Volatility forecasts and the at-the-money implied volatility:
a multi-components ARCH approach
and its relation with market models

Gilles Zumbach

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

gilles.zumbach@riskmetrics.com

Abstract

For a given time horizon ΔT, this article explores the relationship between the realized volatility (the volatility that will occur between t and $t + \Delta T$), the implied volatility (corresponding to at-the-money option with expiry at $t + \Delta T$), and several forecasts for the volatility build from multi-scales linear ARCH processes. The forecasts are derived from the process equations, and the parameters set *a priori*. An empirical analysis across multiple time horizons ΔT shows that a forecast provided by an I-GARCH(1) process (1 time scale) does not capture correctly the dynamic of the realized volatility. An I-GARCH(2) process (2 time scales, similar to GARCH(1,1)) is better, while a long memory LM-ARCH process (multiple time scales) replicates correctly the dynamic of the realized volatility and delivers consistently good forecast for the implied volatility. The relationship between market models for the forward variance and the volatility forecasts provided by ARCH processes is investigated. The structure of the forecast equations is identical, but with different coefficients. Yet the process equations for the variance are very different (postulated for a market model, induced by the process equations for an ARCH model), and not of any usual diffusive type when derived from ARCH.

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

`www.riskmetrics.com`

1 Introduction

The intuition behind volatility is to measure price fluctuations, or equivalently the typical magnitude for the price changes. Yet, beyond the first intuition, volatility is a fairly complex concept, for various reasons. First, turning this intuition into formulas and numbers is partly arbitrary, and many meaningful and useful definitions of volatilities can be given. Second, the volatility is not directly "observed" or traded, but rather computed from time series (although this situation is changing indirectly through the ever increasing and sophisticated option market, the volatility indexes and the options on volatility). For trading strategies, options and risk evaluations, the valuable quantity is the realized volatility, namely the volatility that will occur between the current time t and some time in the future $t + \Delta T$. As this quantity is not available at time t, a forecast needs to be constructed. Clearly, a better forecast of the realized volatility allows to better price options, to make profit on volatility based trades, and to manage better risks in a portfolio.

At a time t, a forecast for the realized volatility can be constructed from the (underlying) price time series. In this paper, multiscales ARCH processes are used. On the other hand, a liquid option market allows to compute the implied volatility, corresponding to the "market" forecast for the realized volatility. On the theoretical side, an "instantaneous", or effective, volatility σ_{eff} is needed to define processes, and the forward variance. Therefore, at a given time t, we have mainly one theoretical instantaneous volatility and three notions of "observable" volatility (forecasted, implied and realized). This paper studies the empirical relationship between these three time series, as a function of the forecast horizon ΔT. There exist already an abundant literature on this topic, and [Poon, 2005] published a book summarizing nicely the available publications ($\sim$100 articles on volatility forecast alone!).

The main line of this work is to model the underlying time series by multi-components ARCH processes, and to *derive* a volatility forecast. This forecast, based only on the underlying, should be close to the implied volatility for the at-the-money (ATM) option. In particular when option data are poor, lacking or not available, such approach allows to obtain a good approximation for the ATM implied volatility. For trading and risk management, the correct pricing of options is clearly an issue, and to have a fall-back solution for the implied volatility surface using a minimal modeling of the underlying is a clear advantage. This article does not address the issue of the full surface, but only the implied volatility for the ATM options, called the *backbone*.

A vast literature on implied volatility and its dynamic already exists. In this article, we will review some recent developments on market models for the forward variance. These models focus on the volatility as a process, and many process equations can be set that are compatible with a martingale condition for the volatility. On the other side, the volatility forecast as induced by a multi-components ARCH process leads also to process equations

2

for the volatility only. These two approaches leading to process for the volatility are contrasted, showing the formal similarity in the structure of the forecasts, but the very sharp difference in the processes for the volatility. If the price time series behave according to some ARCH process, then the implication for volatility modeling is far reaching as the usual structure based on Wiener process cannot be used.

This paper is organized as follow. The required definitions for the volatilities and forward variance are given in the next section. The various multi-components ARCH processes are introduced in sec. 3, and the induced volatility forecasts and processes given in sec 4 and 5. The market models and the associated volatility dynamics are presented in sec. 6. The relationship between market models, options and the ARCH forecasts are discussed in section 7. Section 8 presents an empirical investigation of the relationship between the forecasted, implied and realized volatilities, before the conclusion.

2   Definitions and setup of the problem

2.1   General

We assume to be at time t, with the corresponding information set $\Omega(t)$. The time increment for the processes and the granularity of the data is denoted by δt, and is 1 day in the present work. We assume that there exists an instantaneous volatilities denoted by $\sigma_{\text{eff}}(t)$, which corresponds to the annualized expected standard deviation of the price in the next time step δt. This is a usefull quantity for the definitions, but this volatility is essentially unobserved. In a process, σ_{eff} gives the magnitude of the returns.

2.2   Realized volatility

The *realized volatility* corresponds to the annualized standard deviation of the returns in the interval between t and $t + \Delta T$

$$\sigma^2(t, t + \Delta T) = \frac{1 \text{ year}}{n \ \delta t} \sum_{t < t' \leq t + \Delta T} r^2(t') \tag{1}$$

where $r(t)$ are the (unannualized) returns measured over the time interval δt, and the ratio 1 year$/\delta t$ annualized the volatility. The empirical section is done with daily data and the returns are evaluated over a 1 day interval $\delta t = 1$ day. If the returns do not overlap in the sum, then $\Delta T = n \ \delta t$. At the time t, the realized volatility cannot be evaluated from the information set $\Omega(t)$. The realized volatility is *the* usefull quantity we would like to forecast and to relate to the implied volatility.

2.3 Forward variance

In a continuum time formulation, the *expected cumulative variance* is defined by

$$V(t, t + \Delta T) = \int_{t}^{t+\Delta T} dt' \, E\left[\sigma_{\text{eff}}^2(t') \mid \Omega(t)\right] \tag{2}$$

and the *forward variance* by

$$v(t, t + \Delta T) = \frac{\partial V(t, t + \Delta T)}{\partial \Delta T} = E\left[\sigma_{\text{eff}}^2(t + \Delta T) \mid \Omega(t)\right]. \tag{3}$$

The cumulative variance is an extensive quantities as it is proportional to ΔT. For empirical investigation, it is simpler to work with an intensive quantity as this remove a trivial dependency on the time horizon. For this reason, the cumulative variance is used only in the theoretical part (hence also the continuum definition with an integral), whereas the forecasted volatility is used in the empirical part.

The variance enters into the variable leg of a variance swap, and as such, it is tradable. Related tradable instruments are the volatility indexes like the VIX (but the relation is indirect as the index is defined through implied volatility of a basket of options). Because volatility is becoming tradable, the forward variance should be a martingale

$$E\left[v(t', T) \mid \Omega(t)\right] = v(t, T). \tag{4}$$

For the volatility, this condition is quite weak as it follows also from the chain rule for conditional expectation

$$E\left[E\left[\sigma_{\text{eff}}^2(T) \mid \Omega(t')\right] \mid \Omega(t)\right] = E\left[\sigma_{\text{eff}}^2(T) \mid \Omega(t)\right] \qquad \text{for } t < t' < T \tag{5}$$

and from the definition of the forward variance as a conditional expectation. Therefore, any forecast build as a conditional expectation produces a martingale for the forward variance.

