RESEARCH PAPER
quant.iop.org

QUANTITATIVE FINANCE VOLUME 4 (2004) 70–86
INSTITUTE OF PHYSICS PUBLISHING

Volatility processes and volatility forecast with long memory

Gilles Zumbach

Consulting in Financial Research, Ch. Charles Baudouin 8,
1228 Saconnex d'Arve, Switzerland

E-mail: gilles.zumbach@bluewin.ch

Received 1 April 2003, in final form 19 June 2003
Published 20 October 2003
Online at stacks.iop.org/Quant/4/70 (DOI: 10.1088/1469-7688/4/1/007)

Abstract

We introduce a new family of processes that include the long memory (LM) (power law) in the volatility correlation. This is achieved by measuring the historical volatilities on a set of increasing time horizons and by computing the resulting effective volatility by a sum with power law weights. The processes have two parameters (linear processes) or four parameters (affine processes). In the limit where only one component is included, the processes are equivalent to GARCH(1, 1) and I-GARCH(1). Volatility forecast is discussed in the context of processes with quadratic equations, in particular as a means to estimate process parameters. Using hourly data, the empirical properties of the new processes are compared to existing processes (GARCH, I-GARCH, FIGARCH, ...), in particular log-likelihood estimates and volatility forecast errors. This study covers time horizons ranging from 1 h to 1 month. We also study the variation of the estimated parameters with respect to changing sample by introducing a natural coordinate invariant distance. The LM processes show a small but systematic quantitative improvement with respect to the standard GARCH(1, 1) process. Yet, the main advantage of the new LM processes is that they give a good description of the empirical data from 1 h to 1 month, *with the same parameters*. Their other advantages are that they are efficient to evaluate numerically, that they behave well with respect to the cut-off (i.e. the largest time horizon included in the process) and that they can be extended along several directions.

1. Introduction

The volatility of financial time series enters as an important ingredient in many applications, such as portfolio optimization, risk evaluation or option pricing. In fact, the required input for all these computations is a *forecast* for the mean square volatility between now and time horizons relevant for the evaluation. Yet, in practice, a measure for the historical volatility is often used instead of a forecast. All these financial evaluations will benefit from using a proper volatility forecast.

The possible construction of a volatility forecast is indicated by the lagged correlation of the volatility (see for example figure 4). This positive correlation shows that there is information in the past about the future evolution of the volatility. Yet, the correlation is quantitatively small, of the order of 2–15%, depending on the actual measure of volatility and time horizons. This tells us that there is not much forecastability, namely the largest part of the future volatility cannot be predicted. Another salient feature of the lagged correlation is the slow decay of the correlation with the increasing time lag Δt, according to a power law $1/\Delta t^{\nu}$, with an exponent ν in the range 0.2–0.4. In order to use this long term information in a forecast in the best possible way, a similar long memory (LM) must be used.

 1469-7688/04/010070+17$30.00 PII: S1469-7688(04)61620-4

Another related topic is the modellization of financial time series by various processes. As the volatility memory of empirical data is the single most important feature (the returns show a zero correlation), the modellization tends to focus on the volatility processes. In this direction, the GARCH process and its extensions have been thoroughly explored (see e.g. Bollerslev *et al* 1992, Bera and Higgins 1993). In principle, any volatility process induces a forecast through the computation of conditional averages. In practice, the computations can be carried out only for processes with quadratic equations, like the standard GARCH(1, 1) process. Therefore, for a broad family of processes, the modellization of financial time series is closely related to volatility forecasts. The simplest GARCH(1, 1) and I-GARCH(1) processes are indeed very successful in describing financial time series. Yet, their major shortcoming is to have an exponential decay for the correlation (of the return square), instead of a power law (Ding and Granger 1996).

This deficiency can be cured in several ways. In the FIGARCH process (Baillie *et al* 1996), the LM is brought in by a fractional difference operator. The fractional difference is a mathematical operator taken from time series analysis, but it lacks an intuitive justification originating in the structure of the financial markets. In Ding and Granger (1996), the authors introduce a LM n-component process, which is structurally close to the processes presented in this paper. To avoid the proliferation of constants, they consider the large n limit with random coefficients drawn from a beta distribution. In this limit, they show that the process has the desired LM. The same authors in Granger and Ding (1996) consider other mechanisms to include in the process equations the observed LM. The LM can also be included in a stochastic volatility type process, as done by Breidt *et al* (1998).

In this paper, we introduce a new family of volatility processes that includes, *in a minimal way*, the power law memory of the volatility correlation. This is achieved by measuring the historical volatilities with a set of increasing time horizons, and computing the elementary time step volatility forecast with a simple combination of these historical volatilities. The set of historical volatilities can be seen as a measure of the perception of the market volatility by various participants, from intraday speculators to portfolio managers and pension funds. Yet the goal of the new processes is not to model each market component (which would introduce many choices in the equations, with the related parameters), but to include in a parsimonious way the LM observed in the data. For this reason, we call generically this new family of processes 'LM'. The emphasis is put on simplicity, with processes with only two and four parameters. If one historical volatility only is included, then the processes are equivalent to the GARCH(1, 1) or I-GARCH(1) processes. One goal of this paper is to explore the *structure* of the processes, namely to figure out what are the most important ingredients needed to build a good volatility process and forecast.

A related issue in volatility modelling and forecasting is whether the process equations (for the volatility square) are *linear* or *affine*. With linear process equations, such as I-GARCH, the volatility (square) is a linear combination of the historical volatility (square) and the return (square). For the I-GARCH process, directly rooted in the structure of the equations, the long term mean volatility is fixed by the initial conditions, and the variance of the volatility (square) diverges. A simple example of an affine process equation is that for the GARCH(1, 1) process, which includes in the equation for $\sigma^2(t)$ an additive constant term, usually denoted α_0 or ω. This constant, through a combination of the parameters, fixes the mean volatility (with some constraints on the parameters to ensure the existence of the mean). For the GARCH(1, 1) affine process equations, regardless of the initial conditions, the long term mean volatility converges to a constant related to the process parameters. For these two processes, the above key statistical properties originate in the linear or affine structure of the equations. These two models have been studied rigorously by Nelson (1990), who shows in particular that the probability density of the volatility for the (linear) I-GARCH process converges to a singular distribution with all the mass at zero. This difference between linear and affine equations for the I-GARCH and GARCH processes extends to the presented LM processes, and all the processes studied in this paper come in pairs, with linear and affine versions.

Because the difference between linear and affine processes is rooted in their mathematical structure, it is plausible that the main statistical properties of the LM processes are similar to the I-GARCH and GARCH processes respectively (with possibly some added constraints on the parameter values for the existence of the mean). Yet, proving those properties for the new processes is not elementary. Despite their mathematical difficulties, and possible singular asymptotic properties, linear processes are very interesting for forecasting purposes, precisely because they do not contain the mean volatility (and the corresponding 'coupling constant') among their parameters. This property is the basis for the broad success of the I-GARCH process used in RiskMetric, where a simple process, with one fixed parameter, provides for a good one day volatility forecast for all assets. Indeed, we show that linear LM processes, with only two parameters, provide for better forecasts than GARCH(1, 1). Beside, we have investigated by Monte Carlo simulations the asymptotic properties of the linear and affine LM processes, and argue that they are similar, respectively, to the I-GARCH and GARCH processes.

For the empirical statistical analysis presented in this paper, we used (deseasonalized) hourly data for the foreign exchange USD/CHF. We have used other currency pairs to check that the results are similar. The emphasis is to analyse the data or processes at different time horizons, from 1 h to 1 month, and several methods are used to investigate the various process properties and time horizons. Our goal with the empirical analysis is mainly qualitative, namely to find the *structure* for the process equations that provides for the best description of the data according to a range of criteria. In particular, we show that the linear LM processes, with only two parameters, provide for a parsimonious description of the empirical data up to time horizons comparable to the cut-off used in the process equations.

This paper is organized as follows. The next section sets the notation and common price process, essentially similar to the familiar GARCH set-up. The next four sections introduce the various processes studied in this paper, namely GARCH(1, 1) and I-GARCH(1) (section 3), the LM process LM-ARCH (section 4), FIGARCH (section 5), and I-GARCH(2) (section 6). The section on FIGARCH discusses the difficulties that arise when the fractional difference operator is cut off at a finite value, which should be done in all practical implementations. The relationship between price processes and volatility forecasts is developed in section 7. Then we turn to various empirical studies for the above processes, starting with a description of the data set (essentially, 11 years of hourly data for USD/CHF). This is followed by a study of the cut-off dependence in section 8.2, and of the probability distribution for the volatility in section 8.3. The lagged correlations and cost function landscapes (log-likelihood or forecast errors) are studied in sections 8.4 and 8.5. This section is followed by a comparative study of the log-likelihood and forecast error in sections 8.6 and 8.7. Finally, the robustness of the process parameters with respect to changing the data sample is studied in section 8.8, before the conclusions.

