RISKSize Service

Basic Mode – Methodology

10/01/2025

Contents

Introduction

The aim of this document is to illustrate the methodology used by PROMETEIA to estimate the volatility vector and the correlation matrix for a vast number of risk factors characteristic of financial markets.
The techniques used are compliant with market best practices which by now are widely consolidated and documented.[1]

The daily estimates process could be broken down into three main phases:

- data collection and preliminary processing (if necessary);
- estimate generation;
- output data delivery.

In the beginning of the process a very important choice has to be made, which is the selection of the risk factors to be included in the matrix. As mentioned before, Prometeia 's choice is consistent with that of other international providers. However the preference has been to include those risk factors which best represent the needs of users inside and outside Europe. This section gives a general view - by macro classes – of the risk factors present in the RISKSize matrix.

The first part of this document is dedicated to the description of the inputs used and the output data generated, whereas a complete list of risk factors may be found in the document "RISKSize – Contents". The last part contains a discussion of number of methodological aspects.

Please note that, in order to guarantee constant updating and monitoring, both the perimeter of risk factors, the methodological framnework and the data downloading procedures can be constantly updated and reviewed.

[1] J.P. Morgan, RiskMetricsTM, Technical Document, Fourth Edition, New York, 1996.

Chapter 1

Input Data

1.1 Data source

At the moment, RISKSize makes use of international providers' services and data in order to create and update the database underlying the estimation phase. These data are controlled, corrected and completed (missing, abnormal, non-significant data) by comparing information coming from different data providers and market agents. The following paragraphs describe the most important data sources selected and their in-depth examination.

1.1.1 Money Markets

Money market rates are drawn from various sources, such as local interbank market interest rates, deposit rates or other short term benchmark (e.g. bills).

1.1.2 Government rates and swap curves

In general, interest rate curves for Government rates and Swaps are derived either from intraday YTMs (Yield To Maturity) provided by government securities or from swap rates of the main regulated and OTC markets (input data).

The time choice for data collection is critical as there is a tradeoff between the necessity to synchronize the data collected and to ensure their maximum liquidity.

1.1.2.1 Government rates

In order to better represent the movement of government rates movement under analysis, it is necessary to choose the most liquid government securities for each of the maturities considered. For this purpose, a proprietary benchmark securities list has been set for each maturity and country of issue. The list is constantly updated and monitored.

1.1.2.2 Swap rates

Estimates of the term structure for zero coupon bonds are derived from daily coupon rates in terms of notional amount.

1.1.3 Government Spread curves

RISKSize service contains also the risk factors of the government spread curves such as ESP (Spain), FRF (France), ITL (Italy) and ITLCPI (Italy Inflation Linked) currencies.[1] These government spread curves are computed as the simple difference between the reference buckets of the government zero and the associated Swap zero coupon curve.

1.1.4 Sector spread curves

RISKSize provides sector spread risk factors calculated as the difference between zero-coupon interest rates from the sector curve and from risk-free interest rates. Sector curves are classified according to main currencies (EUR, GBP and USD), industrial sectors (Financial and Industrial) and issuer creditworthiness. The codes identifying risk factors are composed of three letters representing the currency followed by the symbol ".", then three letters representing the industrial sector and two letters representing the issuer creditworthiness. This last part of the code is a short code identifying the subordinated curves and the creditworthiness. This is identified as a synthetic code of values HIGH, MEDIUM, LOW, JUNK and SUBINV.

The preliminary stage of the determination of the spread sector curves market prices of plain[2] bonds quoted on main European exchanges are collected and every security is mapped onto a sector curve based on currency of denomination, industrial sector and creditworthiness of the issuer.

1.1.5 Sector spread curves from regression

In parallel with the sector curves calculated as in Section 1.1.5, sector curves obtained through a cross-sectional regression approach are estimated. To estimate these values, it is assumed that an issue's Z-spread is defined by the contribution of exposures to five mutually independent risk factors:

- geographical area;
- sector;
- creditworthiness;
- seniority;

[1] For the full list of risk factors refer to the document RISKSize Basic – Contents.

[2] In this context, plain bonds are bullet with fixed interest,or multi coupon not subordinated

- issue currency.

Under the assumption of independence between the marginal contributions of the various axes of analysis, the Z-spread of an issue at a specific date can be represented by the following functional form:

$$
\begin{aligned}
Z - Spread = \exp(\text{Intercept} &+ \beta_{area} + \beta_{sector} + \beta_{creditworthiness} \\
&+ \beta_{seniority} + \beta_{currency} + \beta_{timet} + \beta_{timet^2} \\
&+ \beta_{timet \cdot currency} + \beta_{timet \cdot creditworthiness})
\end{aligned}
$$

where

- Intercept represents the value of the intercept
- β represents the exposure to the corresponding risk factor.

Exposures to risk factors are estimated by means of a linear cross-sectional regression in which:

- the dependent variable is the logarithm of the Z-Spreads of the bonds available on the market;
- the explanatory variables are a set of dummy variables each representing an element of the risk factor domain plus the continuous time-related variables.

  In particular, for each Z-Spread, a dummy variable will be set equal to 1 in the case that the issue of the same is characterised by the element identified by the dummy, 0 otherwise.

  The algorithm for estimating sector spreads can therefore be summarised in the following steps:

  1. collection of the closing values of the quoted issues;
  2. valuation of the dummies associated with the quoted issues collected;
  3. estimation of betas representative of risk factor exposures through cross-sectional regression;
  4. estimation of sector spreads from previously estimated betas.

