RISKMETRICS GROUP™

A gentle introduction to the RM 2006 methodology

Gilles Zumbach

RiskMetrics Group
Av. des Morgines 12
1213 Petit-Lancy
Geneva, Switzerland

gilles.zumbach@riskmetrics.com

Abstract

We present the basic concepts used in market risk evaluations, as well as the standard methodologies to compute quantitatively the risk. A new methodology is introduced with the goal to incorporate the state-of-the-art knowledge about financial time series. The performance evaluation of risk methodologies is explained, and the performance measures of the main risk methodologies are compared. The presentation stays at the conceptual level and uses the minimum number of formula needed for clarity.

1 Introduction

The original RiskMetrics methodology was established in 1994. This methodology incorporates in a simple way the key facts on time series and risk. It is robust, can be applied to a wide range of assets, and depends mainly on one parameter. Yet, it has also limitations; for example, the risk horizons are limited to a few weeks. The existing RiskMetrics methodology (RM1994) has also shortcomings, due in part to the advance of our knowledge about financial data. Similarly, the one year "Equal weight" methodology has a comparable set of strengths and weaknesses, resulting in similar performance figures.

In order to improve and extend the existing risk methodologies, we have revisited completely the risk framework, leading to the development of a new methodology called RM2006. Our goals for this new methodology are as follows. First, we want to incorporate the recent knowledge about the generic quantitative behavior of financial time series, in particular the volatility dynamic and the fat tails. Second, we want to evaluate risks from 1 day to 1 year, with a consistent framework. This is particularly important for financial actors with very long time horizons, like insurance firms and pension funds, as a consistent framework allows evaluation of risks from a short term tactical perspective to long term strategic global allocation. At the level of some particular sector or sub-portfolio, one is more interested to tactical risk at horizons from 1 day to 2 weeks. At the department or company level, the focus is shifting to long term stategic risk and global allocation. Having one methodology allows a seamless analysis across time horizons and aggregations. Third, we want to improve quantitatively risk evaluations for short risk horizons. The original risk methodologies are now more than a decade old. The experience that was gained during that lapse of time should allow us to do better. Fourth, we want to keep a robust and universal approach, with as few parameters as possible. Simplicity has clearly been a key factor in the success of the original methodologies, that include zero (Equal weight) or one parameter (RM1994). Today, typical portfolios of large financial institutions can include several thousands of positions, possibly more than 10'000. With such a size, it is clearly not possible to have a number of parameters proportional to the portfolio size. For example, this constraint eliminates all propositions to improve risk measures by using a GARCH(1,1) process. Beside simpicity, a small set of parameters with fixed values is also a good way to avoid overfitting.

In this paper, we introduce our new methodology, as well as the key ideas needed for market risk evaluation. We focus on the main idea, staying at a conceptual level and with the minimum number of formulas. The reader interested to a more in depth presentation can read [Zumbach, 2006b].
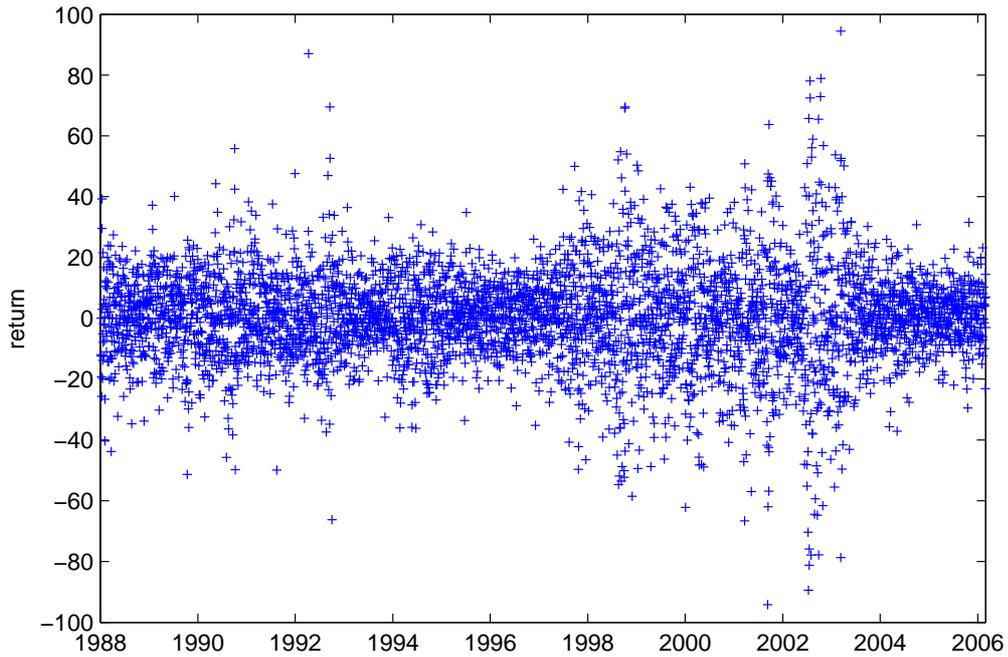
Figure 1: The annualized daily returns for the FTSE 100 index.

