td>
<td>0.0139</td>
</tr>
</tbody>
</table>

Table 4.11: We fix $\tau = 0.012$ coherently with the value obtained in chapter 2 for the VFIAX fund and then compare the parametric and non parametric results.
Chapter 5

Conclusions and Future Research

In this thesis we investigate some methodologies applied to risk attribution analysis. Based on the existing literature, we emphasize the problems encountered in their implementation and proposed some modifications that were supported by the empirical analysis.

In the first three chapters we focus on non parametric approaches and study several methods for identifying the factors that most influence asset or portfolio returns. In Chapter 2 from the empirical analysis performed, it seems that \( \tau \) can be defined as an increasing function of the kurtosis but further investigation is needed in order to model the weight parameter. In Chapter 3, starting from the seminal work of Menchero and Poduri [2008], we obtain a model through which we are able to decompose the return vector in three uncorrelated quantities: the user defined factors, the projected factors and the idiosyncratic error term. It is easy to implement and each risk component has a financial interpretation. In the future, we will apply this methodology to more complex portfolios and for a higher number of custom risk factors considered.

In the last chapter, we propose a new distribution, namely Mixed Tempered Stable. The flexibility of this distribution, supported from our empirical analysis, derives from its capability of having as special cases distributions that are able to fit the tails.

The study of risk attribution using the MixedTS will be object of future research. The definitions of modified VaR Zangari [1996] and modified ES Boudt et al. [2009] that use asymptotic expansions can be a starting point. Another interesting argument is the study of efficient algorithms that generate sparse linear models as the one recently introduced in Zhang et al. [2009]. In a parametric context, with a large number of factors considered, they facilitate numerical implementation of parameters estimation.