1.1.6 Real interest Rate Curves

RISKSize dataset includes "real" risk factors that allow to extract the trend of expected inflation. The values are expressed as "real", i.e. the Interest Rate Swaps (IRS) are deprived of the inflation-linked component. The value of each bucket is calculated from the zero-coupon rates of Interest Rate Swaps and the zero-coupon rates of Inflation-Indexed Swaps. Inflation-Indexed Swaps are extracted at the same time of IRS (as explained in paragraph 1.1.13). In Inflation-Indexed Swaps each part agree to exchange, after M years, a fixed payment with a floating payment that depends on the trend of the Consumer Price Index. The floating leg of the Swap is then inflation-indexed according to the inflation occurring in the interval (0,M).

1.1.7 Exchange stock market indices

The levels used for the estimates are downloaded at approximately 15:30 GMT producing in such a way data that, as explained in the methodological part of this document, reduces the asynchrony between the risk factors present in the matrix. For countries such as Australia, Japan and New Zealand, where the markets' trading activities are already closed at that time, the values used are the closing levels of the trading session.

1.1.8 Sector stock market indices

Sector indices divided by geographical areas have been chosen as risk factors in order to follow more closely and accurately the different equity components potentially present in a generic portfolio. Indeed, an association to a single market index has been deemed too weak, as such an approach does not account for the specific characteristics of different securities in terms of volatility and multivariate correlation structure. Conversely, a breakdown by geographical and sectoral categories highlights these specific characteristics of the various assets.
Codification of indices makes it possible to determine the geographical area and the sector considered. Sector codes used in the RISKSize Service consists of two parts: the first part identifies the geographical area code to which the index belongs; the second part describes the sector of reference as well as the specific market.

1.1.9 Exchange rates

Exchange Rates are downloaded at about 14.00 GMT, which coincides with the hour of government and swap rates data collection.

1.1.10 Benchmark indices for Mutual Funds

The indices provided represent a partition fine enough to classify each mutual fund independently of its nature (equity, balanced, bond or mixed) and geographical position.

1.1.11 Commodities

With reguard to commodities, risk factors included in the matrix are precious metals like gold, silver, platinum as well as other goods like copper and aluminium listed on the London Metal Exchange (LME). Other risk factors concerning the energy market (petrol, oil and natural gas) listed on the New York Mercantile Exchange (NYMEX) are also considered.

The names used to identify these risk factors are made up of two parts. The first part is common to all the commodities and represents the currency of reference that in our case is EUR, followed by a prefix 'C' and a number of characters identifying each commodity.

1.1.12 Hedge Funds

Among risk factors included in the matrix are also Hedge Fund Indices which are set according to their management strategies ("pure strategies"). Such risk factors are usually converted into the base currency. However, if it is necessary to cover hedge funds or funds of hedge funds that follow 100% exchange risk hedging, foreign currency hedge fund indices may be added. The reason for this choice is that a foreign currency index is able to much better explain the variability of the fund than an index converted into the base currency.
Unlike other risk factors' time series, some hedge fund index data are provided with a monthly frequency. This leads to a modification of the estimation methodology for volatility and correlations. A different decay factor is needed. Additional information on this topic may be found in the methodology section.

1.1.13 Data reference time

The following table synthesizes different approaches to input data collection according to different types of securities.

Security type	Observation time	Data availability	Data source
Commodities	At the end of the trading session	Second to last data available	Data quoted on LME and NYMEX
Currencies	14.00 GMT	Last data available	Tick by tick intraday data from major contributors
Hedge Funds	Not applicable	Monthly data	Indexes as average of securities with the same strategy
Equity indices	15.30 GMT for European and US indices. For pacific area indices and for equity sector index data, at the end of trading days	Last data available for Exchange traded indices and the second last data for sector indices.	Major stock indices of the world
Fund indices	At the end of the trading session	Second to last data available	Market indices
Government bonds	8.00 GMT for Asia/Pacific area, 13.00 GMT for the other countries	Last data available	Tick by tick intraday data coming from major contributors.
Money market rates	13.00 GMT or at the end of the trading session.	Last datum available for the main currencies or second to last data available.	Data coming from local interbank market data, local deposit or other short-term benchmarks.
Sector spreads	At the end of the trading session	Second to last data available	Closing data from main corporate bonds exchanges.
Swaps	8.00 GMT for Asia/Pacific area, 13.00 GMT for other countries	Last data available	Tick by tick OTC intraday data of the major contributors

Table 1.1: Data reference time

1.2 Further processing of input data

Market data, coming from the data providers, are not always available for volatility and correlation estimation. While exchange rate and stock index data may be immediately available, this is not always the case for government securities rates with a given maturity though, is implied in the price of those securities. Stock index and sector data as well as exchange levels and commodity prices are thoroughly controlled without any prior processing, whereas other data are subject to preliminary handling.

1.2.1 Money market rates

Daily provided money market rates are expressed according to different calculation conventions given that standards are differentiated by geographical areas. In order to render interest rates comparable, it has been decided to express all of them according to a unique ACT/365 convention by means of a simplified procedure for the rates unavailable in the ACT/365 convention. The conversion applied is an approximation based on a correction factor that accounts for the difference in the number of days in the base. For instance, the level of rates "ACT/360" is multiplied by the fraction 365/360 getting an approximation of the ACT/365 rate. In this way any money market rate is uniformly expressed as an annual compound zero coupon ACT/365 rate.

1.2.2 Government and swap rates

In order to get returns for different maturities present in the RISKSize matrix, first, it is necessary to find zero coupon rates term structure from market data. For each government/swap (nominal) curve it is possible to calculate a standard maturity grid. This grid ranges from a minimum maturity of 3 months to a maximum maturity of 30 years according to data availability. Input data are government securities YTMs for different expirations or par swap rates negotiated in the market.

The following is a description of a number of preliminary calculation for the government/swap curves.