2. The basic structure of the processes

We are considering processes for the price with the following structure:

$$x(t + \delta t) = x(t) + r(t + \delta t) \tag{1}$$

$$r(t + \delta t) = \sigma_{\text{eff}}(t + \delta t)\varepsilon(t + \delta t) \tag{2}$$

$$\sigma^2_{\text{eff}}(t + \delta t) = \tilde{\sigma}^2[\Omega(t), \vartheta, \delta t]. \tag{3}$$

The random variables ε are i.i.d. with $E[\varepsilon(t)] = 0$ and $E[\varepsilon^2(t)] = 1$. This choice enforces that for the unconditional average, $\sigma^2 = E[r^2(t)] = E[\sigma^2_{\text{eff}}(t)]$, where σ is the mean volatility (we assume sufficient restrictions on the parameters so that the processes are covariant stationary). When needed, ε is chosen to be distributed with a Gaussian or a Student-t distribution. The time indices are chosen to emphasize that $\sigma^2_{\text{eff}}(t + \delta t)$ is a forecast for the volatility at time $t + \delta t$. The forecast function $\tilde{\sigma}$ is based on the information set $\Omega(t)$ at time t, and depends on a set of parameters ϑ. This function $\tilde{\sigma}$ depends only implicitly on the time t through the dependence on the information set $\Omega(t)$. The forecasting aspect, with the relation between σ_{eff} and the realized volatility, is explained in section 7 on volatility forecast. For the processes studied in this paper, the function $\tilde{\sigma}^2$ can be linear or affine (for the squares of the variable contained in Ω), leading to the two broad classes of processes.

For the affine process equations, taking the unconditional average of equations (2) and (3), one obtains an equation involving the process parameters

$$\sigma^2 = E[\tilde{\sigma}^2[\Omega(t), \vartheta, \delta t]]. \tag{4}$$

The expectation in the right-hand side can in principle be computed, leading to a function of the process parameters ϑ. By inverting this equation, one of the process parameters can be replaced by the mean volatility (possibly, more constraints must be added on the parameters for the existence of the second moment). As the processes discussed below have quadratic equations, this computation is analytically feasible, and the parametrization of the processes directly includes σ. For the linear process equations, the unconditional expectation of equation (3) leads to an identity, as for the I-GARCH process. In this case, the mean volatility is not fixed by the process parameters, but by the initial conditions.

The various elementary forecast functions $\tilde{\sigma}$ are given in the following sections, in order of increasing complexity. First, the GARCH(1, 1) process is reviewed, introducing a new form for the process equations. Then, the LM processes are introduced as a natural extension of GARCH(1, 1). In section 5, the FIGARCH process is reviewed, and the subtleties related to the fractional difference operator with a finite cut-off are clarified.

3. The GARCH(1, 1) process

The historical (Engle and Bollerslev 1986, Bollerslev 1986) and standard way to write the GARCH(1, 1) volatility equation is

$$\sigma^2_{\text{eff}}(t + \delta t) = \alpha_0 + \alpha_1 r^2(t) + \beta_1 \sigma^2_{\text{eff}}(t). \tag{5}$$

This equation can be rewritten as follows:

$$\sigma_1^2(t) = \mu\sigma_1^2(t - \delta t) + (1 - \mu)r^2(t) \tag{6}$$

$$\begin{aligned}\sigma^2_{\text{eff}}(t + \delta t) &= \sigma^2 + (1 - w_\infty)(\sigma_1^2(t) - \sigma^2)\\ &= (1 - w_\infty)\sigma_1^2(t) + w_\infty\sigma^2\end{aligned} \tag{7}$$

with the three parameters σ, w_∞, μ. The equivalence between the equations is easily proved by expressing equation (7) for σ_1, and then inserting into equation (6). By rearranging the terms, the volatility equation for a GARCH(1, 1) process is obtained, with the coefficient $\alpha_0 = \sigma^2(1 - \mu)w_\infty$, $\alpha_1 = (1 - w_\infty)(1 - \mu)$ and $\beta_1 = \mu$. In this form, the GARCH(1, 1) process appears with one 'internal variable' that can be interpreted as a historical volatility σ_1 measured by an exponential moving average (EMA) at the time horizon $\tau = -\delta t/\ln(\mu)$. The elementary forecast is given by the mean term, plus the difference between the historical volatility and the mean volatility, weighted by $1 - w_\infty$. In the second form, the elementary forecast is a convex combination of the mean volatility σ and the historical volatility σ_1. With these equations, the three parameters have a direct interpretation for the time series, in contrast to the standard α, β parametrization of equation (5).

The correlations for the GARCH(1, 1) process can be computed analytically (see e.g. Ding and Granger 1996); they decay exponentially fast with a characteristic time $\tau_{\text{corr}} = -\delta t/\ln(\mu_{\text{corr}})$ with $\mu_{\text{corr}} = \alpha_1 + \beta_1 = \mu + (1 - w_\infty)(1 - \mu)$. Financial data clearly display a power law decay for the lagged correlation, and this discrepancy is the main shortcoming of the GARCH(1, 1) process. The LM processes introduced in the next section are constructed as an extension of GARCH(1, 1) that remedies this deficiency.

For $w_\infty = 0$, the affine GARCH(1, 1) equations reduce themselves to the linear I-GARCH(1) process

$$\sigma_{\text{eff}}^2(t+\delta t) = \mu\sigma_{\text{eff}}^2(t) + (1-\mu)r^2(t) \tag{8}$$

with one parameter μ. As shown by Ding and Granger (1996), the lagged correlation for the square return of an I-GARCH(1) process also decays exponentially. The widely used RiskMetric formula corresponds to the I-GARCH(1) process, with a given parameter $\mu = 0.94$. In order to cover a wide range of assets across different classes, it is important to have a simple linear process, because the mean volatility does not enter into the parameters. This makes it possible to have a one parameter process that can accommodate all securities. As the lagged correlation for the volatility (or return square) shows a power law with similar exponent for most assets, we can expect to construct a process that incorporates the power law memory, and that describes well the empirical data for most assets with similar parameters. All the processes presented below come in pairs: an affine process and a linear process with two parameters less (the mean volatility σ and the corresponding 'coupling constant' w_∞).

4. The LM volatility processes

In order to build processes with LM, we use a set of (historical) volatilities computed over increasing time horizons. Each historical volatility is measured by an exponential moving average (EMA):

$$\begin{aligned}\tau_k &= l_k\tau_0 \qquad k = 1,\ldots,n\\ \mu_k &= \mu_0^{1/l_k} = \exp(-\delta t/\tau_k)\\ \sigma_k^2(t) &= \mu_k\sigma_k^2(t-\delta t) + (1-\mu_k)r^2(t).\end{aligned} \tag{9}$$

The time horizons τ_k correspond to the characteristic times of the EMA at which the historical volatility is measured. The usual parameters μ_k entering the EMA are related to the EMA timescales τ_k with the second equation. We prefer to work with the time horizons τ_k that have a direct interpretation, whereas the value of the parameters μ_k have none.

We choose the time structure l_k of the process as a geometric series $l_k = \rho^{k-1}$, with the progression of the series chosen to be $\rho = 2$ in this work. The base timescale τ_0 corresponds to the shortest timescale at which a volatility is measured, and is taken as one of the process parameters. The upper cut-off for the measured volatility is at the timescale $\tau_0\rho^n$, namely it increases exponentially with the number of components. This exponential dependence is crucial for applications to high frequency data, because the ratio between the data timescale δt and the timescale of interest can be of several orders of magnitude. The empirical studies in this work have been done with hourly data and mostly with $n = 12$, corresponding to an upper cut-off of 6 months.

The values for the parameters ρ and n have been chosen such that the processes are essentially independent of them, i.e. to be practically close to a continuum of volatility $\rho \to 1$ with an infinite cut-off $n \to \infty$. In fact, the actors present in the financial markets very likely provide a natural cut-off, namely long term participants like pension funds and central banks set the upper limit to a horizon of a few months. Yet, a precise empirical study of this issue is difficult because of the lack of *stationary* empirical financial time series over a length of several decades. Long term studies with daily data can be found in Ding and Granger (1996) and Granger and Ding (1996), but several mechanisms can be invoked to produce the observed long term dependence, as emphasized by these authors.

Empirically, the lagged correlation of the absolute value of the return, or its square, shows a power law decay, from a few hours to a few months (for example, see section 8.4), and this is the salient feature we want to incorporate in a parsimonious way in a volatility process. In this sense, we are not interested in the ultimate long time properties of the lagged correlation, which can be an exponential or a power law decay. In fact, all the processes studied in this paper, including FIGARCH, have a finite cut-off in their actual implementations. Beyond the cut-off, the coefficients of the ARCH(∞) representation have an exponential decay (or are zero for FIGARCH). Therefore, the exact asymptotic results obtained by Davidson (2002) and Zaffaroni (2000) that should be applied to the present models are those relative to the geometric decaying coefficients (i.e. exponential). This is also important with respect to the exact results obtained by Giratis *et al* (2000), which show essentially that to define a stationary process the (asymptotic) memory must decay fast enough.

From a quantitative point of view, at the time horizons of interest, several other issues could be more important than the ultimate asymptotic properties of the processes. A first example is the issue between microscopic or aggregated processes studied in this paper. Another example is the exponent used to measure the volatility. In this paper, we have restricted ourself for analytical reasons to squares, but smaller exponents, like absolute value, might give a better description of the empirical data. For these reasons, we have chosen the cut-off parameter n large enough to be neglected in this study.