At this level, there is a formal analogy with interest rates, with the (zero coupon) interest rate and forward rate being analogous to the cumulative variance and forward variance. Therefore, some ideas and equations can be borrowed from the IR field. For example, on the modeling side, one can write process for the cumulative variance or for the forward variance, the later being more convenient as the martingale condition gives simpler constraints on the possible equations. In this paper, the ARCH path is followed using a multi-scale process for the underlying. The forward variance is computed as an expectation, and therefore the martingale property follows. In section 6, this ARCH approach is contrasted with a direct model for the forward volatility, where the martingale condition has to be explicitly enforced.

2.4 The forecasted volatility

The *forecasted volatility* is defined by

$$\widetilde{\sigma}^2(t, t + \Delta T) = \frac{1}{n} \sum_{t < t' \leq t + \Delta T} E\left[\sigma_{\text{eff}}^2(t') \mid \Omega(t)\right] \tag{6}$$

Up to a normalization and the transformation of the integral into a discrete sum, this definition is similar to the expected cumulative variance.

2.5 The implied volatility

As usual, the implied volatility is defined as the volatility to insert into the Black-Sholes equation so as to recover the market price for the option. The implied volatility $\sigma_{BS}(m, \Delta T)$ is a function of the moneyness m and of the time to maturity ΔT. The moneyness can be defined is various ways, with most definitions similar to $m \simeq \ln(F/K)$, and with F the forward rate $F = Se^{r \, \Delta T}$. The (forward) at-the-money option corresponds to $m = 0$. The *backbone* is the implied volatility at the money $\sigma_{BS}(\Delta T) = \sigma_{BS}(m = 0, \Delta T)$, as a function of the time to maturity ΔT. For a given time to maturity ΔT, the implied volatility as function of moneyness is called the smile.

Intuitively, the implied volatility surface can loosely be decomposed in backbone × smile. The rationale for this decomposition is that the two directions depend on different option features. The backbone is related to the expected volatility until the option expiry

$$\widetilde{\sigma}(t, t + \Delta T) = \sigma_{BS}(m = 0, \Delta T)(t) \tag{7}$$

In the Black-Sholes formula, the volatility appears only through the combination $\Delta T \, \sigma^2$, corresponding to the cumulative expected variance. In the other direction, the smile is the fudge factor to remedy the incomplete modeling of the underlying by a Gaussian random walk. The Black-Sholes model has the key advantage to be solvable, but does not include many stylized facts like heteroscedasticity, fat-tails, or leverage effect. These shortcomings translate into various "features" of the smile.

In principle, the equation 7 should be checked using empirical data. Yet this comparison raises a number of issues, on both sides of the equation. On the left hand side, the variance forecast should be computed using some equations and the time series for the underlying. The forecasting scheme, with its estimated parameters, is subject to errors. On the right had side, the option market has its own idiosyncracies, for example related to demand and supply. Such effect can be clearly observed by computing the implied volatility corresponding to the option bid or ask prices. These points are discussed in more details in sec. 8. Therefore, the equation 7 should be taken only as a first order approximation.

3 Multi-components ARCH processes

3.1 The general setup

The basic idea of a multi-components ARCH process is to measure historical volatilities using exponential moving average on a set of time horizons, and to compute the effective volatility for the next time step as a convex combination of the historical volatilities. A first process along similar line was introduced in [Dacorogna et al., 1998], and this family of processes was throughly developed and explored in [Zumbach and Lynch, 2001, Lynch and Zumbach, 2003, Zumbach, 2004]. A particular simple process with long memory is used to build the RM2006 risk methodology [Zumbach, 2006], with the salient feature to be very parsimonious. One of the key advantage of these multi-components processes is that forecast for the variance can be computed analytically. We will use this property to explore their relations with the option implied volatility.

In order to build the process, the historical volatilities are measured by exponential moving averages (EMA) at time scales τ_k

$$\sigma_k^2(t) = \mu_k \ \sigma_k^2(t - \delta t) + (1 - \mu_k) \ r^2(t) \qquad k = 1, \cdots, n \tag{8}$$

and with decay coefficients $\mu_k = \exp(-\delta t/\tau_k)$. The process time increment is δt, and δt = 1 day in this work. Let us emphasize that the σ_k are computed from historical data, and there is no hidden stochastic processes like in a stochastic volatility model.

The "effective" variance σ_{eff}^2 is a convex combination of the σ_k^2 and of the mean variance σ_∞^2

$$
\begin{aligned}
\sigma_{\text{eff}}^2(t) &= \sum_{k=1}^{n} w_k \ \sigma_k^2(t) + w_\infty \ \sigma_\infty^2 \ = \ \sigma_\infty^2 + \sum_{k=1}^{n} w_k \ \left(\sigma_k^2(t) - \sigma_\infty^2\right) \\
1 &= \sum_{k=1}^{n} w_k + w_\infty
\end{aligned}
\tag{9}
$$

Finally, the price follow a random walk with volatility σ_{eff}

$$r(t + \delta t) = \sigma_{\text{eff}}(t) \ \epsilon(t + \delta t). \tag{10}$$

Depending on the number of components n, the time horizons τ_k and weights w_k, a number of interesting processes can be build. The processes we are using to compare with implied volatility are given in the next subsections.

On general ground, we make the distinction between affine processes for which the mean volatility is fixed by σ_∞ and $w_\infty > 0$, and the linear process for which $w_\infty = 0$. The linear

and affine terms qualify the equations for the variance (i.e. in σ^2). The linear processes are very interesting for forecasting volatility as they have no mean volatility parameter σ_∞ which is clearly time series dependent. However, their asymptotic properties are singular, and affine processes should be used in Monte Carlo simulations. This subtle difference between both classes of processes is discussed in details in [Zumbach, 2004]. As this paper deal with volatility forecasts, only the linear processes are used.

3.2 I-GARCH(1)

The I-GARCH(1) model corresponds to a 1-component linear process

$$
\begin{aligned}
\sigma^2(t) &= \mu\, \sigma^2(t - \delta t) + (1 - \mu)\, r^2(t) \\
\sigma^2_{\text{eff}}(t) &= \sigma^2(t).
\end{aligned}
$$

It has one parameter τ (or equivalently μ). This process is equivalent to the integrated GARCH(1,1) process [Engle and Bollerslev, 1986], and with a given value for μ is equivalent to the standard RiskMetrics methodology. Its advantage is to be the most simple, but it does not capture mean reversion for the forecast (i.e. that forecasts for increasing horizons should converge to a (mean) long term volatility).

For the empirical evaluation, the characteristic time has been fixed *a priori* to $\tau = 16$ business days, corresponding to $\mu \simeq 0.94$.

3.3 I-GARCH(2) and GARCH(1,1)

The I-GARCH(2) process corresponds to a 2-components linear model

$$
\begin{aligned}
\sigma_1^2(t) &= \mu_1\, \sigma_1^2(t - \delta t) + (1 - \mu_1)\, r^2(t) \\
\sigma_2^2(t) &= \mu_2\, \sigma_2^2(t - \delta t) + (1 - \mu_2)\, r^2(t) \\
\sigma^2_{\text{eff}}(t) &= w_1 \sigma_1^2(t) + w_2 \sigma_2^2(t)
\end{aligned}
\tag{11}
$$

It has three parameters τ_1, τ_2 and w_1. Even if this process is linear, it has mean reversion for time scale up to τ_2, with $\sigma_2(t)$ playing the role of the mean volatility.

The GARCH(1,1) process [Engle and Bollerslev, 1986] corresponds to the 1-component affine model

$$
\begin{aligned}
\sigma_1^2(t) &= \mu_1\, \sigma_1^2(t - \delta t) + (1 - \mu_1)\, r^2(t) \\
\sigma^2_{\text{eff}}(t) &= (1 - w_\infty)\, \sigma_1^2(t) + w_\infty \sigma_\infty^2
\end{aligned}
\tag{12}
$$

It has three parameters τ_1, w_∞ and σ_∞. In this form, the analogy between the I-GARCH(2) and GARCH(1,1) processes is clear, with the long term volatility σ_2 playing a similar role as the mean volatility σ_∞.