2   The basic ideas behind the risk methodologies

The basics of the market risk methodologies is rooted in the empirical properties of financial time series. Let us consider for example the time series of the daily returns for the FTSE 100 index, as shown in fig. 1. On this graph, we can observe clearly both key features of empirical data, without using any statistics. First, the heteroskedasticity[1] can be observed, with periods of high volatility and periods of low volatility. The clusters of high and low volatility are also the dominant feature for risk management, as they correspond to periods of high and low risks. Second, the mean annualized volatility σ for this data sample is around 15%. In the same unit, many returns have absolute values above 45%, above 60% or even above 75%, corresponding respectively to events at 3σ, 4σ and 5σ events. This is the signature of a fat tail distribution for the returns, with large events having a larger probability to appear when compared to returns drawn from a Gaussian distribution.

Risk evaluation is tightly related to forecasts as risk is essentially given by the probability of large negative returns in the forthcoming period. The period extends up to the considered risk horizon ΔT, say for example $\Delta T = 10$ days. The desired quantity is a forecast for the probability distribution $\tilde{p}(r)$ of the possible returns r over the risk horizon ΔT. From this probability distribution (pdf), the usual measures for risk can be computed, like the value at risk (VaR) or the

---

[1]Heteroskedastic means that a time series has a non constant variance through time. The American spelling is heteroscedastic, but is less faithful to the Greek root.

3

expected shortfall (ES). In practice, this problem is decomposed into forecasts for the mean and variance of the return probability distribution by using

$$r[\Delta T] = \tilde{\mu}[\Delta T] + \tilde{\sigma}[\Delta T] \cdot \varepsilon \tag{1}$$

The return $r[\Delta T]$ is a random variable corresponding to the possible price changes over the risk horizon ΔT. Risk corresponds to large negative (positive) returns for a long (short) position. The forecast for the mean price change is given by $\tilde{\mu}$ and for the volatility by $\tilde{\sigma}$. These forecasts depend on the risk horizon ΔT. The volatility forecast is a key part as it should capture the heteroskedasticity. Finally, ε is called the residual and corresponds to the unpredictable part. It is a random variable distributed according to a pdf $p_{\Delta T}(\varepsilon)$. The standard assumption is that $\varepsilon(t)$ is an independent and identically distributed (iid) random variable, meaning that the residuals at two different times $\varepsilon(t)$ and $\varepsilon(t')$ are independent and drawn from the same distribution $p_{\Delta T}(\varepsilon)$.

A risk methodology depends mainly on $\tilde{\sigma}$ and $p(\varepsilon)$, and often the mean return $\tilde{\mu}$ is taken to be zero. For example, the RiskMetrics RM1994 methodology uses an exponential moving average scaled by $\sqrt{\Delta T}$ for the volatility forecast, and a Gaussian distribution for the residuals pdf. The "equal weight" methodology is quite similar, but the volatility forecast is computed using one year of historical daily return, scaled by $\sqrt{\Delta T}$.

At a given time t, the return pdf is given by the pdf for the residuals, up to a change of location $\tilde{\mu}$ and size $\tilde{\sigma}$. A subtle point is that even though the residual has a given distribution $p(\varepsilon)$, the *unconditional* distribution for the return is *not* given by $p(\varepsilon)$ because the volatility forecast is time dependent and has a distribution with fat tails. Therefore, the return pdf can have fat tails even with Gaussian residuals.

In order to validate a risk methodology, the above formula 1 is solved for the residual

$$\varepsilon = \frac{r - \tilde{\mu}}{\tilde{\sigma}} \tag{2}$$

Using historical data, the forecasts and the realized returns can be computed, and therefore a time series for the residuals can be obtained. Using these realized residuals, the above hypotheses can be checked, namely that ε is independent and distributed according to $p(\varepsilon)$. In practice, one often tests that ε is uncorrelated, and that at a given risk threshold α, for example $\alpha = 95\%$, the number of exceedance behaves as expected.

The crucial problem for long risk horizons is that back testing becomes *very* difficult to achieve. This is caused by the shrinking sample size for the residuals as the risk horizon increases. Essentially, as ΔT increases, there is not enough data left to compute meaningful statistics. For example, with 15 years of data and a risk horizon of $\Delta T = 1$ month, there are 180 independent residuals. At the 95% VaR, only $\simeq 9$ points should exceed the given threshold. For $\Delta T = 1$ year, only 15 independent data points are left, and $\simeq 0.75$ point should exceed the 95% threshold. Clearly, doing statistics with this kind of sample size is difficult or

impossible. Because of this sample size problem, risk methodologies have been tested essentially only up to 10 days. Let us emphasize that it is a fundamental limitation given by the available time series and the considered risk horizons. In short, it is a road block.