In equation (9), only the return $r = r[\delta t]$ at the smallest time step is used to compute the set of volatilities, and hence the name *microscopic* (Mic) for the related processes. Yet the volatility at the time horizon τ_k can be measured with various formulae: in particular, the return time horizon δt_r can be dependent on τ_k. In general, an empirical volatility depends on two independent parameters, namely the time horizon of the return δt_r and the length δt_σ of the time window on which the returns are averaged. From a modellization perspective, it is reasonable to assume that market participants working at a short time horizon (intraday) use tick-by-tick data, whereas market participants active at long time horizons use daily data, or even economic statistics available at a lower frequency, such as unemployment rate. In order to keep the process equations simple and the number of parameters small, we define an aggregate volatility for which $\delta t_\sigma/\delta t_r = \tau_0$. This choice corresponds roughly to the decreasing frequency of the data used when increasing the characteristic time horizons. This leads to an *aggregated* (Agg) definition of the volatility:

$$\begin{aligned} r[l_k\delta t](t) &= \frac{x(t)-x(t-l_k\delta t)}{\sqrt{l_k}} \\ \sigma_k^2(t) &= \mu_k\sigma_k^2(t-\delta t)+(1-\mu_k)r[l_k\delta t]^2(t) \\ & i=1,\ldots,n. \end{aligned} \tag{10}$$

The return at the timescale $l_k\delta t$ is the usual price difference, scaled from the time horizon $l_k\delta t$ to the time horizon δt using a random walk hypothesis. In this way, all returns $r[l_k\delta t]$ and volatilities σ_k are related to the same fundamental scale δt (another choice is to annualize all the volatilities, but this makes the equations slightly more complicated).

Now the effective volatility σ_{eff} for the LM linear (Lin) processes can be introduced:

$$\begin{aligned} \sigma_{\text{eff}}^2(t+\delta t) &= \sum_{k=1}^{n} w_k\sigma_k^2(t) \\ w_k &= c\rho^{-(k-1)\lambda} = c\left(\frac{1}{l_k}\right)^{\lambda} \qquad \text{with } 1/c=\sum_{k=1}^{n}\rho^{-(k-1)\lambda}. \end{aligned} \tag{11}$$

Depending on the definition of the historical volatility σ_k that is used, the processes LM-Mic-Lin-ARCH(n) (long memory-microscopic-linear-ARCH) or LM-Agg-Lin-ARCH(n) (long memory-aggregated-linear-ARCH) are defined. The weights w_k are given by a simple power law, with the proportionality constant c such that $\sum w_k = 1$. As shown in the next section, the power law decay of the weight w_k, combined with the geometric progression for the time horizon τ_k, induces a power law decay for the memory in an ARCH(∞) representation. This leads to a power law decay for the lagged correlation (of the return square), as checked with Monte Carlo simulations in section 8.4. These two processes have two parameters: a characteristic time τ_0 for the EMA (or equivalently μ_0), and the exponent λ for the decay of the weight. If the weight function w_k decays sufficiently fast and the number of components n is taken large enough, the processes become insensitive to the upper cut-off (i.e. practically $n\to\infty$).

Both above processes are linear, and the mean volatility is fixed by the initial conditions. The modifications needed to introduce an affine term fixing the mean volatility σ are straightforward:

$$\begin{aligned} \sigma_{\text{eff}}^2(t+\delta t) &= \sigma^2+(1-w_\infty)\sum_{k=1}^{n}\chi_k(\sigma_k^2(t)-\sigma^2) \\ &= \sum_{k=1}^{n} w_k\sigma_k^2(t)+w_\infty\sigma^2 \\ \chi_k &= c\rho^{-k\lambda} = c\left(\frac{1}{l_k}\right)^{\lambda} \qquad \text{with } 1/c=\sum_{k=1}^{n}\rho^{-(k-1)\lambda} \\ w_k &= (1-w_\infty)\chi_k. \end{aligned} \tag{12}$$

The 'normalization constant' c is chosen such that $\sum \chi_k = 1$ and $\sum w_k + w_\infty = 1$. Depending on the definition of σ_k (equation (9) or (10)) that is used, the LM microscopic LM-Mic-Aff-ARCH(n) or aggregated LM-Agg-Aff-ARCH(n) affine processes are obtained. Notice that this parametrization implements naturally the separation between magnitude (given by $1-w_\infty$) and memory (given by λ) as advocated by Davidson (2002). Using the above equations, it is easy to check that the unconditional mean volatility square is fixed by σ^2:

$$E[r[\delta t]^2] = E[\sigma_k^2] = E[\sigma_{\text{eff}}^2] = \sigma^2 \tag{13}$$

(we assume covariance stationarity). In equation (12), the mean volatility σ acts similarly to a volatility at an infinite timescale that anchors the mean to σ with a strength given by w_∞. This shows that when the number of components is large, if the parameter w_∞ is small enough, the linear or affine processes are similar and they should deliver a similar forecast for time horizons up to the largest τ_n. Both processes LM-Mic-Aff-ARCH(n) and LM-Agg-Aff-ARCH(n) have four parameters, namely σ, w_∞, τ_0 and λ.

Clearly, for $n=1$ both processes LM-Mic-Aff-ARCH(n) and LM-Agg-Aff-ARCH(n) are identical to the GARCH(1, 1) model. In this sense, the newly defined processes are an extension of the GARCH(1, 1) process, with one more parameter characterizing the decay of the lagged correlation. Similarly, the LM-Mic-Lin-ARCH(1) and LM-Agg-Lin-ARCH(1) reduce to the I-GARCH process for $n=1$.

4.1. ARCH(∞) representation for long memory microscopic processes

It is easy to derive an ARCH(∞) representation for the LM-Mic-*-ARCH processes. For the microscopic process, the EMA equations used to define σ_k can be iterated to give

$$\sigma_k^2(t+\delta t) = (1-\mu_k)\sum_{j=0}^{\infty}\mu_k^j r^2(t-j\delta t). \tag{14}$$

As $|\mu_k|<1$ with the parametrization given in equation (9), the ARCH(∞) representation is valid. This expression can be inserted in the equation for σ_{eff} to give a AR(∞) representation (for r^2) of the LM-Mic-Lin-ARCH process:

$$\begin{aligned} \sigma_{\text{eff}}^2(t+\delta t) &= \sum_{j=0}^{\infty}\alpha(j\delta t)r^2(t-j\delta t) \\ \alpha(j\delta t) &= \sum_{k=1}^{n} w_k(1-\mu_k)\mu_k^j. \end{aligned} \tag{15}$$

A similar equation is obtained for the LM-Mic-Aff-ARCH process. With our choice for the functional dependences of w_k and μ_k, the coefficient $\alpha(j\delta t)$ can be estimated for j large enough. Using equation (9), and shifting the summation index k by -1, we obtain

$$\alpha(j\delta t) = \frac{\sum_{k=0}^{n-1}(1/\rho^k)^\lambda\big(1-\exp\big(-(\delta t/\tau_0)\rho^{-k}\big)\big)\exp\big(-(j\delta t/\tau_0)\rho^{-k}\big)}{\sum_{k=0}^{n-1}\big(1/\rho^k\big)^\lambda}. \tag{16}$$

When the progression of the geometric series ρ is close to 1, and with a large enough cut-off n, the sums can be replaced by integrals:

$$\alpha(j\delta t) = \left\{\int_0^\infty \rho^{-k\lambda} \exp\left(-\frac{j\delta t}{\tau_0}\rho^{-k}\right) \mathrm{d}k - \int_0^\infty \rho^{-k\lambda} \exp\left(-\frac{(j+1)\delta t}{\tau_0}\rho^{-k}\right) \mathrm{d}k\right\} \times \left\{\int_0^\infty \rho^{-k\lambda}\,\mathrm{d}k\right\}^{-1}. \tag{17}$$

The integral in the denominator can be evaluated with the change of variable $x = \rho^{-k}$, and is independent of $j\delta t$. The first integrals in the numerator can be evaluated with the change of variable $x = j\delta t \rho^{-k}/\tau_0$

$$\left(\frac{\tau_0}{j\delta t}\right)^\lambda \int_0^{j\delta t/\tau_0} x^{\lambda-1} \exp(-x)\,\mathrm{d}x. \tag{18}$$

The integrand is maximum at $x = \lambda - 1$, and then decays exponentially fast. When the upper limit $j\delta t/\tau_0$ of the integral is large enough compared to $\lambda - 1$, the integral can be approximated by a constant with value $\Gamma(\lambda)$. The second integral in the numerator is evaluated with the same technique. By combining all the terms, we obtain (up to multiplicative constants)

$$\alpha(j\delta t) \simeq \left(\frac{\tau_0}{j\delta t}\right)^\lambda - \left(\frac{\tau_0}{(j+1)\delta t}\right)^\lambda \simeq \left(\frac{\tau_0}{j\delta t}\right)^{\lambda+1}. \tag{19}$$

Therefore, the autoregressive coefficients of the process decay as a power law with an exponent $\lambda+1$. The above computation shows that these processes are built to have a power law memory, up to the upper cut-off given by τ_n (see the discussion after equation (9)). This computation also allows us to make contact with the FIGARCH process (see the next section), as the coefficients $\alpha(j\delta t)$ play a similar role to the coefficients of the expansion of the fractional difference operator. This shows that for both classes of processes, the ARCH(∞) coefficients have the same asymptotic decay, with $\lambda = d$.