1.2.2.1 Calculation of a one year swap rate

For major currencies the one year swap rate is found using FRA (Forward Rate Agreement). It produces a rate consistent with the buckets of the curve. In order to derive one year swap rate we use a formula based on the assumption of the absence of arbitrage under the simple compounding hypothesis (both money market and forward rates are expressed as simple rates), that is:

$$(1 + r_{6M}t_1)(1 + F_{6x12}t_2) = (1 + s_{1Y}t_3)$$

where:

- r_{6M} is a six month money market rate;
- F_{6x12} is a FRA that starts six months after cut-off date and with a tenor of six months;
- s_{1Y} is a one year swap rate;
- t_1, t_2, t_3 indicate the year to maturity of the rates r_{6M}, F_{6x12}, s_{1Y}

Once the one year swap rate is determined, by solving the previous equation with respect to the variable s_{1Y}, such a rate is converted into annual compound consistently with the convention adopted for all the buckets of the curve.

1.2.2.2 Linear interpolation

The first step in finding zero coupon rates from the input data is to determine the term structure for all the buckets of the maturity grid defined before. This method is called linear interpolation. There many of methodologies used for interest rate interpolation (e.g. by using linear interpolation or by constructing "splines" having different functional forms). In the case of government and swap rates it was decided to use the linear interpolation as it produces a curve consistent with data on benchmark securities whose number is limited and that are well-distributed according to different maturities.

1.2.2.3 Bootstrap

In order to get the zero coupon rates curve it is possible to use, two different approaches: the parametric model or the bootstrap method. In order to deliver as neutral and unbiased information as possible , it was decided to avoid the parametric model for description of the term structure evolution. Bootstrap is an iterative algorithm used to derive a term structure starting from the first rate.

1.2.3 Sector curves

To obtain the sector spread curves, the first step is to drive the term structure of zero-coupon yield curve of each corporate curve sector.
Even in this case, after interpolation, follows bootstrap procedure required to find the zero-coupon rate suitable to obtain spreads. With regards to the interpolation, is not convenient to use the method of linear interpolation used in the case of the government and swap curves, as numerous bonds with an high dispersion of the yield to maturity are available. Between different functional forms proposed in the literature, an exponential variant of the Nelson & Siegel model was chosen to capture term structures with

monotonous trend, with hump or sigmoidal and any inflection points in the distribution of securities of a curve. Furthermore it is a parsimonious model.

Once estimated, the model parameters determine the structure of the theoretical yield to maturity on the due dates of the grid defined above. The next step consists of applying the algorithm of bootstrap for the calculation of the zero-coupon rates. This method is the same as used for government and swaps. Finally, for each relevant bucket, the spread is calculated as the difference between the value of the zero coupon rate obtained previously and the rate of the associated risk-free zero coupon curve.

1.2.3.1 Financial Sub-Investment Grade curve

For the development of the Financial Sub-Investment Grade sector curve, the standard approach described in the previous section, based on the extrapolation of a zero-coupon rate curve from the yield to maturity of listed securities, may not always be robust. In fact, the data relating to this perimeter (European sub-investment grade financial issuers) are very heterogeneous both in terms of yields and the number of listed securities for each issuer.

In order to develop this curve, an approach based on data from the CDS market was adopted.
The methodology used makes it possible to expand the basket of issuers by reducing the weight of each one within it, thus increasing diversification and robustness over time.

1.2.4 Benchmark indices for Mutual Funds

The values referring to these indices as offered by different data providers are expressed as index numbers. Daily returns are then calculated from these values as a logarithm of a level ratio of two subsequent days. The returns thus obtained represent core values for price volatility calculation.

Consequently it becomes possible to monitor daily variations of the index considered. However a problem arises with respect to the indices denominated in a currency different from the base currency. As a matter of fact, the variation in the value of an investor's position on such indices does not depend only on the variation of the index level but also on the daily exchange rate variation. In order to account for this combined effect (level variation + exchange rate variation), it was decided to convert all the indices into the base currency according to the daily exchange rate. In such a way, daily returns, considered as daily log-variations of the new levels, are going to take into account jointly the two sources of variation. Separate consideration of the risk component arising from the exchange rate becomes redundant.

Similarly, an investor willing to use the data provided to calculate the volatility of foreign currency position with respect to a specific index, has to first convert such a position in the base currency using daily exchange rate.

Chapter 2

Output Data

For each risk factor the following data are updated daily: level, price volatility, interest rate volatility (if applicable), correlation with other risk factors and the last year historical return. What follows is an explanation of the estimates produced as well as of the data structure of the matrices available on www.risksize.com.

2.1 Estimates generated

2.1.1 Level

The value of the level present in the RISKSize matrix has different characteristics depending on whether the data are directly observable on the market or rather obtained after a series of preliminary calculations. The first category includes stock indices, exchange rates, money market interest rates and commodities. The second category contains government securitiy rates as well as swap and spread rates. Levels referring to interest rates are expressed in percentage terms.

2.1.2 Price and Yield Volatility

For a number of risk factors the RISKSize matrix contains information regarding both the price and returns volatility. It is worth mentioning that volatility calculated for each risk factor is always based on returns expressed as log variations of the level. However levels of the original factors may be expressed in terms of prices or rates. This fact justifies the adoption of several volatility measures: price volatility and yield volatility. For risk factors such as interest rates and spread rates both volatility measures are provided, while for the remaining risk factors only price volatility is calculated.