3   Our strategy to bridge the gap between 1 day and ΔT

We need an idea to turn around the road block created by the shrinking sample sizes. This idea consist in adding more structures into the problem using a process. The process is taken with a time increment δt of one day. It should capture the essential properties of the financial data, in particular the heteroskedasticity and fat tails. Moreover the structure of the process should involve only linear and quadratic terms, so that we have some analytical tractability. In particular, forecasts can be computed using conditional averages. This is realy the crux of the methodology as it allows us to relate daily data with forecasts at any time horizon. In this way, the process and its parameters can be calibrated for time horizons at which statistics are significant. Then, using the process at the time scale $\delta t = 1$ day, the volatility forecast $\tilde{\sigma}[\Delta T]$ at the risk horizon ΔT can be computed. The forecasts depend only on the process parameters (which are independent of ΔT) and are consistent across risk horizons ΔT. After extensive testing for risk horizons at which there is enough data to compute significant statistics, the structure brought in by the process allows us to reach much longer risk horizons at which our testing ability is limited.

The residuals can then be computed as above, and their properties can be studied. The desired good properties are that the residuals are independent and with the same distribution for all assets, namely that they are iid[2]. Such a study should be done for a large set of time series, and as functions of the risk horizons ΔT. This strategy uses at best the daily data and their properties in order to compute an iid random variable ε at time horizon ΔT.

4   The process

The most salient property of financial time series is that the volatility is time varying and clustered. The cluster properties are measured by the lagged correlation of the volatility. The decay of the lagged correlation quantifies the memory shape and magnitude. This measures the influence of past volatility on forthcomming volatility, and is directly related to our ability to compute a volatility forecast. With empirical data, the lagged correlations decay logarithmically as $1 - \log(\Delta T) / \log(\Delta T_0)$, in the range from 1 day to 1 year (with ΔT_0 or the order of a few years), and for all assets. Intuitively, this means that the

---

[2]There is a general difference between independence and the absence of linear correlation. In practice, we replace the test of independence by the absence of correlations for the residuals and their magnitudes. In the text, we do not make the rigorous distinction between both concepts.
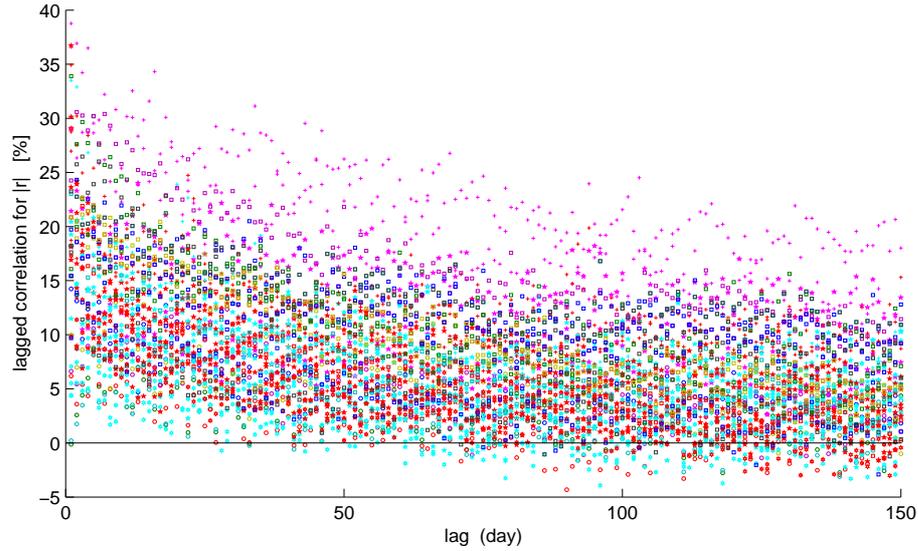
Figure 2: The lagged correlation of $|r|$ for 40 time series. The same color is used for broad asset class, like FX, stock indexes, etc...

memory of the volatility decays very slowly. The universal behavior for the volatility memory is quite remarkable, and is shown on fig. 2. Clearly, the process must capture the long memory of the volatility; for that purpose we need to use a multi-scales long memory extension of I-GARCH called Long-Memory-ARCH (LM-ARCH) process [Zumbach, 2004]. The core idea for this process is to measure the historical volatilities with a set of exponential moving averages (EMA) on a set of time horizons chosen according to a geometric series. These historical volatilities are summed to obtain the effective volatility that influences the magnitude of the returns. Essentially, the feed-back loop of the historical returns on the next random return is identical to the feedback presents in the basic GARCH(1,1) process, but with the modification that it involves the volatilities measured at multiple time scales.