Given a process, the parameters need to be estimated on a time series. GARCH(1,1) is more problematic with that respect because σ_∞ is clearly time series dependent. A good procedure is to estimate the parameters on a moving historical sample, say in a window between $t - \Delta T'$ and t for a fixed span $\Delta T'$. With this setup, the mean variance σ_∞^2 is essentially the sample variance $\sum r^2$ computed on the estimating window. This is a rectangular moving average, similar to an EMA but for the weights given to the past. This argument shows that I-GARCH(2) and (a continuously re-estimated on a moving window) GARCH(1,1) behaves similarly. A detailled analysis of both processes in [Zumbach, 2004] show that they have similar forecasting power, with an advantage to I-GARCH(2).

In this work, we use the I-GARCH(2) process with two parameter sets fixed *a priori* to some reasonable values. The first set is $\tau_1 = 4$ business days, $\tau_2 = 512$ business days, $w_1 = 0.843$ and $w_2 = 0.157$. The second set is $\tau_1 = 16$ business days, $\tau_2 = 512$ business days, $w_1 = 0.804$ and $w_2 = 0.196$. The values for the weights are obtained according to the long memory ARCH process, but with only two given τ components.

3.4 Long Memory ARCH

The idea for a long memory process is to use a multi-components ARCH model with a large number of components but simple analytical form for the characteristic time τ_k and the weights w_k. For the long memory ARCH process, the characteristic times τ_k increase as a geometric series

$$\tau_k = \tau_1 \, \rho^{k-1} \qquad k = 1, \cdots, n \tag{13}$$

while the weights decay logarithmically

$$
\begin{aligned}
w_k &= \frac{1}{C} \, (1 - \ln(\tau_k)/\ln(\tau_0)) \\
C &= \sum_k (1 - \ln(\tau_k)/\ln(\tau_0)) \, .
\end{aligned}
\tag{14}
$$

This choice produces lagged correlations for the volatility that decays logarithmically, as observed in the empirical data [Zumbach, 2006]. The parameters are taken as for the RM2006 methodology [Zumbach, 2006], namely $\tau_1 = 4$ business days, $\tau_n = 512$ business days, $\rho = \sqrt{2}$ and the logarithmic decay factor $\tau_0 = 1560$ days $= 6$ years .
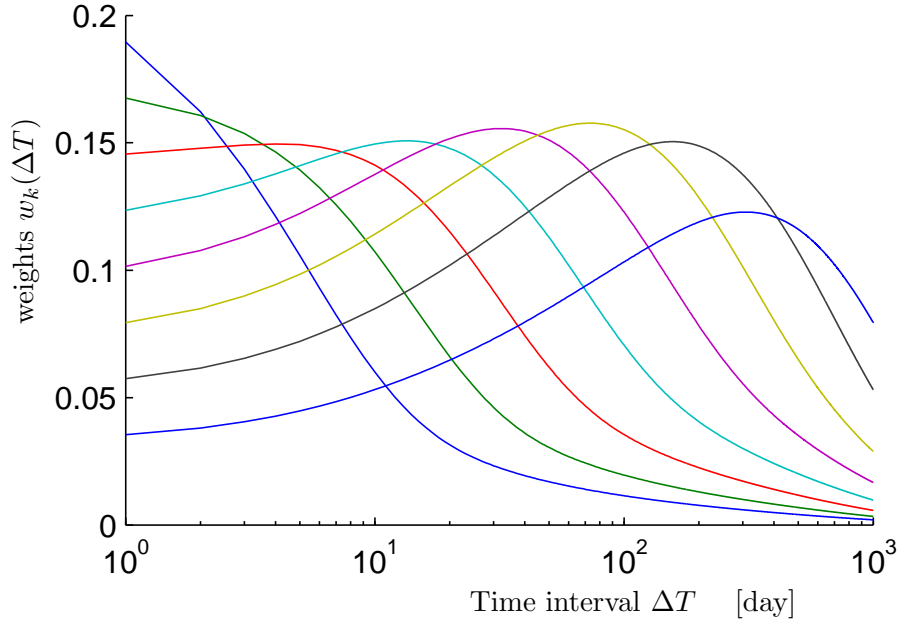
Figure 1: The weights $w_k(\Delta T)$ as function of the forecst horizon ΔT for a long memory process with $w_\infty = 0.1$ and $\tau_k = 2, 4, 8, 16, \cdots, 256$ days . The weights with increasing time horizon τ_k have decreasing initial values and the maximum values going from left to right.

4   Forward variance and multi-components ARCH processes

For multiscales ARCH processes (I-GARCH, GARCH(1,1), long-memory ARCH, etc ...), the forward variance can be computed analytically [Zumbach, 2004, Zumbach, 2006]. The idea is to compute the conditional expectation of the process equations, from which iterative relations can be deduced. Then, some algebra and matrix computations allow to get the following form for the forward variance

$$v(t, t + \Delta T) = E\left[\sigma_{\text{eff}}^2(t + \Delta T) \mid \Omega(t)\right] = \sigma_\infty^2 + \sum_{k=1}^{n} w_k(\Delta T)\left(\sigma_k^2(t) - \sigma_\infty^2\right) \qquad (15)$$

The weight $w_k(\Delta T)$ can be computed by a recursion formula depending on the decay coefficients μ_k and with initial values given by $w_k = w_k(1)$. The equation for the forecast of the realized volatility has the same form but the weights $w_k(\Delta T)$ are different.

Let us emphasize that this can be done for all processes in this class (linear and affine). Moreover, the $\sigma_k^2(t)$ are computed from the underlying time series, namely there is no hidden stochastic volatility to estimate. This makes volatility forecasts particularly easy in this framework.
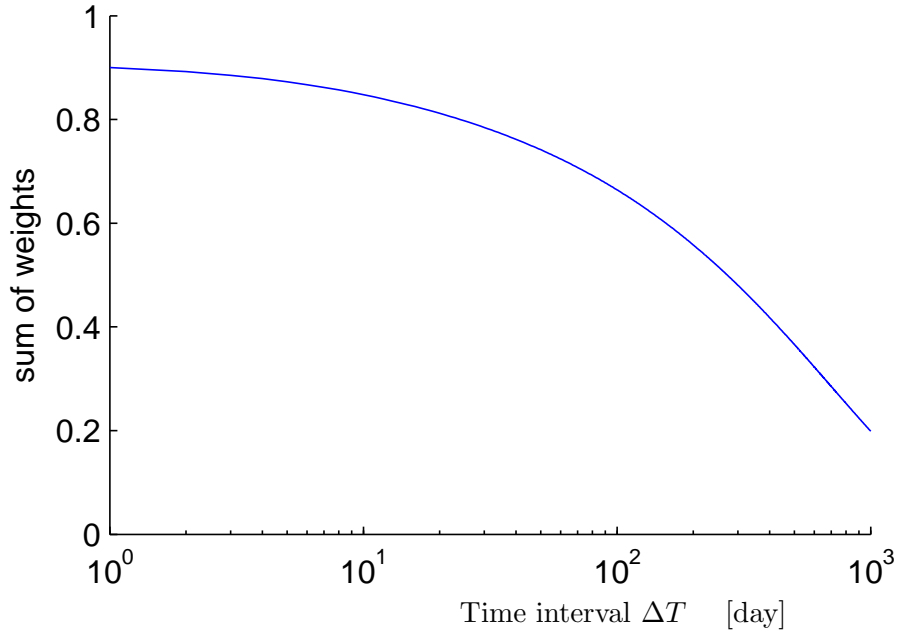
Figure 2: The sum of the weights $\sum_k w_k(\Delta T) = 1 - w_\infty$, for the same parameters as above.