Consider, for instance, an investor exposed to a specific rate with a given maturity (more likely an investor is going to be exposed to multiple maturities according to the cash flows generated by the security but for simplicity an assumption of a zero coupon security bond is made). In order to determine portfolio value variation it is not enough to use the variation of interest rates for the same maturity. First, it is necessary to find prices deriving from such a rate. More specifically, being Z_{10T} 10-year rate at time T, the price of a bond paying 100 in 10 years is given by:

$$P_T = 100 * \frac{1}{(1+Z_{10T})^{10}}$$

If Z_{10T+1} is 10 year rate at time $T+1$; the value of a bond paying 100 at maturity is:

$$P_{T+1} = 100 * \frac{1}{(1+Z_{10T+1})^{10}}$$

It becomes possible to calculate the one day return as log-variation of the two consecutive prices, that is:

$$R_{T+1} = \ln(\frac{P_{T+1}}{P_T})$$

Based on these returns, it is straightforward to calculate volatility by means of an exponentially weighted moving average (EWMA). This is the price volatility value defined above.

Finally, another volatility estimate is provided based on the log-variations of the rates. In this case the returns are calculated as follows:

$$\bar{R}_{T+1} = \ln(\frac{Z_{10T+1}}{Z_{10T}})$$

The next step is the estimation of interest rate variation volatility. Note that although the series of the two variables refer to the same time period ,they cannot be directly compared, being based on different elements.

2.1.3 Historical returns

The column "MEAN" in the output files contains one year historical percentage returns. This value is calculated using levels of the risk factors such as stocks, exchange rates, commodities and benchmark indices whereas for government securities, swaps and money market rates the return includes both price variation and the interest accrued. Technically speaking, the main goal is to produce Total Return and Constant Maturity indices. For further details, consult the methodological section (paragraph 3).

2.1.4 Risk factor encoding

In general, every risk factor present in the matrix is univocally identified by means of a code composed of two parts. The first part identifies the geographical area of the input data. The second part reveals the nature of the risk factor.

2.2 The matrices

Risk factor data are updated on a daily basis. The methodology describing the time and methods of data collection for each risk factors is described in paragraph 2.1.10 above. The matrices vary according to investment horizon chosen (daily or monthly). Both of them are compressed before the transfer to speed up the download via internet according to the ZipTM standard.

The file name is based on the date of matrix creation in the format yyyymmdd (year/month/day) where all the characters are numbers with no interruptions in between. For instance, a matrix with a daily forecasting horizon produced on January, 20 2003and updated on January, 20 2003 is indicated as 20030120.zip. A matrix with a monthly horizon has additional letters "MH" (Monthly Horizon) attached to its name. Each of the two zip files contains two other files, necessary for variance-covariance matrix construction, as well as some other information.

The first file also contains volatilities in addition to historical returns and levels when applicable. The second file includes risk factor correlations arranged in a single column with an indication of the factor considered. The files have different extensions depending on the time horizon of the matrices (.DVF and .DCF for daily horizon and .MVF and .MCF for monthly horizon).

Chapter 3

Methodological Aspects

3.1 Volatility and correlation calculation

3.1.1 Daily returns

The analysis is based on the assumption of normality of risk factor returns calculated using the continuous capitalization method (logarithmic returns). Normality hypothesis is acceptable only if referred to returns and not to levels. Figure 3.1 explain this concept:

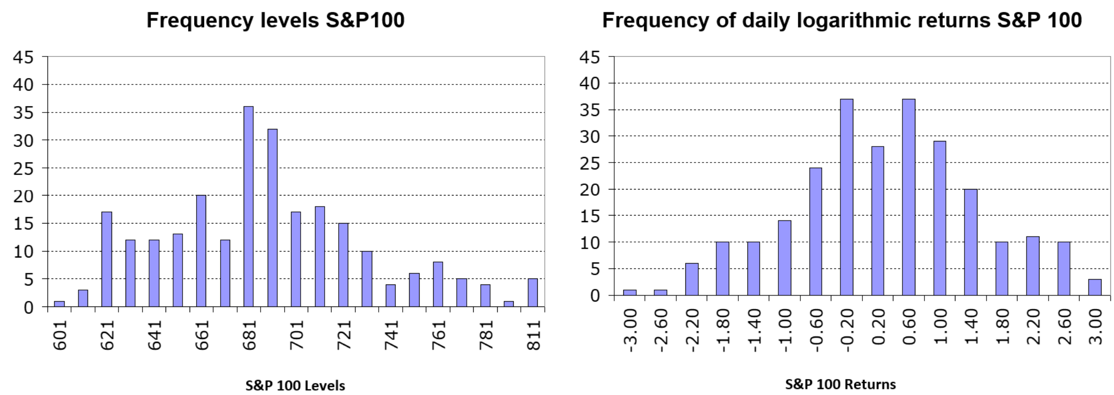

Figure 3.1: Normal distribution of returns

This example presents histograms of daily levels and daily logarithmic returns of the S&P100 index for the year 1999. The typical "bell" shape of the Gaussian distribution is observable also in this case. It is more concentrated around the mean and has thinner tails. The histogram relative to the index levels accurately shows the upward trend present in the year of observation. This result is common to the whole set of risk factors over different time horizons. Logarithmic returns (continuous capitalization) has been preferred to percentage returns (annual compound capitalization) for practical reasons.

Logarithmic returns aggregate nicely across time (total period return is a sum of the sub period returns) whereas percentage returns allow aggregation across assets of a portfolio. When different time horizons are considered it appears more appropriate to choose logarithmic returns.

3.1.2 Mean daily returns

To estimate the variance of each factor's returns as a measure of dispersion of returns around the mean, it is necessary to define the mean value of reference. Figure 3.2 presents a mean estimate of the S&P 100 daily returns over the year 1999 based on a 30 day rolling window. Next to the chart is shown a histogram of the mean distribution over 30 days.