Using Monte Carlo simulations, it can be shown that the lagged correlations decays appropriately for this process. Because the empirical values for the exponent ΔT_0 are similar for all assets, we can choose **the same parameters** for all assets. Our ability to take the same values for the process parameters leads to a very robust methodology. Moreover, if the number of volatility components is one, the process is reduced to the I-GARCH process. Finally, do notice that the process does not include a mean volatility parameter, like in the GARCH(1,1) process. Such parameters are clearly time series dependent and would lead to a much more complex (and fragile) evaluation scheme. Instead, in the LM-ARCH process, the volatility components with the longest time horizons play the role of a mean volatility for the shorter time horizons.

As mentioned earlier, the process has been chosen so that the conditional expectations related to the volatility forecasts can be evaluated analytically. After these computations are done, the desired volatility forecast $\widetilde{\sigma}$ can be expressed
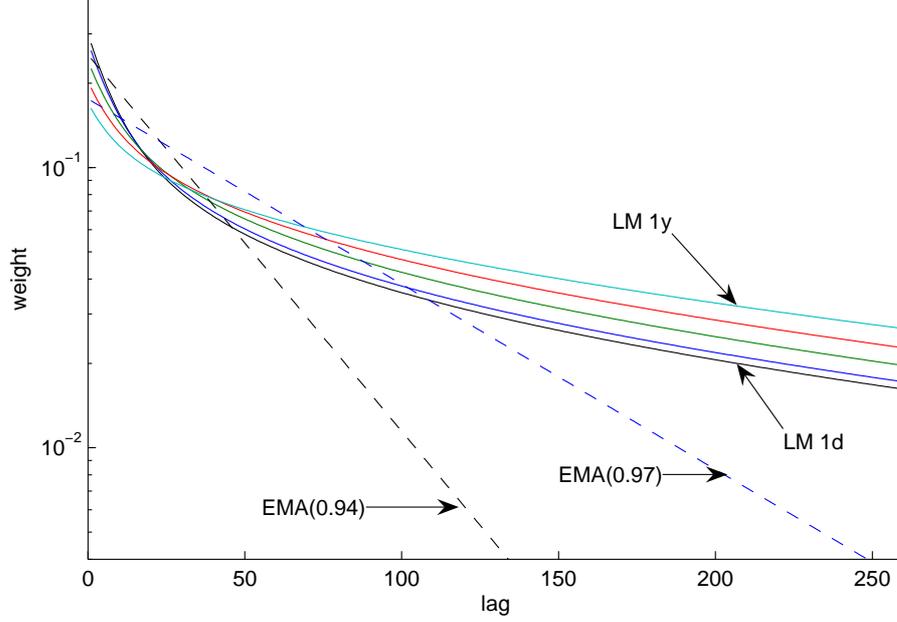
Figure 3: The forecast weights $\lambda(\Delta T/\delta t, i)$ as function of the lag i. The dashed straight lines correspond to the I-GARCH process, for which there is no dependency on $\Delta T/\delta t$. The decay factors are given in the curve labels. The black curved line corresponds to the long memory process for a forecast horizon of 1 day, the colored curve for risk horizon of 5, 21, 65 and 260 days.

as

$$\widetilde{\sigma^2[\Delta T]}(t) = \frac{\Delta T}{\delta t} \sum_i \lambda(\Delta T/\delta t, i)\, r^2(t - i\delta t). \tag{3}$$

