RiskMetrics Group

The RiskMetrics 2006 methodology

Gilles Zumbach

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

gilles.zumbach@riskmetrics.com

Initial version: October 2006
This version: March 2007

Abstract

A new methodology to evaluate market risks is introduced. It is designed to be more accurate than the existing methodologies, and to be able to reach long risk horizons, up to one year. Consistency across risk horizons is obtained by building the methodology using a long memory ARCH process to compute the required forecasts. A large data set covering the main asset classes and geographical areas is used to validate the various sub-components of the methodology. Extensive backtesting using probtiles is done to assess the final performance, as well as the contributions of the various parts. One key quantitative result is that the new methodology applied to a risk horizon of three months is more accurate than the exponential moving average scheme at a risk horizon of one day. This quantitative improvement allows us to analyse risks in a portfolio both at tactical and strategic time horizons.

RiskMetrics Group
One Chase Manhattan Plaza
44th Floor
New York, NY 10005

www.riskmetrics.com

Contents

1 Introduction **4**

2 Methodology **7**

2.1 Sketching the idea 7

2.2 Forecast using processes 10

2.3 The residuals 12

2.4 Annualization 13

3 The data sets **14**

4 Price mapping **14**

5 Probability distribution for the returns **18**

6 Lagged correlations for the volatilities **18**

7 Volatility forecast **24**

8 Lagged correlations for the returns **29**

9 Return forecast **32**

10 Lagged correlations for absolute residuals $|\varepsilon|$ **34**

11 Lagged correlations for the residuals ε **36**

12 Variance for the residuals **36**

13 Mean for the residuals **39**

14 Probability distribution for the residuals **40**

15 Backtesting **45**

15.1 Backtesting the return forecasts 47

15.2 Backtesting the volatility forecasts 47

15.3 Backtesting the residuals pdf 47

15.4 Backtesting the key components of the methodology 48

16 Conclusion **48**

17 Acknowledgments **50**

A Computation of the volatility forecast from the process equations **51**

B Corrections induced by the AR terms **53**

C Computation of the correlations **56**

D FAQ **58**

E Notation **61**

1 Introduction

Nowadays, risk management is an integral part of the financial world. For market risk, the goal is to assess the magnitude of large potential losses in a portfolio due to adverse price fluctuations. One of the accepted basic methodologies was established in a seminal paper [Mina and Xiao, 2001] by researchers at JP Morgan (see also [Mina and Xiao, 2001]). This methodology was used for the internal risk control inside the bank, and its public disclosure raised a strong interest in the financial industry. In turn, it leads to the spin-off of the risk control group, and subsequently to the creation of a whole industry around risk management for various financial institutions. We call this methodology *RM1994*. Another important risk methodology is called the *historical* methodology, and is essentially a resampling of the historical returns. Beyond the basic methodologies, the software tools developed for computing risks have become very sophisticated in order to allow for the diversity of structured products and complex derivatives. Moreover, the dependencies on basic risk factors must be computed for sensitivity analysis or scenario simulations. For example, all the bonds and derivatives are dependent on interest rates, and changing one rate will influence simultaneously many assets in a portfolio.

Since their inception in early nineties, the basic methodologies have stay unchanged. For the RM1994 methodology, the computation relies on the measure of the volatilities and correlations present in the historical data. These quantities are computed by a simple estimator given by an exponential moving average (EWMA). For the historical methodology, one year of equally weighted historical data is used. Using a Gaussian assumption for the residuals allows us to perform Monte Carlo simulations or to compute tail probabilities. The key advantage of these basic methodologies is their simplicity, conceptually and computationally. With the EWMA weighting, the volatility and correlation estimator depends on one parameter, namely the decay factor of the exponential. For assessing risk at a one day horizon, empirical studies show that the decay factor 0.94 provides a good estimate for all assets. Having only one parameter for all time series contributes to the simplicity of the RM1994 methodology and to its wide acceptance.

The need for a volatility forecast in a risk computation is rooted in the heteroscedasticity of the financial time series. Because the volatility is time varying, the estimation of the tail probabilities requires a forecast for the (cumulative) volatility until the considered time horizon ΔT. The RiskMetrics EWMA weighting scheme, scaled by $\sqrt{\Delta T}$, is a simple and effective volatility forecast. Yet, the optimal decay factor must be adjusted to the desired horizon ΔT. This occurs because a long term forecast must use more information from the distant past than a short term forecast. This is one drawback of the current methodology, as the main parameter should be adjusted for each risk horizon.

Since the inception of the original RiskMetrics methodology, our knowledge of financial time series has progressed in several directions.

- The volatility dynamic is better understood. In particular, the long memory of the volatility is observed in all time series. A quatitative measure is given for example by the lagged correlations of the squared returns, and the correlations are observed to decay as a logarithm of the lags. Notice that a GARCH(1,1) process has an exponential decay for the volatility correlations.
- The probability distributions of the returns have fat tails, even at time horizons of several days, and with tail exponents in the range 3 to 5. The implication of this observation on risk estimates is a bit subtle, because of the difference between the unconditional

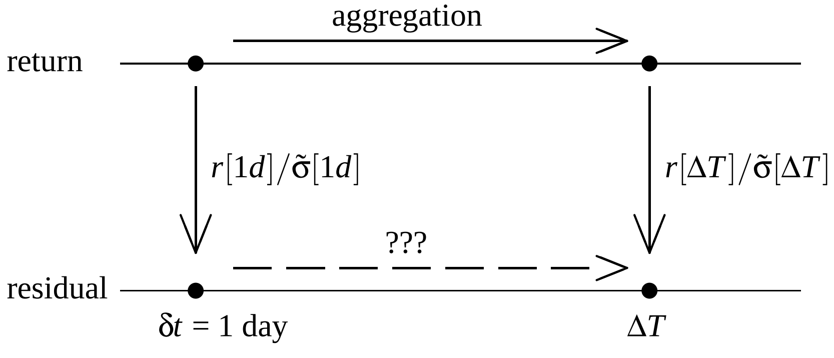
probability distribution of the return and the distribution conditional on the volatility forecast. As explained later, the relevant distribution concerns the conditional distribution, as captured by the residuals $\varepsilon = r/\tilde{\sigma}$. As shown in sec 14, the distributions for the residuals are observed to also have fat tails, albeit with a slightly larger exponent than the unconditional returns. The Gaussian assumption used in the standard methodologies is at odds with these empirical observations. Several authors (see e.g. [Pafka and Kondor, 2001, Bormetti et al., 2006]) have criticise this aspect of the standard methodologies, and have proposed to use fat tail distributions instead of a Gaussian.

- Volatility forecasts have progressed. Particularly relevant to the application in risk evaluations is the connection between processes and forecasts, the distinction between linear and affine processes and the construction of processes with long memory [Zumbach, 2004, Zumbach, 2006b].

The purpose of this paper is to incorporate the state of the art knowledge in risk management, while retaining the simplicity in the methodology that contributed to the success and wide acceptance of the original RiskMetrics scheme. The key idea is to introduce a *process* that leads to simple volatility forecasts, and that incorporates the long memory observed in the empirical data. Moreover, when the process is chosen to be an I-GARCH(1) process, the new scheme reproduces the original EWMA computation. In this sense, the new scheme is an extension of the current accepted RM1994 methodology.

Using a process to infer volatility forecasts has another key advantage: the same process and parameters can be used for every forecast horizon ΔT. In this way, one gains a strong consistency in the methodology, for all horizons. Thanks to this consistency, long term horizons can also be dealt with, even though the amount of independent data is clearly insufficient to validate the forecasts using backtesting. More precisely, our ambition is to be able to compute meaningful risk figures up to a one year horizon. This is of critical interest for pension funds or insurance firms that have a basic business cycle of one year. Yet, the whole financial community will benefit from being able to assess its risk for horizons longer than a few days.

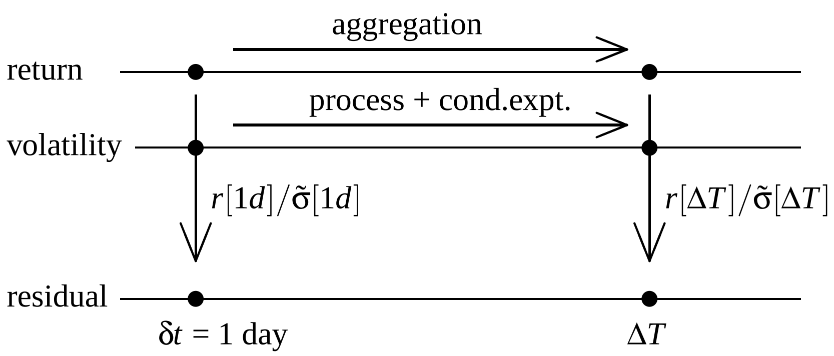
Dealing with risk evaluation for medium to long term horizons is clearly a very difficult topic. The difficulty can be measured by the scarcity of the publications, mostly originating from a group at the ETHZ [Embrecht et al., 2005, McNeil and Frey, 2000] and [Diebold et al., 1998]. The origin of the difficulty can be understood from the following diagram.



On the left side is the one day horizon (denoted $1d$), at which most of the data is available. On the right side is the desired risk horizon ΔT, say for example one year. The top line corresponds to the returns. Moving from one day returns to ΔT days returns can be easily done by aggregation, for example a one year return is a sum of 260 consecutive daily returns.

The bottom line corresponds to the residuals $\varepsilon[\Delta T] = r[\Delta T]/\tilde{\sigma}[\Delta T]$. This is the variable taken to be iid, and for example drawn randomly in a Monte Carlo VaR simulation. Investigating its properties is the key to understanding risk. For example, the forecast $\tilde{\sigma}$ used in computing the residuals must be such that ε is iid, and this needs to be validated by backtesting. It can be done at a daily horizon (vertical left arrow) because there is enough data. Yet, it is not possible to carry the same program on much longer horizons, because the data become too scarce to do the required backtesting: clearly, it is impossible to study forecasts and tail events with only 10 to 50 data points. An alternate route is the bottow dashed arrow, but residuals do not have an aggregation property. To compute risk at long time horizons requires reaching the bottom right corner, but both routes are blocked (lack of data or no aggregation property).

Introducing a process at a daily frequency allows us to resolve this dilemma. The process uses daily data and induces forecasts for any time horizon. The forecasts can be computed analytically and do not contain new parameters. This is very important as the framework can be tested for time horizons where enough data is available; then the consistency brought in by the process allows us to reach time horizons where backtesting is very limited. This key idea can be expressed in the following representation by adding another way to reach the risk horizon ΔT for the volatility forecast.



In [Embrecht et al., 2005, McNeil and Frey, 2000], a somewhat similar line of thought is used. The authors use a GARCH(1,1) process, supplemented with the theoretical aggregation rules for the GARCH(1,1) parameters derived by [Drost and Nijman, 1993], in order to compute a one-step volatility forecast for the aggregated return at the risk horizon ΔT. This approach suffers from a number of drawbacks. First, a GARCH(1,1) process has an exponential memory, and therefore does not correctly capture the slow decay of the lagged correlations. In other words, the volatility forecast does not use at best the available information. A second drawback is that the process cannot easily be changed as the aggregation rules are known only for the GARCH(1,1) process. This makes it difficult to incorporate the long memory, for example. Third, the mean volatility is one of the process parameters and is strongly time series dependent. The GARCH(1,1) process depends on three parameters. The process equations are affine (in the return square and volatility square), and the additive term, usually denoted α_0 or ω, fixes the mean volatility. One can argue about a "one size fits all" approach for the other two parameters, but this will not work for the mean volatility. For a portfolio, one has a parameter estimation problem with a size that grows with the number of time series. By contrast, the I-GARCH(1) process equations are linear instead (in the return square) and depend only on one time decay factor (see page 9 for the difference between linear and affine processes). This decay factor can be chosen so as to give good volatility forecasts for most time series.

A risk methodology should be rooted in the empirical properties of the financial time series. A thorough understanding of the "stylized facts" is needed in order to select the appropriate model and parameters. Because a risk methodology is applied to so many time series, it is essential that the model be “universal”, namely that it captures the generic properties of the data. It should also includes a quantification of the different properties in order to select the most appropriate level of description. To some extend, building a risk methodology is an engineering problem: one has to pick the important parts and neglect smaller or particular effects. For this purpose, large scale empirical studies should be done, the generic features described, and possibly approximations made. This approach contrasts with an academic study, where often a rigorous answer to a well posed question is sought after. For these reasons, a large part of this paper is devoted to empirical studies of an extended set of time series, to check that our description is appropriate. Our aim is not to be rigorous (say in the sense of hypothesis testing), but to be “overall quantitatively correct”.

Building a risk methodology involves many dependent parts related to the forecasts, correlations and residuals. In the end, the empirical results depend on all the parts together, and it is in general not possible to completely extract the various sub-components for individual testing. In other words, a risk methodology cannot be built with a completely constructive approach. The sub-parts that can be isolated are mainly the description of the lagged correlations for the returns and absolute returns, and the forecasts for the return and volatility. From this dependency structure follows the organization of this paper.

The next section introduces the methodology (Section 2.1), the process with the related forecasts (Section 2.2) and the residuals (Section 2.3). The theoretical part is completed by the Appendices A and B which contain the detailed analytical computations. The empirical investigation is done using a set of time series detailed in Section 3 and covering most of the asset classes and world geographic areas. The empirical part is organized roughly around the sequence: price (Section 4) $\rightarrow$ volatility (Section 6 and 7) $\rightarrow$ return (Section 8 and 9) $\rightarrow$ residual (Section 10 to 14). Each section is devoted to one relevant property (lagged correlations, forecasts, variances, pdf, etc...). Then, all the pieces are put together for backtesting in Section 15. The presentation of backtesting is divided in subsections concerning the return forecasts (15.1), volatility forecasts (15.2), residuals pdf (15.3), and how the performances change as the various ingredients are added in the methodology (15.4). The conclusion summarises the new risk methodology and the empirical findings. The appendix A presents the detailled analytical computations related to the long memory ARCH process, and the appendix B the analytical part related to the AR term in the process. The three estimators used to compute correlations are presented in Appendix C, with a short discussion of their main properties regarding robustness and computation time. Finally, a set of frequently ask questions is given in appendix D.

2 Methodology

2.1 Sketching the idea

In this section, we present the key ideas and equations for using quadratic processes for risk evaluations. The detailled derivation of the forecast equations is given in the appendixes that can be skipped in a first reading.

Our strategy to evaluate risk at a time horizon ΔT is:

- To model the data at δt = one day using a (long memory ARCH) process.
- To infer the required mean return and volatility forecasts at the desired risk horizon ΔT

by using the process properties and the aggregation of the returns.

- To use a location/size/shape decomposition for the probability distribution of the returns. The location and size correspond respectively to the mean return and volatility forecasts derived from the process. The random variable appearing in this decomposition is called the residuals.

When building a new risk methodology, the last step is used in the other direction using historical data. The historical returns are standardized at the risk horizon ΔT using the forecasts for the mean returns and volatility, in order to obtain the (historical) residuals. The empirical properties of the residuals can then be studied.

The one business day horizon at which the data are available is denoted by δt. In the risk industry, it is common to work with daily data as they take care in a natural way of the daily and weekly seasonalities[1]. Yet, in our derivation, nothing is specific to this particular time horizon. It would be easy to take a shorter time step for the process, provided that the seasonalities are properly discounted. The desired risk time horizon is denoted by ΔT, and in principle can take any value with $\Delta T \geq \delta t$.

The process follows a standard ARCH set-up:

$$x(t+\delta t) = x(t) + r[\delta t](t+\delta t) \quad (1)$$

$$r[\delta t](t+\delta t) = \mu_{\text{eff}}(t) + \sigma_{\text{eff}}(t)\, \varepsilon(t+\delta t) \quad (2)$$

where

- **$x(t)$:** the transformed price. For most securities, x is the logarithm of the price $x(t) = \ln(p(t))$, but the transformation is slightly different for bonds (see section 4).
- **$r[\delta t]$:** the return for the time horizon δt = 1 day.
- **μ_{eff}:** the mean return in the next time increment. In most process studies, this term is zero, because the empirical lagged correlation for the return is negligible. As we will see later in sec. 8, neglecting this term is not a valid assumption, particularly for interest rates.
- **σ_{eff}:** the effective volatility in the next time period δt. This is the key term that models the heteroscedasticity of the financial time series. We consider the class of processes where σ^2_{eff} is a linear function of the past squared returns

$$\sigma^2_{\text{eff}}(t) = \sum_{i \geq 0} \lambda(i)\, r^2(t - i\delta t) \qquad \lambda(i) > 0 \quad (3)$$

  with $\sum_{i \geq 0} \lambda(i) = 1$.
- **ε:** iid random variable with distribution $p(\varepsilon)$. The distribution $p(\varepsilon)$ must be such that $E[\,\varepsilon\,] = 0$ and $E\left[\,\varepsilon^2\,\right] = 1$.

The most important property of the process is embedded in the weights $\lambda(i)$. When the weights decay logarithmically

$$\lambda(i) \simeq 1 - \frac{\ln(i\,\delta t)}{\ln(\tau_0)} \quad (4)$$

[1] A seasonality is a statistical pattern that repeat itself after a given time interval. In financial time series, they occur mainly with a daily and weekly frequency, corresponding respectively to the daily activity cycle (day/night) and weekly activity cycle (no financial transactions during the week-end). The intra-day volatility behavior is dominated by these two strong seasonalities and the heteroskedasticity can be observed only after discounting them.

the lagged correlation of $|r|$ decays also logarithmically. This is precisely the long memory observed in the financial time series we want to capture. On the other hand, when the weights decay as an exponential $\lambda(i) \simeq \exp(-i\,\delta t/\tau)$, the process reproduces the EWMA used in the current RiskMetrics methodology. Finally, a constant weight $\lambda(i) = 1/i_{\text{max}}$ up to the cutoff $i_{\text{max}} = 260$ reproduces the "Equal Weight" model which uses one year of equally weighted historical returns. In this way, our framework includes all the currently accepted methodologies [2].

The long memory process we consider is naturally expressed in a different form. A set of (historical) volatilities σ_k are evaluated on geometric time horizons τ_k:

$$\begin{aligned} \tau_k &= \tau_1\, \rho^{k-1} \qquad k = 1, \cdots, k_{\text{max}} \\ \mu_k &= \exp\left(-\,\delta t/\tau_k\right) \\ \sigma_k^2(t) &= \mu_k\, \sigma_k^2(t-\delta t) + (1-\mu_k) r^2(t). \end{aligned} \tag{5}$$

Essentially, σ_k is an exponential moving average of the return square, with a characteristic time τ_k. The effective volatility is obtained as a sum over the historical volatilities, with weights that decay logarithmically

$$\sigma_{\text{eff}}^2(t) = \sum_{k=1}^{k_{\text{max}}} w_k\, \sigma_k^2(t) \tag{6}$$

$$w_k = \frac{1}{C}\left(1 - \frac{\ln(\tau_k)}{\ln(\tau_0)}\right). \tag{7}$$

The normalization constant C is chosen such that $\sum_k w_k = 1$. The model is conveniently parameterized by the three time intervals τ_0 (logarithmic decay factor), τ_1 (lower cut-off) and $\tau_{k_{\text{max}}}$ (upper cut-off). The parameter ρ does not influence the properties of the process for $\rho \simeq 1$, and we take $\rho = \sqrt{2}$. In essence, this model is a sum of EWMA (exponential moving average) over increasing time horizons. This structure captures in a simple way the multiscale structure of the markets, with market participants acting mainly at intra-day, daily, weekly and monthly time horizons. By unwinding the EWMA in the definition of σ_k, and inserting in eq. 6, the form 3 is obtained with

$$\lambda(i) = \sum_{k=1}^{k_{\text{max}}} w_k\, (1-\mu_k)\, \mu_k^i$$

An important distinction in the analytical form of the volatility process is the affine versus linear structure. A linear volatility process has a structure of the form 3, where the volatility square is a linear combination of the past returns square. In an affine process, this structure is modified by an additive term that fixes the mean volatility

$$\sigma_{\text{eff}}^2(t) = w_\infty \sigma^2 + (1-w_\infty) \sum_{i\geq 0} \lambda(i)\, r^2(t - i\delta t). \tag{8}$$

The parameter σ is equal to the unconditional mean volatility, and w_∞ is a "coupling constant" that fixes the convex combination between a constant mean volatility and the auto-regressive

[2] This is not the most general formulation of quadratic volatility model, as we include only one day returns at equal time. A model with different return time horizons, but still at equal time, is given by

$$\sigma_{\text{eff}}^2(t) = \sum_{i\geq 0} \sum_{j\geq 1} \lambda(i,j)\, r[j\delta t]^2(t - i\delta t)$$

and is explored in [Zumbach, 2004]. We restrict ourself to the form 3 in order to keep efficient the numerical computations of the forecast. When returns at longer time horizons are used in the process, the volatility forecasts cannot anymore be expressed as in eq. 12

volatility term. For a given memory structure $\lambda(i)$, the linear and affine processes can be build. For example, with an exponential memory, the I-GARCH(1) is the linear model whereas the GARCH(1,1) is the affine version. The number of parameters differs by 2, corresponding to σ and w_∞. Clearly, the mean volatility parameter σ is time series dependent. For risk estimation on large scale, it is not possible to have a number of parameters proportional to the number of assets in a portfolio. In order to avoid this bad estimation problem, we have to use linear model. This criterion eliminates the (affine) GARCH(1,1) process and its generalizations. As processes, the asymptotic properties of the linear models are not well defined (see [Nelson, 1990] and [Zumbach, 2004] for a detailled discussion of this point). Yet, as we use the process equations to derive a forecast, this long term asymptotic subtlety is irrelevant for our purpose.

2.2 Forecast using processes

In principle, any process induces a forecast through conditional expectations. For quadratic processes, the integral implied in the conditional expectation

$$E\left[\ r^2[\delta t](t') \mid \Omega(t)\right] \qquad t' > t$$

can be computed analytically. Because the class of processes we consider is given by iterative equations (from the states at t, they give the states at $t+\delta t$), iterative equations are obtained which express the forecast at t' as a function of the forecast at $t'-\delta t$. These equations can be iterated until $t'=t$. If the scheme is straight forward, the actual analytical computations are a bit involved because of the multiple terms in the process equations. The details of these computations are given in the appendix A.

Inferring the forecasts from a process has two key advantages. First, the forecast inherits its properties from the process. For example, if the process is short or long memory (i.e. with an exponential, power law or logarithmic form for λ_i), the forecast will have the same memory structure. This is the reason why the structure of the process should capture as well as possible the main stylized facts observed in the empirical data. Second, the forecast depends only on the process parameters, namely there is no additional parameter regardless of the forecast horizon ΔT. This strategy ensures a strong consistency across the forecast horizon ΔT, where moreover the information at the daily horizon is used at its best. Because of this consistency, inferring the forecast from a process at the daily horizon allows us to reach long risk horizons, where backtesting of an unconstrained forecast would be very difficult or even impossible.

For risk evaluations at the time horizon ΔT, we are interested in the distribution of the return $r[\Delta T]$ at this horizon. The return at the time interval ΔT is given simply by the aggregation of the daily returns

$$r[\Delta T](t+\Delta T) = \sum_{t<t'\leq t+\Delta T} r[\delta t](t') \tag{9}$$

Because of the heteroscedasticity of the volatility, a key quantity is the forecasted volatility at horizon ΔT:

$$\widetilde{\sigma^2}[\Delta T](t) = E\left[\ r^2[\Delta T](t+\Delta T) \mid \Omega(t)\right] \tag{10}$$

where the notation emphasizes that this is a forecast computed at t for the volatility of the price change in the next time period $r[\Delta T](t+\Delta T)$. Using the process equations and the aggregation of the daily return, the right hand side can be computed explicitely as a function of the quantities in the information set $\Omega(t)$. This step is the crux of the method as it relates the forecasts at $\Delta T = n\ \delta t$ to δt, using the process properties.

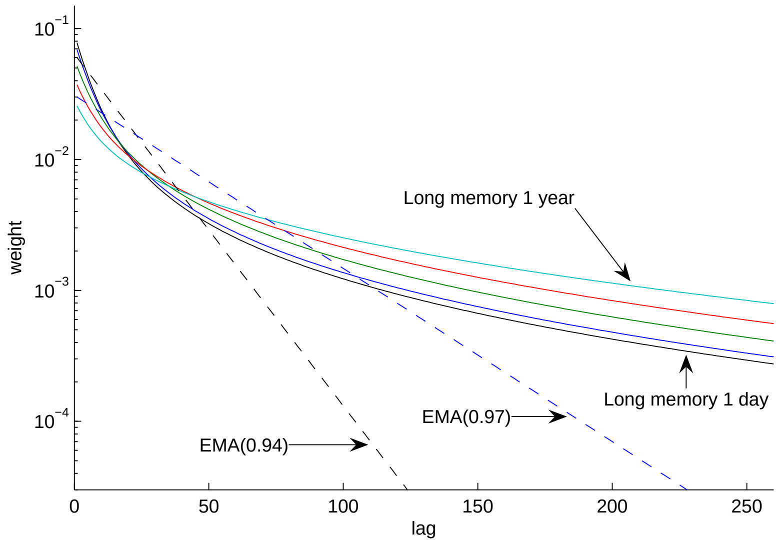


Figure 1: The weight $\lambda(n,i)$ versus the lag i for the I-GARCH and long memory processes. The curves with labels "long memory 1 day" and "long memory 1 year" correspond to the long memory process, for a forecast horizon of one day and one year repectively.