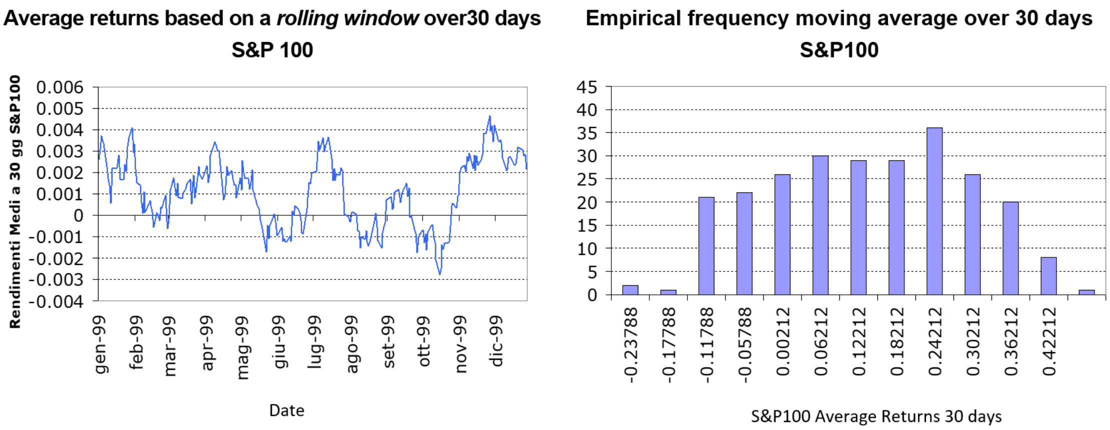

Figure 3.2: Daily returns

The oscillation of the mean around zero is evident. This result is quite reasonable since, when considering daily returns, the increase in the value of the index turns out to be insignificant over such a small period of time.

In the short-term the mean return has a negligible impact as it is dominated by the volatility in terms of absolute values. Therefore a zero mean assumption has been made to estimate volatility and correlation. The result of such a choice is a small increase of the estimated variance especially during the periods characterized by a strong trend in the underlying series. As a matter of fact, it is always possible to break down the variance with respect to a constant (zero in this case) into effective variance and distortion. The latter is nothing else but a squared difference between the true mean and the constant. This approach ensures stability in the estimates over time and represents, prudential attitude particularly appropriate for market risk measurement and control.

3.1.3 Exponentially weighted approach

The choice of the number of observations as well as of the weighting method to be used is very important since all the parameters must be carefully calibrated in order to generate an estimate consistent with future risk factor volatility. This choice must account promptly for the changes in the market conditions. At the same time it must be able to guarantee a certain level of stability of the volatility estimates in order to reflect only relevant risk factor variations and not just temporary oscillations. Equally weighted method attributes the same importance to all the observations of the period producing a "smoothed" measure of volatility. This may causes problems as recent events affecting markets are going to have a deferred impact on volatility estimates.

On the contrary, an approach attributing a higher importance to recent observations while giving little weight to the past ones may cause the over-fitting problem as it underestimates the past history and leads to overreaction to the new information. RISKSize matrix adopts the exponential weighted approach with a different decay factor depending on the time horizon considered (daily/monthly). This approach attributes higher importance to the past history when monthly estimates are concerned while, at the same time, maintaining the sensitivity to recent events at a high level.

Figure 3.3 is a graphical representation of the standard deviation of S&P100 daily returns distribution for the period between the beginning of January 1999 and February 2001. The value displayed is expressed in percentage terms and is consistent with the 95% percentile of the Standard Gaussian Distribution being already multiplied by 1.645 (in the same way as in the RISKSize matrix). The series estimated are based on an equally weighted and on a exponentially weighted methods. It becomes clear that exponentially weighted volatility estimates are much more sensitive to market conditions.

The figure beside the chart shows a sequence of exponential weights calculated using a decay factor λ = 0.94 and a one day variance-covariance matrix.

The following equations give a rigorous explanation of how volatility and correlation are calculated.

- volatility for risk factor (j):

$$\sigma_j = \sqrt{\frac{\sum_{t=1}^{T} \lambda^{t-1} \times r_{j,t}^2}{\sum_{t=1}^{T} \lambda^{t-1}}}$$

- covariance between two risk factors (i, j):

$$\sigma_{i,j} = \frac{\sum_{t=1}^{T} \lambda^{t-1} \times r_{i,t} \times r_{j,t}}{\sum_{t=1}^{T} \lambda^{t-1}}$$

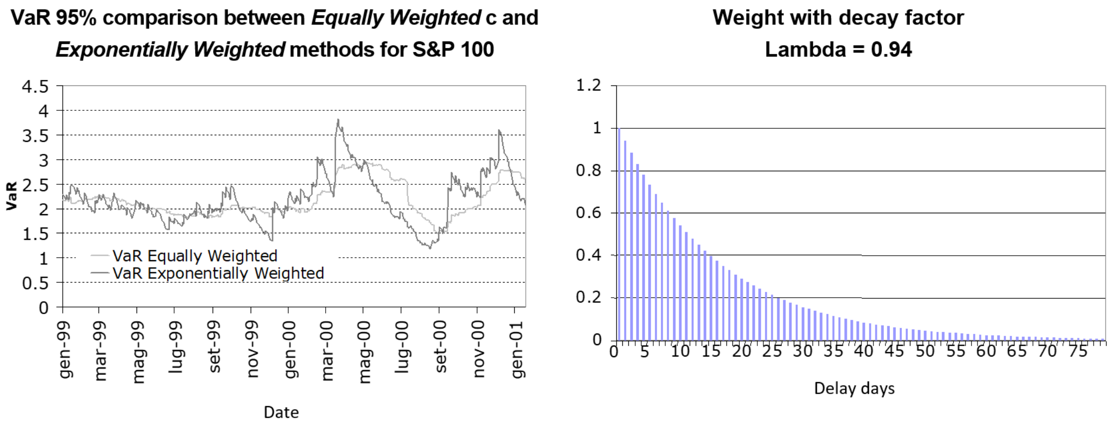

Figure 3.3: Decay factor

where both the decay factor λ and the number of observations considered are results of a specific choice.