The weights $\lambda(\Delta T/\delta t, i)$ obey the sum rule

$$\sum_i \lambda(\Delta T/\delta t, i) = 1. \tag{4}$$

This key formula can be understood intuitively as follows. The ratio $\Delta T/\delta t$ is the forecast horizon expressed in days. The leading term for the forecast is given by $\widetilde{\sigma} \simeq \sqrt{\Delta T/\delta t}$. This is the "usual" square root scaling for the volatility with the time horizon. This term originates in the diffusive behavior of the prices, which is captured by the base random walk characteristic of our process. The next term $\sum_i \lambda\, r^2$ is a measure of the past volatility constructed as a weighted sum of the past return square. The weights $\lambda(\Delta T/\delta t, i)$ are derived from the process equations, and depend both on the lag i and on the forecast horizon $\Delta T/\delta t$. These weights are plotted on fig. 3, for the I-GARCH process and the long memory ARCH process. Intuitively, we expect that a short term forecast depends more on the very recent past, whereas a long term forecast depends more on the distant past. In the figure, we see how the weights induced by the long memory process follow just this expected behavior depending on the forecast horizon $\Delta T/\delta t$.
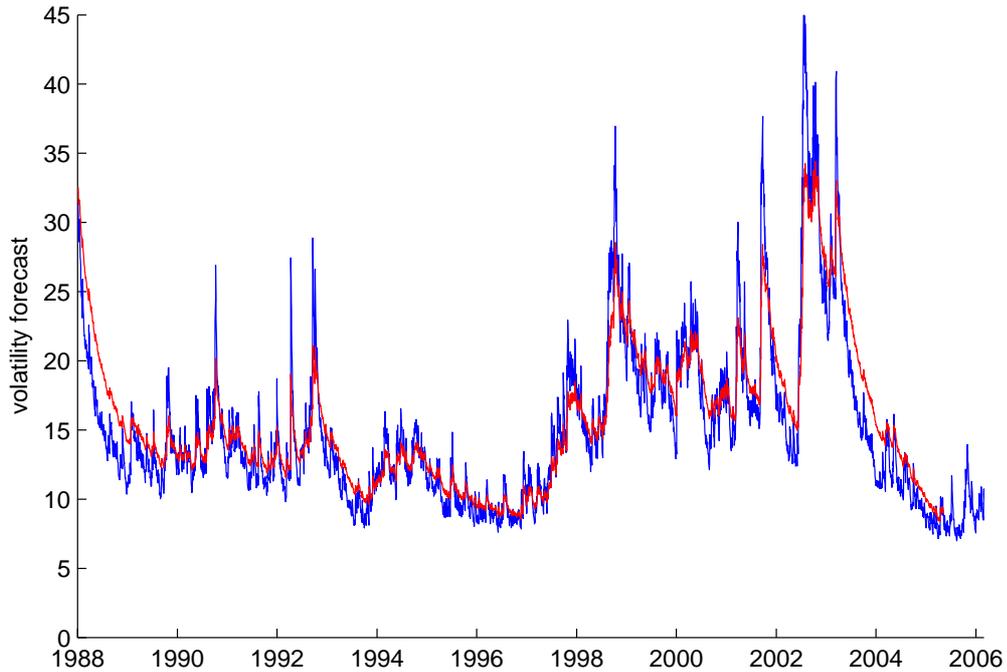
7

Figure 4: The volatility forecast for the FTSE 100, for a time horizon of one day (blue) and one year (red). The volatilities are annualized.

Fig. 4 shows on the same graph the 1 day and 1 year volatility forecasts. The difference in the dynamic between the forecast is very clear: the one day forecast adjusts very rapidly to the changing market conditions whereas the one year forecast has a smoother evolution. The quality of the volatility forecasts is the major determinant factor for a risk methodology. Another interesting feature is that even at a one year horizon, the volatility forecast has a substantial dynamic, with a ratio of 3 to 4 between the forecasts during high and low volatility periods. This shows that even for such long risk horizons, neglecting this dynamic by using a very long term mean volatility is a poor approximation of the market behavior.

Regarding the mean return forecast $\tilde{\mu}[\Delta T]$, the usual assumption is to neglect this term. Our empirical investigations have shown that it is not correct, particularly for interest rates and stock indexes. For interest rates, the yields can follow a downward or upward trend for very long periods, of the order of a year or more. These long trends introduce correlations, equivalent to some predictability in the rates themselves. Similarly, stock indexes follow an overall upward trend related to interest rates. These effects are quantitatively small, but they introduce clear deviations from the random walk with ARCH effect on the volatility. Therefore, we introduce autoregressive terms in the process equations and we derive their effects in perturbations in the LM-ARCH process. The autoregressive coefficients are essentially related to correlations, and we evaluate them on the last 2 years of data. From these coefficients, the mean return
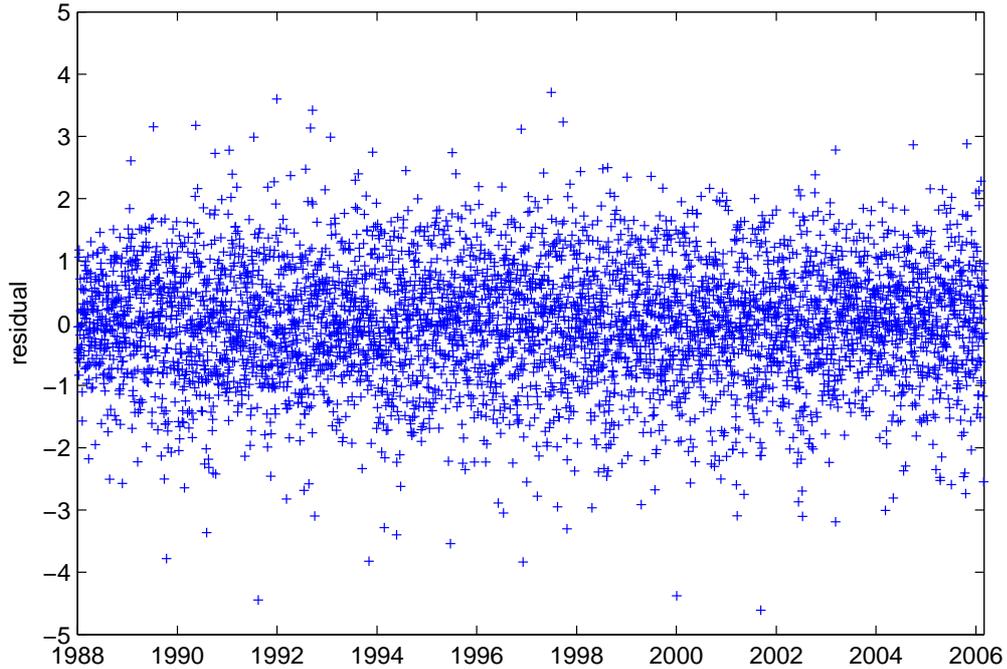
Figure 5: The daily residuals for the FTSE 100 index.