The equations are simpler when the autoregressive term $\mu_{\text{eff}}(t)$ is absent. In this case,

$$E\left[\, r[\delta t](t')\; r[\delta t](t'') \mid \Omega(t)\right] = 0 \qquad \text{for} \quad t' \neq t'', \tag{11}$$

and the cross terms cancel after the expansion of the square. Then, the volatility forecast can be expressed in the simple form

$$\widetilde{\sigma^2}[\Delta T](t) = \frac{\Delta T}{\delta t} \sum_{i \geq 0} \lambda(n,i)\; r^2(t - i\,\delta t) \tag{12}$$

with $\Delta T = n\ \delta t$. The coefficients λ obey $\sum_{i\geq 0}\lambda(n,i) = 1$ for all n, and can be computed by recursion equations. This form for the forecast shows the strong similarity between the one step forecast as used in the process definition in eq. 3, and the n-step forecast. The leading term in eq. 12 is given by $\widetilde{\sigma} \simeq \sqrt{\Delta T/\delta t}$, namely by a "square root law" of the risk horizon. This leading term originates in the diffusive nature of the process, namely the (logarithmic) price process follows a random walk. Then, the coefficient $\lambda(n,i)$ brings in the corrections due to the particular memory structure of the process.

The current RiskMetrics methodology, with an EWMA weighting, can be cast in the current processes framework by taking an I-GARCH process, for which

$$\lambda(i) = \frac{1-\mu}{1-\mu^{i_{\max}}}\ \mu^i \tag{13}$$

(the above long memory process is reduced to this form when one component $k_{\max} = 1$ is used). In this case, as shown at the end of Appendix A, the conditional expectation for the forecast becomes

$$\lambda(n,i) = \lambda(i). \tag{14}$$

namely there is no n dependency. In other words, the volatility forecasts for all time horizons are given by the volatility as computed with the single EWMA, scaled by $\sqrt{\Delta T/\delta t}$.

Figure 1 shows the weights $\lambda(n,i)$ for an I-GARCH and a long memory process, in semi-logarithmic scales. The black and blue dashed lines correspond to the I-GARCH process, with no dependency on the forecast horizon. The group of curved lines corresponds to the long memory process, where the black line gives $\lambda(i)$ (i.e. the one day forecast), and the colored curves corresponding to the forecast horizons $n = 5$, 21, 65 and 260 days. We see clearly that with increasing forecast horizons, the weights decrease on the recent past and increase on the distant past. This behavior is fairly intuitive, as a long term forecast should use more information from the distant past, whereas a short term forecast is dominated by the recent events.

The detailled empirical study shows that the autoregressive term μ_{eff} must be included in order to construct a process that captures well the properties of the financial time series. Our theoretical approach is to consider μ_{eff} as small, and treat it in perturbations. This introduces corrections for the volatility forecast, as well as a non zero forecast $\widetilde{r}[\Delta T]$ for the return at the risk horizon ΔT. The analytical computations related to μ_{eff} are detailled in the appendix B. The key point is that, because the process includes only linear and quadratic terms, we are able to evaluate analytically the conditional expectations needed to compute return and volatility forecasts.

2.3 The residuals

With the forecasts for the return $\widetilde{r}[\Delta T](t)$ and volatility $\widetilde{\sigma^2}[\Delta T](t)$ computed at t, the basic formula to evaluate market risk is

$$r[\Delta T](t) = \widetilde{r}[\Delta T](t) + \gamma[\Delta T] \sqrt{\widetilde{\sigma^2}[\Delta T](t)}\, \varepsilon(t). \tag{15}$$

The residual ε is a random variable with a given distribution $p_{\Delta T}(\varepsilon)$. The distribution is such that $E[\varepsilon] = 0$ and $E[\varepsilon^2] = 1$. The scale factor $\gamma[\Delta T]$ is a fixed function with a weak dependency on ΔT given by eq. 17 below, and essentially $\gamma[\Delta T] \simeq 1$. On the left hand side, $r[\Delta T](t)$ is a random variable, dependent on ε. From the returns distribution, computed for example with a Monte Carlo simulation, the risk measures can be estimated, like VaR and expected shortfall. The risk measures can also be computed analytically (for one time series!), for example with a normal distribution. In this formula, the forecasted return and volatility act respectively as location and size factors. Therefore, the probability distribution of the returns, *conditional* on $\widetilde{r}$ and $\widetilde{\sigma}$, is identical to the probability distribution of the residuals. We are using the shorthand for the volatility forecast $\widetilde{\sigma} = \left\{\widetilde{\sigma^2}\right\}^{1/2}$.

The most important part in this formula is the forecast for the volatility $\widetilde{\sigma}$. This term should capture the heteroscedasticity of the financial time series, so that the pdf of the residuals is stationary. This is expressed by the fact that $p_{\Delta T}(\varepsilon)$ depends on the risk horizon, but not on t, and only the volatility forecast will make the risk larger in periods of high volatility. The $\widetilde{\sigma}$ term emphasizes the tight relation between volatility forecasts and risk evaluation.

The volatility forecast is itself a time series, with a non trivial distribution. Essentially, all volatility forecasts are based on the (magnitude of the) past returns, and therefore, they will share some properties with $p(r)$. In particular, the volatility forecast has also a fat tail distribution (say similar to the $|r|$ distribution). Because the above formula is essentially $r = \widetilde{\sigma}\, \varepsilon$ and all the terms are time series, the *unconditional* distribution of the returns $p(r)$ is different from the distribution of the residuals $p_{\Delta T}(\varepsilon)$. Intuitively, because the distribution of the volatility forecasts captures some of the fat tails, the distribution of the residuals should have less fat tails than the distribution of the returns. The question is whether the volatility forecast captures all the fat tails and the residuals have a Gaussian distribution, or whether part of the fat tails for the returns originate in the residual distribution. We will investigate this question in sec. 14

using empirical data. Let us emphasize again the important difference between the conditional and unconditional distribution of the returns. The distribution specified in a risk methodology is the distribution of the residuals $p_{\Delta T}(\varepsilon)$. It is identical to the conditional distribution of the returns (up to shift and magnitude factors), but not to the unconditional distribution of the returns (because of $\widetilde{\sigma}(t)$).

The formula 15 is used to evaluate the forthcoming risk at t. In order to establish and test a risk methodology, historical data is used and the formula is solved for ε, leading to

$$\varepsilon[\Delta T](t) = \frac{r[\Delta T](t+\Delta T) - \widetilde{r}[\Delta T](t)}{\gamma[\Delta T]\,\sqrt{\widetilde{\sigma^2}[\Delta T](t)}}. \tag{16}$$

On the right hand side, the return $r[\Delta T](t+\Delta T)$ corresponds to the historical return at the time $t+\Delta T$, or to the realized return at t. This formula is similar in structure to the "standardization" of a time series by removing its sample mean and dividing by the sample standard deviation; the key difference is that we use the respective forecasts at t. The motivation is to discount the realized return by the available information at t in order to build a random variable ε that is well behaved, namely iid. The pdf $p_{\Delta T}(\varepsilon[\Delta T])$ can be computed empirically, and compared to different models for $p_{\Delta T}$. Then, using an analytical model for $p_{\Delta T}$, the $\text{VaR}_p[\Delta T]$ can be computed, as well as other risk measures.

Notice that the formulas 15 and 16 are very similar to the basic one step return increment $r[\delta t]$ used in our process definition. The main difference is that the definitions 15 and 16 are expressed for the risk horizon ΔT instead of δt.

The probability distributions of the residuals must be such that $E\left[\,\varepsilon^2[\Delta T]\,\right] = 1$. On the other hand, the volatility forecast $\widetilde{\sigma}$ is computed from the process equations in our approach, and does not contains any free parameter. In order to be able to respect the condition $\langle\varepsilon^2\rangle = 1$ when computed from empirical data according to the formula 16, a "scale factor" depending on ΔT has to be introduced. This is the role of the scale function $\gamma[\Delta T]$ given by

$$\gamma[\Delta T] = 1.06 + 0.008\,(\ln(\Delta T/\delta t))^2\,. \tag{17}$$

Its form and parameters has been established on a subset of 47 time series corresponding to fairly liquid assets (these time series were shorter than the ones used in both final data sets). The form was found by plotting $\langle\varepsilon^2[\Delta T]\rangle$ as a function of ΔT for each time series (and with $\gamma = 1$), and finding an overall decent approximation. When used on the final ICM and G10 data sets, no change were made as the initial formula works very well.

To summarize our approach, the goal is to obtain an iid random variable ε at time horizon ΔT. For this purpose, we build the needed forecasts using a process set at the daily time step δt. These forecasts depend only on the process and have no additional parameters; they are used to remove the predictable part of the mean and the variance of the price changes.

2.4 Annualization

Mean returns and volatilities have a scaling dependency with respect to ΔT, namely $\mu_{\text{eff}} \sim \Delta T$ and $\sigma \sim \sqrt{\Delta T}$. As we want to study risk at different risk horizons, it is better to remove this dependency. The common convention is to scale all quantities to a reference time horizon of one year, a procedure called annualization. The typical numbers that one has in mind are indeed annualized volatility, and for most (free floating) liquid assets it is in the range of 5 to 30% per year. For quantities computed at a scale ΔT, the corresponding annualized quantities are

$$\begin{aligned}\sigma_a &= \sqrt{\frac{1\text{ year}}{\Delta T}}\,\sigma \\ r_a &= \sqrt{\frac{1\text{ year}}{\Delta T}}\,r. \\ \mu_{\text{eff},a} &= \frac{1\text{ year}}{\Delta T}\,\mu_{\text{eff}}.\end{aligned} \tag{18}$$

The annualized quantities can be directly compared at various time horizons, as the leading random walk scaling has been discounted. For example, the forecast for the annualized volatility has no leading scaling with $\sqrt{\Delta T/\delta t}$, but only the constant factor $\sqrt{\frac{1\text{ year}}{\delta t}}$. In the following empirical sections, because we want to systematically compare results at different time horizons, all the quantities are annualized, but the a subscript has been dropped.

3 The data sets

The empirical investigations are done using two data sets. First, the "ICM" (International Capital Market) data set is a fairly large ensemble of time series covering major asset classes and world geographic areas. This 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 length of the time series are of at least of 1200 days, most of them between 2000 to 5000 days. The second set, called G10, contains 58 time series for the G10 countries, and covering commodities, FX, stock indexes and interest rates. Details for these two sets are given in [Zumbach, 2006a].

In the scatter plots, the symbols are as follows: stock indexes: green square; FX: blue circle; CDS: magenta triangle (pointing up); Commodity: black triangle pointing right; Stock: black triangle pointing down; IR: essentially red with different symbols according to the maturities. For the maturities, the symbols are as follows: 1 day: 'x'; 1 month: '+'; 1 year: five points star; 10 years: six points star. For the geographic locations, the colors are as follows: magenta for USA and canada; pink for Australia and New Zealand; red otherwise.

4 Price mapping

The standard assumption in finance is to model the *logarithm* of the price $x = \ln(p)$ by a random walk, possibly with a non trivial structure for the volatility, for example with an ARCH process. Considering that the logarithm of the price is the "good" variable is rooted in the economic invariance under the multiplication by a positive number of all the prices. In other words, for a given currency, if all the prices are multiplied by the same constant, nothing changes. After the transformation $p \to x = \ln(p)$, the logarithmic price differences do not depend on an overall price multiplication. The processes used in finance are build only on the logarithmic price changes r, but not on the prices p. As a result, the correlation between price p and volatility is zero, in agreement with the empirical results. Otherwise, should the volatility be computed from price differences (without logarithms!), then there would be a positive correlation between prices and volatility because large prices lead to larger price changes, and hence to larger volatility.

The situation is different for bonds and interest rates. These securities allow to trade the time value of money and have boundaries on their possible values. For example, the value of a bond must be between zero and its par value, or the interest rates must be positive. Because there is no theoretical invariance argument, it is much less clear which transformation leads to the "good" variable that follows a random walk. In this case, one has to relies on empirical investigations. Our criterion is as described in the previous paragraph, namely to minimize the correlation between the prices (or the yields) and the volatility, with the volatility computed with differences of the transformed variables. Similarly, we want also to minimize the correlation between mapped price (or mapped yield) and volatility. The underlying idea is that the volatility must depend only on differences, and not on the yield level (or on the mapped yield level).

As a candidate for the "good" variable, the bond price itself is inappropriate, because it must obey two constraints, and because it has a strong dependency on the time to maturity. The interest rate y, or yield, for a given maturity δT is a better candidate, and must obey[3] the single constraint $y[\delta T](t) \geq 0$. We have investigated empirically a few simple transformations for y, for the interest rates in our test set. The transformations are *none* ($x = y$), *log* ($x = \ln(y)$), *sqrt* ($x = \sqrt{y}$) and *linLog* ($x = \ln(y_0 + y)$). The return r are computed by differences of x, and the volatility is computed by a sum of returns squared over a time period ΔT. Similarly, the variable x or y is averaged over the same time interval ΔT. Then, we compute the correlation $\rho(y, \sigma)$ between yield y and volatility, and the correlation $\rho(x, \sigma)$ between mapped yield x and the volatility. Both correlations are computed for each yield time series, and the results are displayed below in the form of a probability density $p(\rho)$ for the correlation (i.e. the empirical probability in our test set).

The parameter ΔT has the following influence on the estimation of the correlation: the number of independent yields decreases as $1/\Delta T$, but the variance of the volatility estimator decreases as $1/\sqrt{\Delta T}$. Essentially, with increasing ΔT, the volatility is estimated more accurately, but the number of independent yields decreases. Because of the difference of the exponents, it is better to use a small ΔT. This has been checked empirically, and the results are presented below for ΔT = 1 day (the results are similar for larger ΔT). The correlations $\rho(y, \sigma)$ or $\rho(x, \sigma)$ show very small differences, and the figures 2 and 3 are given for the mapped yield x. For comparison, the same computation is done for foreign exchanges and stock indexes, and shown on fig. 3. In this case, the base variable is the price p, with the transformation *none* ($x = p$), and log ($x = \ln(p)$).

For the transformation *none*, the interest rate y shows a positive yield-volatility dependency for large y. On fig. 2, this appears as essentially positive correlations (red curve). This dependency can be understood intuitively because large interest rates have larger changes. This explanation is validated by a more detailled investigation of the results, where large correlations are related to large yields. This shows that the yield is not a "good" variable. A simple solution consist in using a logarithmic transformation $\ln(y)$. For this *log* transformation, there is a (price-volatility) dependency for small values of y: this correlation originates in the boundary $y = 0$ and in the finite tick size that leads to large relative changes for small y values. The "boundary effect" leads to negative correlations, as shown by the blue curve on fig. 2. Therefore, the logarithm of the yield is also not a "good" variable.

Both explanations point to a transformation that is linear for small y and logarithmic for large y. A simple transformation with these properties is $\ln(y_0 + y)$, where y_0 fixes the cross-over from

[3]The constraint $y[\delta T](t) \geq 0$ can even be violated by minute amounts, as this occurred for the Japanese Yen.

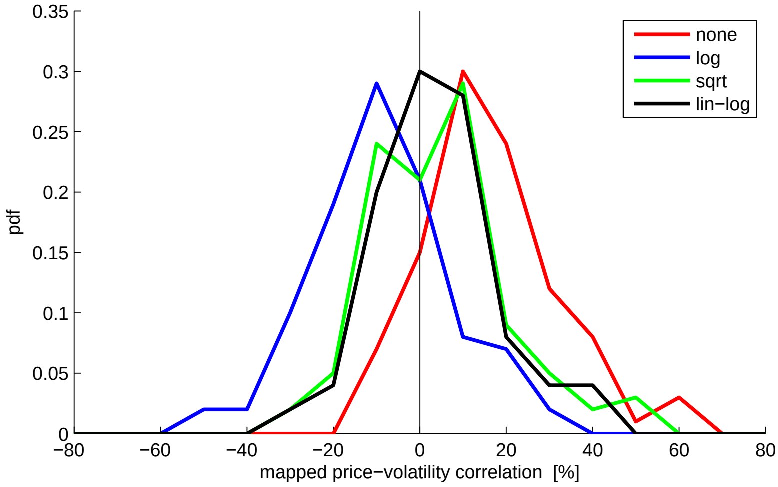


Figure 2: Probability density for the correlation between mapped yield and volatility.

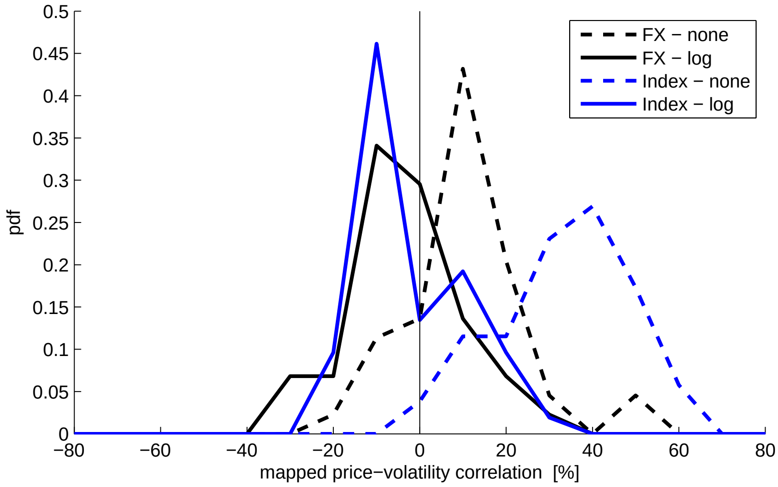


Figure 3: Probability density for the correlation between mapped price and volatility, for foreign exchanges and stock indexes.

linear to logarithmic. Computations for different values y_0 show that a value around $y_0 \simeq 4\%$ leads to a minimal price-volatility dependency. The parameter y_0 is not critical as there is only a weak dependency of the results with respect to y_0. The corresponding probability density for the correlation is plotted with the black curve, which is essentially symmetric around zero. A more detailled investigation shows no particular dependency for large or small yields. This shows that the *linLog* transformation leads to a “good” variable.

A transformation with similar properties is the square root $x = \sqrt{y}$. Statistically, its pdf does not differ from the *linLog* transformation, showing that it is also effective at neutralizing the dependency between yield and volatility. Yet, we felt that there is no intuitive explanation for why this particular form should be used, whereas the arguments given above for the *linLog* transformation make sense with respect to the market behavior.

Notice that the statistics for one time series are not very good because the interest rates have a slow evolution. Even within a 15 years period, they tend to stay in a limited range, and many currencies have had very low interest rates in the last decade. Added difficulties are the high correlations between different maturities as well as the limited number of currencies with free and liquid fixed income market. All these limitations preclude an accurate determination of the optimal transformation. For the rest of this work, we use the *linLog* transformation $x = \ln(y_0 + y)$ for interest rates, with $y_0 = 4\%$. A value for y_0 around a few percents is plausible, and the final risk estimates are very weakly sensitive to this values.

The comparison figure for the FX and stock indexes is also interesting. The *none* transformation (i.e. computing the volatility with price differences!) shows clear positive correlations, as expected. For the FX, the *log* transformation makes symmetric the pdf for the correlations. Yet, for the stock indexes with the *log* transformation, the correlation probability is clearly skewed on the negative side. This is related to the asymmetry between up and down moves, where large negative moves (stock index crashes) are followed by high volatility periods. This dependency leads to a negative correlation, as observed on the graph.

Finally, the modelization of interest rate processes is related to the above change of variable. A simple one factor model for the short rate can be writen as

$$dy = (y_0 - y)\frac{\delta t}{\tau} + \sigma\, s(y)\, dw.$$

The first term $(y_0 - y)\ \delta t/\tau$ introduces a mean reversion at a time horizon τ, the second term introduces a random component through a Wiener process dw. Depending on the particular functional form for $s(y)$, various models can be writen. Standard choices are $s(y) = 1$ (Vasicek) and $s(y) = \sqrt{y}$ (Cox, Ingersoll and Ross). A change of variable $x = x(y)$ can be chosen so that the process is written as

$$dx = m(x)\frac{\delta t}{\tau} + \sigma\, dw \tag{19}$$

namely the random term appears as a simple additive component, but the drift factor $m(y)$ is more complex. This leads to a differential equation $\partial_y x(y) = 1/s(y)$, which can be solved to find the appropriate change of variables. For the Cox-Ingersoll-Ross model, the transformation is given by $x = 2\sqrt{y}$. Therefore, the above analysis for the yield-volatility correlation shows that the *sqrt* transformation that is implicit in the Cox-Ingersoll-Ross process is effective at decoupling yield and volatility. The analysis also suggests building an interest rate model with $s(y) = y_0 + y$, leading to $x = \ln(y_0 + y)$. This model will be as effective at decoupling yield and volatility. This line of thought with variable changes suggests writing the mean reverting term for x as a simple form (i.e. $m(x) = x_0 - x$), but with a more complex form for the equivalent drift $m(y)$ in the equation for y.

For risk evaluations, say with Monte Carlo simulations, the yield processes is simulated. In our framework, the simulations are done using eq. 19, and $m(x)/\tau$ should be estimated. The form for $m(x)/\tau$ is set indirectly through the lagged correlations for the returns. Yet, the empirical investigation in sec. 8 (page 29) shows a more complex situation than a simple one factor model for interest rates. In particular, $m(x)/\tau$ is trend following, and a mean reversion shows only for time horizons of the order of one year or above.

5 Probability distribution for the returns

The key concern for market risk evaluation is the probability distribution for large price changes, namely the tails behavior of $p(r)$. Because the mean volatility of the various time series can be very different, it is better to use $r' = r/\,\mathrm{stdDev}(r)$ as the variable for such a study, with

$$\mathrm{stdDev}^2(r) = \frac{1}{n}\sum_t r^2(t) \tag{20}$$

and where n is the number of terms in the sum. Notice that this standardization of the returns uses the full sample information. The tail behavior is conveniently studied by the empirical cumulative probability density $\mathrm{cdf}(r)$. For negative residuals, the tail corresponds to the convergence of the cdf toward zero, whereas for positive residuals, the tail is given by the convergence toward one. In order to have similar figures for both tails, we plot the cdf versus $-r'$ for the negative tail, and $1-\mathrm{cdf}$ versus r' for the positive tail.

The resulting figures for the G10 data set are given in fig. 4 and 5 for the negative and positive tails respectively. Clearly, the distribution at one day is well described by a distribution with fat tails, but badly by a Gaussian. With increasing time horizons, we can observe the slow convergence toward a Gaussian. Concurently, the sizes of the data samples diminish, and it becomes increasingly difficult to make clear assertions about the tail behavior for $p(r)$.

There is a common confusion in this field between the probability distribution of the returns and of the residuals. Let us emphasize that our approach to risk is based on the residuals ε, and the important distribution for the risk methodology is $p(\varepsilon)$. It is however interesting to compare $p(r)$ with $p(\varepsilon)$, and we will return on this topic in sec. 14.

6 Lagged correlations for the volatilities

The dominant feature of financial time series is the heteroscedasticity, or volatility clustering. The clustering is quantified by the lagged correlation of some measures of volatility, and this lagged correlation decays slowly. The non zero lagged correlation means that there is information in the past about the future volatility, and that forecasts for the volatility can be computed. It is therefore important to understand and quantify in details the available information in order to build good forecasts $\tilde{\sigma}$. Indeed, volatility clustering and the related forecast is at the core of our approach of risk measurement. Without volatility clustering, no volatility forecast is possible, and the only possible approach to risk management would be through an unconditional return probability distribution.