Volatility values are consistent with the 95% percentile of the Standard Gaussian Distribution being already multiplied by 1.645. Moreover, monthly volatility is calculated using daily volatilities scaled to a monthly value with the help of a square root of time (25 days).

In conclusion, with σ volatility calculated on a daily basis, the following relations are used for calculations :

- Volatility given in the daily RISKSize matrix: $1.645 \times \sigma_j \times 100$
- Volatility given in the monthly RISKSize matrix: $1.645 \times \sigma_j \times \sqrt{25} \times 100$

3.1.4 Estimation methodology for hedge fund indices

Before estimating volatilities and correlations using the EWMA approach, it must be revised according to the availability of the underlying historical series at a monthly frequency.

1. First of all, it is necessary to verify whether, even for low frequencies, volatility clustering effect is present justifying the GARCH approach expressed in its simplest form GARCH(1,1) as follows:

$$\sigma^2_{t|t-1} = \omega + \alpha\epsilon^2_{t-1} + \beta\sigma^2_{t-1|t-2}$$

2. Once this effect is detected it is essential to evaluate the possibility of using a particular I-GARCH (Integrated GARCH) that has the following analytical form:

$$\sigma^2_{t|t-1} = (1-\lambda)\epsilon^2_{t-1} + \lambda \times \sigma^2_{t-1|t-2}$$

3. The parameters of the I-GARCH process must be estimated. The process is going to be approximated by a variance EWMA during the estimation phase.

Prometeia has revised all the logical steps described above finding that for all the Hedge Fund strategies present in the matrix:

1. GARCH effect is present for the frequencies considered
2. in most cases this effect may be expressed as I-GARCH;
3. estimation of I-GARCH parameters has produced an optimal decay factor λ = 0.8 for a variance EWMA process.

As an example, General Hedge Fund index has been considered for the period between March 1994 and June 2003 (110 monthly data points).

1. The table presents the results of the GARCH (1,1) estimation and as it emerges, there is a GARCH effect (parameters α and β are significant at 10% , 5% and 1% level);

Dependent Variable: Hedge Fund General Index, Method: ML – ARCH (Marquardt)

	Coefficient	Std. Error	z-Statistics	Prob
C	$1.62E-06$	$1.64E-06$	0.10	0.92
α	0.29	0.10	2.96	0.00
β	0.76	0.06	12.38	0.00
Log likelihood	246.62			

Table 3.1: Decay factor estimation for monthly hedge fund indices

2. After the GARCH effect has been detected it becomes possible to estimate the I-GARCH process. The results are shown in the following table:

I-GARCH: model Wald test to verify non stationarity of the GARCH Model

Null Hypothesis:		$C = 0$	
		$\alpha + \beta = 1$	
Chi-square	1.06	$P - Value$	0.59

Table 3.2: Test di Wald

Considering that these findings confirm the soundness of the I-GARCH implementation, it is possible to introduce constraints to the model in order to get an estimate for the β parameter. This parameter could be interpreted as an optimal decay factor used for exponential weighting.

3. Based on the estimate made, a decay factor value of 0.8 is obtained. The results of this estimate are shown in the following table:

Constrained estimation on General Hedge Fund Index, 110 monthly data

Mean log-likelihood parameters	Estimates	Gradient	Std Error (-10.5243)
α	0.20	-0.31	0.05
β	0.81	-0.31	0.05

Table 3.3: Regression model estimation

3.1.5 Missing data processing

The problem of missing data processing is a very delicate issue with significant impact on the quality of volatility and correlation estimates produced.

For various reasons it may happen that returns for one or more days (even consecutive) are not available. This problem is worse than missing price level data since it leads to twice as many returns being unavailable. Moreover, for correlations, missing data for one asset makes it impossible to use data available for other risk factors.

RISKSize takes a neutral attitude towards missing data processing by avoiding data replacement, in standard conditions, through an estimation algorithm. It is inevitable that such algorithms make a number of assumptions regarding the missing data which affects the neutrality of the estimates. So, taking into consideration that information is missing for a specific asset and date, volatilities and correlations are estimated by using the available observations whose weights are properly rescaled.

In case of exogenous events that do not depend on Prometeia (e.g. September 11th, 2001), the following procedure will be followed. We will wait until 3:45 PM CET, which is the deadline by which we will try to retrieve all the necessary data. If, by that time, it will not be possible to retrieve all or part of the data, the following actions will be taken: The data related to the curves (government, swap and inflation swap) and the monetary rates will be replicated starting from the previous day's data, effectively generating a zero return for the day. The data related to the stock indices and exchange rates will be considered at the most recent available date.

3.1.6 Time discrepancy in data collection

Data asynchrony is of particular importance while estimating the correlation structure of risk factors. Given the diversity in the world's time zones and markets' working hours, last prices of different stock exchanges are registered at a different time. This generally causes an underestimation of the correlation as the prices in a market closing later may have moved in an independent way from the prices in a market already closed.

During a working day, a lot of markets open and close at a different time. In particular, East Asian markets open one day before the European and American markets (according to GMT) and close before other markets open (the case of Japan and Australia). For instance, it is reasonable to assume that the last prices of Japanese and Australian markets may influence European opening prices. Whereas the last prices of European markets may influence those of the US markets. Similarly, today's last prices of the US market may in turn have an impact on the opening of Asiatic markets tomorrow triggering a sort of chain drive.