forecast $\tilde{\mu}[\Delta T]$ is computed, as well as the correction to the volatility forecast.

The above description gives the main idea used to compute the needed forecasts. Yet, substantial research is still needed to build robust algorithms, tested on hundreds of financial time series originating from around the world. The set of time series includes the major foreign exchange rates, stock indexes and interest rates, covering all the major economies. The various sub-problems are studied as such, for example the volatility forecast is evaluated using appropriate measures of accuracy. We take care to validate the process and its (fixed) parameters, as well as to incorporate the autoregressive terms in a non parametric way. The readers interested in more details can see [Zumbach, 2006b, Zumbach, 2006a]. After we having good grasp of the process and the required forecasts, we can move to the study of the residuals.

5 Empirical investigation of the the residual properties

With a methodology to compute the forecasts for the return and volatility, the residuals can be computed using historical data and the formula 2. Fig. 5 shows an example for the 1 day residuals for the FTSE 100. The comparison with fig. 1 is particularly impressive, and shows that the heteroskedasticity is correctly discounted, at least to the naked eye.
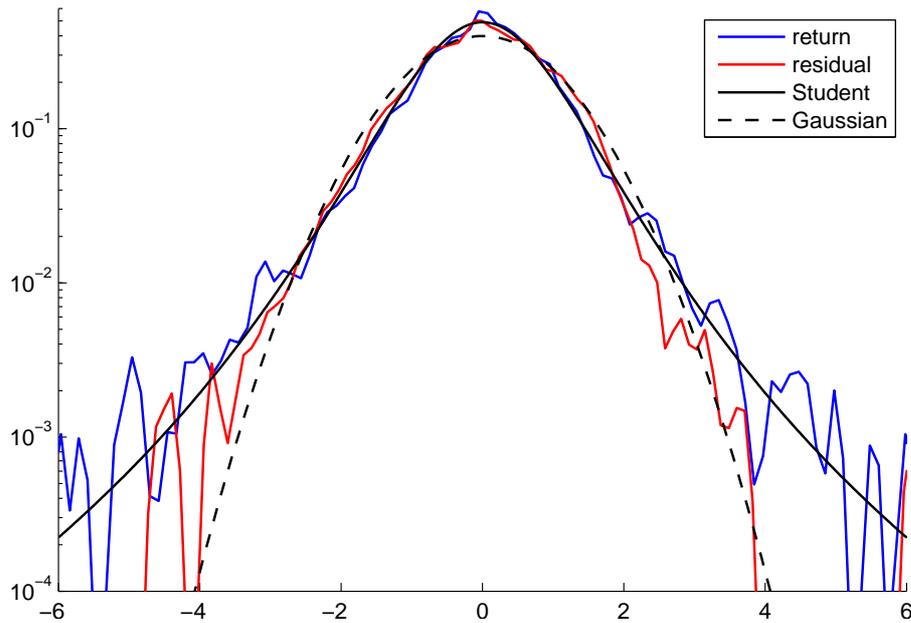
9

Figure 6: Probability distributions for the daily returns and residuals. The returns are normalized so as to have a unit variance. The Student distribution has 5 degrees of freedom, and has been rescaled to have a unit variance.