In order to select the appropriate process that captures accurately the properties of the financial data, the empirical volatility memory needs to be characterized in more details. This is done typically by studying the lagged correlation of the absolute value of the daily return $|r[\delta t]|$,

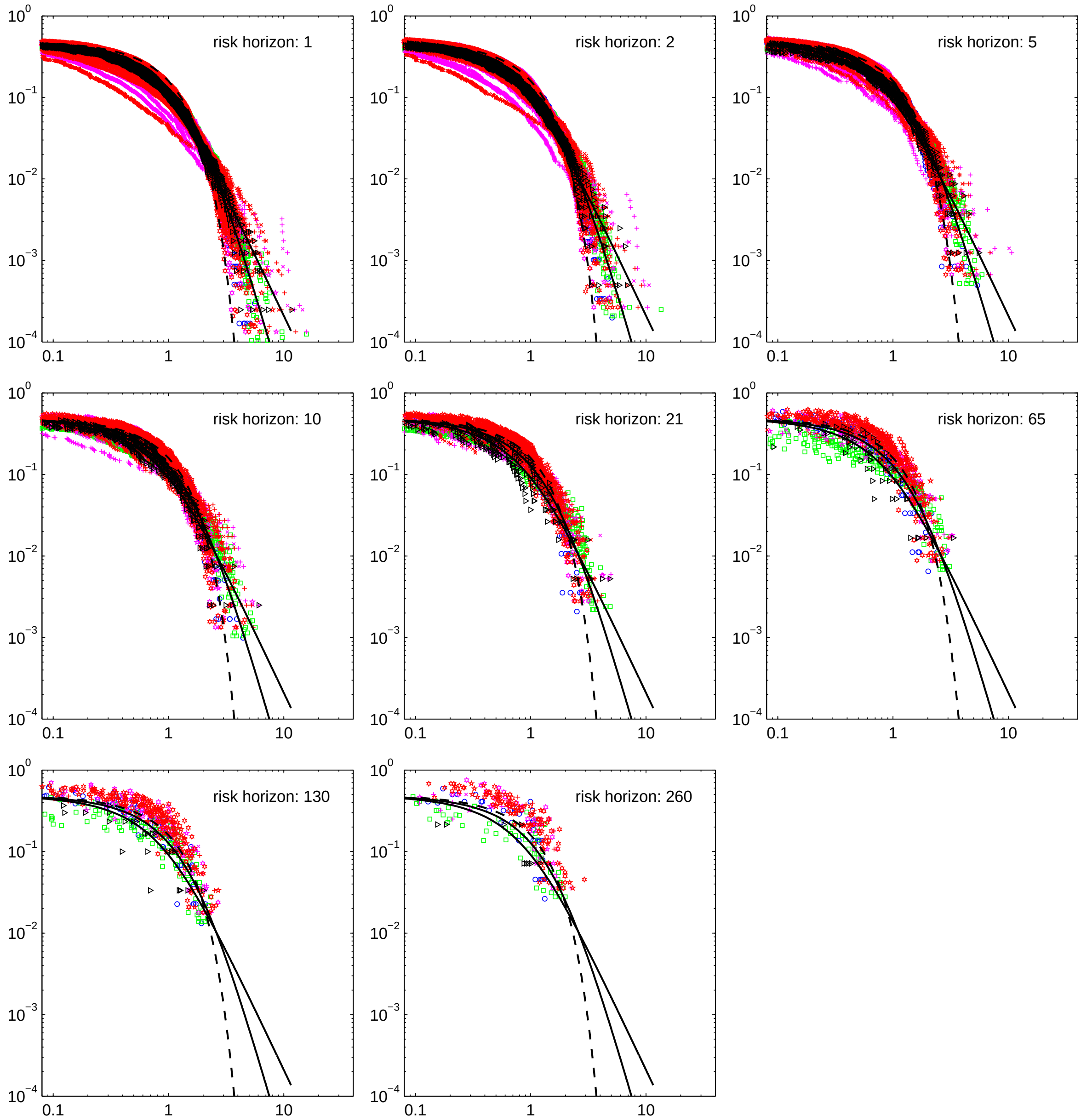


Figure 4: The cumulative probability density $\mathrm{cdf}(r)$ versus $-r/\,\mathrm{stdDev}(r)$ for the negative tails. The time series are the G10 data set. The solid line corresponds to a Student distribution with 3 and 5 degrees of freedom (rescaled to have a unit variance); the dashed line corresponds to a standard normal distribution.

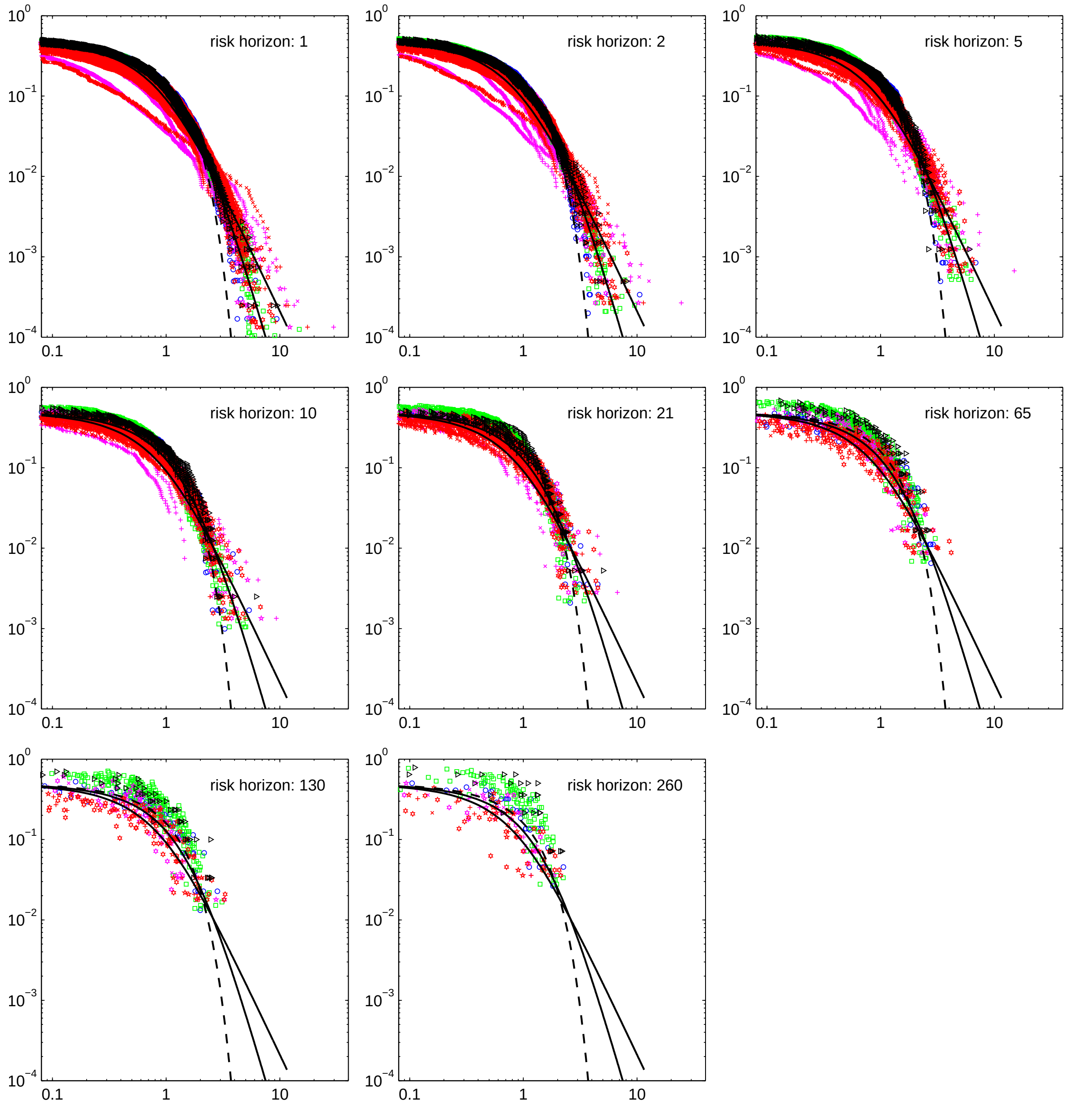


Figure 5: The cumulative probability density $1 - \text{cdf}(r)$ versus $r/\text{stdDev}(r)$ for the positive tails. The time series are the G10 data set. The solid line corresponds to a Student distribution with 3 and 5 degrees of freedom (rescaled to have a unit variance); the dashed line corresponds to a standard normal distribution.

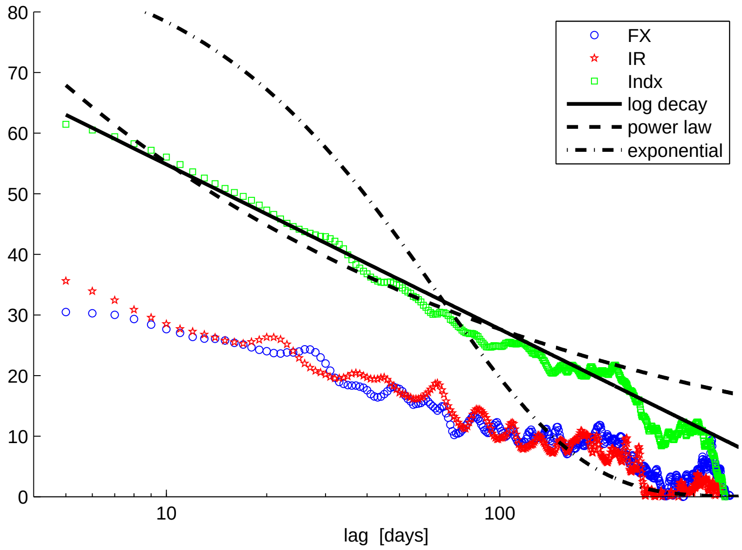


Figure 6: Lagged correlation for the 10 days realized volatility, and averaged over the G10 countries. The parameters for the theoretical curves are as follows: log decay: ρ_0 = 82%, τ_0 = 4 years; power law: ρ_0 = 110%, $\nu = 0.3$; exponential: ρ_0 = 90%, τ_0 = 65 days.

however such a study can be done with any measure of volatility, and/or more robust measures of dependency. The difficult task is to pinpoint the behavior for large lags, as the size of the independent sample is shrinking. In order to find the most convincing answer, we have combined two improvements over the plain lagged correlation of the daily absolute return.

The first improvement is to select a better volatility estimator. There is a trade-off between the variance of the volatility estimator, the size of the (independent) volatility sample used to compute the correlation, and the range for the lags. The absolute value of the daily returns, or one day volatility, is a very poor volatility estimator, but leads to the largest data sample and range for the lags. The realized volatility for a few days is a better volatility estimator, but the independent sample and the lag range are smaller. In this trade-off, we found that the optimal is between 5 days to 1 month realized volatility, and the figures below are presented for the 10 days realized volatility.

The second improvement is to pool together different time series. The error for the computed correlation for one time series of 15 years can be estimated as follows. The memory for the volatility has a slow decay, but an equivalent exponential decay could be around 3 months (see fig. 6). The number of independent data points can be estimated by $N =$ 180 months/3 months = 60 points. The error on the correlation is of the order of $1/\sqrt{N} \sim 8\%$. At a lag of one year, the lagged correlations are between 5 to 30%. Clearly, the error due to the sample size is large, and preclude to have a clear view for the memory decay. Using a robust correlation estimator helps in reducing the prefactor for the statistical error. The only other way to reduce the statistical error is to average the lagged correlations over a set of time series (assuming a similar decay). For this purpose, we have computed simple means, with equal weights, of the lagged correlations for the different asset classes (FX, IR, stock indexes, ...).

On fig. 6, we plot the lagged correlation for the 10 days realized volatility, aggregated by asset classes, for the G10 countries. Fig. 7 is the same graph, but averaged over all the time series in the ICM data set. For this second graph, because of the equal weighting scheme, and because of the respective number of time series in the G10 and ICM data sets, the largest contributions originate in the non G10 countries. Similar results are obtained with other robust volatility estimators and other correlation estimators. A good simple description of the data is given by

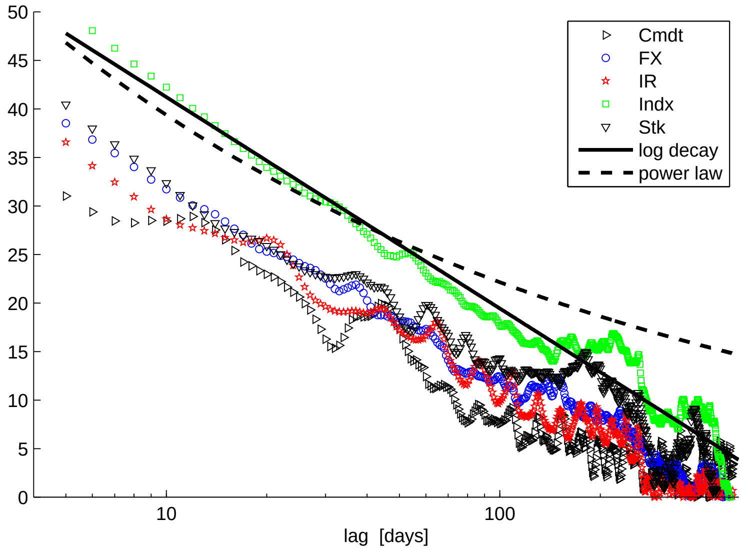


Figure 7: Lagged correlation for the 10 days realized volatility, and averaged over the world ICM data set. The parameters for the theoretical curves are as follows: log decay: ρ_0 = 63%, τ_0 = 3 years; power law: ρ_0 = 70%, $\nu = 0.25$.

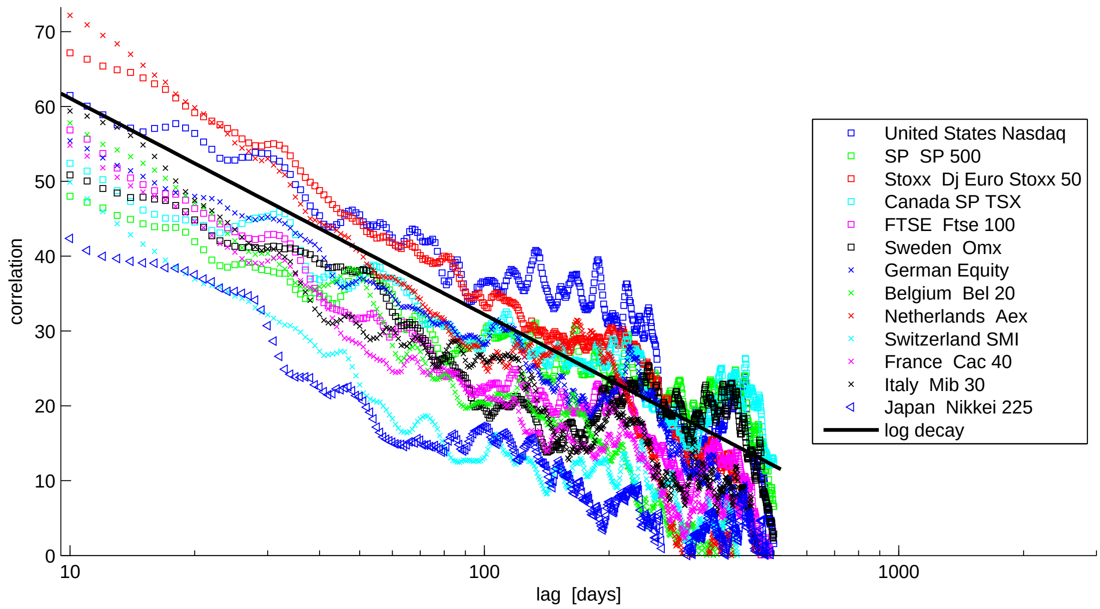


Figure 8: Lagged correlation for the 10 days realized volatility, for the stock indexes in the G10 countries. The parameters for the theoretical curve are as follows: log decay: ρ_0 = 90%, τ_0 = 5 years.

a log decay

$$\rho_\tau = \rho_0 \,\left(1 - \frac{\ln(\tau/\delta t)}{\ln(\tau_0/\delta t)}\right) \tag{21}$$

with τ_0 of the order of 3 to 6 years and ρ_0 depending on the time series. This analytic form seems to be valid up to 1 to 2 years, possibly with a faster decay for larger lags. The usual analytical description for the lagged correlations is done with a power law

$$\rho_\tau = \rho_0 \,\left(\frac{\delta t}{\tau}\right)^{\nu}. \tag{22}$$

We find that this second form gives consistently an inferior description of the empirical data, for all the volatility estimators and correlation estimators. A strong indication for the validity of the log decay is that the aggregation for different asset classes (i.e. with completely unrelated time series) shows consistently that a log decay is superior to a power law. The aggregated curve for the stock indexes is likely the best one, as each index is already an aggregation over many stocks, and the individual indexes are fairly independent in the G10 set. On the contrary, there are more redundant time series in the IR and FX sets (all the European countries have similar behaviors). The results for most individual time series are also consistent with a log decay, and inconsistent with a power law decay, albeit the noise is larger. For example, on fig. 8 is displayed the lagged correlations for the stock indexes for the G10 countries. To summarize, we consistently find that the best description of the volatility memory is given by a logarithmic decay, for time tags between a few days to one year. Moreover, the decay range τ_0 has a very similar value $\tau_0 \simeq 5$ years for all time series. This suggest that a similar mechanism creates this long memory, and that one decay factor can be used to model all time series.

Notice that this analytical description of the data goes against the usual claim that the lagged correlations for the volatility decays as a power law. For $\nu \ln(\tau/\delta t) \ll 1$, both analytical descriptions are related by

$$\rho_0 \,\left(\frac{\delta t}{\tau}\right)^{\nu} \simeq \rho_0 \,\left(1 - \nu \ln(\tau/\delta t)\right). \tag{23}$$

The relationship between the parameters is $\nu = 1/\ln(\tau_0/\delta t)$, and with $\tau_0 = 4$ years we obtain $\nu \simeq 0.14$. This explains the consistent small values for the exponent ν reported from power law estimates of empirical data. Yet, beware that on the graphs 6 to 9, the above approximation is valid up to $\simeq$32 days. A better fit with a power law decay is obtained with larger exponents (e.g. for the graphs $\nu = 0.3$), but the best value for the exponent is dependent of the selected domain for the lags. For comparison purposes, we have also plotted on fig. 6 an exponential decay $\rho_0 \exp(-\tau/\tau_0)$ (corresponding for example to the I-GARCH(1) and GARCH(1,1) processes). The exponential decay is easily excluded as a good description of the empirical time series.

The usual (Pearson) correlation is fragile with respect to large events, and therefore not the most appropriate for variables with fat tails distributions. We have investigated two other robust measures of dependency, namely the Kendall's Tau ρ_τ and ρ_{SSD}. The definitions for the correlation estimators are given in Appendix C. The robust correlation based on the standardized sums and differences ρ_{SSD} proves to give a less noisy answers than the usual linear correlation, and still with a decent computational time. The Kendall's Tau ρ_τ gives very similar values compared to ρ_{SSD}, but with a much longer computational time (it scales as $O(n^2)$ instead of $O(n)$). Using these robust estimators lead to the same conclusions as with the usual correlation, namely a logarithmic decay is always a better description of the decay of the

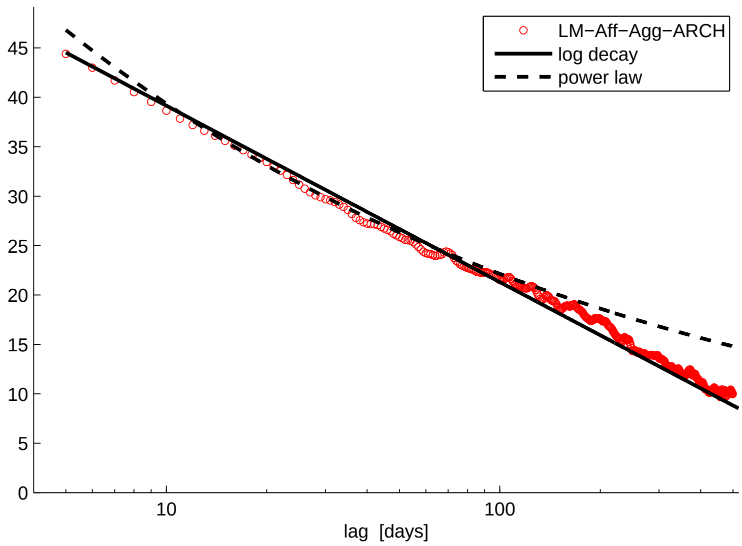


Figure 9: Lagged correlation, evaluated with the ρ_{SSD} estimator, for the 10 days realized volatility, for a Long Memory - Affine - Microscopic - ARCH process. The simulation length is of 1000 years. The parameters for the theoretical curves are as follows: log decay: ρ_0 = 57%, τ_0 = 6 years; power law: ρ_0 = 80%, $\nu = 0.45$;.

lagged correlation. The major difference is that the empirical curves are less noisy, and the parameter values for the analytic description are slightly different.

This finding has strong implications on the choice of the volatility process. The memory kernel, as specified by the weights w_k, should be chosen so as to reproduce the empirical memory. The lagged correlation for the LM-ARCH process cannot be computed analytically, but can be estimated by Monte Carlo simulations (using the affine version of the process, see [Zumbach, 2004]). For the simulation, the parameters for the process are choosen according to the values found in the next section 7. Fig. 9 displays the lagged correlation for the 10 days volatility, for the long memory process using daily returns. The agreement is clearly very good, showing that this process capture the correctly the decay of the volatility lagged correlations.

Beyond the precise analytical description of the generic decay for the volatility lagged correlation, the difference between a power law and logarithmic decay is quantitatively small on the accessible range of lags. Using backtesting for the two long memory specifications shows indeed very small differences. For the consistency of the methodology with the empirical stylized fact, we decide to use a logaritmic decay for the process.

7 Volatility forecast

The volatility forecast is the single most important part in a risk methodology. With the process set-up, the forecast depends only on the process parameters. These parameters are the logarithmic decay factor τ_0, the lower cut-off τ_1, and the upper cut-off $\tau_{\max} = \tau_{k_{\max}}$. A-priori values for them could be as follows. The logarithmic decay factor τ_0 has a value estimated from the empirical lagged correlation of $|r|$ or of the volatility; the estimates are in the range 3 to 6 years. The lower cut-off τ_1 should be of the order of one to a few days: there are more information in the most recent past, but a small lower cut-off puts strong weights on the few last returns resulting in a noisy estimator. The upper cut-off $\tau_{\max}$ should be in the range of a few months to a few years. The origin of the heteroscedasticity is the memory of the traders and the strategies they follow; the longest time horizon for market participants are probably of a few months to one year (likely pension funds and central banks). A direct estimate using empirical data proves difficult because of the increasing noise for very large lags, yet most lagged correlations for the volatility seems to decay faster than a logarithmic decay above one year. These three values give a first plausible estimate for the parameters. Notice also the advantage of our choice for the parameters, as we can have a fairly intuitive perception of their meaning and respective values. We denote the three dimensional vector of parameters by $\theta = (\tau_0, \tau_1, \tau_{\text{max}})$.

The process allows us to compute a volatility forecast at horizon ΔT that should be compared with the realized volatility. The realized volatility (or quadratic variation) is defined by

$$\sigma^2_{\text{real}} = \sum_{t+\delta t \le t' \le t+\Delta T} r^2[\delta t](t') \tag{24}$$

The annualized volatility is obtained by multiplying by 1 year$/\Delta T$. The sum contains $\Delta T/\delta t$ terms, leading to a noisy estimator of the realized volatility for short risk horizons. This poor realized volatility estimator for short risk horizon implies that the measures of quality for the forecast (defined below) becomes worst for short horizons.

Straight forward measures of the forecast accuracy are L_1 and L_2 distances. In our context, we want to compare forecasts for time series with very different volatility, and it is therefore more appropriate to use relative measures of the forecast accuracy. To define the relative accuracy, we use the sample mean volatility as the reference forecast. This leads to the following measures of distance

$$L_{1,\text{rel}} = \frac{\langle |\tilde{\sigma} - \sigma_{\text{real}}| \rangle}{\langle |\langle \sigma \rangle - \sigma_{\text{real}}| \rangle} \tag{25}$$

$$L^2_{2,\text{rel}} = \frac{\langle (\tilde{\sigma} - \sigma_{\text{real}})^2 \rangle}{\langle \left(\sqrt{\langle (\sigma^2 \rangle} - \sigma_{\text{real}} \right)^2 \rangle} \tag{26}$$

where $\langle \cdot \rangle$ denotes the sample average. With these normalizations, a perfect forecast has a distance of zero to the realized value, and a forecast as good as the sample mean has a distance of 1.

The volatility forecast accuracies have been investigated in the parameters space in order to find a good value $\theta^\star$ for the parameters. The goal is to find one set of parameters that can produce good forecasts for all assets and time horizons. The results are summarized in the graphs 10, 11 and 12, showing one dimensional cuts along the parameter axis through the point $\theta^\star$. Good overall optimal parameter values $\theta^\star$ are $\tau_0 = 1560 = 6$ years, $\tau_1 = 4$ days and $\tau_{\text{max}} = 512$ days. Fig. 10 shows the dependency on the logarithmic decay factor. The horizontal axis gives the values for the logarithmic decay factor τ_0, while the other parameters are at the $\theta^\star$ values. The vertical axis is the ratio

$$L_{2,\text{rel}}(\theta)/L_{2,\text{rel}}(\theta^\star). \tag{27}$$

This quantity measures the forecast accuracy compared to the forecast using the reference parameters $\theta^\star$. The curves correspond to the different forecast horizons ΔT; values below 1 indicate that the forecast can be improved by changing the parameter value compared to $\theta^\star$. The set of curves shows that less that 0.5% of the performance forecast can be gained by optimizing the logarithmic decay factor as a function of the forecast horizon. This small gain shows that the same parameter value for τ_0 can be used for all time horizons.