For a multi-component ARCH process, the intuition for the forecast can be understood from a graph of the weights $w_k(\Delta T)$ as function of the forecast horizon ΔT as given in Fig. 1. For short forecast horizon, the volatilities with the shorter time horizons dominate. As the forecast horizon get larger, the weights of the short term volatilities decay while the weights of the longer time horizons get larger. The weight for a particular horizon τ_k peaks at a forecast horizon similar to τ_k, for example the Burgundy curve corresponds to $\tau = 32$ days and its maximum is around a similar value. The figure 2 shows the sum of the volatility coefficients $\sum_k w_k = 1 - w_\infty$. This shows the increasing weight of the mean volatility as the forecast horizon get longer. Notice that this behavior corresponds to our general intuition about forecasts, namely short term forecasts depend mainly on the recent past while long term forecasts need to use more informations from the distant past. The nice feature of the multi-components ARCH process is that the forecast weights are derived from the process equations, and that they have a similar content compared to the process equations (linear or affine, one or multiple time scales).

5   The induced volatility process

The multi-components ARCH processes are stochastic processes for the return, in which the volatilities are convenient intermediate quantities. It is important to realize that

10

the volatilities σ_k and σ_{eff} are useful and intuitive in formulating a model, but they can be completely eliminated from the equations. An important advantage of this class of process is that the forward variance $v(t, t + \Delta T)$ can be computed analytically. Going in the opposite direction, we want to eliminate the return, namely to derive the equivalent process equations for the dynamic of the forward variance induced by a multi-component ARCH process. This will allow us to make contact with some models for the forward variance that are available in the literature and presented in the next section.

The eq. 8 for σ_k can be rewritten as

$$
\begin{aligned}
d\sigma_k^2(t) &= \sigma_k^2(t) - \sigma_k^2(t - \delta t) && (16) \\
&= (1 - \mu_k) \left\{ -\sigma_k^2(t - \delta t) + \epsilon^2(t)\, \sigma_{\text{eff}}^2(t - \delta t) \right\} \\
&= (1 - \mu_k) \left\{ \sigma_{\text{eff}}^2(t - \delta t) - \sigma_k^2(t - \delta t) + (\epsilon^2(t) - 1)\, \sigma_{\text{eff}}^2(t - \delta t) \right\}
\end{aligned}
$$

The equation can be simplified by introducing the annualized variances $v_k = 1y/\delta t\; \sigma_k^2$, $v_{\text{eff}} = 1y/\delta t\; \sigma_{\text{eff}}^2$ and a new random variable χ with

$$
\chi = \epsilon^2 - 1 \qquad \text{such that} \qquad E\left[\,\chi(t)\,\right] = 0, \quad \chi(t) > -1. \qquad (17)
$$

Assuming that the time increment δt is small compared to the time scales τ_k in the model, the following approximation can be used

$$
1 - \mu_k = \frac{\delta t}{\tau_k} + O(\delta t^2). \qquad (18)
$$

In the present derivation, this expansion is used only to make contact with the more usual continuous time form, but no term of higher order are neglected. Exact expressions are obtained by replacing $\delta t/\tau_k$ by $1 - \mu_k$ in the equations below.

These notations and approximations allows to write the equivalent equations

$$
dv_k = \frac{\delta t}{\tau_k} \left\{ v_{\text{eff}} - v_k + \chi\, v_{\text{eff}} \right\} \qquad (19a)
$$

$$
v_{\text{eff}} = \sum_k w_k\, v_k + w_\infty v_\infty \qquad (19b)
$$

The process for the forward variance is given by

$$
dv_{\Delta T} = \sum_k w_k(\Delta T)\, dv_k \qquad (20)
$$

with $dv_\tau(t) = v(t, t + \Delta T) - v(t - \delta t, t - \delta t + \Delta T)$.

11

The content of Eq. 19a is the following. The term $\delta t \left\{ v_{\text{eff}} - v_k \right\} / \tau_k$ gives a mean reversion toward the current effective volatility v_{eff} at a time scale τ_k. This structure is fairly standard, except for v_{eff} which is given by a convex combination of all the variances v_k. Then, the random term is unusual. All the variances share the same random factor $\delta t \, \chi / \tau_k$, which has a standard deviation of order δt instead of the usual $\sqrt{\delta t}$ appearing in Gaussian model.

An interesting property of this equation is to enforce positivity for v_k through a somewhat peculiar mechanism. The equation 19a can be rewritten as

$$dv_k = \frac{\delta t}{\tau_k} \left\{ -v_k + (\chi + 1)v_{\text{eff}} \right\} \tag{21}$$

Because $\chi \geq 1$, the term $(\chi + 1)v_{\text{eff}}$ is never negative, and as $\delta t \, v_k(t - \delta t)/\tau_k$ is smaller than $v_k(t - \delta t)$, this implies that $v_k(t)$ is always positive (even for a finite δt). Another difference with the usual random process is that the distribution for χ is not Gaussian. In particularly if ϵ has a fat-tail distribution, as seems required in order to have a data generating process that reproduce the properties of the empirical time series, the distribution for χ also has fat tails.

The continuum limit of the GARCH(1,1) process was already investigated by [Nelson, 1990]. In this limit, GARCH(1,1) is equivalent to a stochastic volatility process where the variance has its own source of randomness. Yet Nelson constructed a different limit as above because he fixes the GARCH parameters α_0, α_1 and β_1. The decay coefficient is given by $\alpha_1 + \beta_1 = \mu$ and is therefore fixed. With $\mu = \exp(-\delta t/\tau)$, fixing μ and taking the limit $\delta t \to 0$ is equivalent to $\tau \to 0$. Because the characteristic time τ of the EMA go to zero, the volatility process becomes independent of the return process, and the model converges toward a stochastic volatility model. A more interesting limit is to take τ fixed and $\delta t \to 0$, as in the computation above. Notice that the computation is done with a finite time increment δt; the existence of a proper continuum limit $\delta t \to 0$ for a process defined by eq. 19b to 20 is likely not a simple question.

Let us emphasize that the derivation of the volatility process as induced by the ARCH structure involves only elementary algebra. Essentially, if the price follows an ARCH process (one or multiple time scales, with or without mean σ_∞), then the volatility follows a process according to 19. The structure of this process involves a random term of order δt and therefore it cannot be reduced to a Wiener process. This is a key difference from the processes used in finance that were developed to capture the price diffusion.

The implications of eq. 19 are important as they show a key difference between ARCH and stochastic volatility processes. This has clearly implication for option pricing, but also for risk evaluation. In a risk context, the implied volatility is a risk factor for any portfolio that contains options, and it is likely better to model the dynamic of the implied volatility by a process with a similar structure.

6   Market model for the variance

In the literature, the models for the implied volatility are dominated by stochastic volatility processes, essentially assuming that the implied volatility "has its own life", independently of the underlying. In this vast literature, a recent direction is to write processes directly for the forward variance. Recent papers in this direction include [Buehler, 2006] and [Bergomi, 2005], and a presentation by [Gatheral, 2007]. In this direction, we present here simple linear processes for the forward variance, and discuss the relation with a multi-components ARCH in the next section.

The general idea is to write a model for the forward variance

$$v(t, t + \Delta T) = G(v_k(t); \Delta T) \tag{22}$$

where G is a given function of the (hidden) random factors v_k. In principle, the random factors can appear everywhere in the equation, say for example as a random characteristic time like τ_k. Yet, Buehler has showed that strong constraints exist on the possible random factors, for example forbiding random characteristic time. In this paper, only linear model will be discussed, and therefore the random factor appears as a variance v_k.