Several extensions can be made to the LM processes by introducing a more flexible parametrization. One extension is to choose each coefficient $w_k \in [0, 1]$ as a free parameter (instead of taking them according to an analytical function with a few parameters). This would introduce one parameter per volatility component, therefore increasing the difficulty for the estimation of the parameters and the risk of overfitting. The HARCH process introduced in Müller (1995) and Dacorogna *et al* (1998) is of this type (the lag structure is chosen slightly differently and with a progression factor of 4, in order to reduce the overfitting). Instead, we prefer to restrict ourselves to a simple analytical form for the time horizons and for the coefficients. Another possibility is to choose the characteristic times τ_k of the volatility components σ_k according to the structure of the market, namely to capture the behaviour of the intraday market makers, daily traders, portfolio managers or long term pension funds. This direction is followed in Zumbach and Lynch (2001) and Lynch and Zumbach (2003), where information about the structure of the market is obtained by measuring the response function for the change of volatility. Through this response function, quantitative measures of the actual timescales and magnitudes of each component are obtained, quantities that can be used to estimate the process. Although this leads to a better description of the data, the process and its estimation are significantly more complex. In this paper, we have restricted ourselves to the simplest processes that describe the LM observed in the empirical data.

5. FIGARCH (p, d, m)

The fractionally integrated GARCH process has been introduced by Baillie *et al* (1996) in order to include a LM in the volatility. The idea of the derivation is to note that, often, in empirical fits of GARCH processes, the estimated coefficients are such that $\sum_{i=1} \alpha_i + \beta_i \simeq 1$ (see however Zumbach (2000) for potential pitfalls in such diagnositics). This may indicate that the GARCH process is mis-specified, and an integrated process should be used instead.

The practical implementation of the FIGARCH process presents an important difficulty related to the cut-off of the fractional difference operator, as already pointed out by Teyssière (1996) and Chung (1999). As these references are not readily available, a simple derivation of the FIGARCH process is presented below, with the cut-off problem explained. This derivation allows us to fix the notation, and to give an unambiguous name to the two resulting fractional processes.

Consider a GARCH(p, q) process, with the volatility equation

$$\begin{aligned}\sigma^2_{\mathrm{eff}}(t) &= \omega + \alpha(L)r^2(t) + \beta(L)\sigma^2_{\mathrm{eff}}(t)\\ &= \omega + \beta(L)\sigma^2_{\mathrm{eff}}(t) + (1 - \beta(L) + \alpha(L) + \beta(L) - 1)r^2(t).\end{aligned} \tag{20}$$

The terms $\alpha(L)$ and $\beta(L)$ are polynomials of degree q and p, respectively, of the lag operator L, with the property $\alpha(0) = \beta(0) = 0$. In this section, to ease the notation and make contact with the literature, the time indices have been translated backward by δt compared to equation (1). If $\alpha(L) + \beta(L)$ has a unit root, we can use the substitution

$$1 - \alpha(L) - \beta(L) \to \phi(L)(1 - L)^d \tag{21}$$

with $\phi(L)$ a polynomial of degree $m = \max(p, q) - 1$, and $\phi(0) = 1$. A term $1 - L$ would introduce a unit root, yet the volatility process is clearly stationary. Therefore, the exponent d tames down the unit root, and introduces the LM. The new term $(1 - L)^d$ is a fractional difference operator, defined through a Taylor expansion in the lag operator

$$\begin{aligned}\delta_d(L) &= (1 - L)^d = \sum_{j=0} \delta_{d,j} L^j = 1 - dL + \cdots\\ \delta_{d,j} &= \frac{(-1)^j \Gamma(d+1)}{\Gamma(j+1)\Gamma(d-j+1)}.\end{aligned} \tag{22}$$

The fractional difference operator obeys the important identity $\delta_d(1) = 0$ which will play an important role below. The coefficient of the Taylor expansion can be computed easily with the recursion

$$\delta_{d,j+1} = \frac{j-d}{j+1}\delta_{d,j}$$
$$\delta_{d,0} = 1. \tag{23}$$

For large j, the coefficients $\delta_{d,j}$ decay as $j^{-(d+1)}$. This power law decay of the operator coefficients induces the LM of the process. Using the above substitution, the volatility process becomes

$$\sigma_{\text{eff}}^2(t) = \omega + \beta(L)\sigma_{\text{eff}}^2(t) + (1 - \beta(L) - \phi(L)\delta_d(L))r^2(t). \tag{24}$$

By taking the average of this equation and using $E[r^2] = E[\sigma_{\text{eff}}^2]$ and $\delta_d(1) = 0$, we obtain $\omega = 0$ and the mean volatility is not determined (we assume, throughout the paper, that the parameters are such that the processes are covariant stationary). Therefore, the FIGARCH(p, d, m) volatility process is given by the linear equation

$$\sigma_{\text{eff}}^2(t) = \beta(L)\sigma_{\text{eff}}^2(t) + (1 - \beta(L) - \phi(L)\delta_d(L))r^2(t). \tag{25}$$

Moreover, the positivity of the left-hand side imposes that all the coefficients in the expansion of $1 - \beta(L) - \phi(L)\delta_d(L)$ are positive, giving further restriction on the coefficients of β and ϕ.

The practical implementation of this equation is quite subtle, because the expansion of the fractional difference operator has to be cut off at some upper limit $j_{\max}$. The cut-off of the infinite sum implies that the identity $\delta_d(1) = 0$ is no longer true. Let us emphasize that the violation of this identity can be important for realistic values of the parameters: for example taking $d = 0.25$ and $j_{\max} = 1000$, we obtain $\delta_d(1) = 0.145$. For the differential operator *with a finite cut-off*, taking again the average of equation (24) and using $E[r^2] = E[\sigma_{\text{eff}}^2]$, we obtain that ω fixes the mean volatility:

$$E[\sigma_{\text{eff}}^2] = \frac{\omega}{\phi(1)\delta_d(1)}. \tag{26}$$

Therefore, we define the affine FIGARCH process, or Aff-FIGARCH(p, d, m), by the equation

$$\sigma_{\text{eff}}^2(t) = \sigma^2\phi(1)\delta_d(1) + \beta(L)\sigma_{\text{eff}}^2(t) + (1 - \beta(L) - \phi(L)\delta_d(L))r^2(t) \tag{27}$$

and the affine parameter σ fixes the mean volatility. On the other hand, in order to define a linear process Lin-FIGARCH from the FIGARCH equation (25), with a finite cut-off in the fractional difference operator, the identity $\delta_d(1) = 0$ has to be enforced. This is done by defining

$$\tilde{\delta}_d(L) = 1 + \gamma(d)\sum_{j=1}^{j_{\max}} \delta_{d,j}L^j \tag{28}$$

and adjusting the constant $\gamma(d)$ so that $\tilde{\delta}_d(1) = 0$. Notice that the process Lin-FIGARCH has only one parameter less than Aff-FIGARCH, namely the mean volatility, and not two as with the other processes. This occurs because the weight of the mean volatility is controlled by $\delta_d(1)$ that is not taken as a free parameter, but given implicitly by $j_{\max}$ which is fixed *a priori*. Finally, let us add that the process that is discussed in the original literature (Baillie *et al* 1996) is the Aff-FIGARCH process.

In the empirical section, we have considered only the Aff-FIGARCH$(1, d, 0)$ and Lin-FIGARCH$(1, d, 0)$ processes, derived from GARCH$(1, 1)$. These processes have respectively three parameters (σ, β, d) and two parameters (β, d). Moreover, the positivity of the volatility imposes the further restriction $d > \beta$.

Finally, notice that the amount of computation required to evaluate one iteration of the FIGARCH process is dominated by the fractional difference, and the computational time grows linearly with the cut-off of this operator. This behaviour contrasts sharply with the above LM processes whose computational time grows as the logarithm of the cut-off. For the computations performed for this paper, the evaluation of the FIGARCH process requires typically 10–100 times more resources than for the LM processes (the same scaling is true with respect to the memory requirement, but today this is not an issue). The good scaling with respect to the upper cut-off of the LM processes is an important advantage compared to FIGARCH, particularly when dealing with high frequency data.

6. The I-GARCH(2) process

When written in the form (7), the GARCH$(1, 1)$ equation calls for the natural extension

$$\begin{aligned}\sigma_1^2(t) &= \mu_1\sigma_1^2(t - \delta t) + (1 - \mu_1)r^2(t)\\ \sigma_2^2(t) &= \mu_2\sigma_2^2(t - \delta t) + (1 - \mu_2)r^2(t)\\ \sigma_{\text{eff}}^2(t + \delta t) &= w\sigma_2(t)^2 + (1 - w)\sigma_1(t)^2(t).\end{aligned} \tag{29}$$

This process is linear, with two components. As the equations are very similar to those for I-GARCH(1), we name it I-GARCH(2). It has three parameters: τ_1, τ_2 and w (or μ_1, μ_2, w). This process can be seen as a natural extension of I-GARCH(1). Because of its simplicity, this process is included in the systematic empirical study. Notice that Ding and Granger (1996) introduced a similar two component process (with an affine term).