The correlation matrix takes into account all this information and models its effects through a system of equations containing latent variables. The solution implies an increase in the simultaneous correlation calculated on the available data (correlation is calculated using data of the day before trying to account for the information transmission effect). In the previous example the correlation is calculated using data coming from Australian and American markets with one day delay.

Table 3.4 reports simultaneous and delayed correlations between the Japanese (NIKKEI 225) and US (S&P 100) stock market indices. These indices, labelled as JPY.SE and USD.SE, are registered daily in two distinct moments in time having a 14 hours difference. Past experience has proved that the correlation between these two factors is much higher that zero. Our findings are presented in the table below and as one may see the estimates differ significantly:

Though the simultaneous correlation is negative, the correlation calculated with one day delay is surprisingly higher than 50%, corresponding exactly to the 14 hours delay

Risk factor	Type of simultaneous correlation	Estimation value
JPY.SE(t) vs USD.SE(t)	Simultaneous correlation	-0.12
JPY.SE(t) vs USD.SE(t-1)	1 day delayed correlation	0.58
JPY.SE(t-1) vs USD.SE(t)	1 day delayed correlation	-0.01

Table 3.4: Matrix correlation (based on the data going from 24/6/2002 to 25/09/2002)

in the registration of the USD.SE data. Based on these results it was decided to map all the risk factors according to the time of data registration. Correlation values for each couple of risk factors are based both on the simultaneous covariance and on the delayed one. A different weighting coefficient is applied to each covariance. However this procedure is particularly delicate as the structure of the covariance matrix changes, leading to an uncertainty regarding its fundamental properties such as the property of being positive semi-definite. Therefore weighting coefficients are chosen carefully so that these properties are preserved.

3.2 Historical Return Calculation

The RISKSize matrix also contains risk factor historical returns over the last year of observation. Such returns are expressed in percentage terms. The period considered contains one year of data, 365 calendar days (366 for leap years) starting from the reference day of matrix generation. Certain degree of tolerance is allowed while determining initial and final (most recent) values. Let us suppose that the reference matrix was created on 13/11/2001. The returns are realized starting on 13/11/2000. It may happen that some prices are missing for the period under consideration. In this case it was decided to report a return anyway using the price of the nearest date, with a limit of 15 business days preceding the reference day and 10 business days after it. In such a way if there is no price available on 13/11/2001 it is possible to use the price of 12/11/2001 and so on until reaching the limit of 10 days. When no data for the initial date are available on 13/11/2001 the value of 14/11/2001 is taken or that of the next available day but without breaching the limit of 15 days.

If no data are available for this range of tolerance "NC" (not calculated) is entered into the matrix by default. The methodology used to calculate historical returns is different for factors such as government, swap and money market rates that require consideration of both the price variations and the interest accrued while holding the instrument that in turn depends on interest rate curve variations.

In order to correctly account for these variations when calculating risk factor returns it is necessary to find total return that incorporates both components. The value found in such a way is a constant maturity index by definition, and could be interpreted as the return an investor is going to realize by buying today a zero coupon bond and selling it tomorrow at market price, iterating this strategy over one year.

This index is calculated as follows:

$$P_T = \frac{1}{(1+r_T)^1}$$

where, P_T is the price of a zero coupon bond with one year maturity and r_T is one year spot rate.

Let us assume that the bond is going to be sold the next day at a new market price whose variation depends on two factors. The discount period is reduced by one day and market rates may have moved in the meantime. The new market rate is that of the spot rate curve at time $T+1$ for synthetic zero coupon bonds having a maturity of 364 calendar days. As a rule, such a rate may well be approximated by one year spot rate r_{T+1}. The new price then becomes:

$$\bar{P}_{T+1} = \frac{1}{(1+r_{T+1})^{\frac{364}{365}}}$$

The compound return of such an operation is calculated as:

$$R_{T+1} = \log(\frac{\bar{P}_{T+1}}{P_T})$$

This return may be expressed as a sum of two components: capital gain/loss and interest accrued. By denoting the price of a zero coupon bond evaluated at time T+1 and having a maturity of one year as P_{T+1}, the return at time $T+1$ becomes:

$$R_{T+1} = \log(\frac{P_{T+1}}{P_T}) + (\frac{1}{365})\log(1 + r_{T+1})$$

where the first component is given by a capital gain/loss of a one year zero coupon bond and the second one represents the interest accrued in a holding period of one day. In order to find the annual return, daily continuous returns are summed up and converted in compound returns expressed in percentage terms.

Chapter 4

Market data checks

Prometeia data management daily activity involves a series of checks from market data. The main activities regard:

- checks of the correct acquisition of each input data from different data providers during the working day, for each market data type;
- data quality checks on the downloaded data. The checks can be summerized as follows, for each risk factor:
  - checks of data availability;
  - check of persistent data, as repeated values erroneously provided by data vendors;
  - checks if data is above certains thresholds, and it automatically triggers the needs of a specific monitoring. The thresholds depend on marked data type and their variability and are constantly monitored.
- check of the values of estimated market data;
- check of the risk measures estimated from market data, in order to trigger the daily production of data feeds.

During each calendar day, every step automatically depends on the positive outcome of the preceding step. This allows the automatic delivery of the data feeds at the end of each day.

The following sections report the checks for each market data type.

4.1 Interest rates

For government rates, monetary rates, swap rates and inflation swap rates, automatic logs check the outcome of downloads from data providers and automatic alerts are triggered if some data are absent.

These checks monitor:

- the acquisition of each monetary rates, yields to maturity and par swap rates;
- the check of missing data, potentially comparing with other providers;
- the presence of national holidays to explain missing data.