The dynamic for the random factor v_k are given by processes

$$dv_k = \mu_k(v) \; dt + \sum_{\alpha=1}^{d} \sigma_k^\alpha(v) \; dW^\alpha \qquad k = 1, \cdots, n. \tag{23}$$

The processes have d sources of randomness dW^α, and the volatility $\sigma_k^\alpha(v)$ can be any function of the factors.

As such, the model is essentially unconstraint, but the martingale condition 4 for the forward variance still has to be enforced. Through standard Ito calculus, the variance curve model together with the martingale condition lead to a constraint between $G(v; \Delta T)$, $\mu(v)$ and $\sigma(v)$

$$\partial_{\Delta T} G(v; \Delta T) = \sum_{i=1}^{n} \mu_i \; \partial_{v_i} G(v; \Delta T) + \sum_{i,j=1}^{n} \sum_{\alpha=1}^{d} \sigma_i^\alpha \sigma_j^\alpha \; \partial_{v_i, v_j}^2 G(v; \Delta T) \tag{24}$$

A given function G is say to be compatible with a dynamic for the factors if this condition is valid. The compatibility constraint is fairly weak, and many processes can be written for the forward variance that are martingale. As already mentionned, we consider only functions G that are linear in the risk factors. Therefore, $\partial_{v_i, v_j}^2 G = 0$, leading to first order differential equations that can be solved by elementary techniques. For this class of models, the condition does not involve the volatility $\sigma_k^\alpha(v)$ of the factor, which therefore can be chosen freely.

6.1 Example: one factor market model

The forward variance is parameterized by

$$\begin{aligned}
G(v_1; \Delta T) &= v_\infty + w_1(\Delta T)(v_1 - v_\infty) \\
w_1(\Delta T) &= w_1 \, e^{-\Delta T/\tau_1}
\end{aligned} \tag{25}$$

which is compatible with the stochastic volatility dynamic

$$dv_1 = -(v_1 - v_\infty) \, \frac{dt}{\tau_1} + \gamma \, v_1^\beta \, dW \qquad \text{for} \quad \beta \in [1/2, 1]. \tag{26}$$

The parameter w_1 can be chosen freely, and for identification purpose the choice $w_1 = 1$ is often made. Because G is linear in v_1, there is no constraint on β. The value $\beta = 1/2$ corresponds to the Heston model, $\beta = 1$ to the log-normal model. This model is somewhat similar to the GARCH process, with one characteristic time τ_1, a mean volatility v_∞, and the volatility of the volatility (vol-of-vol) γ. This model is not rich enough to describe the empirical forward variance dynamic, which involve multiple time scale.

6.2 Example: two factors market model

The linear model with two factors

$$\begin{aligned}
G(v; \Delta T) &= v_\infty + w_1(\Delta T) \, (v_1 - v_\infty) + w_2(\Delta T) \, (v_2 - v_\infty) \\
w_1(\Delta T) &= w_1 \, e^{-\Delta T/\tau_1} \\
w_2(\Delta T) &= \frac{1}{1 - \tau_1/\tau_2} \left(-w_1 \, e^{-\Delta T/\tau_1} + (w_1 + w_2) \, e^{-\Delta T/\tau_2} \right)
\end{aligned} \tag{27}$$

is compatible with the dynamic

$$\begin{aligned}
dv_1 &= -(v_1 - v_2) \, dt/\tau_1 + \gamma \, v_1^\beta \, dW_1 \\
dv_2 &= -(v_2 - v_\infty) \, dt/\tau_2 + \gamma \, v_2^\beta \, dW_2.
\end{aligned} \tag{28}$$

The parameters w_1 and w_2 can be chosen freely, and for identification purpose the choice $w_1 = 1$ and $w_2 = 0$ is often made. Notice the similarity of the equation 27 with the Nelson-Siegel-Svensson parameterization for the yield curve.

The linear model can be solved explicitely for n-components, but the ΔT dependency in the coefficients $w_k(\Delta T)$ becomes increasingly complex. It is therefore not natural in this approach to create the equivalent of a long-memory model with multiple time scales.

7 Market models and options

Assuming a liquid option market, the implied volatility surface can be extracted, and from its backbone, the forward variance $v(t, t + \Delta T)$ is computed. At a given time t, given a market model $G(v_k(t); \Delta T)$, the risk factors $v_k(t)$ are estimated by fitting the function $G(\Delta T)$ on the forward variance curve. It is therefore important for the function $G(\Delta T)$ to have enough possible shapes to accommodate the various forward variance curves. This estimation procedure for the risk factors gives the initial condition $v_k(t)$. Then, the postulated dynamics for the risk factors induce a dynamic for G, and hence of the forward variance.

Notice that in this approach, there is no relation with the underlying and its dynamic. For this reason, the possible processes are weakly constrained, and the parameters need to be estimated independently (say for example the characteristic times τ_k). Another drawback of this approach is to rely on the empirical forward variance curve, and therefore a liquid option market is a prerequisite.

Our choice of notations makes clear the formal analogy of the market model with the forecasts produced by a multi-component ARCH process. Except for the detailled shapes of the functions $w_k(\Delta T)$, the equations 15 and 27 have the same structure. They are however quite different in their spirits as the v_k are computed from the underlying time series in the ARCH approach, whereas in a market model approach the v_k are estimated from the forward variance curve obtained from the option market. In other word, ARCH leads to a genuine forecast based on the underlying, whereas market model provides for a constraint fit of the empirical forward curve. Beyond this formal analogy, the dynamic for the risk factors are quite different as the ARCH approach leads to the unusual eq. 19a whereas market models use the familiar generic Gaussian process in eq. 23.

8 Comparison of the empirical implied, forecasted and realized volatilities

As explained in sec. 4, a multi-components ARCH process provides us with a forecast for the realized volatility, and the forecast is directly related to the underlying process and its properties. At a given time t, there is three volatilities (implied, forecasted and realized) for each forecast horizon ΔT. Essentially, the implied and forecasted volatilities are forecasts for the realized volatility. In this section, we investigate the relationship between these three volatilities and the forecast horizon ΔT. When analyzing the empirical statistics and comparing these three volatilities, several factors should be kept in mind.

1. For short forecast horizons ($\Delta T =$ up to 10 days), the number of returns in ΔT is

small and therefore the realized volatility estimator (computed with daily data) has a large variance.

2. The forecastability decreases with increasing ΔT.

3. The forecast and implied volatilities are "computed" using the same information set, namely the history up to t. This is different from the realized volatility, computed using the information in the interval $[t, t + \Delta T]$. Therefore, we expect the distance between the forecast and implied to be the smallest.

   At a more detailed level, the information set for the implied volatility is richer, because traders use intra-day information which helps building better forecasts, particularly for short risk horizons. This contrasts with all the present ARCH forecasts that are computed using only daily close prices. From this difference on their actual information sets, the implied volatility can be expected to provide for a better forecast of the realized volatility.

4. The implied volatility has some particular idiosyncracies related to the option market, for example supply and demand, or the liquidity of the underlying necessary to implement the replication strategy. Similarly, an option bears a volatility risk, and a related volatility risk premium can be expected. These particular effects could bias the implied volatility upward.

5. From the raw options and underlying prices, the computations leading to the implied volatility are complex, and therefore error prone. This data quality problem is inherent to the original data provider and the option market, and is a reflect of the difficulty to compute clean and reliable implied volatility surfaces. For stocks, the problem is made more difficult because of the dividents, the corporate events and the smaller liquidity. For this reason, we present only the figures corresponding to two of the most liquid option markets. The results have been checked with other FX rates, stock indexes and stocks, and are essentially valid for all underlyings.