7. Volatility forecast

A volatility process can be used to compute a forecast for the volatility. More precisely, at time t, given the information set $\Omega(t)$, the forecast $\mathcal{F}[\sigma_{\text{eff}}^2]$ for the time $t + \Delta t$ is given by the conditional average

$$\mathcal{F}[\Delta t; \sigma_{\text{eff}}^2](t) = E[\sigma_{\text{eff}}^2(t + \Delta t)|\Omega(t)] \tag{30}$$

with $\Delta t = k\delta t$. Then, using the process equations iteratively, the right-hand side can be expressed as a function of the variables contained in the information set $\Omega(t)$. Note that this is possible by simple analytical means because the above

Table 1. Salient features of the log-likelihood estimate and forecast error estimate.

	Log-likelihood	Forecast error
Hypothesis	Distributional assumption on the residues	Choice of the realized volatility and distance functional
Sensitivity to the different timescales	Mainly at the smallest timescale δt	Mainly at all timescales up to Δt
Can be evaluated for	Any processes	Quadratic process equations and forecast models

equations are quadratic in the returns. General non-linear processes can be easily written, for example by replacing the square by a power p, but this analytical tractability is then lost. The mean forecast volatility between t and $t + \Delta t$ is given by

$$\bar{\mathcal{F}}[\Delta t; \sigma^2_{\text{eff}}](t) = \frac{1}{m} \sum_{j=1}^{m} \mathcal{F}[j\delta t; \sigma^2_{\text{eff}}](t) \tag{31}$$

with $\Delta t = m\delta t$. The forecast of mean volatility is very interesting for process estimations: on historical data, the realized volatility can be computed and the parameters of the process be optimized to give the best forecast for the realized volatility. The realized volatility can be defined for example as

$$\sigma^2_{\text{real}}[\Delta t](t) = \frac{1}{m} \sum_{j=1}^{m} r^2[\delta t](t + j\delta t). \tag{32}$$

This is indeed an interesting quantity to forecast as it is relevant for portfolio management, option pricing and risk assessment. A simple L^2 distance $d[\Delta t, \theta]$ gives a convenient functional to measure the forecast error:

$$d^2[\Delta t, \theta]\left(\sqrt{\bar{\mathcal{F}}[\sigma^2_{\text{eff}}]}, \sigma_{\text{real}}\right) = \sum_t \left(\sqrt{\bar{\mathcal{F}}[\Delta t, \sigma^2_{\text{eff}}](t)} - \sigma_{\text{real}}[\Delta t](t)\right)^2. \tag{33}$$

This quantity is usually called the RMSE (root mean square error). The forecast depends on the process parameter θ, and therefore the distance functional $d[\theta]$ can be minimized to produce the best forecast. This procedure is an alternative to the usual estimate of process parameters by the maximization of a log-likelihood functional.

If the emphasis is on producing a volatility forecast, then estimating the parameters by minimizing the forecast error is appropriate, whereas if the emphasis is on the data generating process, a log-likelihood estimate is better suited. The salient features of the log-likelihood estimate and forecast error estimate are summarized in table 1.

In particular, when optimizing on the forecast error, no distributional assumption on the residues need be made, but the explicit form of the realized volatility and of the forecast error (i.e. the distance functional between the realized and forecast volatility) needs to be chosen. For this paper, we restrict ourselves to a quadratic realized volatility and to the above L^2 (or RMSE) distance.

For the LM-Mic-Aff-ARCH(n) and LM-Mic-Lin-ARCH(n) processes, the iterative equations for the conditional average take a very simple form that is derived in this section. Using the process equations for the LM-Mic-Aff-ARCH process, we obtain

$$E[\sigma^2_k(t + j\delta t)|\Omega(t)] = \mu_k E[\sigma^2_k(t + (j-1)\delta t)|\Omega(t)] + (1 - \mu_k) E[\sigma^2_{\text{eff}}(t + j\delta t)|\Omega(t)] \tag{34}$$

$$E[\sigma^2_{\text{eff}}(t + (j+1)\delta t)|\Omega(t)] = \sigma^2 + \sum_k w_k \{E[\sigma^2_k(t + j\delta t)|\Omega(t)] - \sigma^2\}. \tag{35}$$

Introducing the new variables

$$\begin{aligned} \delta_k(j) &= E[\sigma^2_k(t + j\delta t))|\Omega(t)] - \sigma^2 \\ \gamma(j) &= E[\sigma^2_{\text{eff}}(t + j\delta t)|\Omega(t)] - \sigma^2, \end{aligned} \tag{36}$$

the conditional average equations are reduced to

$$\begin{aligned} \delta_k(j) &= \mu_k \delta_k(j-1) + (1 - \mu_k)\gamma(j) \\ \gamma(j+1) &= \vec{w}' \cdot \vec{\delta}(j) \end{aligned} \tag{37}$$

where $\vec{w}$ is the vector of weights w_k, and similarly for $\vec{\delta}$ and $\vec{\mu}$. For the LM-Mic-Lin-ARCH(n), the computations are identical except for the term σ^2 which is absent. We can introduce the diagonal matrix $M_{k,p} = \delta_{k,p}\,\mu_k$, with $\delta_{k,p}$ the Kronecker symbol ($\delta_{k,p} = 1$ if $k = p$, zero otherwise). Equations (37) can be combined:

$$\vec{\delta}(j) = \{M + (\vec{1} - \vec{\mu})\vec{w}'\}\vec{\delta}(j-1) \tag{38}$$

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

$$\vec{\delta}(j) = \{M + (\vec{1} - \vec{\mu})\vec{w}'\}^j \vec{\delta}(0) \tag{39}$$

and $\vec{\delta}(0)$ is in the information set. This expression relates $\mathcal{F}[\sigma^2_{\text{eff}}]$ linearly to the $\sigma^2_k(t)$. For γ, equations (37) can be expressed as

$$\gamma(j+1) = \vec{w}'(j) \cdot \vec{\delta}(0) \tag{40}$$

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

$$\begin{aligned} w'(j+1) &= w'(j)\{M + (\vec{1} - \vec{\mu})\vec{w}'(0)\} \\ \vec{w}(0) &= \vec{w}. \end{aligned} \tag{41}$$

Therefore, the coefficients $\vec{w}(j)$ can be evaluated *a priori*, and the forecast for the effective volatility computed by a simple scalar product. This computation is interesting for two purposes. First, it allows us to evaluate very efficiently the forecasts for the processes LM-Mic-Aff-ARCH(n) and LM-Mic-Lin-ARCH(n). Secondly, the effective weights $\vec{w}(j)$ can be used in non-quadratic forecasting models, as a parsimonious

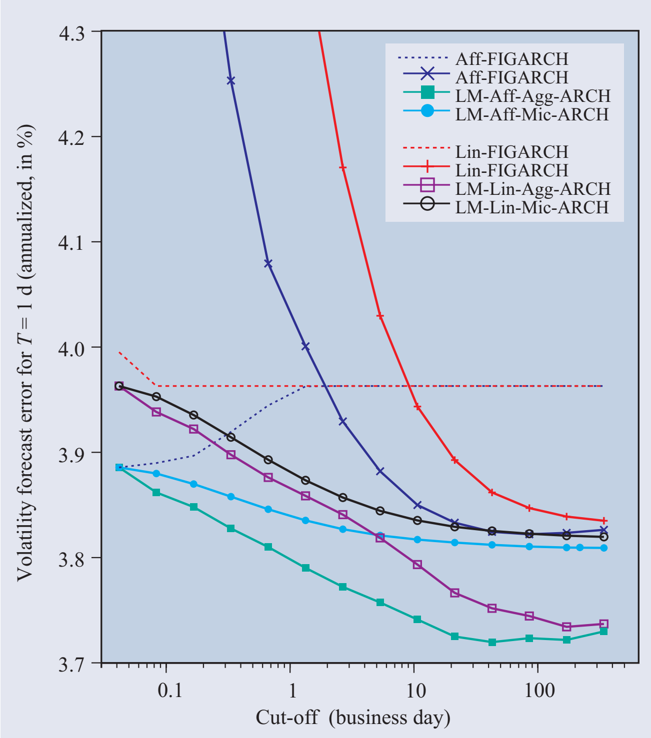

Figure 1. The one day forecast error as a function of the cut-off time.

parametrization of the weightings that are used for each component.

For the LM-Agg-Aff-ARCH(n) and LM-Agg-Lin-ARCH(n) processes, the forecast cannot be expressed in a simple scalar product. This happens because the returns at different time horizons enter into the equations, and therefore the forecast depends on the recent price history. Yet the evaluation of the conditional average for the volatilities is fairly straightforward when computed iteratively, and the numerical implementation of the equations, even though cumbersome, possesses no particular difficulty.