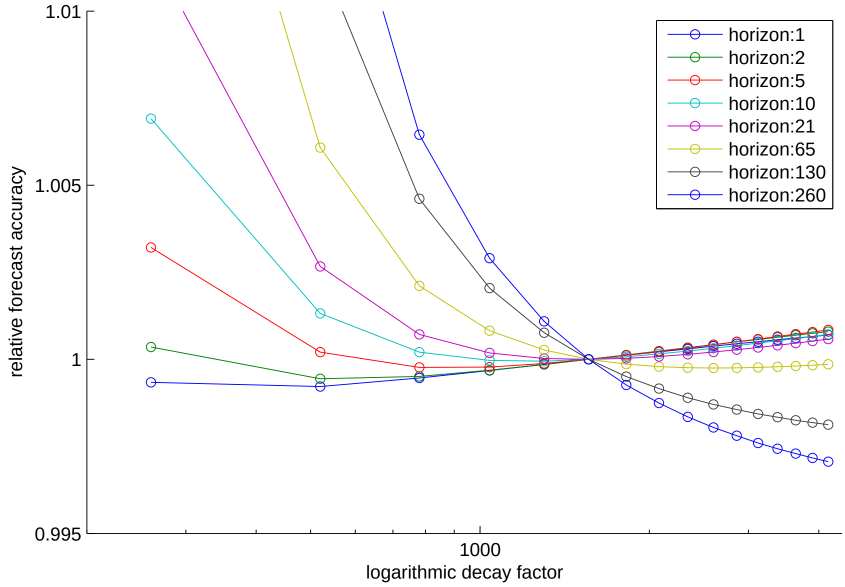


Figure 10: The relative volatility forecast performance as a function of the logarithmic decay factor, for various forecast horizons. The curves are normalized to the reference point $\theta^\star$.

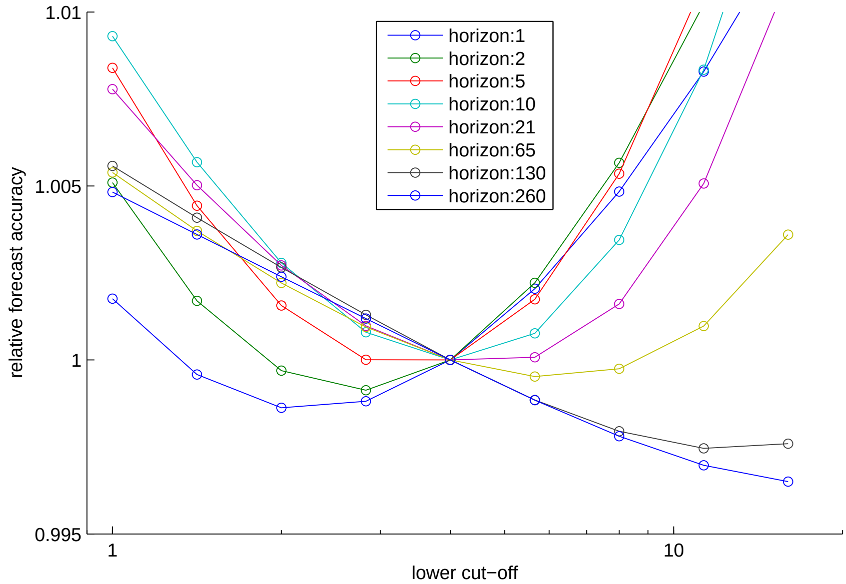


Figure 11: The relative volatility forecast performance as a function of the lower cut-off, for various forecast horizons. The curves are normalized to the reference point $\theta^\star$.

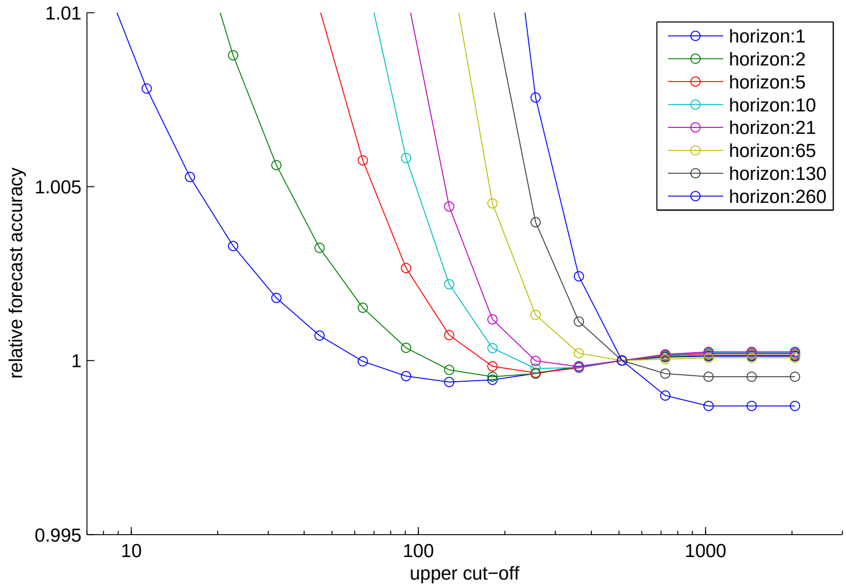


Figure 12: The relative volatility forecast performance as a function of the upper cut-off, for various forecast horizons. The curves are normalized to the reference point $\theta^\star$.

The same analysis is performed with respect to the lower and upper cut-offs, as shown on fig. 11 and 12. The forecast performance has a similar sensitivity with respect to these two parameters, and fine tuning these parameters leads to improvement of less than 0.5%.

As a comparison, the same computation is done for the RM1994 methodology. In this case, the parameter space for the I-GARCH process is one dimensional and corresponds to the characteristic time of the exponential. Fig. 13 displays the same set of curves as for the RM2006 methodology. Notice that the vertical scale is expanded by a factor 40 compared to the graphs for RM2006. The reference point $\theta^\star$ is taken at 16 days, and corresponds to the usual value $\mu = 0.94$. This value is nearly optimal for forecast horizons up to 10 days. For longer horizons, it is better to use larger characteristic times, and the optimal value increases with the forecast horizon. Clearly, for longer horizons, the performance gain obtained by taking the optimal parameter value can be quite large.

The figures 10 to 13 show the volatility performance relative to the reference points. The volatility performance $L_{2,\text{rel}}(\theta^\star)$ at the reference point $\theta^\star$, as a function of the forecast horizon, is showed on fig. 14. Between two days and three months, the relative performance of the volatility forecast is better than the in-sample mean. At shorter horizons, the forecast accuracy decreases due to the poor estimator for the realized volatility. For example, at one day, the realized volatility is computed with only one return, leading to a very large variance for the realized volatility estimator. This is probably the origin of the apparent lower performance at short time horizons. Yet, going beyond this simple argument would require to measure the realized volatility using high frequency data. Notice also that the forecast given by the long memory process is consistently better than the I-GARCH process.

All the graphs are for the $L_{2,\text{rel}}$ measure of forecasting performance. Very similar results are obtained for the $L_{1,\text{rel}}$ measure, possibly with a slightly different optimal point $\theta^\star$. As for the precise measure of the forecast quality, this analysis depends on the set of time series. Clearly, going beyond the broad picture would be overfitting the data sample. The important point is that one set of parameters is able to deliver a good forecast for all time horizons and all time series, and that the values for the parameters make intuitive sense.

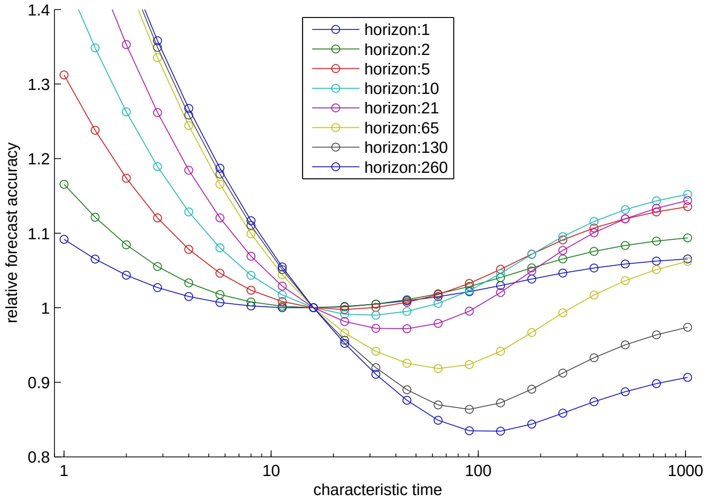


Figure 13: For the RM1994 methodology, the relative volatility forecast performance as a function of the EWMA characteristic time, for different forecast horizons. The curves are normalized to the reference point $\tau = 16$ days.

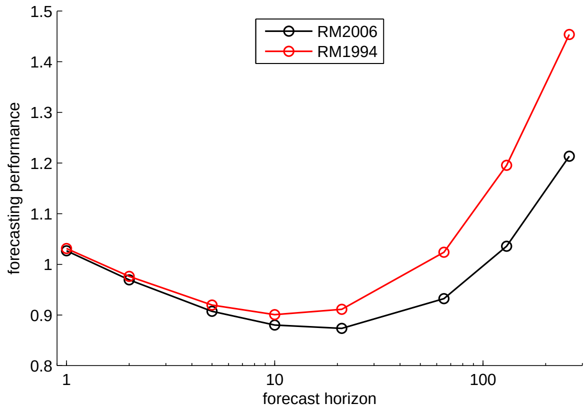


Figure 14: The volatility forecast performance as function of the forecast horizon ΔT (smaller values are better).

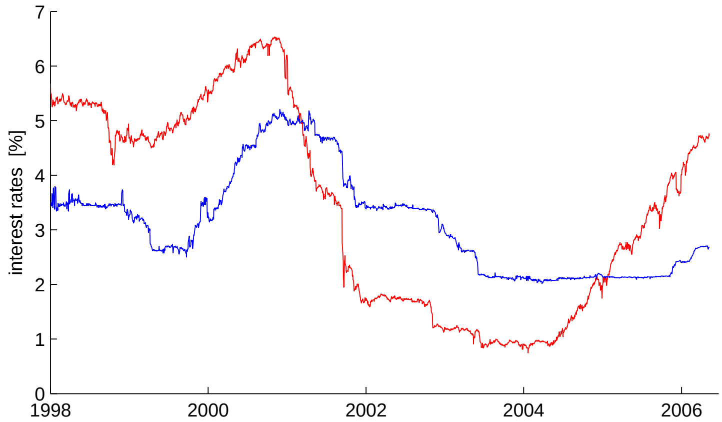


Figure 15: The short term interest rates for USD (red curve; 1 month government debt benchmark) and Euro (blue curve; 1 month swap).

8 Lagged correlations for the returns

For a liquid and free floating asset, the lagged correlations of the returns have to be zero (up to statistical noises). If this is not the case, the correlations can be used to set a profitable trading strategy, until the correlations disappear. This simple and powerfull argument is used to neglect the return forecast in eq.15 or in the process equation 2.

They are however two points of concern with setting $\tilde{r} = 0$ from the start. First, stocks and stock indexes should earn in average the risk free rate of return, plus possibly a risk premium. This indicates that for these time series, we should include a long term positive mean return, or a long term lagged correlation. Second, short term interest rates are set by central banks, namely they are not free floating. Their decisions are often quite predictable within a time horizon of the order of one month. For example, the short term interest rates for USD and Euro are given on fig. 15: there are clear trends that can extend for more than a year. Clearly, these time series are not pure random walks, and the distinctive trends correspond to correlations between lagged returns. Therefore, the trading argument given in the previous paragraph should be somewhat tonned down: the lagged correlations for the returns must be small.

We apply the same technique that is used for the volatility to the returns, namely study the return lagged correlation in order to search for information in the past data, construct various forecasts built on this information, and consider the impact of these forecasts on the risk evaluation. The key difference is that the available information to be extracted from the data is much smaller, and therefore this study is much more difficult. To make a simple analogy, a Gaussian random walk would be the zero-th order model for the price, adding the heteroscedasticity is the first order correction, while the correlation for the returns is the second order correction. As we will show in sec. 15.1, the larger contribution induced by the lagged correlation of the returns is through the correction to the volatility forecast (as derived in eq. 58). The return forecast has a much smaller impact, mainly on the lagged correlation for the residuals.

In order to enhance the signal for the lagged correlation of the returns, the figures reported here are for an equally weighted mean of the (robust) lagged correlation of various time series in the G10 data set. Fig. 16 is for the one day returns: the lagged correlations are at the noise level. The only noticeably feature concern interest rates, where the correlations are mostly positive for lags larger than a few days. Fig. 17 is for the monthly returns: the lagged correlations seem at the noise level for foreign exchanges and interest rates. Yet, the lagged correlations are fairly large for interest rates, at a distance of 3 to 5 σ from zero. For the usual linear

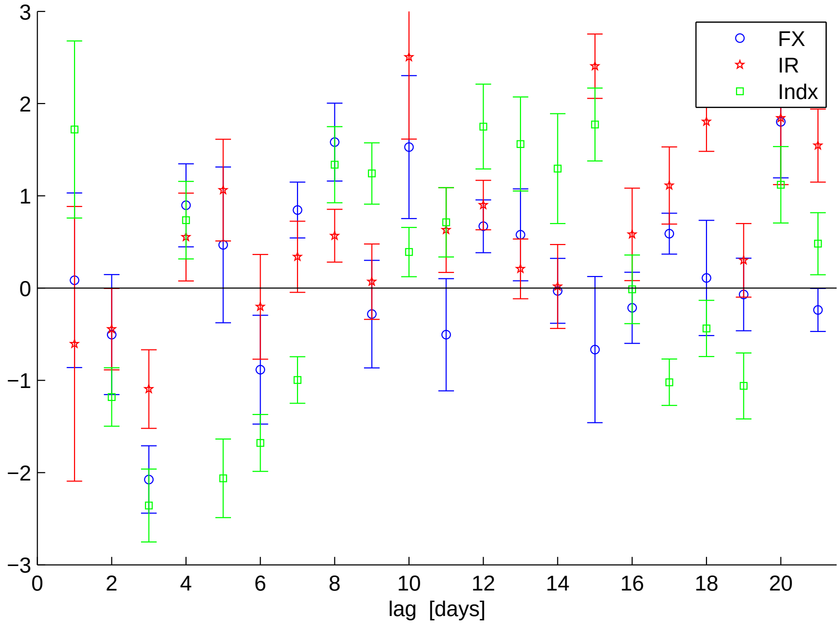


Figure 16: The robust lagged correlations ρ_{SSD}, in %, of the daily returns in the G10 data set, averaged for each asset classes. The error bars give the empirical standard deviation for the mean.

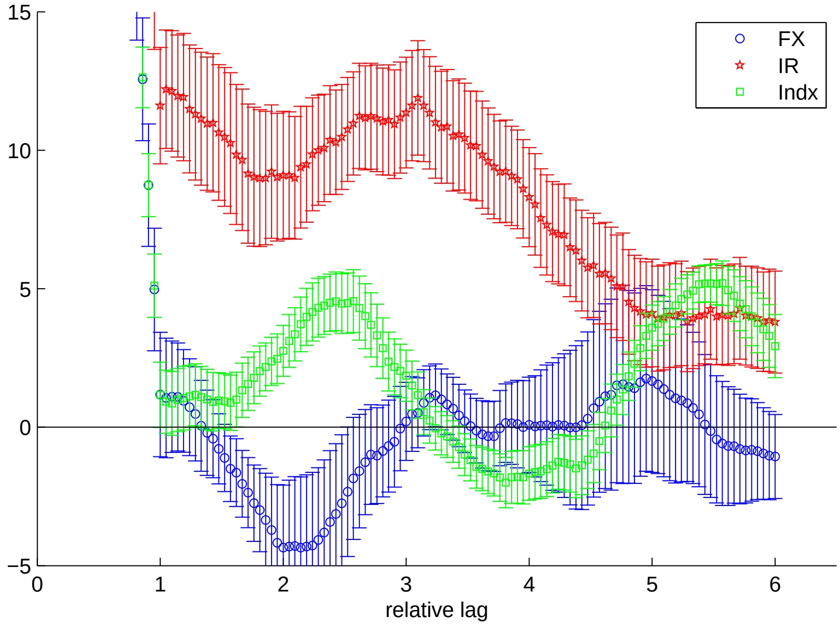


Figure 17: The robust lagged correlations ρ_{SSD}, in %, of the monthly returns in the G10 data set, averaged for each asset classes. The error bars give the empirical standard deviation for the mean. The x-axis is the relative lag $\tau_r = \tau/\Delta T$ (with τ the lag and ΔT the return horizon). For $\tau_r < 1$, the returns overlap; at $\tau_r = 1$ is the first point without overlap for the variable.

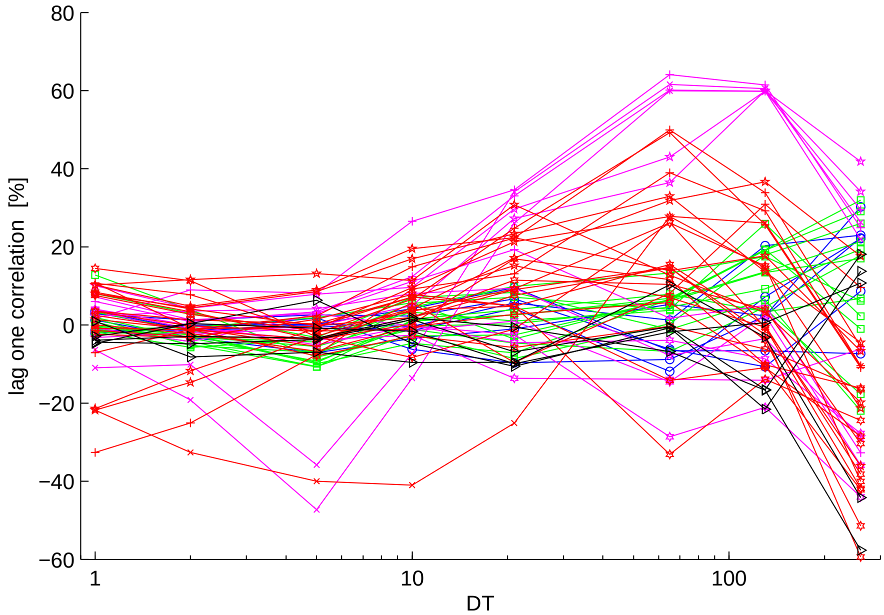


Figure 18: The lagged correlations $\rho_{\Delta T}$, at lag ΔT, in %, of the returns $r[\Delta T]$, in the G10 data set. The colors are blue for FX, green for stock indexes, and red for IR.

correlation, the figures are similar, with the lagged correlations of the monthly returns for the interest rates of the order of 12%. This is getting large!

The analytical computations of the corrections due to the return lagged correlations in the process are detailled in appendix B, and the mechanism behind these larger correlations can be understood from eq. 60. Beyond the analytical formula, the intuition is that the correlations originate likely in the central bank decisions for the short term rates. Essentially, small but positive one day correlations add up, creating larger correlations for longer time intervals. The best summary of the effect is visualized by the "lag one" correlations, namely the correlation between $r[\Delta T](t)$ and $r[\Delta T](t+\Delta T)$, as a function of ΔT. On fig. 18, the different behavior of the IR is clearly visible, in particular the correlations are essentially *positive* for lags ΔT between 10 days and 3 months. This goes against a simple mean reversion term that leads to negative lagged correlations. Another interesting observation is the sharply decreasing correlations for the longest time intervals. This should be due to the long term mean reversion of the interest rates, and the figure is roughly consistent with a mean reversion time of the order of one to a few years. Overall, this figure shows the complexity of the lagged correlation of the returns for interest rates, and that different mechanisms take place at different time horizons. Most of the existing data generating processes for IR use a simple mean reversion term, and therefore are not able to capture the observed multiscales behaviors. The conclusion of this analysis is that the lagged correlations of the returns cannot be neglected, but more fundamental work would be needed to gain a deeper understanding and to build finer models.

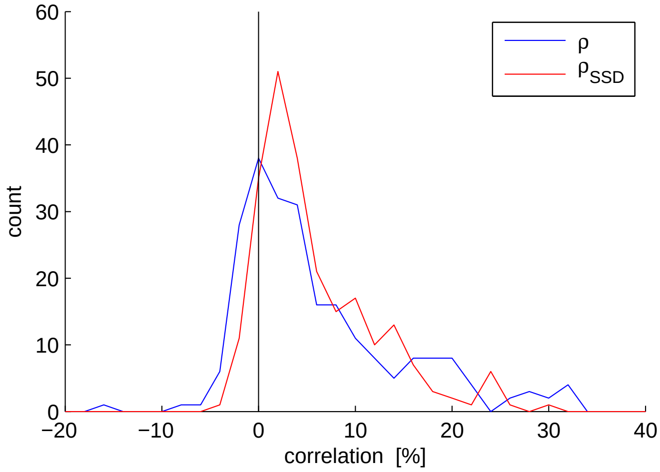


Figure 19: The density histogram for the correlations ρ and ρ_{SSD} between the forecasted and realized returns. The time series are the ICM data set.

9 Return forecast

To define relative distances for the return forecast, we use a zero forecast as benchmark. This choice leads to the definitions

$$L_{1,\mathrm{rel}} = \frac{\langle|\tilde{r} - r_{\mathrm{real}}|\rangle}{\langle|r_{\mathrm{real}}|\rangle} \tag{28}$$

$$L^2_{2,\mathrm{rel}} = \frac{\langle(\tilde{r} - r_{\mathrm{real}})^2\rangle}{\langle r^2_{\mathrm{real}}\rangle} \tag{29}$$

As already discussed for the definition of the residuals, the realized return at time t is the historical return at time $t + \Delta T$. With these choices, $L_{\mathrm{rel}} = 0$ means a perfect forecast and $L_{\mathrm{rel}} = 1$ is a forecast as good as $\tilde{r} = 0$.

The $L_{2,\mathrm{rel}}$ is related to the linear correlation (up to factors $\langle\tilde{r}\rangle$ and $\langle r_{\mathrm{real}}\rangle$ that are neglected): by expanding the square and denoting by $p = \langle\tilde{r}^2\rangle / \langle r^2_{\mathrm{real}}\rangle$, the optimal value for p is $p = \rho$. For this optimal value, we have the relation $L^2_{2,\mathrm{rel}} = 1 - \rho^2$; for example a 20% correlation implies $L_{2,\mathrm{rel}} = 0.98$, namely a small departure from 1.

Other measures of quality are the correlations between forecasted and realized returns, computed either with a simple linear correlation ρ or the robust correlations ρ_τ or ρ_{SSD}. We tried many possibilities for the forecast, but none proved clearly superior to the others. Often, some measures of quality can be improved but at the expense of other measures of quality. After testing many variations on the return forecasting formula, including the analysis per asset types (IR, FX, Index, etc...), we set on the one given at the end of appendix B.

Fig. 19 gives the correlations between forecasted and realized returns. The asymmetry on the positive side indicates that we are clearly capturing some information in the forecast. Yet, the relative distances plotted on fig. 20 show that in the majority of the cases, we do worst than a nil forecast $\tilde{r} = 0$.

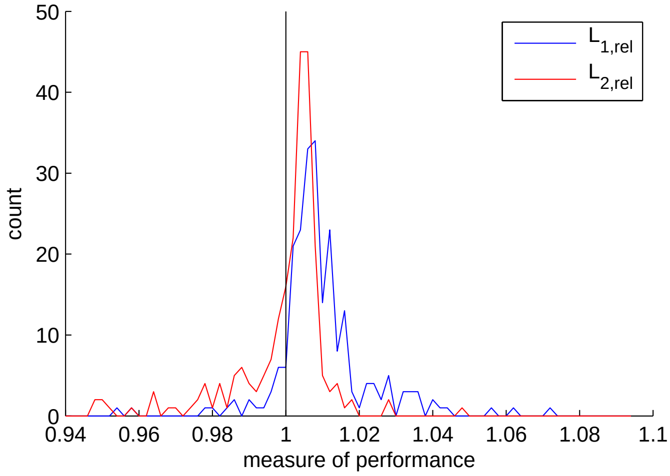


Figure 20: The density histogram for the relative distances $L_{1,\text{rel}}$ and $L_{2,\text{rel}}$ between the forecasted and realized returns. The time series are the ICM data set.