6. The options are traded for fixed maturity time, whereas the convenient volatility surface is given for constant time to maturity. Therefore, some interpolation and extrapolation need to be done. As exchanged traded options are defined with one maturity per month, it is difficult to get reliable implied volatility for time to maturity smaller than one month.

7. The ARCH based forecasts are dependent on the choice of the process and the associated parameters.

8. As the forecast horizon increases, the dynamic of the volatility get slower and the actual number of independent volatility points decreases (as $1/\Delta T$). Therefore, the statistical uncertainty on the statistics are increasing with ΔT.
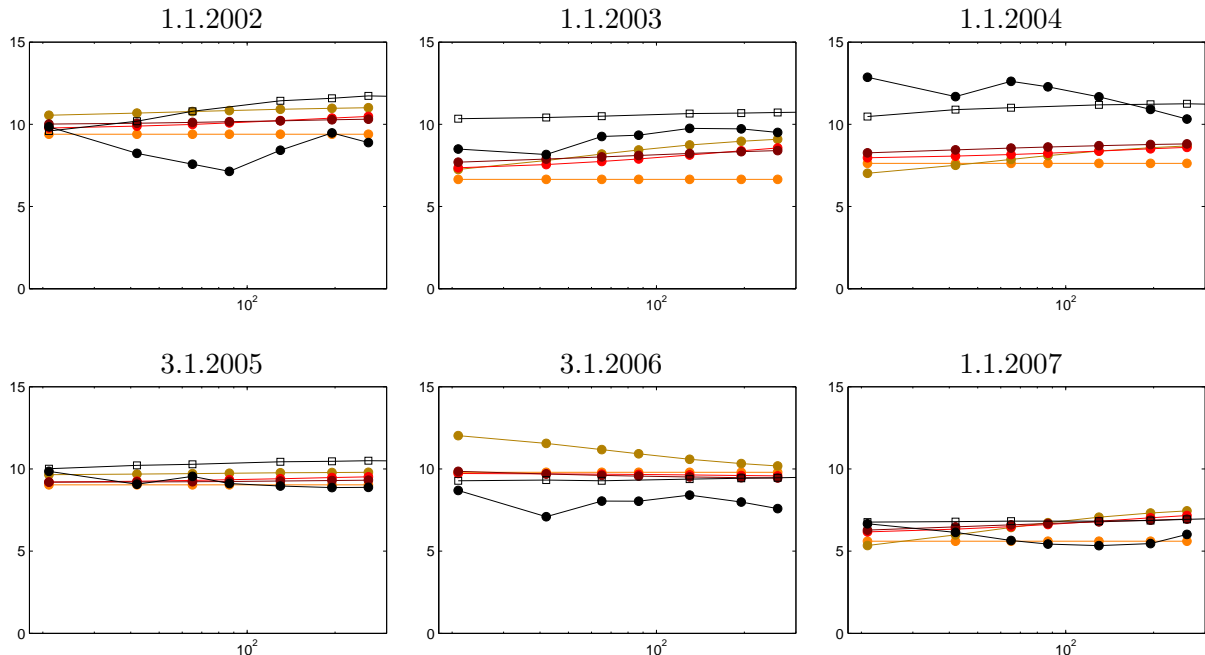
Figure 3: The volatilities at the beginning of the years 2002 to 2007, for EUR/USD. The black curve with square symbols is the realized volatility, the black curve with full circle symbols is the implied volatility, and the color curve with full circle symbols is the forecast according to the various ARCH processes (with the same colors as below). The vertical axis gives the annualized volatility in %, the horizontal axis the forecast time interval ΔT in day.

Because of the above points, each volatility has some peculiarities, and therefore we do not have a firm anchor point to base our comparison. Given that we are on a floating ground, our goals are fairly modest. Essentially, we want to show that a process with one time scale is not good enough, and that the long-memory process provides for a good forecast with an accuracy comparable to the implied volatility. The processes used in the analysis are I-GARCH(1), I-GARCH(2) with two set of parameters and LM-ARCH. The equations for the processes are given in sec. 3, with the values for the parameters.

The best way to visualize the dynamic of the three volatilities would be to use a movie of the $\sigma[\Delta T]$ time evolution. Unfortunately, the present analogic paper does not allow for such medium, and we present instead 6 snapshots for EUR/USD in Figure 3. Overall, the realized volatility has a weak term structure, although the global level changes significantly with time. The implied volatility has more structures as function of the time to maturity, but this seems not always appropriate. The term structures for the ARCH forecasts are in line with the implied volatility, with essentially a weak term structure. The I-GARCH(1) process has a constant term structure, and this explains why its forecasting performances are indeed very good compared to more complex processes. Beyond a qual-

17

itative assessement of the term structure, the various forecasts for the realized volatility
are difficult to rank, but clearly the ARCH forecasts are close to the target and compare
well with the implied volatility.

The statistics are presented for two time series, the USD/EUR foreign exchange rate and
the DAX stock index. The time series for the volatilities are shown on fig. 4 for a 3 months
forecast horizon. The time series are not very long ($\sim$10 years for USD/EUR, $\sim$6 years
for DAX). This clearly makes statistical inferences difficult, as the effective sample size
is fairly small. The lagging behavior of the forecast and implied volatility with respect
to the realized volatility is clearly observed. For the DAX, the data sample contains an
abrupt drop in the realized volatility at the beginning of 2003. This pattern was difficult
to capture for the models with long term mean reversion.

For the statistics, all the horizontal and vertical scales are identical, and the colors are
fixed for a given process. The graphs are presented for the mean absolute error (MAE)

$$\text{MAE}(x, y) = \frac{1}{n} \sum_t |x(t) - y(t)| \tag{29}$$

where n is the number of term in the sum. Other measures of distance like root mean
square error, or the MAE for $\ln(\sigma)$, give very similar figures.

The overall relationship betwen the three volatilities can be understood on figure 5. The
pair of volatilities with the closest relationship is the implied and forecasted volatilities,
because they are build upon the same information set. The distance with the realized
volatility is larger, with similar values for implied-realized and forecast-realized. This
shows that it is quite difficult to assert which one of the implied and forecasted volatility
provides for a better forecast of the realized volatility. All the distances have a global
U-shape form as function of ΔT. This originates in the points 1 and 2 above, which
leads to a minimum between 2 to 6 months for the distances. The distance is larger for
shorter ΔT because of the bad estimator for the realized volatility, and larger for longer
ΔT because of the decreasing forecastability. The time structures of the ARCH processes
impact the distances between the forecasted and implied volatility (dotted line), and the
relation between process structure and forecast quality discussed in the next paragraph.