The next key point consists in studying the empirical probability density $p_{\Delta T}(\varepsilon)$ of the residuals ε. Fig. 6 displays the probability distribution of the returns and residuals, for the FTSE 100 used in fig. 1 and 5. The empirical data show clearly that a Gaussian distribution can be excluded as it does not provide enough tails. On the other hand, a Student distribution gives a good description of $p_{\Delta T}(\varepsilon)$. In principle, the residuals distribution can depend on the risk horizon ΔT. In practice, the distribution is essentially independent of ΔT, and the same number of degrees of freedom $\nu = 5$ can be taken for all time horizons. This choice for the residual distribution completes the overall description of the RM2006 methodology.

6   Back testing

Even if the direct figures comparison between fig. 1 and 5 is impressive, quantitative measures of the risk accuracy need to be built. It is essential for long risk horizons (both above figures are for 1 day, the easiest horizon!). Our goal is to compare quantitatively different risk horizons and various methodologies.

For this purpose, we introduce a function $\delta(z)$, called *relative exceedance fraction*, that measures the difference between the actual and expected relative number of exceedances. The argument z, called *probtile*, corresponds to the cumulative density function of the return, and is such that $0 \leq z \leq 1$. It is directly related to the risk threshold by $z = 1 - \alpha$. For a perfect risk methodology, we

must obtain $\delta(z) = 0$, namely at all risk thresholds, the actual relative number of exceedances agree with z. The main advantage of this back testing scheme is that the whole distribution is tested (and not only a choosen risk level). [3]

In order to have an scalar measure of performance for one time series, we define

$$d_p = (p+1)\, 2^p \int_0^1 dz\, |\delta(z)|\, \left(z - \tfrac{1}{2}\right)^p.$$

Essentially, the integral measures the overal departure from $\delta(z) = 0$, and where the weights given to the extremes can be chosen by the exponent p. Low values for d_p are better. As we are interested in financial risk at the 95% or higher, we take large values for the exponent p. The constant in front of the integral cancels the leading p dependency of the integral for $|\delta(z)| = $ constant. In this way, the number d_p can be directly interpreted as a weighted measure of discrepency between theoretical and actual relative exceedances.

The above measure of performance d_p can be computed for various time series, at a given risk horizon ΔT. In order to assess the global quality of a given methodology, we need to average the quality measures d_p on a test set of time series. This allows us to obtain overal quality measures $d_p(\Delta T)$ that can be compared for different methodologies. The figure 7 shows that the risk estimates are improved by a factor ~ 3 when using the new methodology. Another way to analyze these curves is that the new RM2006 methodology at a 6 months risk horizon is as accurate as the existing methodologies at a 10 days risk horizon. This is clearly a very large gain in term of the risk horizons that can be used. As mentioned above, we also expect that the residuals are iid. A similar procedure can be used to compute the lagged correlations for ε and $|\varepsilon|$, or for $z - 1/2$ and $|z - 1/2|$. The most important measures are for $|\varepsilon|$ and $|z - 1/2|$, as this is mainly sensitive to the discounting of the heteroskedasticity. For the price changes, the lagged correlations of $|r|$ have values in the 5% to 30% range, with a broad dispersion. These numbers are a direct measure of the volatility clustering, and the starting point for constructing a risk measure based on $r/\tilde{\sigma}$.

---

[3]The main idea used in back testing is the following: the theoretical Student distribution allows to map the empirical residuals ε to a random variable $z \in [0,1]$ given by the corresponding cdf

$$z = \mathrm{cdf}(\varepsilon) \qquad z \in [0,1]$$

where $\mathrm{cdf}(\cdot)$ is the cumulative density function corresponding to a Student distribution with 5 degrees of freedom. In more generality, for a complex portfolio with non linear positions, a risk methodology gives a forecast for the return distribution $\tilde{p}(r)$. The probability to observe a given quantile r, called "probtile", is defined by

$$z = \int_{-\infty}^{r} dr'\, \tilde{p}(r').$$

These two definitions are equivalent for a simple time series.

If the risk methodology captures correctly the behaviour of the financial time series, the empirical probability distribution $p(z)$ for z must be uniform on $[0,1]$ (and iid). For a uniform pdf, the corresponding cdf is linear. Therefore, we define the function $\delta(z)$ that measures the departure from a linear cdf by

$$\delta(z) = \mathrm{cdf}(z) - z.$$

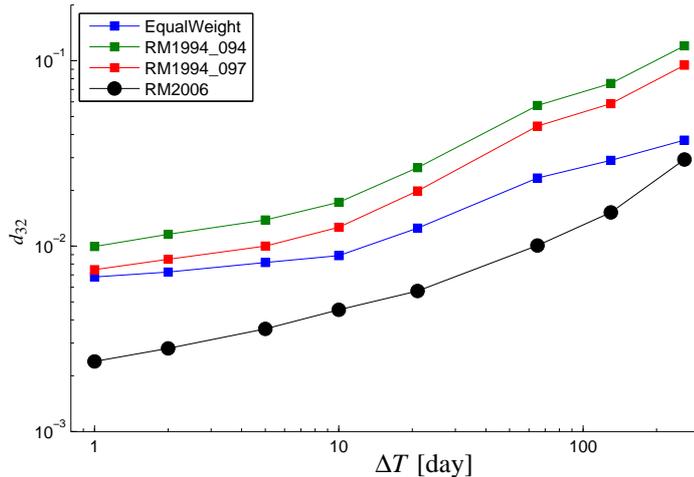where $\mathrm{cdf}(z)$ is the empirical cumulative distribution for the probtiles z.