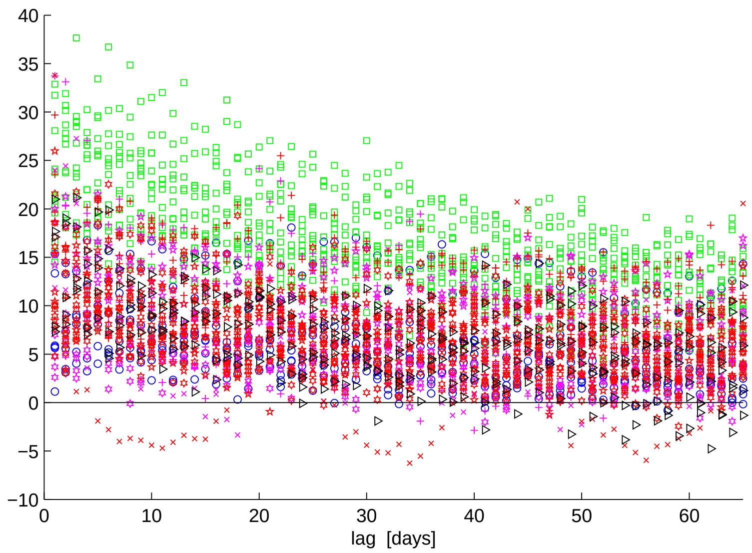


Figure 21: The lagged correlation for the daily absolute returns $|r[\delta t]|$, for the G10 data set.

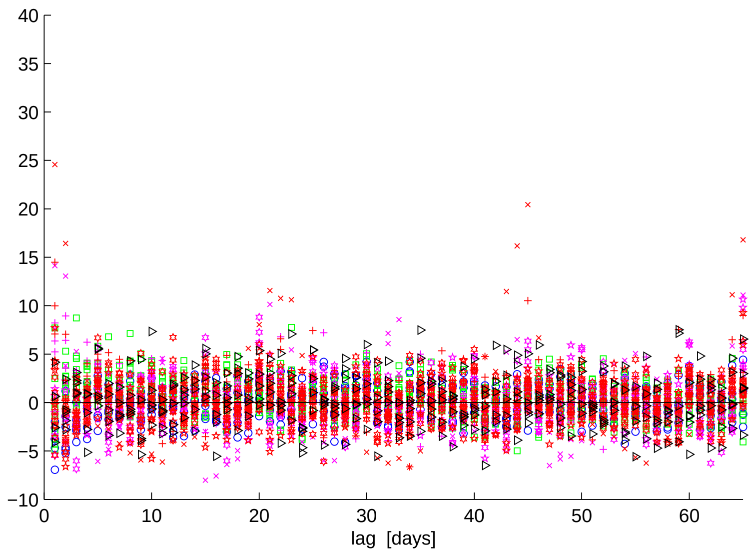


Figure 22: The lagged correlation for the daily absolute residuals $|\varepsilon[\delta t]|$, for the G10 data set.

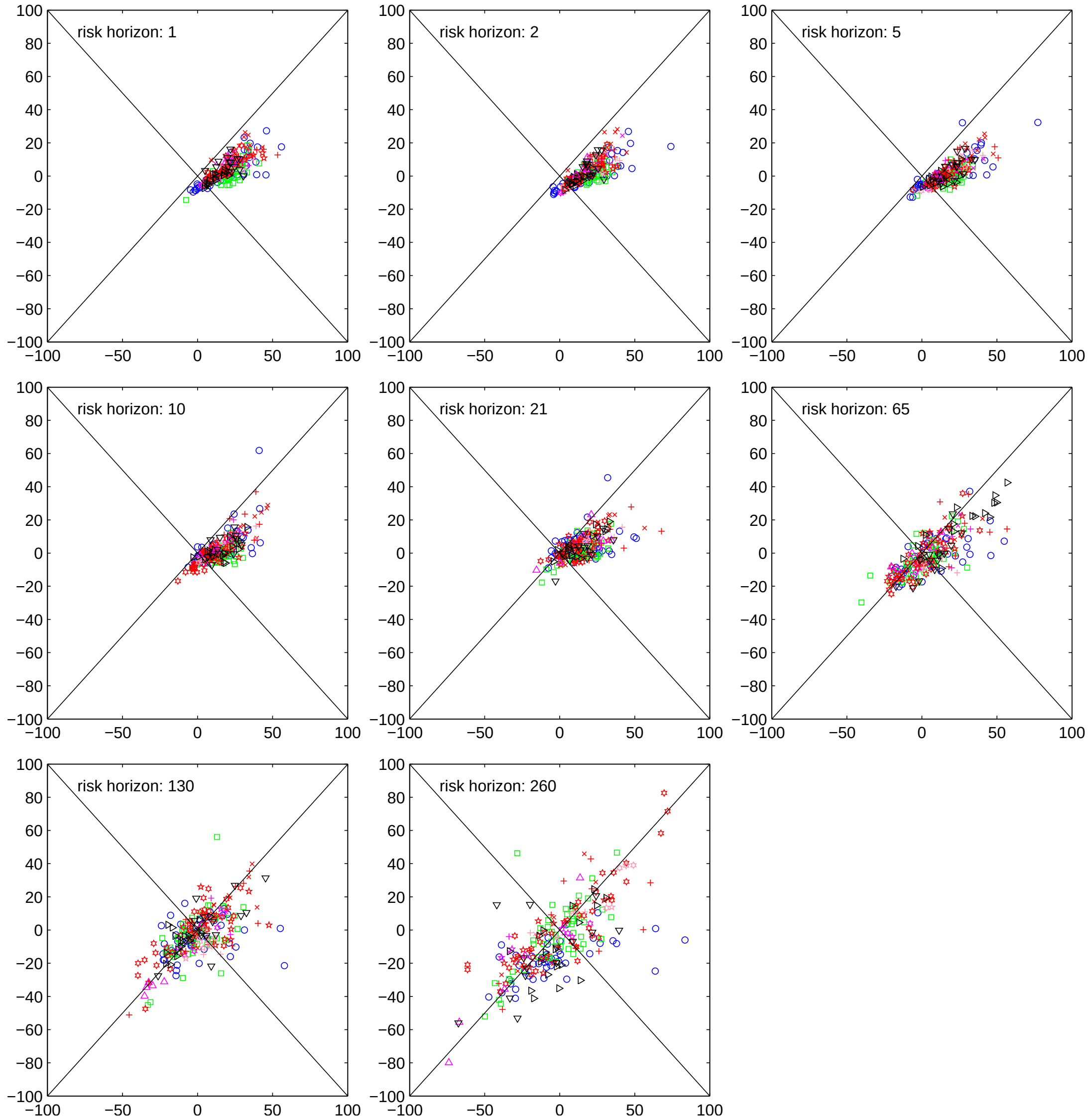


Figure 23: The “lag one” correlation for the absolute residuals $|\varepsilon[\Delta T]|$ (vertical axis) versus the “lag one” correlation for the absolute returns $|r[\Delta T]|$ (horizontal axis). One point corresponds to one time series in the ICM data set. Each panel corresponds to one risk horizon ΔT.

10 Lagged correlations for absolute residuals $|\varepsilon|$

In empirical financial time series, the largest deviations from a simple random walk is due to the heteroskedasticity. This motivates the present approach to risk, and in particular the key definitions 15 and 16 used to discount the expected volatility in the forthcoming risk horizon period. The effectiveness of the approach can be observed directly by comparing fig. 21 and 22. They show respectively the lagged correlations for the daily absolute returns and daily absolute residuals, for all the time series in the G10 data set. On fig. 21, the slow decay of the volatility memory is very clear, as well as the difference in correlation levels between different asset classes like equity indexes (green) or interest rates (red). The same computation but for the residuals shows essentially no remaining lagged correlations, as is visible on fig. 22. At closer inspection, one may notice two features. First, for a few interest rates, there is a monthly seasonality due to central bank decisions. This is visible only for short maturity rates (daily (x) and monthly (+)), at lags around 22, 44 and 65 business days. Second, at a lag of one day, there are small remaining correlations, particularly for interest rates.

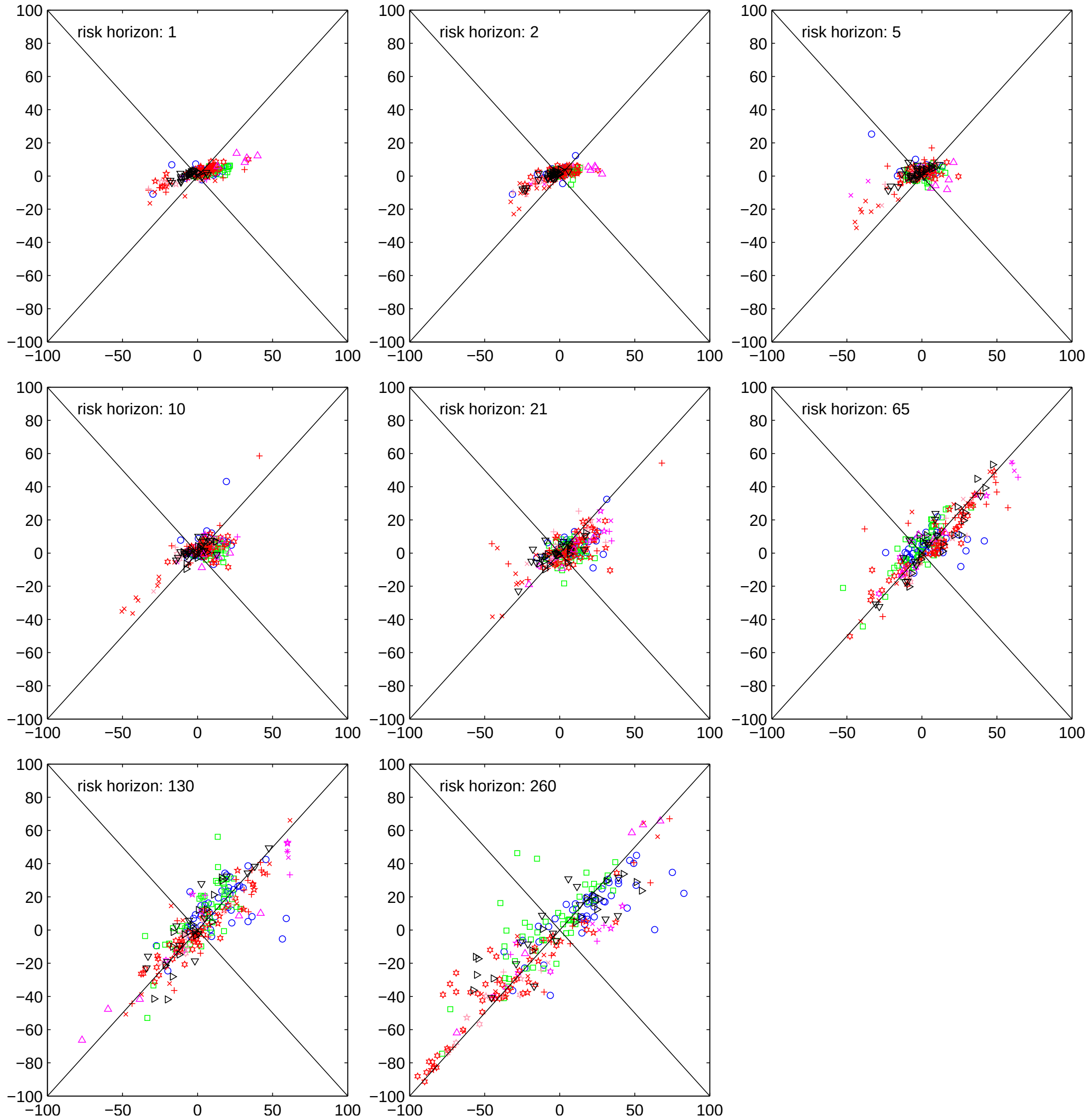


Figure 24: The "lag one" correlation for the residuals $\varepsilon[\Delta T]$ (vertical axis) versus the "lag one" correlation for the returns $r[\Delta T]$ (horizontal axis). One point corresponds to one time series in the ICM data set. Each panel corresponds to one risk horizon ΔT.

This last observation prompted us to study in more details the "lag one" correlations, namely the lagged correlations of the returns and residuals at risk horizon ΔT, at a lag of one risk horizon or ΔT days. Figure 23 shows the "lag one" correlations for the absolute residuals versus the "lag one" correlations for the absolute returns. A perfect risk methodology (applied on a infinitely long data set) should show points only along the horizontal axis, corresponding to non zero correlations for the returns but zero correlations for the residuals. A good risk methodology would show points in the east and west sectors, corresponding to smaller correlations for the residuals than for the returns. The figures show that up to a month, the heteroskedasticity is well discounted by the risk framework. Above one month, the efficiency of the volatility discounting decreases, but the statistical noise increases, so that it is difficult to draw strong conclusions.

11 Lagged correlations for the residuals ε

The lagged correlations for the returns has been investigated in sec. 8, with the important outcome that correlations at intermediate time horizons cannot be neglected (see fig. 18). The way to introduce return correlations in our process framework is straightforward, as detailled in appendix. B. Fig. 24 shows the "lag one" correlations for the residuals versus the "lag one" correlations for the returns. The improvements at risk horizons up to one week are very clear, but disappear quickly with longer risk horizons. The mostly negative lagged correlations at ΔT = 1 year for interest rate (i.e. the long term mean reversion) is also very clear. Despite the poor performance of the return forecasts, the backtesting of the RM2006 methodology shows that it increases the performance measured by the relative exceedance fraction $\delta(z)$ (see sec. 15).

12 Variance for the residuals

As discussed in sec. 2.3, the probability distribution of the residuals is such that $E[\varepsilon^2] = 1$. This condition fixes the scale of the distribution, and it becomes $\langle\varepsilon^2\rangle = 1$ on empirical data. It is very important for a risk methodology to follow this scale condition. Systematic deviations from the equality correspond to systematic over or under estimation of the risks. Indeed, these deviations are the factors that limit most the accuracy of a risk methodology. For $\langle\varepsilon^2\rangle >$ 1, the relative exceedance fraction $\delta(z)$ used in back testing (see [Zumbach, 2006a]) shows systematic deviations from zero with a shape corresponding to a too large variance. Intuitively, a too large forecasted variance leads to a too large forecasted risk, and the scalar error measures like d_p grow. Therefore, it is important to correct the systematic factors affecting the variance.

Beyond the long memory for the volatility and the Student distribution for the residuals, the RM2006 methodology includes two terms directly relevant for the scale condition. The first one corrects for the non zero lagged correlations for the returns and is given by eq. 59. The second term is the $\gamma[\Delta T]$ term included in the definition of the residuals and given in eq. 17. Because the volatility forecast is completely specified by the process, this term is a free adjustable function used to enforce the scale condition of the residuals.

A methodology including only a long memory process and Student distribution for the residuals is used for computing fig. 25 (i.e. none of the corrections above are included). With increasing risk horizons ΔT, the foreign exchanges (blue) are doing well, the stock indexes (green) show a systematic upward trend, but the interest rates (red) are quite scattered. Moreover, the quantitative deviations from 1 are quite large, leading to substantial misestimates of the risks.

The bulk of the deviations are due to correlations between the returns. Positive lagged correlations between the returns lead to larger volatilities at longer time horizons, as is clearly visible for many interest rates. The correction 59 derived from the process equations in appendix B leads to a better situation, as shown in fig. 26. Yet, there is a systematic upward trend with increasing risk horizons, as is clearly visible by the deviation from the horizontal black line. The role of the scale factor correction $\gamma[\Delta T]$ consist in absorbing this systematic deviation. A fit was done on the mean of the standard deviations, but on a smaller and shorter subset of the G10 data. This leads to formula 17, which proves to work well enough on the larger data sets. The correction can be seen at work on fig. 27 for the ICM data set, where no systematic deviation on the vertical axis can be observed.

The final results with both corrections included are shown in fig. 28 for the ICM data set.

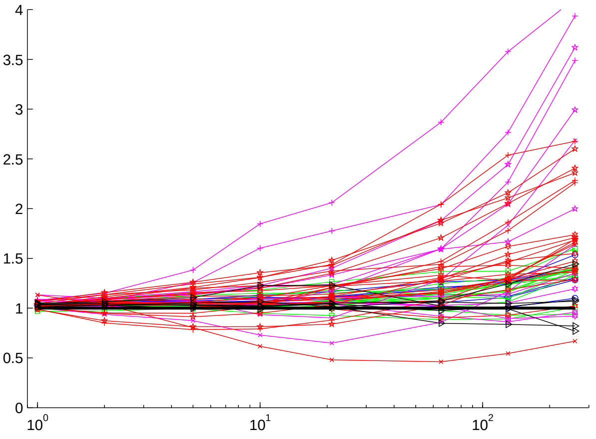


Figure 25: The standard deviations of the residuals $\varepsilon[\Delta T]$ versus ΔT, for the G10 data set, without corrections for the lagged correlation (eq. 59 on the volatility) and without scale factor $\gamma[\Delta T]$. The horizontal black line is a guide for the eyes sets at 1.

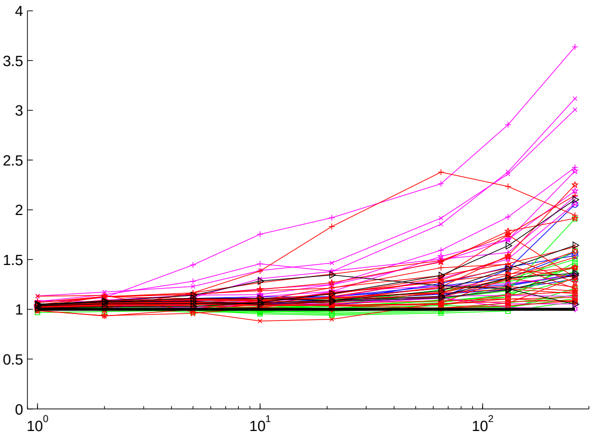


Figure 26: The standard deviations of the residuals $\varepsilon[\Delta T]$ versus ΔT, for the G10 data set, without scale factor $\gamma[\Delta T]$.

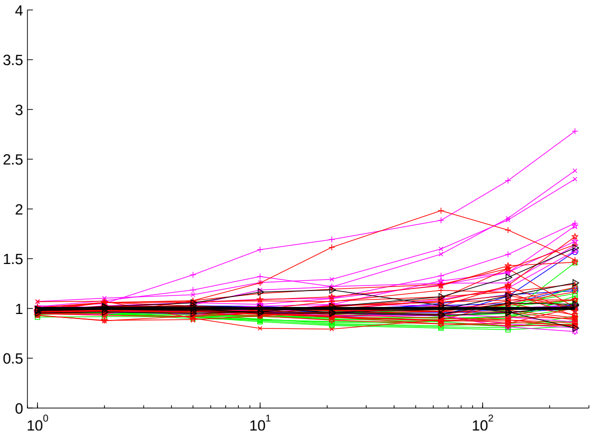


Figure 27: The standard deviations of the residuals $\varepsilon[\Delta T]$ versus ΔT, for the G10 data set. Both corrections are included.

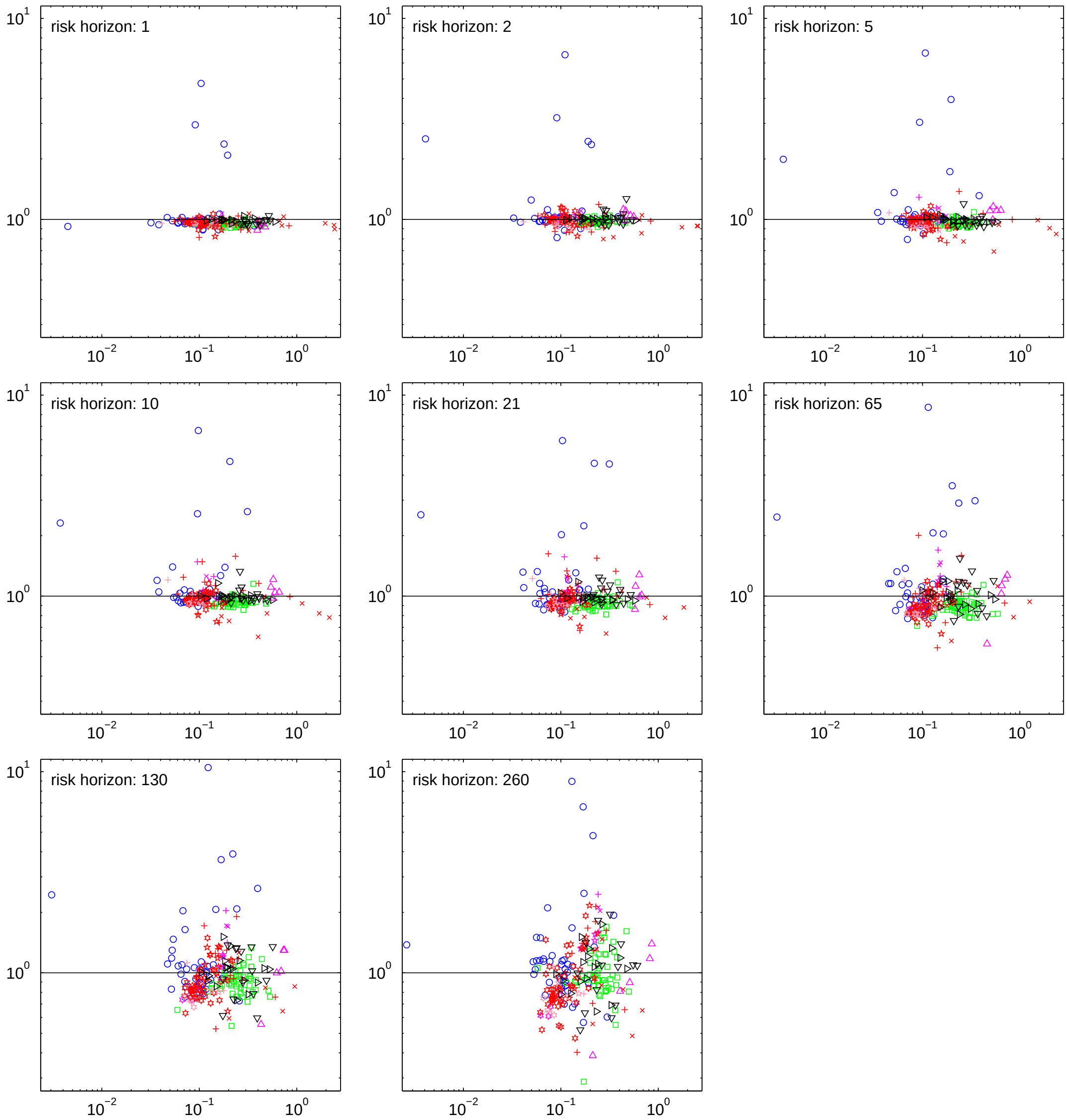


Figure 28: The standard deviation for the residuals $\varepsilon[\Delta T]$ (vertical axis) versus the standard deviation for the returns $r[\Delta T]$ (horizontal axis), in log-log scales. One point corresponds to one time series in the ICM data set. Each panel corresponds to one risk horizon ΔT.

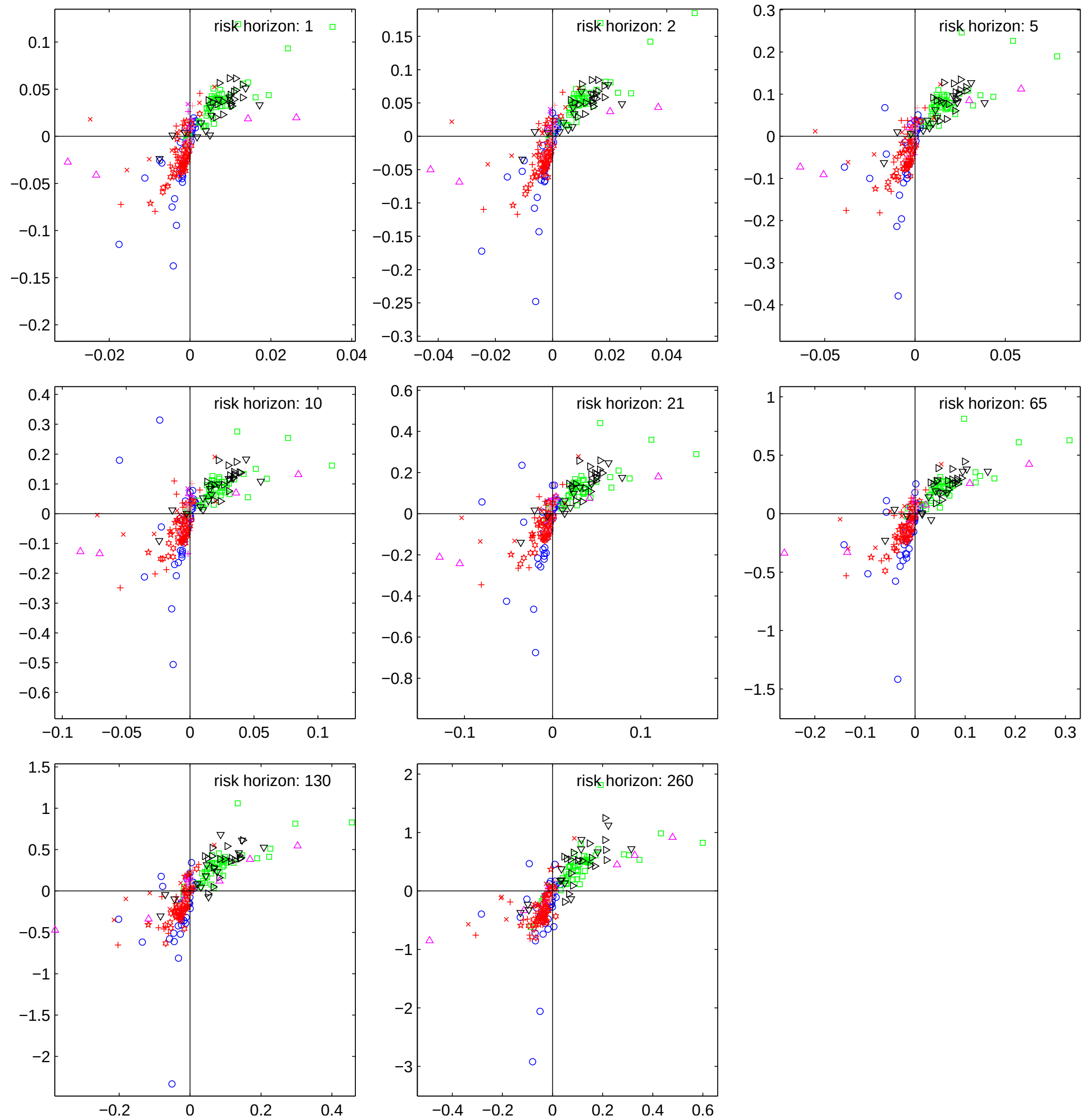


Figure 29: The mean for the residuals $\varepsilon[\Delta T]$ (vertical axis) versus the mean for the returns $r[\Delta T]$ (horizontal axis). One point corresponds to one time series in the ICM data set. Each panel corresponds to one risk horizon ΔT, with the graph limits scaled by $\sqrt{\Delta T}$.