The figure 6 shows the distances for given volatility pairs, depending on the process used to
build the forecast. The forecast-implied distance shows clear difference between processes
(left panels). The I-GARCH(1) process is lacking mean reversion, an important feature
of the volatility dynamic. The I-GARCH(2) process with parameter set 1 is handicapped
by the too short characteristic time for the first EMA (4 days); this leads to a noisy
volatility estimator and subsequently to a noisy forecast. The same process with a longer
characteristic time for the first EMA (16 days, parameter set 2) shows much improved
performance up to a time horizon comparable to the long EMA (512 days). Finally,
the LM-ARCH produces the best forecast. As the forecast becomes better (1 time scale
$\to$ 2 time scales $\to$ multiple time scales), the distance between the implied and forcasted
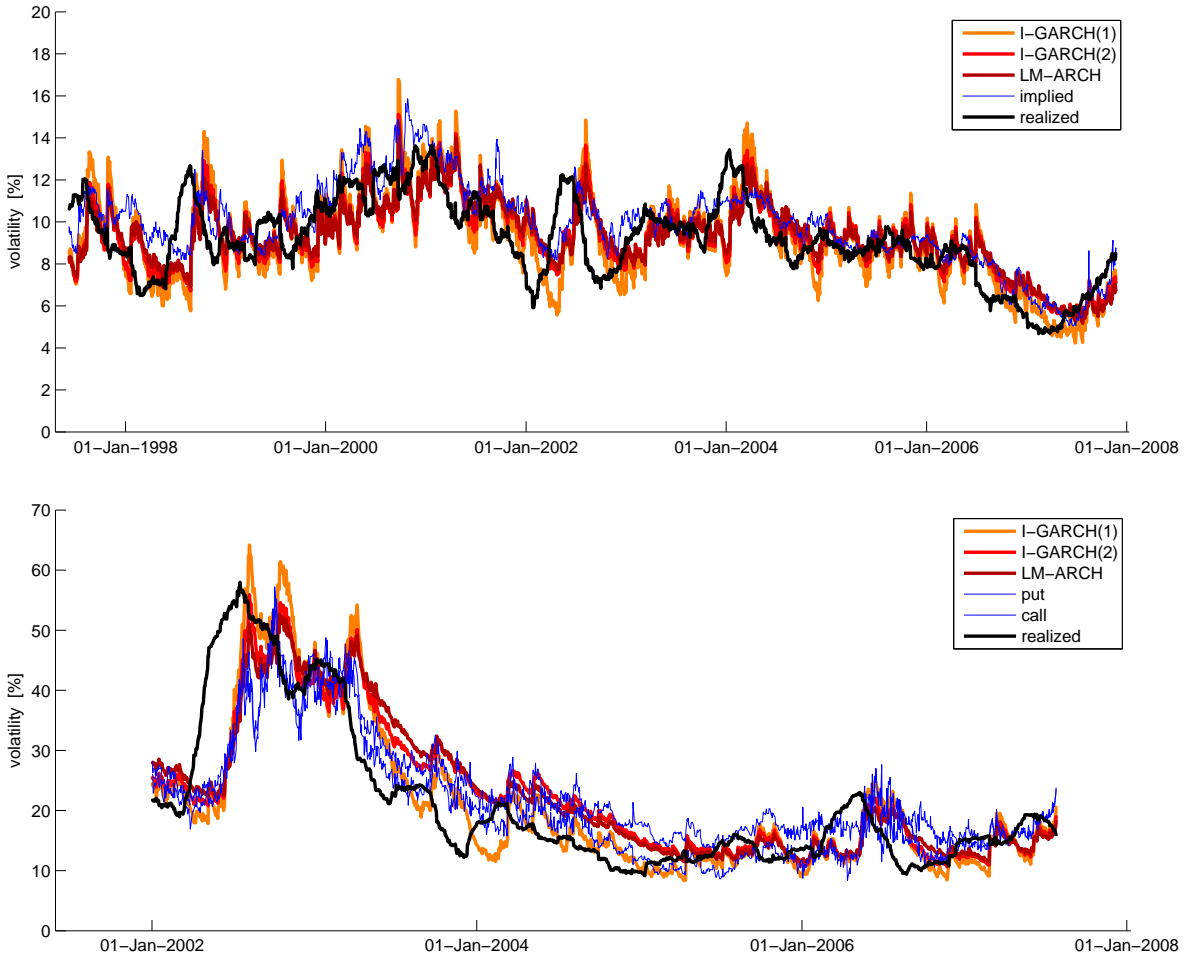
Figure 4: The volatilities time series for the USD/EUR (upper panel) and DAX (lower panel), for a 3 months forecast horizon. For the DAX data, the implied volatility is given for the put and call options (blue curves).
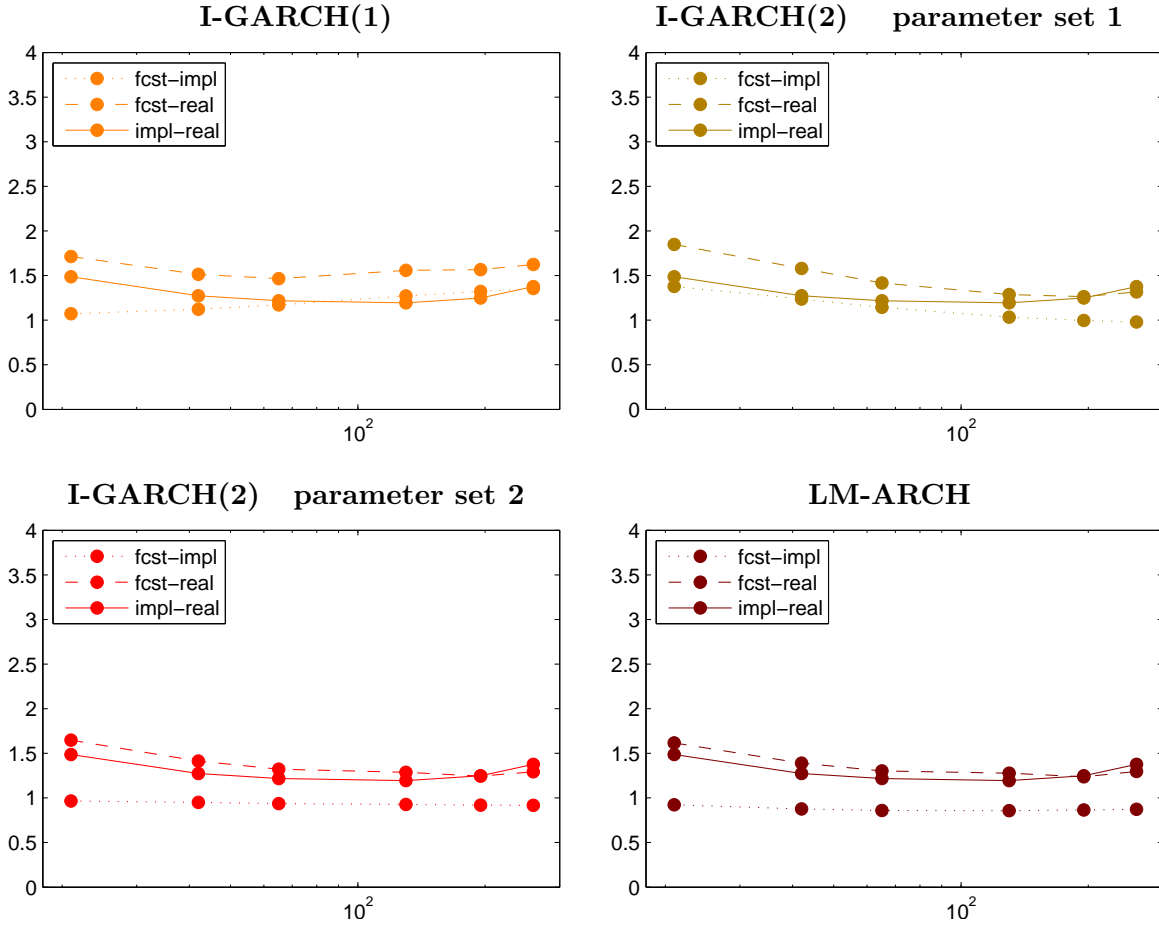
Figure 5: The MAE distances between volatility pairs for different forecasts: I-GARCH(1) (upper left, red), I-GARCH(2) parameters 1 (upper right, blue), I-GARCH(2) parameters 2 (lower left, blue) and LM-ARCH (lower right, black). The vertical axis gives the MAE for the annualized volatility in %, the horizontal axis the forecast time interval ΔT in day. The data is EUR/USD.
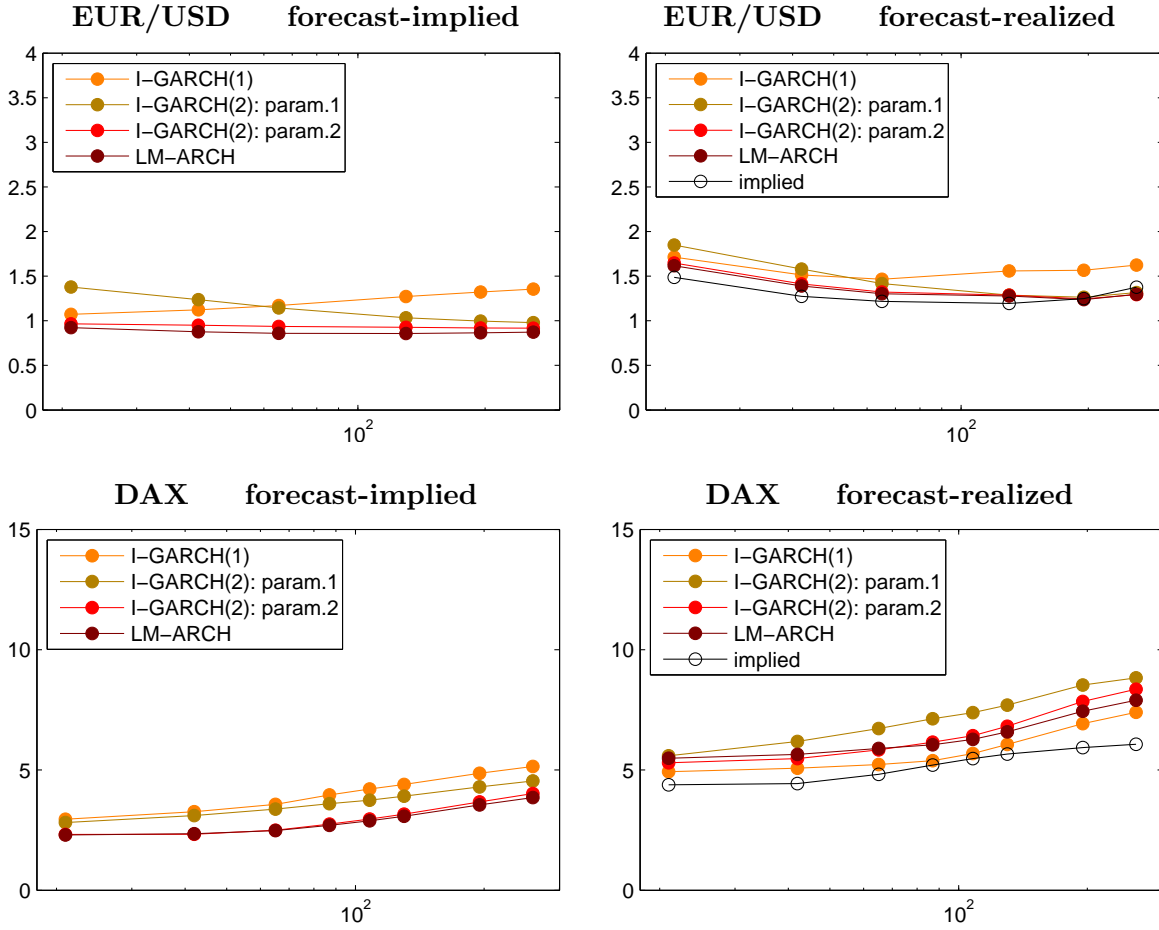
Figure 6: The MAE distances between volatility pairs: forecast-implied (left) and forecast-realized (right). The upper figures are for EUR/USD, the lower figure for the DAX stock index. The vertical axis gives the MAE for the annualized volatility in %, the horizontal axis the forecast time interval ΔT in day.