8. Empirical results

8.1. The data set

For all the empirical results presented in this section, the data set is a regular time series for USD/CHF, sampled hourly in business time. More precisely, we use the recent dynamic business timescale as developed in Breymann *et al* (2000). Similarly to the familiar daily business timescale, this timescale contracts periods of low activity (night, week-end) and expands periods of high activity. The activity pattern during the week is related to the measured volatility, averaged on a moving sample of six months. Holidays and daylight saving time are taken into account. Returning to the data sample used in this work, the high frequency tick-by-tick USD/CHF data are sampled using a linear interpolation, every 1 h 24 min on a dynamical timescale. A very short EMA is taken before sampling in order to attenuate the tick-by-tick incoherent price formation noise (Corsi *et al* 2001). The sampling time interval of 1 h 24 min corresponds to 7/5 of 1 h, and is such that, on average, 120 points per week are taken. Essentially, during the business week from Monday to Friday, it corresponds to one point per hour, and no sampling points are taken during the week-end. Therefore, we will use the (imprecise) wording 'hourly' data when referring to this data set. The data set is computed from 1 January 1989 to 1 July 2000. The year 1989 is used for the build-up of the processes (i.e. the data for 1989 are input into the volatility processes so that they build their internal states and forget the initial conditions, but the cost functions do not include these data). The following 10.5 years of data are used for the various studies. A similar data set has been computed for USD/JPY, and is used to check that the empirical results are 'generic'.

8.2. Dependence with respect to the cut-off

The LM and FIGARCH processes have an upper cut-off for the memory induced respectively by the finite number of components and the length of the Taylor expansion for the fractional difference. For this first study, we would like to choose the cut-off so that the process is essentially independent of it. Figure 1 shows the forecast error for a 1 day forecast for the LM-*-ARCH and *-FIGARCH processes (where * is a string of characters) as a function of the cut-off time interval. For the LM processes, the cut-off is given by the longest volatility time horizon $\delta t_{\text{cut-off}} = \delta t \rho^n$, and for the *-FIGARCH processes, the cut-off is given by the truncation of the fractional derivative $\delta t_{\text{cut-off}} = \delta t j_{\max}$. The two 'accumulation' points on the left, with a cut-off at 1 h and with forecast errors of 3.96 and 3.89, correspond to the linear and affine processes with one component for the memory, namely to the GARCH and I-GARCH processes respectively. Clearly, the LM-*-ARCH processes improve regularly when increasing the cut-off, with a saturation at roughly three months. Moreover, when the cut-off is large enough, the linear processes become better than (affine) GARCH(1, 1). The picture for the fractional GARCH processes is more intricate, because two solutions can exist. The 'lower' one corresponds to a parameter d in the range 0.1–0.4, and it is the 'good' solution. A 'high' solution also exists, with $d \simeq 1$. From the figure, the cut-off has to be high enough so that the 'good' solution exists and corresponds to the minimum of the forecast error. Moreover, the 'good' solution exists only if the cut-off is large enough, and it bifurcates from the side of the parameter's domain with a high value for the forecast error. In other words, for the FIGARCH process, the cut-off *must* be taken large enough. This bad behaviour with respect to the cut-off, combined with the linear scaling in computational effort, is an important drawback of the FIGARCH processes. The above pattern is generic, and a similar figure is obtained for other forecast horizons, for a log-likelihood cost function, or for other currency pairs. The only significant difference is that for longer forecast horizons, the cut-off at which the saturation occurs for the LM processes increases, as can be expected.

8.3. Probability distribution for the volatility

In this section, we investigate empirically the probability distribution for the daily volatility, as obtained for various

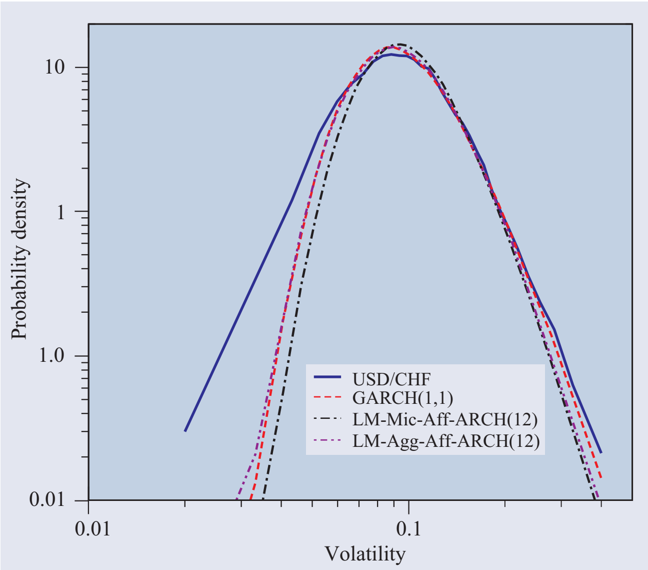

Figure 2. The volatility probability density function for some affine processes, in log–log scales. The mean volatility for all processes is the same as for the empirical data: $\sigma = 11\%$. The parameters are for GARCH(1, 1): $\alpha_1 = 0.0267$, $\beta_1 = 0.966$ or $w_\infty = 0.218$, $\tau_1 = 1.2$ day; for LM-Mic-Aff-ARCH(12) and LM-Agg-Aff-ARCH(12): $w_\infty = 0.123$, $\tau = 0.37$ day, $\lambda = 0.3$. The residues ε are drawn from a Student distribution with 4.5 degrees of freedom.

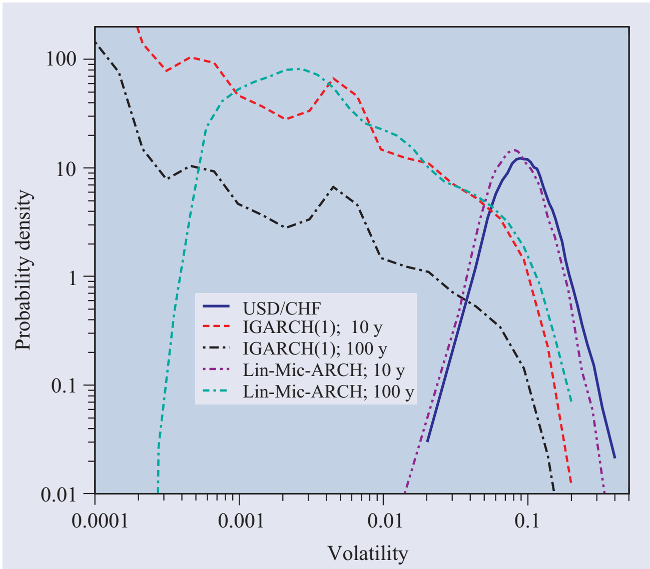

Figure 3. The volatility probability density function for some linear processes, in log–log scales. The initial value for the volatility for all processes is the same as for the empirical data: $\sigma = 11\%$. The parameters are for I-GARCH(1): $\tau_1 = 5.6$ day; for LM-Mic-Lin-ARCH(12): $\tau = 0.37$ day, $\lambda = 0.3$.

processes and for the foreign exchange rate USD/CHF. For the processes, the PDF is computed by a Monte Carlo simulation, with a time increment of $\delta t = 1$ (business) hour. The daily volatility is computed with

$$\sigma^2(t) = \sum_{t-1d<t'\leqslant t} r^2[\delta t](t'). \tag{42}$$

The simulation length is over 100 years (or over 10 years when specified for some linear processes). For the affine processes, the resulting probability density is drawn in figure 2. The parameters for the processes correspond to the values obtained by log-likelihood estimates. Clearly the agreement between the PDF for the empirical data and all the processes is excellent. The main disagreement occurs for low values of the volatility. This may be a genuine effect, or could be explained by external causes, such as an imperfect deseasonalization for holidays.

The results for the linear I-GARCH and LM-Mic-Lin-ARCH processes are shown in figure 3. For both processes, the simulation is done for lengths of 10 and 100 years. The figure shows clearly that the PDF is *not* stationary for the linear processes. Indeed, it is very likely that for all these processes, the PDF converges to a singular distribution with all the mass at zero (see Nelson (1990), Davidson (2002) for related processes where this can be proved). Yet, for the LM processes, if the simulation length is smaller than the cut-off time horizon, the empirical PDF is very close to the empirical one. Therefore, for 'short' Monte Carlo simulations not exceeding a time horizon of a year, for example in stress testing or scenario analysis, the LM linear processes can very well be used, despite their singular asymptotic properties.

8.4. Lagged correlation

Empirically, it is well documented that the lagged correlation for any measure of volatility decays as a power law with an exponent in the range 0.1–0.4. The measure of volatility can be $|r|$, r^2 or a definition similar to equation (32). The GARCH(1, 1) process has a lagged correlation function that decays as an exponential, and this is a clear shortcoming of this process. Particularly for forecasting, its memory decays too fast. Yet, as locally a power law can be approximated quite well by an exponential, the forecasting performances of GARCH(1, 1) are nevertheless good. The LM*ARCH processes have been constructed to have power law correlations (for large lag). This is checked using Monte Carlo simulations for the LM-Mic-Aff-ARCH process using a value $\lambda = 0.3$. The simulated lagged correlation is also compared to USD/CHF empirical data in figure 4, and with a simple power law with exponent 0.4. Clearly, the agreement is excellent.