The standard deviations for the returns are reported on the horizontal axis, while the standard deviations of the residuals are given on the vertical axis. The difference of volatility between the broad asset classes is clearly visible, for example stock indexes (green) are more volatile that interest rates (red). The standard deviations for the residuals are clearly clustered around 1, with obviously larger deviations for larger risk horizons ΔT. The clear outliers correspond to the foreign exchanges (blue circle) for Turkey, Brazil, Kazakhstan, Hong Kong Dollard and Philippines. All these currencies were at some point strongly regulated by central banks, and have experienced sudden and disruptive events. Clearly, any random walk model performs badly on such time series.

13 Mean for the residuals

The same computations but for the mean are reported on fig. 29. In this case, we want to check that $\langle \varepsilon \rangle = 0$. Most interest rates went down during the last decade, resulting in negative mean values for the corresponding returns. On the other hand, stocks and stock indexes raise in the analysis period, resulting in mostly positive mean values for the returns. This systematic difference leads to the separation along the horizontal direction. There is no separation in the vertical direction except at the longest risk horizons, showing the effectiveness of the methodology in removing the means for the residuals. Without corrections for the short term correlations and the long term mean returns, the means essentially fall on a straight line, with the same segregation between asset classes for the mean of the returns and the residuals.

14 Probability distribution for the residuals

A key ingredient in a risk methodology is the probability density $p_{\Delta T}(\varepsilon)$ for the residuals, and in particular, the tail behavior corresponding to large events. This is conveniently studied by the empirical cumulative probability density $\mathrm{cdf}(\varepsilon)$. For negative residuals, the tail corresponds to the convergence of the cdf toward zero, whereas for positive residuals, the tail is given by the convergence toward one. In order to have similar figures for both tails, we use the same mapping as in sec. 5.

The resulting figures for the G10 data set are given in fig. 30 and 31 for the negative and positive tails respectively. The salient feature of these graphs is the very good data colapse that the RM2006 methodology gives. This shows that one analytical form for the residual pdf should be enough to characterize all assets. The single red curve that departs from the others corresponds to the Euro swap rate at one day. A close examination of the raw time series shows frequent changes in the 0.5% to 1% range, probably due to competing data source and a less liquid market. These frequent large changes induce a lack of small changes, visible for small ε on the graphs.

At a one day risk horizon, the Gaussian distribution can be easily excluded as it clearly misses the fat tails of the empirical data. On the other hand, a Student distribution with 5 degrees of freedom provides for a good global description of the cdf, including the tails. Changing the number of degrees of freedom ν for the Student distribution shows that $\nu = 3$ gives clearly too much tails whereas $\nu = 8$ gives not enough tails. These rough bounds for the tail exponent are confirmed in the section 15.3 on backtesting.

With longer risk horizons, the amounts of data in the tails diminish, and it becomes increasingly difficult to make a clear distinction between different analytical descriptions. At 3 months and above, it is not possible to differentiate between the Student with $\nu = 5$ and a Gaussian. Up to one month, the Student with a fixed number of degrees of freedom $\nu = 5$ seems to provide for a good description of the residual pdf. One possibility to improve the description of the empirical pdf is to include a weak dependency with respect to ΔT in ν. Notice that we cannot use an analytical hint for $\nu(\Delta T)$, as there is no simple aggregation property for the residuals. Instead, we use a Monte carlo simulation for a ARCH process, sampled daily, and with a simulation length of 1460 years. The process is a market component model [Zumbach and Lynch, 2001, Lynch and Zumbach, 2003], with intra-day, daily, weekly and montly time horizons. This process is probably the best available model today in term of replicating the observed statistical properties of the empirical data. The parameters are estimated on USD/CHF, but parameters for other curency pairs are similar. The process is simulated with a time increment of 3 minutes, and the prices for the residual computations are taken daily. The resulting residual distributions are shown on fig. 32, together with the empirical cdf for the six foreign exchange rates in the G10 data set. At a risk horizon of one day, the tail for the process is well described by a Student distribution with $\nu = 7$, and this agree

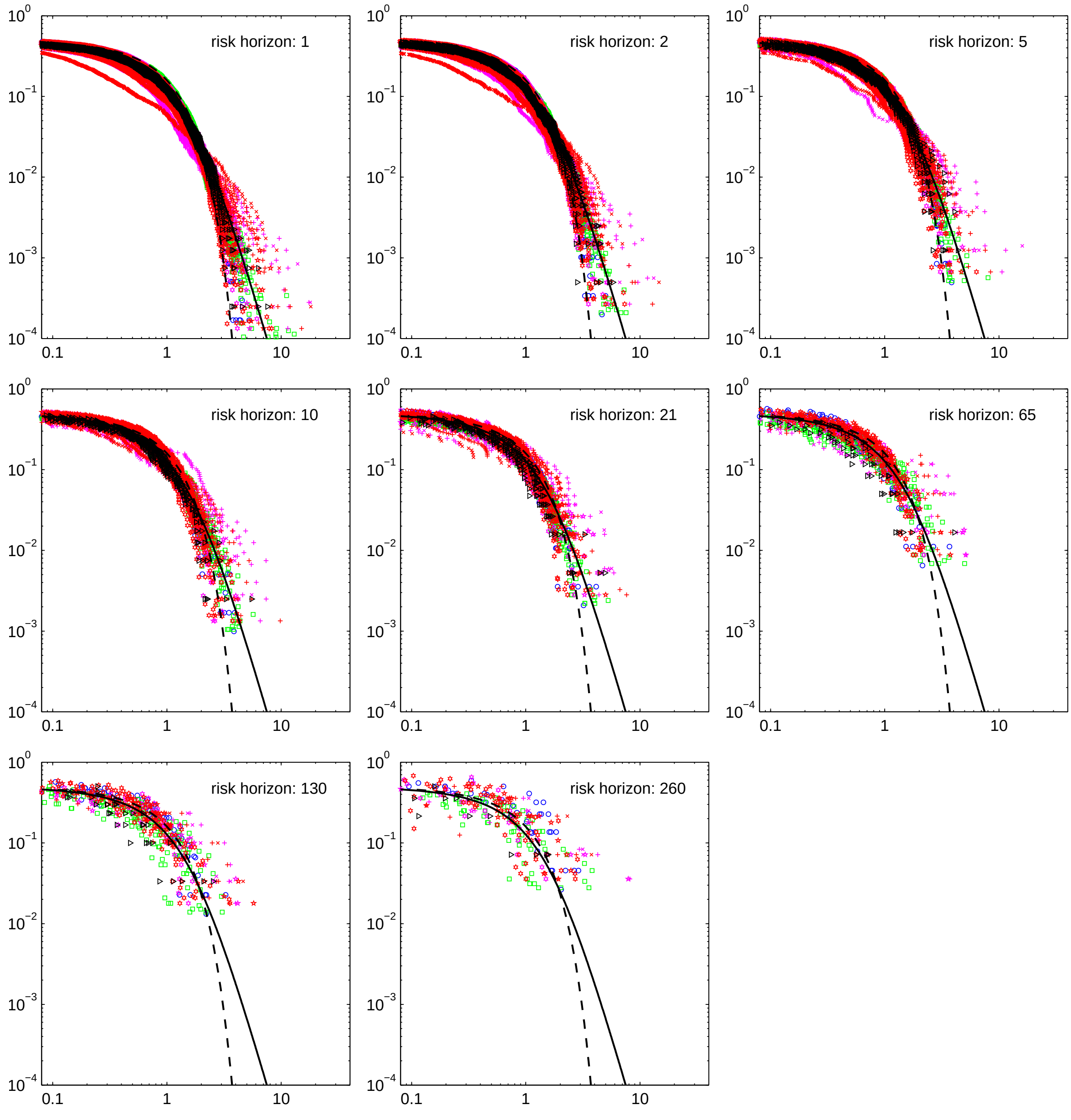


Figure 30: The cumulative probability density cdf(ε) versus −ε for the negative tails. The time series are the G10 data set. The solid line corresponds to a Student distribution with 5 degrees of freedom (rescaled to have a unit variance); the dashed line corresponds to a standard normal distribution.

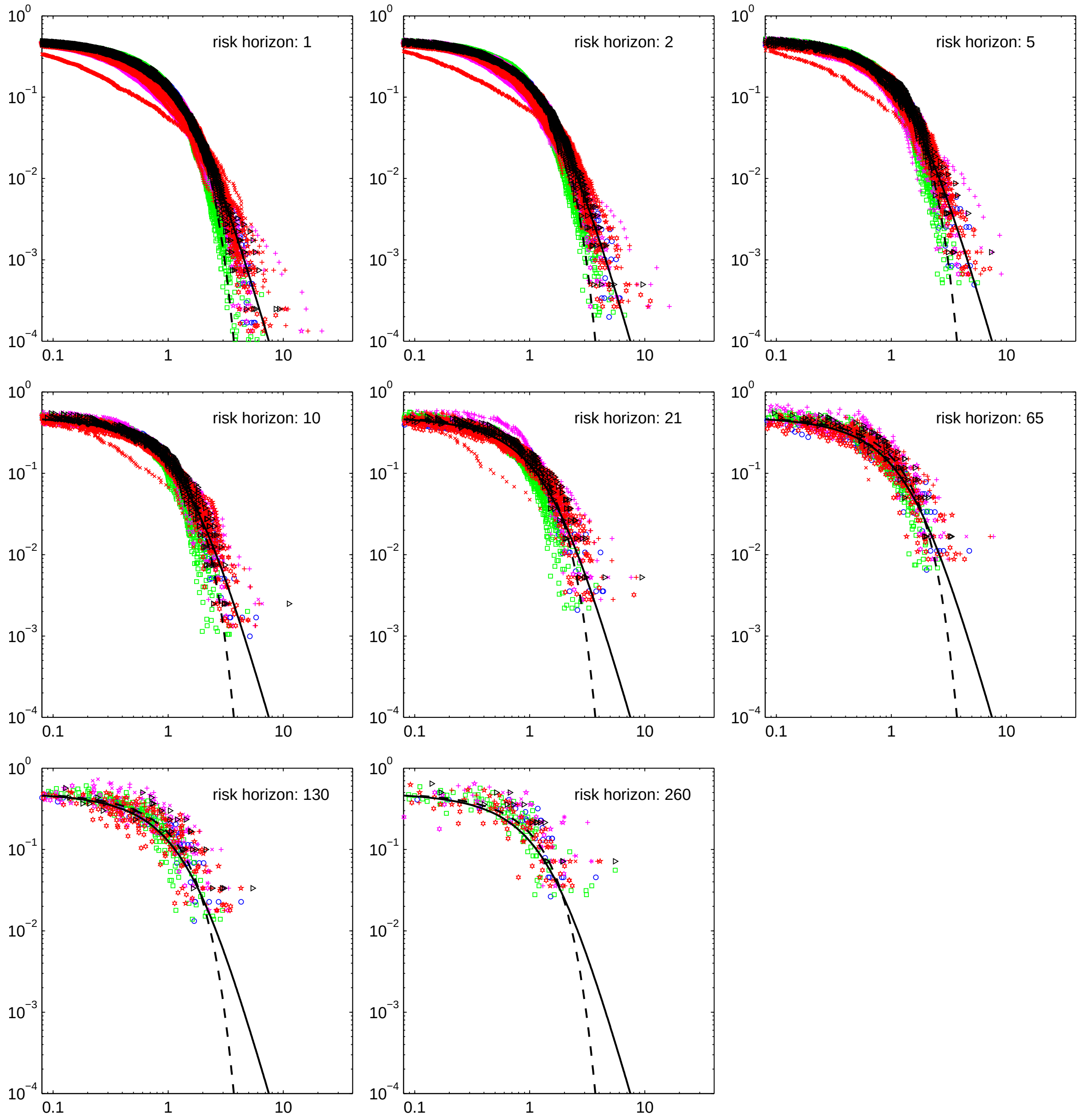


Figure 31: The cumulative probability density $1 - \mathrm{cdf}(\varepsilon)$ versus ε for the positive tails. The time series are the G10 data set. The solid line corresponds to a Student distribution with 5 degrees of freedom (rescaled to have a unit variance); the dashed line corresponds to a standard normal distribution.

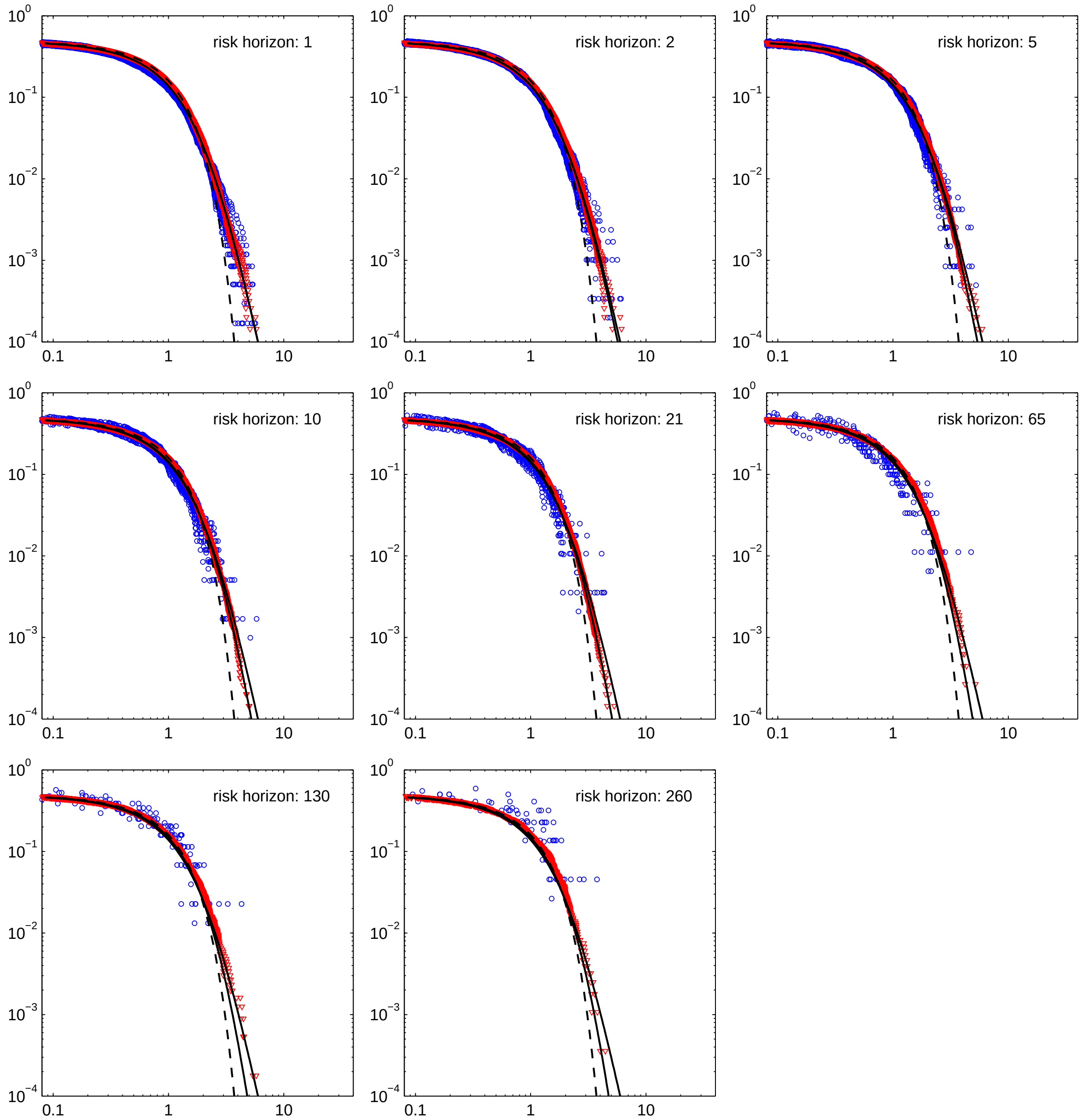


Figure 32: The cumulative probability density versus the residuals for a market component ARCH process (red triangle) and the six foreign exchange rates (blue circle) in the G10 data set. Both tails are plotted, with the mapping used in fig. 30 and 31. The dashed curve corresponds to a Gaussian cdf, the two continuous lines to Student distributions.

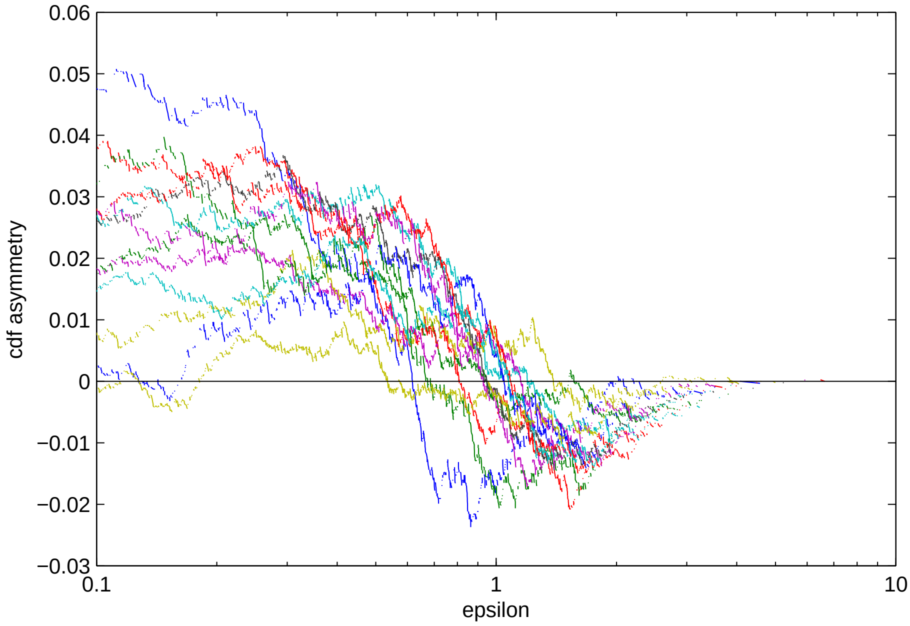


Figure 33: The cdf asymmetry for the stock indexes in the G10 data set.

with the empirical distribution for the FX rates (although their dispersion is fairly large). This tail estimate is slightly thinner than the $\nu = 5$ estimate made on the ICM data set, likely due to the large liquidity and absence of regulations for the major FX rates. At longer risk horizons, the graphs include two Student distributions with $\nu = 7$ and $\nu = 7 + \ln(\Delta T)$ (and a Gaussian plotted with a dashed line). Even with such a long and clean Monte Carlo data set, it is not clear what is the best description of the residual pdf for long risk horizons. With the length of the empirical time series we use, there is no advantage in adding one more parameter in the empirical description of the residual pdf. As there is no clear benefit in using a more complex description with respect to ΔT, we decide to keep a fixed number of degrees of freedom $\nu = 5$, independent of ΔT.

The first direct comparison between figures 30 and 31 shows no obvious asymmetry between the tails. At closer inspection, small but clear differences can be observed for stock indexes. The pdf asymmetry is measured by

$$(1 - \mathrm{cdf}(\varepsilon)) - \mathrm{cdf}(-\varepsilon). \tag{30}$$

This quantity is plotted for the stock indexes in the G10 data set in fig. 33. The systematic deviations from zero show that stock indexes experience frequent small increases (gains) and infrequent large decreases (losses). The difference between large and small is given essentially by the current forecast for the volatility, namely by $\varepsilon \simeq 1$. This same behavior is observed for individual stocks, but no systematic asymmetry is observed for the other asset classes. Although clear, the asymmetry is quantitatively small: for $\varepsilon = \pm 1$, the cdf is of the order of 0.2 and the asymmetry of the order of 0.01 to 0.02.

Because the asymmetries are small and observed only for stocks and stock indexes, we decide to neglect them in the analytical description of the residual distribution. This decision simplifies considerably the parameters evaluation of the residuals pdf. In principle, univariate and multivariate generalized hyperbolic distributions can be constructed (see e.g. [McNeil et al., 2005] and references therein). This family can accommodate asymmetric distributions, however with the drawback that the number of parameters are at least proportional to the number of assets (depending on the simplifying assumptions). More work is needed to find an overal description depending only on the asset classes, with one asymmetry parameter for stocks and stock indexes and symmetric distributions otherwise. We leave this topic for further research, as it involves a multivariate analysis of the residual pdf.

The comparison between $p(r/\,\mathrm{stdDev}(r))$ (fig. 4 and 5) and $p(\varepsilon)$ (fig. 30 and 31) shows no important difference. The tails of the residuals are slightly thinner at one day, but seem more persistent for increasing ΔT. This shows that the volatility forecast (used to derive the residuals from the returns) captures part of the large events. The curve collapse is also slightly better for the residuals, mainly in the core of distribution. Yet, a key distinction between the figures is that the returns are standardized using the full sample standard deviation, whereas the residuals $\varepsilon(t)$ are computed only using information up to t. This key difference makes the residuals fit to use in a day-to-day procedure, whereas the normalization of the return using the standard deviation of a sample up to t would give worse results.

15 Backtesting

The analysis so far concerns the various pieces that enter into the new methodology. The strategy was to isolate as much as possible the various components, and to test and justify them. Yet, the final methodology contains all the parts that interact non linearly, essentially due to the definition 16 of the residual. Moreover, the real test for a risk methodology is how it performs with respect to risk evaluations! The goal of this section is to analyse the performance of the methodology, to understand where the improvements originate from, and to compare with one well established methodology.

The backtesting framework follows [Zumbach, 2006a] and is based on the probtiles corresponding to the residuals. The main analysis tool is the relative exceedance fraction $\delta(z)$ that shows the difference between the actual and expected exceedances fraction, for a given probtile. Then, convenient norms are set on this function, with the property that lower values correspond to better performances, and with an emphasize on the tails that can be chosen. We use the d_0 and d_{32} norms as measures of performance, with d_0 measuring the overall exceedance difference and d_{32} capturing the exceedance differences in the far tails. Similarly, the lagged correlations are computed from the probtiles. The reader should consult [Zumbach, 2006a] for the details of the backtesting framework, as well as for a systematic comparison between various risk methodologies.

In order to give a scale for the performance measures, we include two benchmark methodologies in the figures below. One benchmark is the straight forward RiskMetrics exponential moving average methodology, with a decay factor $\lambda = 0.97$. The corresponding curves are labelled “RM1994_097” on the graphs. Another benchmark is the “long memory + Student” which includes a volatility forecast derived from a process with long memory as in appendix A, and a Student distribution for the residuals. No other contribution is included, leading to a simple extension of the RM1994 methodology. This benchmark is labeled “LMPlusStudent”. The other methodologies correspond to the full RM2006 methodology, but with one of its components removed. This allows us to measure the impact of each contribution in the final risk estimates. The three subsections below present the impact on the RM2006 methodology of the different terms for the return forecasts, for the volatility forecasts, and for the residual pdf. A fourth subsection compares the overall impact of the main ingredients used in the final methodology.