volatilities decreases. For EUR/USD, the mean volatility is around 10% (the precise value depending on the volatility and time horizon), and the MAE is in the 1 to 2% range. This shows that in this time to maturity range, we can build a good estimator of the ATM implied volatility based only on the underlying time series.

The distance forecast-realized is larger than the forecast-implied volatility (right panel), with the long memory process giving the smallest distance. The only exception is the I-GARCH(1) process applied to the DAX time series, due to the particular abrupt drop in the realized volatility at early 2003. This shows the limit of our analysis due to the fairly small data sample, and longer time series for implied volatility are required to gain more statistical power. Given the limited sample size, a cross sectional study over 9 other time series shows consistent results.

9   Conclusion

The "ménage à 3" between the forecasted, implied and realized volatilities is quite a complex affair, where each participants have their own character. The salient outcome is that the forecasted and impled volatilities have the closest relationship, while the realized volatility is more distant as it incorporates a larger information set. This picture is dependent to some extend on the quality of the volatility forecast: the multi-scale dynamic of the long memory ARCH process is seen to capture correctly the dynamic of the volatility, while the I-GARCH(1) process is not rich enough in its time scale structures. This conclusion falls in line with the risk methodology developed in [Zumbach, 2006], where the same long memory process is shown to capture correctly the lagged correlation for the volatility.

The connection with the market model for the forward variance shows the parallel in the structure of the volatility forecasts provided by both approaches. However, their dynamics are very different (postulated for the forward volatility market models, induced by the ARCH structure for the multi-components ARCH processes). Moreover, the volatility process induced by the ARCH equations is of a different type than the usual price process, because the random term is of order δt instead of $\sqrt{\delta t}$ used in diffusive equations. This emphasize a fundamental difference between price and volatility processes. A clear advantage of the ARCH approach is to deliver a forecast based only on the properties of the underlying time series, with a minimal number of parameters that need to be estimated (none in our case as all the parameters correspond to the values used in [Zumbach, 2006]). This point brings us to a nice and simple common framework to evaluate risks as well as a good approximation for the implied volatilities of at-the-money options.

The natural extension of this work is to study the whole implied volatility surface. As the backbone is essentially under control, the perpendicular direction needs to be studied, namely the volatility smile should be related to the underlying behavior. Due to the

heteroscedasticity, any multi-component ARCH process will capture some (symmetric) smile. Moreover, fat tail innovations will make the smile stronger, as the process becomes increasingly distant from a Gaussian random walk. Yet, adding an asymmetry in the smile, as observed for stocks and stock indexes, requires to enlarge the family of process to capture asymmetries in the distribution of returns. This is left for further work.

References

[Bergomi, 2005] Bergomi, L. (2005). Smile dynamics ii. *Risk*, 18:67–73.

[Buehler, 2006] Buehler, H. (2006). Consistent variance curve models. *Finance and Stochastics*, 10:178–203.

[Dacorogna et al., 1998] Dacorogna, M. M., Müller, U. A., Olsen, R. B., and Pictet, O. V. (1998). Modelling short-term volatility with GARCH and HARCH models. *published in "Nonlinear Modelling of High Frequency Financial Time Series" edited by Christian Dunis and Bin Zhou, John Wiley, Chichester*, pages 161–176.

[Engle and Bollerslev, 1986] Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5:1–50.

[Gatheral, 2007] Gatheral, J. (2007). Developments in volatility derivatives pricing. Presentation at "Global derivative", Paris, May 23.

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

[Nelson, 1990] Nelson, D. (1990). Arch model as diffusion approximation. *Journal of Econometrics*, 45:7–38.

[Poon, 2005] Poon, S.-H. (2005). *Forecasting financial market volatility*. Wiley Finance.

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

[Zumbach, 2006] Zumbach, G. (2006). The riskmetrics 2006 methodology. Technical report, RiskMetrics Group. Available at www.riskmetrics.com.

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