However, the situation is more subtle in the intraday range. Figure 5 shows the same data, but using a logarithmic axis for the time lag. The logarithmic axis enhances the intraday region, where the discrepancies occur. In particular, there is a clear difference between the empirical data and a theoretical power law, while the LM-Mic-Aff-ARCH agrees quite well with the empirical curve. Similar studies on other currency pairs show the same pattern, with the empirical curve showing a depression around 12 h and a steeper slope at shorter time horizons. This structure originates in the detailed composition and dependence of the financial market, and such features are absent from the simple LM processes. A detailed description of the market components is required to include such features, as for example in the process developed in Zumbach and Lynch (2001). The same study has been done for the LM-Agg-Aff-ARCH process, with very similar results.

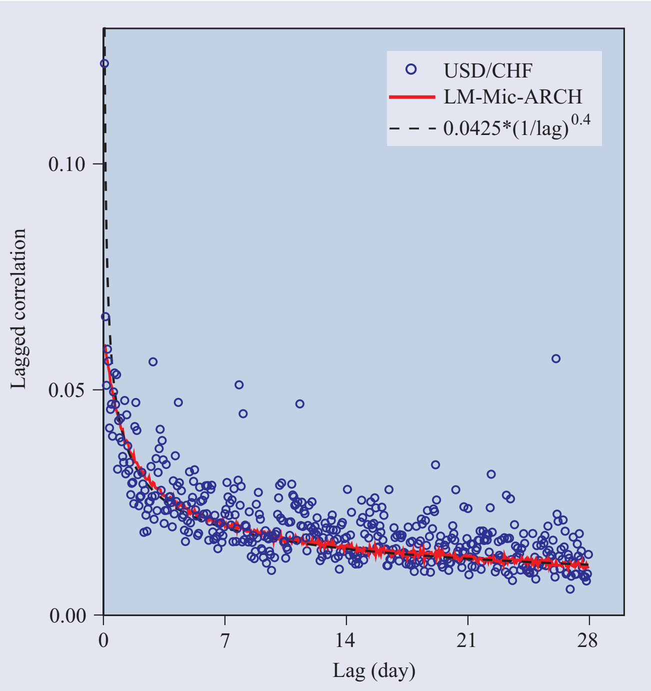

Figure 4. The lagged correlation for the LM-Mic-Aff-ARCH process. The process includes time ratio l_k from 1 to 8192 with a progression $\rho = 2$ (i.e. 14 partial volatilities σ_k), and the parameters are $\sigma = 0.12$, $w_\infty = 0.03$, $\tau_0 = 0.55$ day, $\lambda = 0.30$. The process is simulated with $\delta t = 1$ h 24 min (i.e. 1 business hour), for a length of 10^5 steps. The simulation is repeated 50 times, and the averages and standard deviations computed.

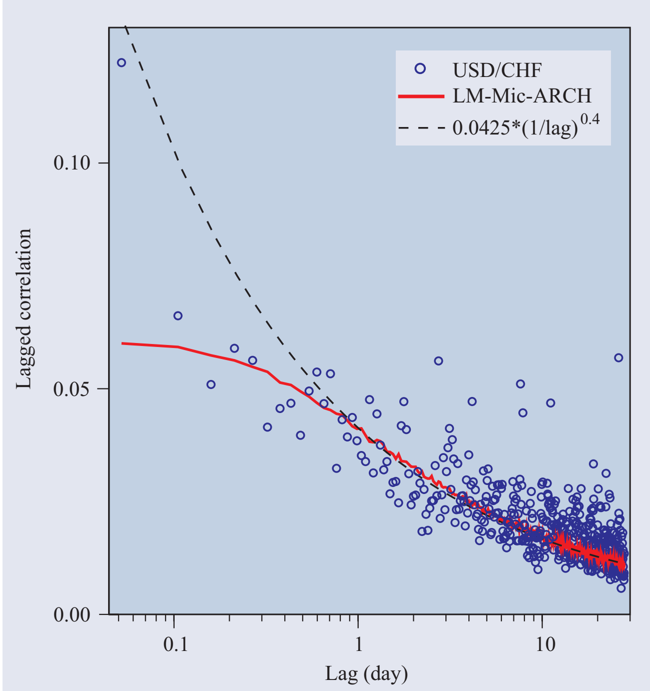

Figure 5. The lagged correlation for the LM-Mic-Aff-ARCH process with a logarithmic axis for the lag. The data are as for figure 4.

8.5. Cost function landscape

The processes studied in this paper have very few parameters, therefore it is possible to draw pictures of the cost functions (log-likelihood and volatility forecast error) in the parameter spaces. For the processes GARCH(1, 1), Aff-FIGARCH and LM-Mic-Lin-ARCH, the three sets of figures 6, 7 and 8 display the log-likelihood at $\delta t = 1$ (business) hour and volatility forecast error for $T = 1$ day, 1 week and 1 month. The LM-Agg-Lin-ARCH process shows an almost identical cost function landscape as LM-Mic-Lin-ARCH, and therefore is not shown. It should be added that the amount of computation required to produce such figures is very large, as the cost function must be estimated for a large number of parameter values.

The most important common feature of these graphs is the overall similarity between log-likelihood and volatility forecast errors. This means that the parameters obtained from a log-likelihood estimate will produce good forecasts, at least for time horizons not too large compared to δt. Let us now analyse these pictures in detail for each process. For the GARCH(1, 1) process shown in figure 6, the parameter τ_{corr} shifts systematically to higher values for increasing time horizons. This is the response of the exponential decay which adjusts further out to the power law by increasing the correlation time. As particular cases, these figures have on the left axis the process with constant volatility (i.e. a simple geometric random walk), and on the right axis the I-GARCH(1) process. Therefore, we can see the quantitative improvement when changing the process from random walk or I-GARCH(1) to GARCH(1, 1).

For the Aff-FIGARCH process, the cut-off is chosen to be large enough so that the global minimum is the relevant solution. For the log-likelihood cost function, a proper (local) minimum exists in the upper right corner at $d \sim 0.95$. Moreover, the two minima are separated by a saddle. Because at a saddle the derivatives are zero, it can fool a minimization algorithm to convergence, leading to a value for $d \sim 0.75$ that is too large. Therefore, one should be careful when optimizing FIGARCH processes to take a cut-off large enough, and to search for the global minimum. The global minimum, in the lower left corner, shifts from $d = 0.25$ to 0.45 with increasing time horizons.

For the LM-Mic-Lin-ARCH process, there is a well defined valley of nearly degenerate solutions. As can be expected, the minima shifts to larger values of $\ln(\tau_0)$ when increasing the time horizon of the cost function. Yet, a point at $\lambda \sim 0.3$ and $\ln(\tau_0) \sim -1$ gives a near optimal solution for all time horizons, from 1 h to 1 month. In our opinion, this is the major advantage of the LM processes, as they provide for a good description of the data with the *same* parameters over a large range of time horizons. It indicates that the LM processes capture well the structure of the market and of the data generating process. This opens the way to forecasting the volatility over a broad range of time horizons with the same set of parameters. In the case of linear processes, there are only two.