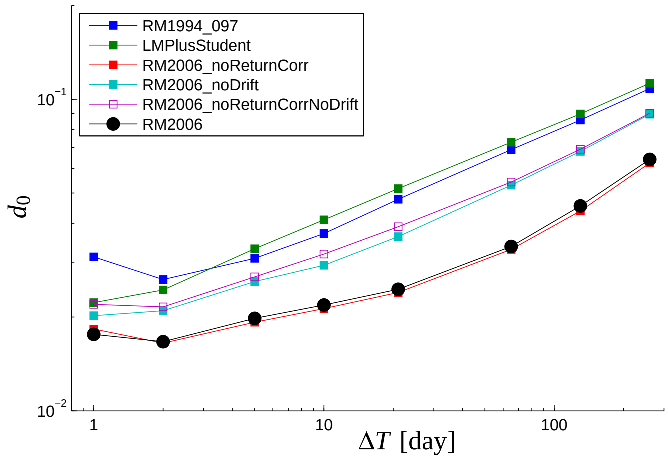


Figure 34: The aggregated error measure d_0 (computed with a geometric mean), between one day and one year, for the ICM data set, and for various return forecasts.

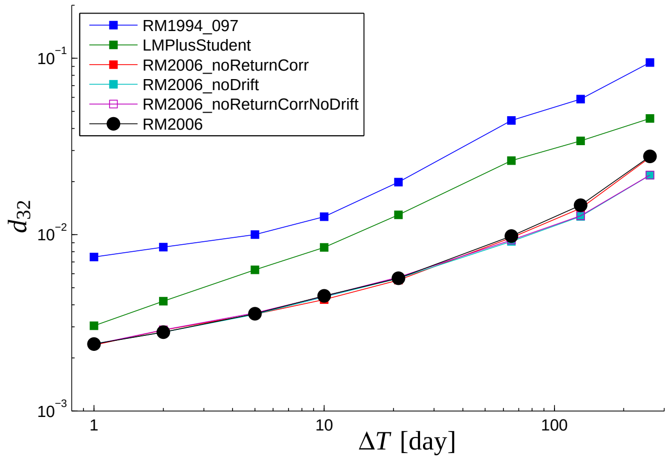


Figure 35: As for fig. 34, but for d_{32}.

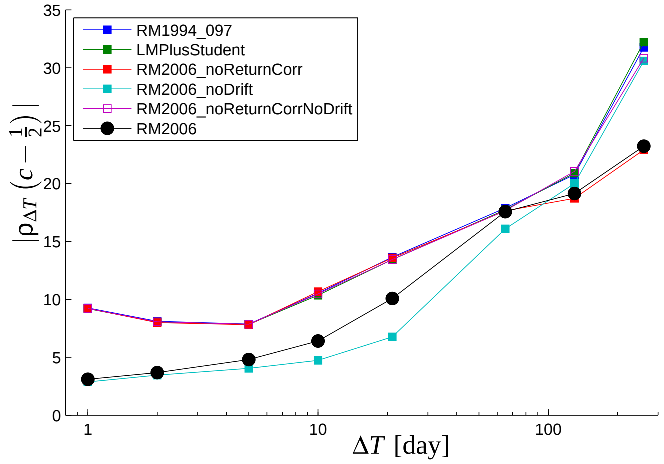


Figure 36: The aggregated lagged correlation of the probtiles (computed with an arithmetic mean), between one day and one year, for the ICM data set, and for various return forecasts.

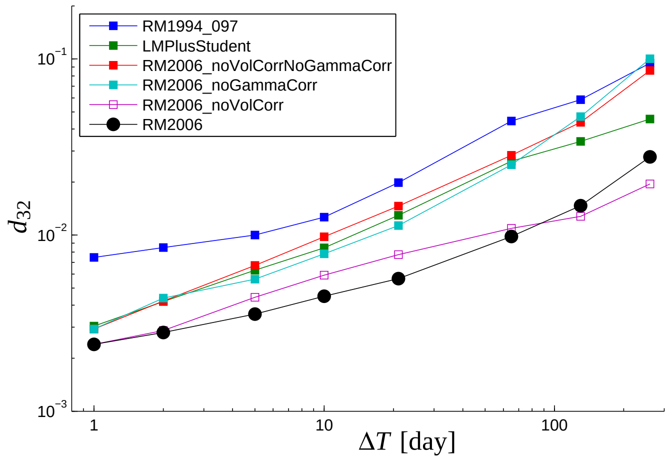


Figure 37: The aggregated error measure d_{32} (computed with a geometric mean), between one day and one year, for the ICM data set, and with various volatility forecasts.

15.1 Backtesting the return forecasts

The influence of the return forecasts can be assessed on fig. 34, 35 and 36. Fig. 34 shows the impact of the return forecasts on the overall exceedance fraction. Clearly, the long term drift is a key component (as removing it decreases substantially the performances). Fig. 35 shows an error measure focused on the far tails for the same methodologies. In this case, the story is quite different, as the return forecasts does not contribute at all to the performances. This can be easily understood, as large events are much larger than the forecasted returns. As other performance measures d_p are selected, the impact of the mean drift decreases gradually as p increases. Therefore, including the drift ensures optimal performances at all risk levels. Another measure of performance is given by the lagged correlation of the probtiles, measuring the tendency of events to cluster. Fig. 36 shows that the key contribution is given by the return forecast originating in the lagged correlations of the returns.

The final picture about the return forecasts is that they improve the methodology, but the different components (lagged correlation or long term drift) matter for different measures of performances. In order to have optimal performances with respect to all criteria, both terms for the return forecasts should be included.

15.2 Backtesting the volatility forecasts

The volatility forecast is a very important part of a risk methodology as it should discount efficiently the heteroskedasticity. The figure 37 analyzes the contributions of the terms that influence the volatility forecasts. For the exceedance differences in the tails, as measured by d_{32}, the key contribution is the scale factor $\gamma[\Delta T]$. It shows that it is very important quantitatively to follow the size condition $E\left[\,\varepsilon^2\,\right] = 1$.

15.3 Backtesting the residuals pdf

The influence of the probability distribution $p(\varepsilon)$ of the residuals can be seen on fig. 38. The RM2006 methodology, but with a Gaussian distribution, is labeled “RM2006_nuInfty” (dashed black line). Clearly, a Student distribution improves the performances, but the sensitivity with

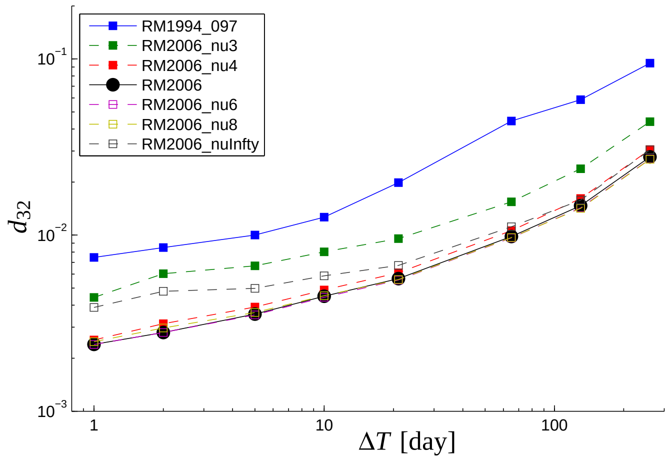


Figure 38: The aggregated error measure d_{32} (computed with a geometric mean), between one day and one year, for the ICM data set, and for various residual distributions.

respect to the number of degrees of freedom ν is very weak in the region $4.5 \leq \nu \leq 8$. The measures for the correlations of the probtiles and the correlation of the absolute probtiles show a very weak dependency of the residual pdf, for all distributions. This shows that it is important to use a distribution with fat tails, but the final methodology is weakly sensitive to the details of the distribution.

15.4 Backtesting the key components of the methodology

As a quick summary, the RM2006 methodology is based on an ARCH process as a model for the data generating process. The key components for the process and risk methodology are long memory for the volatility, correction for the scale of the residual (i.e. $\gamma[\Delta T]$), Student distribution for residuals, and corrections for the lagged correlations between returns. This contrasts with the standard RM1994 methodology, based on an exponential memory, no correction for the residuals scale, Gaussian distribution, and no auto-correlations between returns. Figure 39 shows the contribution of the pieces used to construct the RM2006 methodology. The labels correspond to the following risk methodologies, adding successively the different components: LMPlusGaussian: long memory + Gaussian residues; LMPlusStudent: long memory + Student residues; LMPlusStudentPlusGamma: long memory + Student residues + scale correction; RM2006: long memory + Student residues + scale correction + lagged correlations corrections. We see that, depending on the error measure, each part contributes to the improved performance. In particular, the lagged correlation corrections is an important component.

16 Conclusion

We started this work with two simple ideas. First, a process should be used to compute forecasts in order to be able to reach long risk horizons. Second, we should use a long memory process and a Student distribution for the residuals, according to the recent progresses in financial time series modelization. These two ideas correspond to the computations given in appendix A. This improves the performances with respect to the existing methodologies, but the gains can be disappointingly small according to some error measures, for example d_0 (see

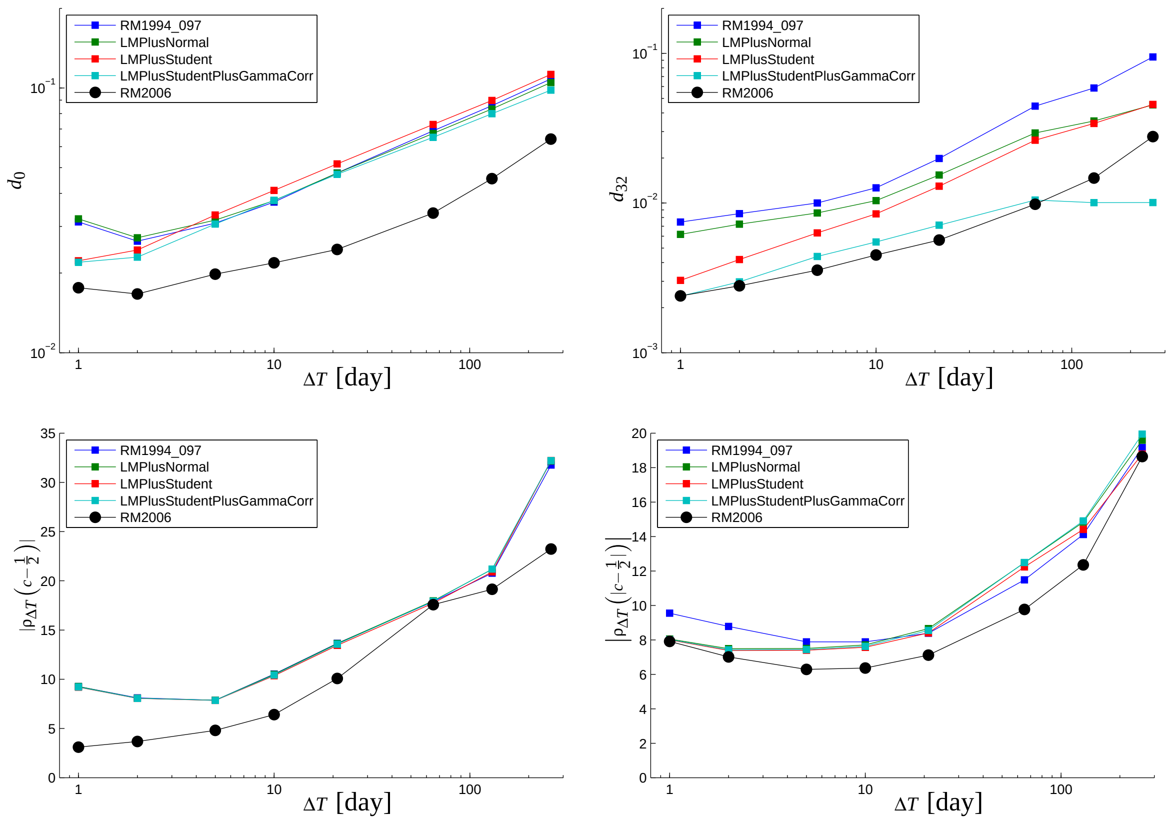


Figure 39: The main aggregated error measures, between one day and one year, for the ICM data set. The methodologies incorporate successively the main ingredients of the RM2006 methodology.

fig.35 and fig.39). Although these basic ideas are good, they are not good enough to lead to a breakthrough.

The key to understand the limitation of this simple extension of the standard methodology is the weak correlations between lagged returns. Following the textbook argument, these correlations should be very small otherwise they are arbitrage away. Indeed, our computations of the lagged correlations for the daily returns is in agreement with this idea. Possibly, rigorous testing of the null hypothesis of zero correlations can be rejected, but the correlations are at the noise level (see the graph 16) and the statistical machinery of hypothesis testing should be brought in. If these correlations are nill, then the best return forecast is also zero. However, there are other better signatures for these correlations: lagged correlations of returns at longer time horizons (e.g. monthly in fig. 17) and the standard deviation of the (uncorrected) residuals (see fig. 25). The inclusion of lagged correlations between the returns in the process equations is straight forward, leading to the analytical computations detailled in appendix B. The inclusion of the resulting terms substantially improves the performance of the risk methodology. Therefore, the new RM2006 methodology can be summarized as based on a *ARCH-like process* with *long memory* + *Student distribution* + *residuals scale correction* + *lagged correlations between returns*. All the ingredients contribute to the performances, albeit possibly at different risk horizons or according to different performance measures. At the end, the new RM2006 is clearly a more complicated methodology than the simple exponential moving average of RM1994. This is the price to pay for the increased accuracies and the possibility to reach long risk horizons.

This paper is devoted to the construction of the RM2006 methodology. The approach consist is separating as much as possible the different parts, and to justify and test them separately. At the very end, the influence of the various parts are quantified using the backtesting methodology. Intentionally, we have not included a backtesting comparison with the existing methodology. The opposite approach is taken in the companion paper [Zumbach, 2006a] where many methodologies are compared, but treated as black boxes. In one sentence, the key result of the systematic backtesting comparison is that the RM2006 methodology improves the risk measures by a factor 1.5 to 3, for all risk measures considered, and for all risk horizons. This is a large gain, which allows to have better risk estimates at short risk horizons and to be able to compute meaningful figures at long risk horizons.

It is interesting to reflect on the "Amendment to the Capital Accord to incorporate market risks" [BIS, 1992] in view of our understanding of financial time series and risk methodologies. The particular point in case is given in section B4, item d (page 40). This point reads:

> The choice of *historical observation period* (sample period) for calculating value-at-risk will be constrained to a minimum length of one year. For banks that use a weighting scheme or other methods for the historical observation period, the "effective" observation period must be at least one year (that is, the weighted average time lag of the individual observations cannot be less than 6 months).

This last sentence can be translated in the condition

$$m_1(n) = \sum_{i\geq 0} \lambda(n,i)\, i \quad \geq \quad 125 \text{ day} \qquad \forall n \tag{31}$$

where m_1 stands for the first moment, and n is the risk horizon ΔT expressed in day. With an exponential weighting scheme with characteristic time τ, as for the RM1994 methodology, the first moment of the weight is $m_1 = \tau$. With the standard accepted value $\mu = 0.94$, this corresponds to 16 business days. Clearly, this value is very far to meet the BIS criterion. For the RM2006 methodology, a few values are $m_1(1)$ = 44 days, $m_1(21)$ = 61 days and $m_1(260)$ = 100 days; none of them meet the BIS criterion. Methodologies that fullfil this BIS criterion are the Equal weight and the historical methodologies. Yet, they put too much weights on the distant past, leading to a sub-optimal discounting of the volatility clusters [Zumbach, 2006a]. Together, this shows that this BIS condition is too stringent for an optimal risk evaluation.

As a direction for further works, a conceptually clean univariate framework as this one should help progressing in the more complex multivariate risk analysis. This contribution uses only a univariate analysis, eventhough many time series are used for the empirical analysis. The extension of the current analytical framework to multivariate processes is natural, as the equations involve only linear and quadratic terms. Yet, the multivariate empirical performances should be investigated in details. Intuitively, the multi-variate correlations should be better estimated with a long memory kernel, as more returns are included in the computations. Going beyond this argument requires substantial work that we leave for further research.

17 Acknowledgments

This work benefited from many discussions with my colleagues at RiskMetrics. Special thanks are due to Chris Finger, Fabien Couderc, Daniel Straumann, Tarik Garidi and Pete Benson.

A Computation of the volatility forecast from the process equations

In [Zumbach, 2004], the volatility forecast for a slightly more general class of processes is given. The process we are using is called in this reference "Long Memory Microscopic Linear ARCH" process, or LM-Mic-Lin-ARCH. For completeness, we give here the derivation of the volatility forecast for this particular long memory process.

Using the process equations for the LM-Mic-Lin-ARCH process 5 and 6, the conditional expectations are given by the recursive equations:

$$
\begin{aligned}
E\left[\, \sigma_k^2(t+j\delta t) \mid \Omega(t)\right] &= \mu_k E\left[\, \sigma_k^2(t+(j-1)\delta t) \mid \Omega(t)\right] \\
&\quad + (1-\mu_k) E\left[\, \sigma_{\text{eff}}^2(t+(j-1)\delta t) \mid \Omega(t)\right] \qquad (32)\\
E\left[\, \sigma_{\text{eff}}^2(t+j\delta t) \mid \Omega(t)\right] &= \sum_k w_k E\left[\, \sigma_k^2(t+j\delta t) \mid \Omega(t)\right] \qquad (33)
\end{aligned}
$$

for $j \geq 1$. Introducing the new variables

$$
\begin{aligned}
\delta_k(j) &= E\left[\, \sigma_k^2(t+j\delta t) \mid \Omega(t)\right] \qquad (34)\\
\gamma(j) &= E\left[\, \sigma_{\text{eff}}^2(t+j\delta t) \mid \Omega(t)\right],
\end{aligned}
$$

the conditional average equations are reduced to

$$
\begin{aligned}
\delta_k(j) &= \mu_k \delta_k(j-1) + (1-\mu_k)\gamma(j-1) \qquad (35)\\
\gamma(j) &= \vec{w}' \cdot \vec{\delta}(j)
\end{aligned}
$$

where $\vec{w}$ is the vector of weights w_k, and similarly for $\vec{\delta}$ and $\vec{\mu}$. We can introduce the diagonal matrix $M_{k,k'} = \delta_{k,k'}\ \mu_k$, with $\delta_{k,k'}$ the Kronecker symbol ($\delta_{k,k'} = 1$ if $k = k'$, zero otherwise). Both equations 35 can be combined

$$
\vec{\delta}(j) = \left\{M + (\vec{1} - \vec{\mu})\vec{w}'\right\} \vec{\delta}(j-1) \qquad (36)
$$

where $\vec{1}$ denotes the constant vector $\vec{1}_k = 1$. This equation can be iterated j times

$$
\vec{\delta}(j) = \left\{M + (\vec{1} - \vec{\mu})\vec{w}'\right\}^j \vec{\delta}(0) \qquad (37)
$$

and $\vec{\delta}(0)$ is in the information set. This expression relates $E\left[\, \sigma_k^2 \mid \Omega(t)\right]$ linearly to the $\sigma_k^2(t)$. For γ, the Eq. 35 can be expressed as

$$
\gamma(j) = \vec{w}'(j) \cdot \vec{\delta}(0) \qquad (38)
$$

with the coefficients $w_k(j)$ given by the recursive equation

$$
\begin{aligned}
\vec{w}'(j) &= \vec{w}'(j-1)\left\{M + (\vec{1} - \vec{\mu})\vec{w}'(0)\right\} \qquad (39)\\
\vec{w}(0) &= \vec{w}.
\end{aligned}
$$

Therefore, the coefficients $\vec{w}(j)$ can be evaluated *a priori*, and the forecast for the effective volatility computed by a simple scalar product. Using the property $\sum_k w_k = 1$ and the above definitions, it is easy to show that

$$
\sum_{k=1}^{k_{\max}} w_k(j) = 1 \qquad \text{for } j \geq 0. \qquad (40)
$$

We can now express the above forecast equations in the form 3 and 12. The iterative equation for σ_k^2 can be unwinded in order to express the volatility with the lagged returns and exponential weights

$$\sigma_k^2(t) = (1-\mu_k) \sum_{i=0}^{\infty} \mu_k^i \, r^2(t - i\delta t) \tag{41}$$

In practice, the sum over the lags needs be cut-off at some $i_{\max}$, and the formula becomes

$$\sigma_k^2(t) = \sum_{i=0}^{i_{\max}} \frac{1-\mu_k}{1-\mu_k^{i_{\max}}} \, \mu_k^i \, r^2(t - i\delta t) \tag{42}$$

This form is introduced in the definition 6 for σ_{eff}^2 to obtain the equation 3 with the weights

$$\lambda(i) = \sum_{k=1}^{k_{\max}} w_k \, \frac{1-\mu_k}{1-\mu_k^{i_{\max}}} \, \mu_k^i \tag{43}$$

As the process is defined with the constraint $\sum_{k=1}^{k_{\max}} w_k = 1$, the coefficients $\lambda(i)$ obey

$$\sum_{i=0}^{i_{\max}} \lambda(i) = 1. \tag{44}$$

The desired volatility forecast is

$$\widetilde{\sigma^2}[\Delta T](t) = E\left[\, r^2[\Delta T](t+\Delta T) \mid \Omega(t)\right] = \sum_{j=0}^{n-1} E\left[\, \sigma_{\text{eff}}^2(t + j\delta t) \mid \Omega(t)\right] \tag{45}$$

where the cross term in the expansion of $r^2[\Delta T]$ vanishes because the autoregressive term $\mu_{\text{eff}}(t)$ is set to zero. The same substitution as in the forecast equations above leads to eq. 12 with

$$\lambda(n,i) = \sum_{k=1}^{k_{\max}} \; \frac{1}{n} \sum_{j=0}^{n-1} \; w_k(j) \; \frac{1-\mu_k}{1-\mu_k^{i_{\max}}} \; \mu_k^i \tag{46}$$

with $w_k(j)$ given by the iterative equation 39. The iterative equation and forecast coefficients can be evaluated numerically very easily. Using the property $\sum_k w_k(j) = 1$, we obtain

$$\sum_{i=0}^{i_{\max}} \lambda(n,i) = 1 \qquad \text{for } \; n \geq 0. \tag{47}$$

When only one component $k_{\max} = 1$ is taken in the LM-ARCH process, the equations are reduced to an I-GARCH process. The coefficients w_k are degenerate to $w = 1$. The equation 39 becomes $w(j+1) = w(j)$, and the solution is $w(j) = 1$. Inserting in eq. 46, we obtain

$$\lambda(n,i) = \frac{1-\mu}{1-\mu^{i_{\max}}} \, \mu^i = \lambda(i). \tag{48}$$

Therefore, for the I-GARCH process, there is no dependency on n in the forecast coefficients λ. This shows that the RiskMetrics EWMA formulation corresponds exactly to use an I-GARCH process. In this sense, our new formulation using processes is a natural extension of the existing methodology.

Notice that the process can be defined in a more general form using only eq. 3 for a given set of coefficient $\lambda(i)$ (i.e. without using a LM-ARCH process with a set of underlying EWMA). This formulation gives a broader class of process where the coefficients $\lambda(i)$ can be freely specified, for example by a functional form. In this class of processes, the resulting forecast equation for the volatility is identical to eq. 12, but the equations leading to $\lambda(n,i)$ are more complicated. Yet, a LM-ARCH process already provides for a convenient and intuitive parameterization of the coefficients $\lambda(i)$, and this process is accurate enough for our purpose.

B Corrections induced by the AR terms

The analysis of the empirical data shows that an auto-regressive term for the returns is important and introduces many corrections in the equations. In this section, we derive these corrections in perturbation, assuming that the leading term is given by the ARCH volatility, and that the auto-regressive terms are smaller. With an auto-regressive (AR) term for the returns, the process equations become:

$$x(t+\delta t) = x(t) + r[\delta t](t+\delta t) \tag{49}$$

$$r[\delta t](t+\delta t) = \mu_{\text{eff}}(t) + \sigma_{\text{eff}}(t)\,\varepsilon(t+\delta t) \tag{50}$$

$$\mu_{\text{eff}}(t) = \sum_{q=0}^{q_{\max}} \mu(q)\, r(t-q\delta t) \tag{51}$$

$$\sigma^2_{\text{eff}}(t) = (w_\infty - \beta)\,\sigma^2 + (1-w_\infty) \sum_{i\geq 0} \lambda(i)\, r^2(t-i\delta t). \tag{52}$$

The auto-regressive coefficient for the lag q return is $\mu(q)$ (be carefull not to confuse $\mu(q)$ with the decay coefficient of the EWMA μ_k). In the volatility equation for σ_{eff}, the first term in the right hand side fixes the mean volatility. The parameter w_∞ sets the ratio between the mean volatility σ and the ARCH term. The lagged correlation for the returns is defined as

$$\begin{aligned} \rho_q &= \frac{E\,[\; r[\delta t](t)\; r[\delta t](t+q\delta t)\;]}{\sigma^2} \\ \sigma^2 &= E\left[\; r^2[\delta t]\;\right] \end{aligned} \tag{53}$$

where σ is the mean volatility of the one day returns. The term β is a function of the AR parameters given by

$$\beta = \sum_{q,q'} \mu(q)\; \rho_{q-q'}\; \mu(q'). \tag{54}$$

and is of order $O(\mu^2)$. The above equation for the volatility σ_{eff} is with an affine term that fixes the mean volatility. To obtain a linear process, we take $w_\infty = \beta \simeq O(\mu^2)$.