Besides, reports are automatically generated to check interest rates. They allow to:

- check persistent data, as repeated values erroneously provided by data vendors, for the following downloaded data:
  - levels of monetary rates;
  - yield-to-maturity of benchmark instruments;
  - par swap rates of swap contracts.
- mark the variations in the level of monetary rates if they are above the threshold, in absolute and relative terms;
- check for each government and swap curve the daily variation of yields to maturity/par swap rate of single securities/swap contracts that are used to bootstrap curves;
- check for each government and swap curve the daily variation of zero-coupon government and swap rates estimated with bootstrapping;
- check the market data underlined in the monitoring procedures and potentially substitute them;
- check the Price Volatility variations above threshold levels compared with the previous day, in absolute and relative terms. The Price Volatility checks regard each time series of the risk factors derived, i.e.:
  - monetary rates;
  - zero-coupon government rates;
  - zero-coupon swap rates;
  - zero-coupon real rates;
  - zero-coupon government spreads.

4.2 Exchange rates, indices and commodities

For exchange rates, equity indices, fund indices, hedge fund indices and commodities, automatic logs check the download procedures from data providers and automatic alert warn in case of missing data.

These checks aim to verify:

- the download of levels of each market data every day;
- the availability of missing data on other data providers, to proceed with integrations;
- the presence of national holidays to explain missing data.

Subsequently, for these risk factors automatic report show sime alerts to:

- check if thresholds are triggered, in relative terms depending in the instrument type, compared to the previous day;
- check the marked data underlined by the monitornig procedures and potentially substitute/eliminate it;
- check persistent data, as repeated values erroneously provided by data vendors;
- check the variation of Price Volatility estimated from the time series of these risk factors above threshold levels compared to the previous day, in absolute and relative terms.

4.3 Spread sector curves

In the process of elaboration of the term structures of sector curves, automatic logs verify:

- the acquisition of the prices of bond issuances and of the necessary info to associate securities to each curve;
- the calculation of yields to maturity.

Then automatic reports support the control of the term structures of zero-coupon rates of each sector curve. These reports allow to:

- check the daily estimation of each curve;
- use a dispersion diagram "'yield to maturity-maturity"' to monitor *outliers* to eliminate (*clearing*). If a security is not representative, it is possible to investigate the substitution of:

  - the contributor or the market (if available);
  - the security with another with similar features (if available).

- check for each sector curve the daily variation of yields to maturity of the instruments used to bootstrap the curves;
- check for each sector curve the daily variation of zero-coupon rates estimated with bootstrapping and used in the calculation of spreads;
- check for each sector curve the daily variation of zero-coupon spreads versus the risk-free rate, if they are above the threshold, in absolute and relative terms;
- check the variation of Price Volatility of time series of the spreads above thresold levels compared with the previous day, in absolute and relative terms.

Chapter 5

Incident management and continuous improvement

The performance of RISKSize service are monitored with the periodic check of Service Level Agreements (SLA) defined in the agreements, with regards to:

- production and delivery of Data Feeds;
- output availability regarding single risk factors in accordance with methodologies and technical aspects reported in the documentation.

In the context of analytics production, data processing and data management, the control system plays a crucial role in promptly identifying any anomalies or malfunctions in the various phases of the process, from the collection of raw data to their final processing and to the delivery of outputs.

A key element of the system is represented by the continuous improvement process, based on the detection and management of incidents and subsequent remediation activities. In the event of an incident being detected, the system provides a series of actions aimed at immediately managing the situation, following a structured procedure which aims to progressively strengthen the control system.

Accidents are classified into four different degrees of severity, depending on the impact detected. When an incident occurs, a contingency resolution is implemented, proportionate to the degree of severity, to minimize the impact and quickly restore normal system functioning.

After the temporary resolution, an in-depth analysis of the causes is conducted, with the aim of identifying the roots of the problem and developing an intervention plan for a permanent solution. Incident management involves the drafting of a detailed report that documents:

- the description of the incident;

- the contingency actions adopted;
- the remediation plan to implement a stable solution.

Incident reports are made available to the governance structure responsible for overseeing the data processing process. These reports are analyzed during the planning of development activities to identify improvements in operational procedures and policies.

The incident analysis process is oriented towards the collection of useful evidence to support a continuous improvement approach. This approach allows operational procedures and controls to be progressively refined, ensuring an increase in efficiency and a constant reduction in the number of accidents.

Chapter 6

Vendors qualification (info data provider)

The process of activating and maintaining suppliers, in particular data providers for services offered by Prometeia Data Analysis Service (DAS) provides a structured path and guidelines to continuously ensure the quality of data underlying the processing and therefore greater efficiency of DAS services, with assessments that include both new suppliers and existing suppliers.

The process begins with a scouting phase, during which possible suppliers capable of satisfying specific needs are identified, followed by a preliminary collection of information to define the main characteristics of the service.

The decision regarding the activation of a data provider requires an in-depth study of specific requirements, such as type and coverage of the data, supply methods, quality of the data samples, compliance with the required criteria, as well as the evaluation of the economic aspects and the comparison with the solutions already in use. Similarly, active suppliers are re-evaluated periodically (at least once a year), updating their qualification with respect to the quality of the service received, economic competitiveness and any changes in operational needs.

The evaluation system involves the attribution of scores in the various areas subject to analysis which lead, according to predetermined thresholds, to the overall qualification of the supplier. The analysis criteria also include incident assessments.

Among the various tasks, the DAS teams periodically monitor the market of info data providers and supplier offers, thus ensuring that the choice of suppliers maintains high standards of precision, reliability and support for production processes, and can effectively respond to new needs or to improved conditions.