Figure 7: The overall quality measure d_{32} for the four main methodologies. The date set contains a total of 233 times series, divided into commodities (18), foreign exchange (44), stock indexes (52), stocks (14) from France and Switzerland, CDS spreads (Credit Default Swap) on US firms (5), interest rates (100) with maturities at 1 day, 1 month, 1 year and 10 years. The time series are taken from all geographic areas.

The lagged correlations for $|z - 1/2|$ are displayed on fig. 8. This figure shows the improvement given by the new methodology, as well as the inferior results of the "equal weight" scheme due to the incorrect weighting of the past returns.

7 Remarks and conclusions

All the above statistical tests show the consistent improvements provided by the new RM2006 methodology. Yet, it comes with a price which is the added complexity of the methodology. If the main idea is quite straight forward and appears as a natural extension of the existing methodologies, an essential contribution to the overal final performances is made by the discounting of the small returns correlations. This part has only been alluded to in the present paper. It introduces its own set of analytical calculations in the process setup as well as non parametric statistical estimates in the actual implementation. All of these factors contribute to the final increase of performance and complexity of the new scheme.

Because of the observed heteroskedasticity of the financial time series, risk is tightly related to volatility forecasts. The best volatility forecast is obtained using all the returns, as this preserves the entire information. Our scheme using a process with daily increments, $\delta t = 1$ day, allows us to extract the existing information from the past within a clean framework. The long memory kernel weights optimally this information, whereas an exponential (rectangular) kernel emphasizes too much the near-by (distant) past. On the other hand, any scheme
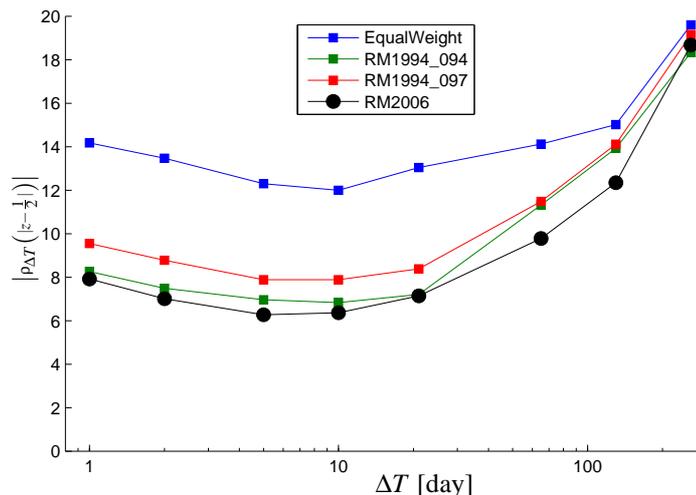
Figure 8: The overall lagged correlation of $|z - 1/2|$ for the four main methodologies. For a given risk horizon ΔT, the lagged correlation $\rho_{\Delta T}$ at lag ΔT is computed for each time series. Then, we compute the mean of $|\rho_{\Delta T}|$ on our test set of time series.

that uses returns on longer time horizons is losing information, and therefore leads to inferior forecasts. For example, using monthly or yearly data to forecast the one year volatility is essentially throwing away most of the information.

Finally, the idea of using a process to set the market risk framework allows us both to reach long risk horizons and to have consistent risk estimates across horizons. This is important to analyze a portfolio at different level of details versus different risk horizons. At the tactical level, say for each trading desks, one can analyze and optimize the short term risk of every positions. As one moves to a coarser level in an organization, one becomes more interested in strategic allocation and at longer risk horizons, say for example to assess the overal fraction of equities versus bonds. The new RM2006 methodology allows us to have one consistent framework for the tactical and the strategic risk analysis over a broad range of time horizons.

References

[Zumbach, 2004] Zumbach, G. (2004). Volatility processes and volatility forecast with long memory. *Quantitative Finance*, 4:70–86.

[Zumbach, 2006a] Zumbach, G. (2006a). Back testing risk methodologies from 1 day to 1 year. Technical report, RiskMetrics Group.

[Zumbach, 2006b] Zumbach, G. (2006b). The riskmetrics 2006 methodology. Technical report, RiskMetrics Group.