Using the usual properties for ε, it is easy to show that for the above process equations, the mean volatility is given by $E\left[\; r^2[\delta t](t+\delta t)\;\right] = \sigma^2$. The mean effective volatility is given by $E\left[\; \sigma^2_{\text{eff}}(t)\;\right] = \sigma^2\;(1-\beta)$; it differs from the mean volatility by a term of order μ^2.

The one day return correlations can be evaluated using eq. 50 and 51:

$$\begin{aligned} E\,[\; r(t)\; r(t+j\delta t)\;] &= \sum_q \mu(q) E\,[\; r(t)\; r(t+(j-1-q)\delta t)\;] + E\,[\; r(t)\; \sigma_{\text{eff}}\; \varepsilon(t+j\delta t)\;] \\ &= \sum_q \mu(q)\; \sigma^2\; \rho_{j-1-q} \end{aligned}$$

We obtain an equation for the correlations

$$\rho_j = \sum_{q=0}^{q_{\max}} \mu(q)\ \rho_{j-1-q} \tag{55}$$

with $\rho_0 = 1$. The solution for ρ can be expanded in μ, and the leading term is

$$\rho_j = \mu(j-1) + O(\mu^2) \qquad j = 1, \cdots. \tag{56}$$

This equation shows that the empirical lagged correlations provide a direct estimate of $\mu(j)$. With financial data, the computation of the lagged correlations using the usual text book estimator proves to be problematic. This is due to the fat tail distribution of the returns, and to the possibly unclean raw data for less traded securities. Clearly, a robust estimator is needed, and we use the ρ_{SSD} correlation estimator described in appendix C. When computed with the one day return over the last 2 year of data, the robust estimator still shows the trace of large events. Particularly annoying is when a large event drop out of the 2 year windows, creating an abrupt change in the correlation. To mitigate this effect, the robust correlation is computed with weighted returns, and the weights implement a simple linear decay in the 2 years windows. When estimating the correlation at the time t, the return $r(t')$ is weighted with

$$w(t') = 1 - \frac{t-t'}{2 \text{ year}} \qquad \text{for} \qquad t - 2 \text{ year} \leq t' \leq t. \tag{57}$$

The weights have the advantage to put more emphasize on the recent past, in line with the volatility estimator.

The process equations are parameterized[4] so that the mean one day volatility $E\left[\, r^2[\delta t] \,\right]$ is set by σ^2. For risk evaluations, the volatility at the horizon ΔT is needed. This volatility is different from the usual scaling $\sqrt{\Delta T/\delta t}\ \sigma$ because of the AR term. The mean n day volatility $\sigma[n\delta t]$ can be estimated using the aggregation of the returns:

$$\begin{aligned}
\sigma^2[n\delta t] &= E\left[\, r^2[n\delta t](t) \,\right] = E\left[\left(\sum_{j=0}^{n-1} r[\delta t](t - j\delta t) \right)^2 \right] \\
&= \sigma^2 \sum_{j,j'} \rho_{j-j'} = \sigma^2 \left(n + 2 \sum_{j=1}^{n-1} \sum_{j'<j} \rho_{j-j'} \right) \\
&= \frac{\Delta T}{\delta t}\ \sigma^2 \left(1 + \frac{2}{n} \sum_{j=1}^{n-1} (n-j) \rho_j \right).
\end{aligned}$$

This leads to the formula

$$\sigma^2[n\delta t] = E\left[\, r^2[n\delta t](t) \,\right] = \frac{\Delta T}{\delta t}\ \sigma^2 \left(1 + 2 \sum_{j=1}^{n-1} (1 - \frac{j}{n}) \rho_j \right). \tag{58}$$

The last contribution in the right hand side gives the correction due to the AR term. This computation shows that the volatility forecasts needed for risk estimations must be corrected by the same factor. In practice, the correction term is evaluated directly by

$$\langle \frac{r^2[n\delta t](t)}{\sum_{j=0}^{n-1} r^2[\delta t](t - j\delta t)} \rangle \tag{59}$$

[4]We are deriving the effect of the autoregressive term in a process set-up. In order to have well defined asymptotic properties, the mean volatility must be fixed by an appropriate constant, in our case σ. Therefore, in this section, we use an affine process [Zumbach, 2004] (affine in σ^2_{eff} and r^2). At the end of the computation, we set $w_\infty = \beta$ so that the parameter σ disappears from the equation, and a linear (in the square) process is obtained. This is legitimate, as our goal is to compute forecasts, for which well defined asymptotic properties are not needed. In this way, we have no mean volatility parameter, like σ in the final equations.

where the empirical average $\langle\cdot\rangle$ is computed over the period $t-2$ years to t. In the risk computations, the volatility forecast $\widetilde{\sigma^2}[\Delta T](t)$, as computed in the previous section, is multiplied by this factor.

In the empirical investigation, the lagged correlations for the lag ΔT of the returns $r[\Delta T]$ is of particular interest. This term can be evaluated for our process

$$\begin{aligned} E\left[\, r[n\delta t](t)\, r[n\delta t](t+n\delta t)\,\right] &= \sum_{j,j'=0}^{n-1} E\left[\, r(t+(n-j)\delta t)\, r(t-j'\delta t)\,\right] \\ &= \sigma^2 \sum_{j,j'=0}^{n-1} \rho_{n-j+j'} \;=\; \sigma^2 \sum_{j=1}^{2n-1} (n-|n-j|)\,\rho_j. \end{aligned}$$

The desired correlation is therefore

$$\rho_n[n\delta t] = \frac{E\left[\, r[n\delta t](t)\, r[n\delta t](t+n\delta t)\,\right]}{E\left[\, r^2[n\delta t](t)\,\right]} = \frac{n\sigma^2}{\sigma^2[n\delta t]} \sum_{j=1}^{2n-1} \left(1-\left|1-\frac{j}{n}\right|\right)\rho_j. \tag{60}$$

This equation shows that even when the one day correlations ρ_j are small, their effects can accumulate and lead to larger correlations at longer time horizons. In particular, the empirical one day lagged correlations can be below the statistical significance threshold, yet they lead to significant correlations at longer time horizons.

Finally, the auto-regressive term induces a non zero forecast for the return. Let us define

$$a(j) = E\left[\, r(t+j\delta t) \mid \Omega(t)\right] \tag{61}$$

In order to simplify the computations, let us define $\mu(q)=0$ for $q>q_{\mathrm{max}}$. Using the process equation 50, we obtain

$$a(j+1) \;=\; \sum_{j'=0}^{j-1} \mu(j')\, E\left[\, r(t+(j-j')\delta t) \mid \Omega(t)\right] + \sum_{j'=j}^{\infty} \mu(j')\, r(t+(j-j')\delta t)$$

from which a recursion equation for a is obtained:

$$a(j+1) = \sum_{j'=0}^{\infty} \mu(j+j')\, r(t-j'\delta t) + \sum_{j'=0}^{j-1} \mu(j')\, a(j-j') \tag{62}$$

With $\mu \ll 1$, we have $a=O(\mu)$ and the second sum can be neglected. The desired forecast μ_{eff} is given by

$$\begin{aligned} \mu_{\mathrm{eff}} &= E\left[\, r[\Delta T](t+\Delta T) \mid \Omega(t)\right] \;=\; \sum_{j=1}^{n} a(j) \\ &= \sum_{j=0}^{\infty} \sum_{j'=0}^{n-1} \mu(j+j')\, r(t-j\delta t) \end{aligned}$$

and can be set in the same form as the volatility forecast

$$\begin{aligned} \mu_{\mathrm{eff}} &= \sum_{j=0}^{q_{\mathrm{max}}} \mu(n,j)\, r(t-j\delta t) \\ \mu(n,j) &= \sum_{j'=0}^{n-1} \mu(j+j') \end{aligned} \tag{63}$$

As explained above, the coefficients $\mu(j)$ are evaluated using the robust estimator ρ_{SSD} for the lagged correlations (computed with weighted one day returns). Then, the return forecast is computed using the formula 63. The number of lags to consider has been investigated empirically. An important factor in this choice is related to short term interest rates. Typically, the central banks set the overnight interest rates and modify the rate on a monthly basis. This human intervention is particularly obvious for the dollar rate and creates strong lagged correlations up to a month. For this reason, we choose $q_{\max} = 23$ days. An empirical investigation for this parameter shows that this is a good choice.

Long term drifts are also important, particularly for stocks and stock indexes. A measure for the long term drift, scaled at $\delta t = 1$ day, is given by

$$\mu_{\text{drift}} = \frac{\delta t}{2\text{ years}} \; (x(t) - x(t - 2\text{ year}) \tag{64}$$

where this estimator uses the last 2 years of data. The drift appropriates for a forecast at risk horizon ΔT should be scaled by $\Delta T/\delta t$. In back testing, this estimator appears to be fairly good. This estimator can be rewritten as

$$\mu_{\text{drift}} = \frac{1}{n} \sum_{t-2\text{ year} \le t' \le t} r(t') \tag{65}$$

where n is the number of days in the sum. The final forecast for the expected return is the sum of the short term auto-regressive part and of the long term drift. As we are using both estimations, we want to avoid the double counting of the drift. Therefore, when a drift is also used, the auto-regressive part is modified by

$$\mu(n, j) = \sum_{j'=0}^{n-1} \mu(j + j') - \frac{\delta t}{2\text{ year}}. \tag{66}$$

C Computation of the correlations

By default, the correlation is evaluated by the usual Pearson product of standardized variables. Yet, for random variables with fat tails distributions, this estimator has the disadvantage of being very sensitive to extreme events. To mitigate this effect, we have tried two other estimators of dependency. After evaluation of the three correlation estimators on financial time series, we decided to used the ρ_{SSD} estimator as it provides for a good compromise between robustness and computational efficiency.

A robust correlation measure is given by the Kendall's Tau estimator τ (as described for example in the Numerical Recipes [Press et al., 2002]). In [Lindskog, 2000], the author proves that for elliptical distribution, the relation

$$\tau = \frac{2}{\pi} \sin^{-1} \rho \tag{67}$$

holds between the linear correlation and the Kendall tau. Therefore, we use the Kendall Tau correlation estimator

$$\rho_\tau = \sin\left(\frac{\pi\,\tau}{2}\right). \tag{68}$$

This estimator is robust, but the evaluation time grows as $\sim n^2$. For large scale computations, this is a clear drawback even with today computers.

[Gnanadeskian and Kettenring, 1972] have introduced another robust estimator for dependency. In a finance context, it has been used for example by [Lindskog, 2000] to study correlations in elliptical distributions with fat tail marginals. The estimator is based on standardized sums and differences, and is called ρ_{SSD}. For a time series x, the standardized time series is

$$\tilde{x} = \frac{x - \bar{x}}{\sigma_x} \tag{69}$$

with $\bar{x}$ the mean of x and $\sigma_x^2 = \mathrm{Var}(x)$ the variance of x. Then, the usual (Pearson) correlation can be written as

$$\rho(x,y) = \tilde{x} \cdot \tilde{y} \ = \ \frac{\mathrm{Var}(\tilde{x}+\tilde{y}) - \mathrm{Var}(\tilde{x}-\tilde{y})}{\mathrm{Var}(\tilde{x}+\tilde{y}) + \mathrm{Var}(\tilde{x}-\tilde{y})} \tag{70}$$

A robust estimator for the correlation is obtained by replacing the usual variance estimator by a robust one. In this work, we have used the mean absolute deviation (MAD) as a more robust estimator than the variance

$$\mathrm{MAD}(x) = \frac{1}{n} \sum_{i=1}^{n} |x_i - \mathrm{median}(x)| \tag{71}$$

We define the τ_{SSD} dependency measure by

$$\tau_{SSD}(x,y) = \frac{\mathrm{MAD}(\tilde{x}+\tilde{y}) - \mathrm{MAD}(\tilde{x}-\tilde{y})}{\mathrm{MAD}(\tilde{x}+\tilde{y}) + \mathrm{MAD}(\tilde{x}-\tilde{y})}. \tag{72}$$

Empirically, this estimator is close to the Kendall Tau, and therefore a similar relationship with the usual linear correlation seems plausible. We define the robust measure of correlation based on the standardized sum and difference as

$$\rho_{SSD} = \sin\left(\frac{\pi\, \tau_{SSD}}{2}\right). \tag{73}$$

This estimator is robust (but less robust than the Kendall Tau), and the computational time grows as $\sim n$ (because the MAD evaluation requires computing the median, and this can be done in O(n) time as a complete ordering is not needed [Press et al., 2002]). As 2 years of data is already large $n \simeq 520$, this estimator proves to be substantially faster than the Kendall Tau.

D FAQ

What is this *memory decay* for the volatility?
The volatility of the financial time series changes with time, namely there are periods of high volatility and periods of low volatility. This is the *heteroskedasticity*, meaning that the variance is not constant but is clustered. Thus, the volatility shows some persistence, or "memory". To quantify the persistence, the lagged correlation of the volatility is computed. The decrease of the correlation measures how and how fast the clusters of volatility can appear and change. This lagged correlation is a characterization of the "memory" of the volatility.

This memory is crucial for risk estimation. The important point to realize is that a risk evaluation is a forecast, namely an estimate of the possible financial losses between now and some future date. Beside, the magnitude of the risk is directly proportional to the volatility. Because of the memory of the volatility, forecasting the volatility is possible, and hence a risk estimation.

Why a single decay factor can be used in the RM1994 methodology The fact that an EWMA with a "universal" decay factor gives good forecasts for every security can be understood with the following argument. To a decay factor of $\mu = 0.94$ corresponds an exponential with a characteristic time $\tau = -1/\ln(\mu) \simeq 16.2$ business day. The memory for the volatility –as measured for example by the lagged correlation of the absolute return– decays slowly (as a logarithm). Therefore, most of the information comes from the most recent past, and the forecast should put the largest weights on the few last returns. On the other hand, using only a few returns leads to a fairly noisy estimator, and a better statistic requires to use as many returns as possible. Therefore, there is a trade off between where most of the information lies and the variance of the forecast estimator. The optimal trade off is obtained for an intermediate horizon, in the range of 10 to 40 days. This argument is generic, as the only financial time series property used is the decay of the lagged correlations. This decay is fairly similar for all time series and therefore, the same value can be used for all securities.

What is the optimal decay factor for the RM1994 methodology?
The optimal decay factor depends on the risk horizon: the longer the risk horizon, the closer the decay factor should be to 1. For the volatility forecasts, the optimal values for the characteristic time τ can be read from figure 13 and converted to the decay factor using $\mu = \exp(-1/\tau)$ with τ expressed in days. Backtesting the RM1994 methodology as done in [Zumbach, 2006a] indicates slightly larger optimal values.

For risk evaluation at long time horizons, is it better to use returns with longer time intervals, say for example weekly or monthly returns?
The short answer is that, with the new methodology, it is always better to use daily returns. For the other methodologies, the long answer involves two lines of arguments related to the minimal sufficient statistics and to the correlation between returns.

The "minimal sufficient statistics" answers the following question. Let us estimate a given statistical quantity, say for example the long term mean of the return. What is the minimal amount of information that should be given to be able to compute, without loss of accuracy, the desired statistical quantity? This minimal information is called "minimal sufficient statistics". For the example of the long term mean, only the start and end prices are sufficient. This is a considerable reduction of the data set as only three values are needed (the two prices and the time interval). For the volatility, the answer is that all the (absolute values of the) returns should be given. Any reduction of the data set is loosing information. In other words, the returns at the highest frequency should be used. This implies that the volatility (and its forecasts) should be computed with daily returns (even more efficient would be to use intra-day prices).

Yet, the whole situation is not so simple because of the lagged correlations between returns. These correlations create serial dependencies, which implies that $E\left[\, r^2[\delta t] \,\right] \neq E\left[\, r^2[\Delta T] \,\right]$. This inequality means that the variance of the returns is dependent on the risk horizon, an effect that can be observed on fig. 25. For a given risk horizon ΔT, the relevant quantity is a forecast for $E\left[\, r^2[\Delta T] \,\right]$. Therefore, the aggregated return $r[\Delta T]$ should be used so that the serial correlations are included in the computations.

As both arguments go in the opposite directions, the optimal value for the return time horizon should be somewhere between δt = 1 day and the risk horizon ΔT. As the lagged correlations are small, the optimal is likely close to one day. Beyond this simple argument, an answer to the above question would require a specific quantitative study.

In the new RM2006 methodology, the lagged correlations between the returns are computed and discounted explicitely. Therefore, only the argument about minimal sufficient statistics applies, and it is always better to use one day returns in the computations. Using any longer time interval for the returns leads to a loss of accuracy.

Why not use a GARCH(1,1) process to improve the risk methodology?
The GARCH(1,1) process is parameterized by 3 values, typically denoted by $(\alpha_0, \alpha_1, \beta_1)$ or by (ω, α, β). A combination of these parameters fixes the mean volatility, namely

$$E\left[\, r^2 \,\right] = E\left[\, \sigma^2 \,\right] = \frac{\omega}{1 - \alpha - \beta}. \qquad (74)$$

Because the mean volatility $E\left[\, r^2 \,\right]$ is strongly asset dependent, the parameters cannot be chosen with the same values for all time series. For a universe with n assets, a multivariate GARCH(1,1) should have at least n parameters, and possibly $O(n^2)$ parameters. This would be very impractical and very fragile.

Another reason is related to the decay of the correlation for a GARCH(1,1) process. The decay of the correlation for the volatility can be computed analytically for this process, and is found to be exponential. On the other hand, the empirical time series have a much slower logarithmic decay. Therefore, the volatility forecasts derived from a GARCH(1,1) process are not optimal. Both reasons lead us to use a long memory process, as explained in sec. 2.1

Why not use a Student distribution with the RM1994 methodology?
You can, but the improvement is not that large. To improve significantly the RM1994 methodology, several modifications must be done. For example, the fig. 39 shows a comparison between RM1994, Long memory + Student, and RM2006.

Beyond the underlying risk methodology, why is the risk evaluation of actual securities so complicated?
This paper discusses risk methodologies, namely given one "simple" asset, it shows how to compute a risk estimate. A "simple" base time series is called a *risk factor*. They corresponds essentially to linear assets, like equities or foreign exchange currencies. Yet, today's financial world uses derivatives, possibly with very complex structures. These derivatives should be priced as a function of the underlying risk factors, and these contracts can be fairly sophisticated. Another example of "derivative" are bonds: the underlying simple risk factor is the interest rate curve as function of the maturity. The various bonds, with the coupons occuring at various time point, should be priced from the underlying interest rates and the cash flows stream. More complex derivatives can be for example an option on a swap for interest rates (i.e. a swaption). Moreover, some trading strategies involve long and short positions in

derivatives and the underlying assets. These kinds of strategies can expose the discrepancies between the option pricing used in the market and in the risk evaluation. Therefore, the option pricing used in risk evaluation has to follow closely the best practice used in the market. This is a first level of complexity.

A second level of complexity is related to the actual details of the financial contracts. For example, a simple bond pays coupons according to a given schedule. In order to price this bond from the interest rates curve, the actual coupon schedule must be known. Therefore, it is not enough to have long time series for the prices of the risk factors, the *terms and conditions* of the actual contracts should also be available.

A third level of complexity is related to the detailed analysis of the risks inside a portfolio: beyond the global risk figure, one also would like to understand where the risks originate from. For example, an analysis according to the currencies or industrial sectors can give a diagnostic that will help reducing the risk exposure. Insights about more global changes can be obtained by scenario analysis, or “what if” analysis (what if oil prices rise? What if the USD interest rates rise? etc...). Possibly, hedges can be added in a portfolio following such an analysis.

At the end, the techniques presented in this paper are the foundation of risk evaluations, yet they represent only a tiny fraction of the software in today applications developed for risk management. A large part of such softwares correspond to the first and third points above, whereas the second point requires painstakingly maintaining large data bases feeded by many data sources. Moreover, today users expect convenient point-and-click tools to analyse and understand the risks of their positions. Eventhough the new RM2006 methodology can appear as being quite complex, it is just a small part in a much larger machinery.

E Notation

δt: The elementary time step for the process, in our case $\delta t = 1$ day.

ΔT: The risk horizon (i.e. a time interval).

n: The number of days for the risk horizon $n = \Delta T / \delta t$.

j: Index for the days in the future $1 \leq j \leq n$.

p: The price time series.

r: The return time series, with $r[\Delta t](t) = p(t) - p(t - \Delta t)$.
The returns at a one day horizon are abbreviated by $r[\delta t] = r$.

ε: The residual time series $\varepsilon[\Delta T]$. See eq. 15 and 16.

$p_{\Delta T}(\varepsilon)$: The probability distribution for the residuals. In principle, the shape of the distribution can depend on ΔT.

$\gamma[\Delta T]$: The scale factor such that $\langle \varepsilon^2[\Delta T] \rangle = 1$. See eq. 16 and 17.

σ_{eff}: The "effective" volatility, as given by a particular process.

k: The component index for the multiscales long memory ARCH model $1 \leq k \leq k_{\text{max}}$.

σ_k: The volatility measured by an EWMA (exponential moving average) at the time horizon τ_k.

μ_k: The decay coefficient for the k-th EWMA with characteristic time τ_k in a multi-time scales ARCH process $0 < \mu_k < 1$ and $\mu_k = \exp(-\delta t / \tau_k)$.

w_k: The weight for the k-th component in a multi-time scales ARCH process. The coefficients must obey $1 \leq w_k \leq 1$ and $\sum_k w_k = 1$.

τ_0: The decay coefficient for the weights w_k with a logarithmic decay form.

$\lambda(i)$: In the computation of the effective volatility, the weight for the lag i return square. This corresponds to the weights for a 1-step forecast $\lambda(1, i)$.

$\lambda(n, i)$: In the computation of the n steps forecasted volatility, the weight for the lag i return square.

i: The index for the days in the past $0 \leq i \leq i_{\text{max}}$ for the volatility computation.

μ_{eff}: The autoregressive term $\mu_{\text{eff}}(t) = \sum_{q=0}^{q_{\text{max}}} \mu(q)\, r(t - q\delta t)$.

$\mu(q)$: The coefficients for the expansion of the autoregressive term (beware not to confuse $\mu(q)$ with the decay coefficients μ_k of the EWMA).

q: The index for the days in the past for the auto-regressive terms $0 \leq q \leq q_{\text{max}}$.

ρ_j: The lagged correlation for the one day return, for the lag $j\delta t$.

$\rho_j[\Delta t]$: The lagged correlation for the Δt days return.

$y[\Delta t]$: The yield (or interest rate) for a time-to-maturity Δt, typically related to a bond with maturity at time $t + \Delta t$.

References

[BIS, 1992] BIS (1992). Amendment to the capital accord to incorporate market risks. Technical report, Bank for International Settlement, Basel Committee on Banking Supervision. Updated in Novembre 2005.

[Bormetti et al., 2006] Bormetti, G., Cisana, E., Montagna, G., and Nicrosini, O. (2006). Risk measures with non-gaussian fluctuations. Technical report. available at: arXiv:physics/0607247.

[Diebold et al., 1998] Diebold, F., Gunther, T., and Tay, A. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39:863–883.

[Drost and Nijman, 1993] Drost, F. and Nijman, T. (1993). Temporal aggregation of garch processes. *Econometrica*, 61:909–927.

[Embrecht et al., 2005] Embrecht, P., Kaufmann, R., and Patie, P. (2005). Strategic long-term financial risks: single risk factors. *Computational Optimization and Applications*, 32:61–90.

[Gnanadeskian and Kettenring, 1972] Gnanadeskian, R. and Kettenring, J. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrika*, 28:81–124.

[Lindskog, 2000] Lindskog, F. (2000). Linear correlation estimation. Technical report, RiskLab, ETH, Zürich, Switzerland.

[Lynch and Zumbach, 2003] Lynch, P. and Zumbach, G. (2003). Market heterogeneities and the causal structure of volatility. *Quantitative Finance*, 3:320–331.

[McNeil and Frey, 2000] McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7:271–300.

[McNeil et al., 2005] McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quatitative Risk Management*. Princeton University Press.

[Mina and Xiao, 2001] Mina, J. and Xiao, J. (2001). Return to riskmetrics: The evolution of a standard. Technical report, Risk Metrics Group. available at: http://www.riskmetrics.com/pdf/rrmfinal.pdf.

[Nelson, 1990] Nelson, D. (1990). Stationarity and persistence in the garch(1,1) model. *Econometric Theory*, 6:318–34.

[Pafka and Kondor, 2001] Pafka, S. and Kondor, I. (2001). Evaluating the riskmetrics methodology in measuring volatility and value-at-risk in financial markets. *Physica A*, 299:305–310.

[Press et al., 2002] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (2002). *Numerical Recipes in C++*. Cambridge University Press.

[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). Relevance of volatility forecasting in financial risk management. Risk Management. Academic Press.

[Zumbach and Lynch, 2001] Zumbach, G. and Lynch, P. (2001). Heterogeneous volatility cascade in financial markets. *Physica A*, 298(3-4):521–529.