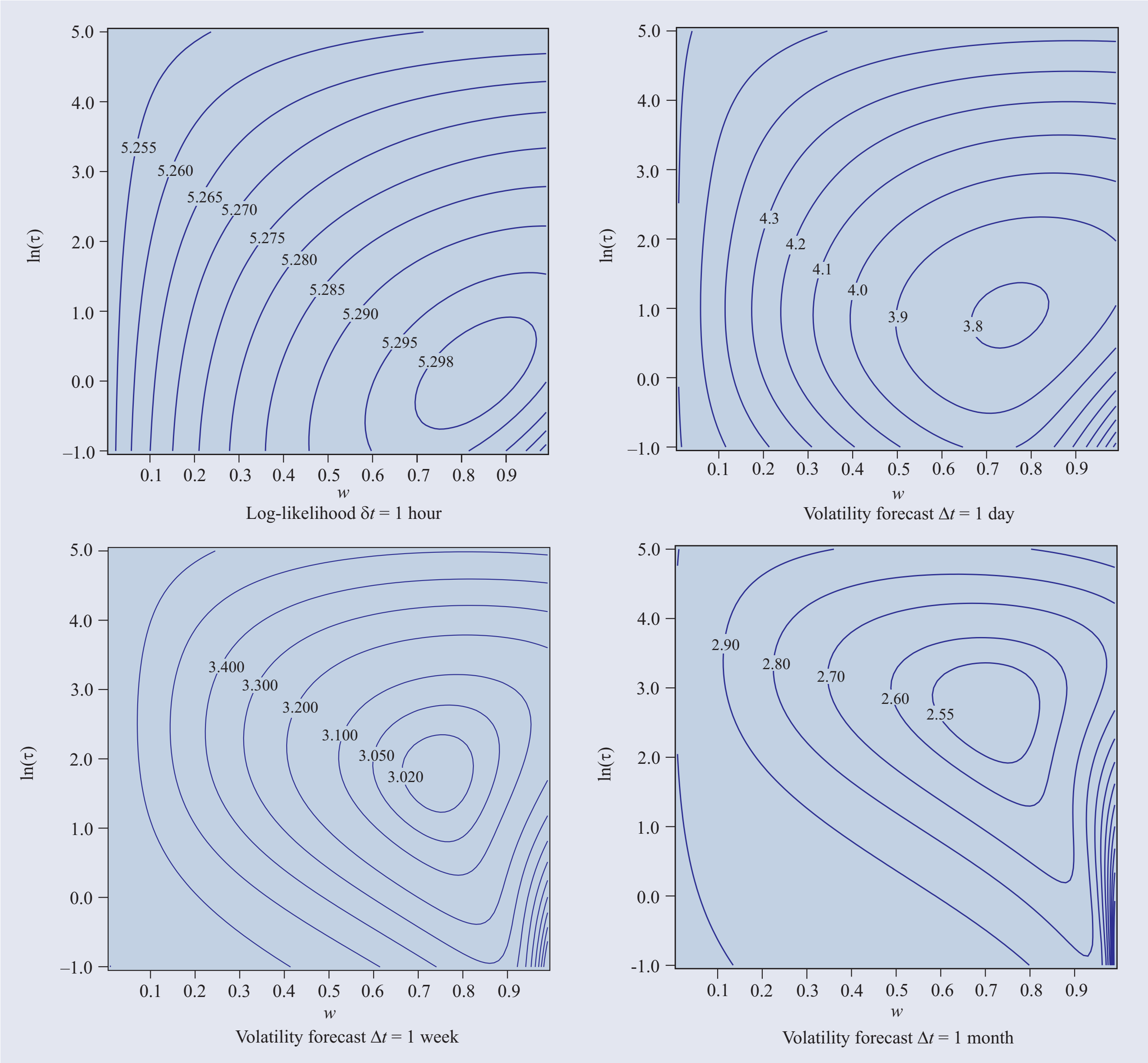

Figure 6. The cost function landscape for the GARCH(1, 1) process. These figures are a cut at constant volatility σ in the three-dimensional parameter space, with $\sigma^2 = \sum r^2$ the empirical volatility.

8.6. Log-likelihood estimate

In this section, we want to compare the log-likelihood $\mathcal{L}$ estimates corresponding to the various processes. The results given in table 2 are relative to the GARCH(1, 1) process, in per cent; namely the values in the first column are

$$100 \times \left(\frac{\mathcal{L}(\text{process})}{\mathcal{L}(\text{GARCH}(1,\ 1))} - 1 \right). \tag{43}$$

The residuals are taken with a Student-t distribution, and the number of degrees of freedom ν is estimated jointly with the process parameters. For all the processes, the optimal value of ν is between 4 and 5, clearly indicating strong fat tails for the residues. The major result is that the improvement over GARCH(1, 1) and GARCH(2, 2) for all the processes with a LM is quite small, respectively of the order of 0.06% and 0.05%. Let us emphasize that our goal in this section is not to set a formal log-likelihood ratio test to decide which process is best (on a particular data set), but to show that all processes that incorporate the LM are better than GARCH(1, 1), although the quantitative improvement is disappointingly small.

On the other side, the process called PastVolatility is defined by the effective volatility σ_{eff} being the historical volatility measured on the given time interval (1 day, 1 week, 1 month). This is equivalent to a rectangular memory kernel. Regardless of the chosen time interval, the results are clearly inferior to GARCH(1, 1). This shows that a continuously decaying memory is an important ingredient for a good volatility process.

The lagged correlations of the square residuals are measured with the Ljung–Box test at 6 h, 1 day and 5 days. All

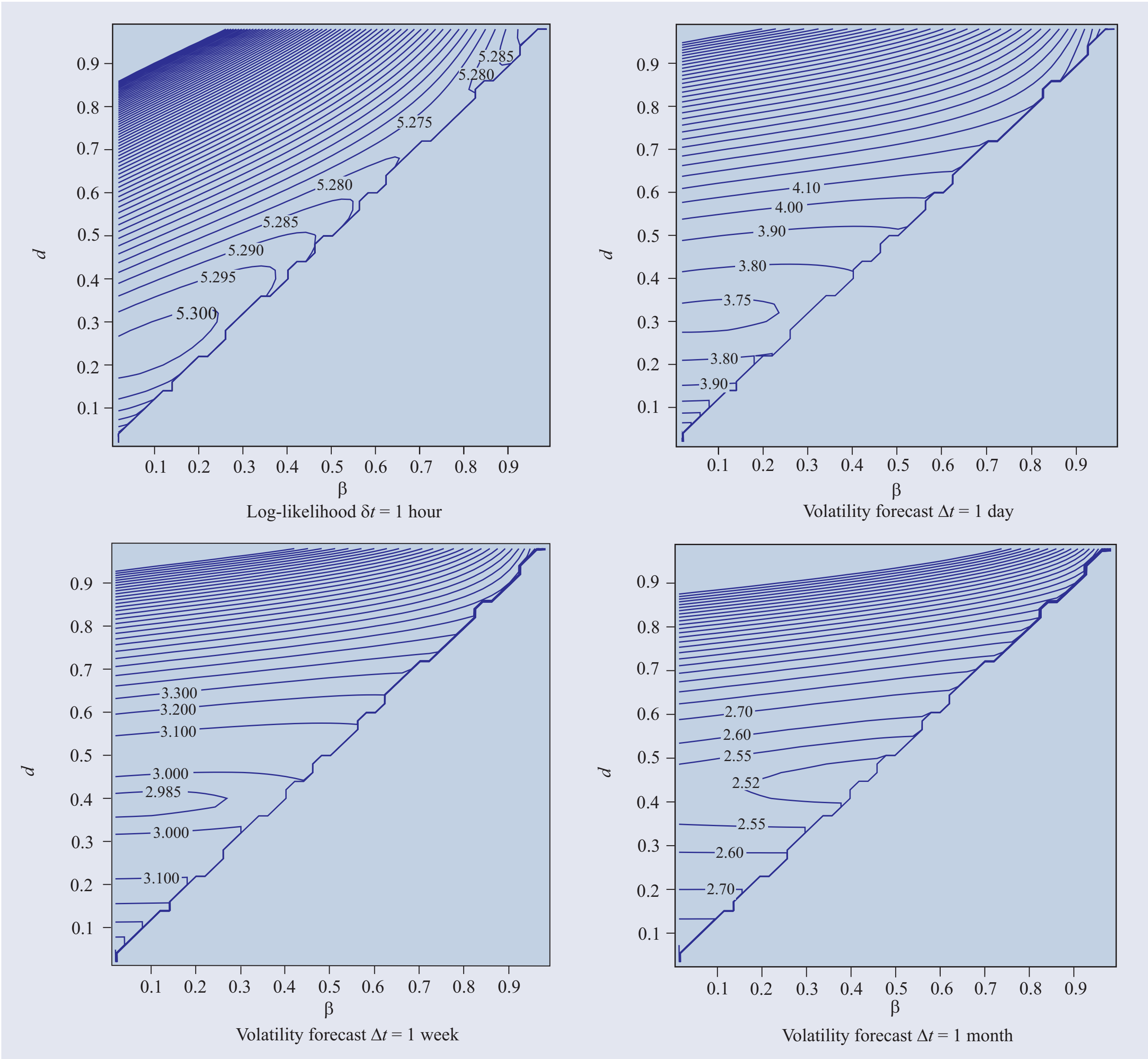

Figure 7. The cost function landscape for the Aff-FIGARCH(1, d, 0) process.

the LM processes are fairly good in whitening the residuals, with a clear advantage to the microscopic processes.

8.7. Out of sample forecast error

As developed in section 7, a forecast $\bar{\mathcal{F}}[\Delta t; \sigma^2_{\mathrm{eff}}]$ for the realized volatility can in principle be computed for any time horizon Δt, and the process parameters can be estimated to minimize the (in-sample) forecast error. For the empirical study, we have restricted ourselves to a one day time horizon $\Delta t = 1d$. There are also many possibilities regarding the choice of in-sample and out-of-sample studies; we present here two choices:

- *Full in-sample study.* 0.5 year of build-up, and from 1 January 1990 to 1 January 2000 for the parameter estimation and in-sample measure.
- *Continuous optimization.* A moving sample for estimating the parameters (0.5 year for the build-up, 5 years in-sample), followed by an out-of-sample forecast for the next day. This corresponds to continuously re-estimating the parameters over the last 5.5 years, before making the out-of-sample forecasts. Needless to say, this procedure is fairly costly in terms of computational power, and requires very efficient estimation algorithms.

The forecast error in equation (33) is reported in table 3, relative to GARCH(1, 1), in %:

$$100 \times \left(\frac{d\,[\text{process}]}{d\,[\text{GARCH}(1,\ 1)]} - 1 \right). \tag{44}$$

Overall, the results are consistent with the log-likelihood estimate. The past volatility provides for a bad forecast, regardless of the historical depth (with a optimum around

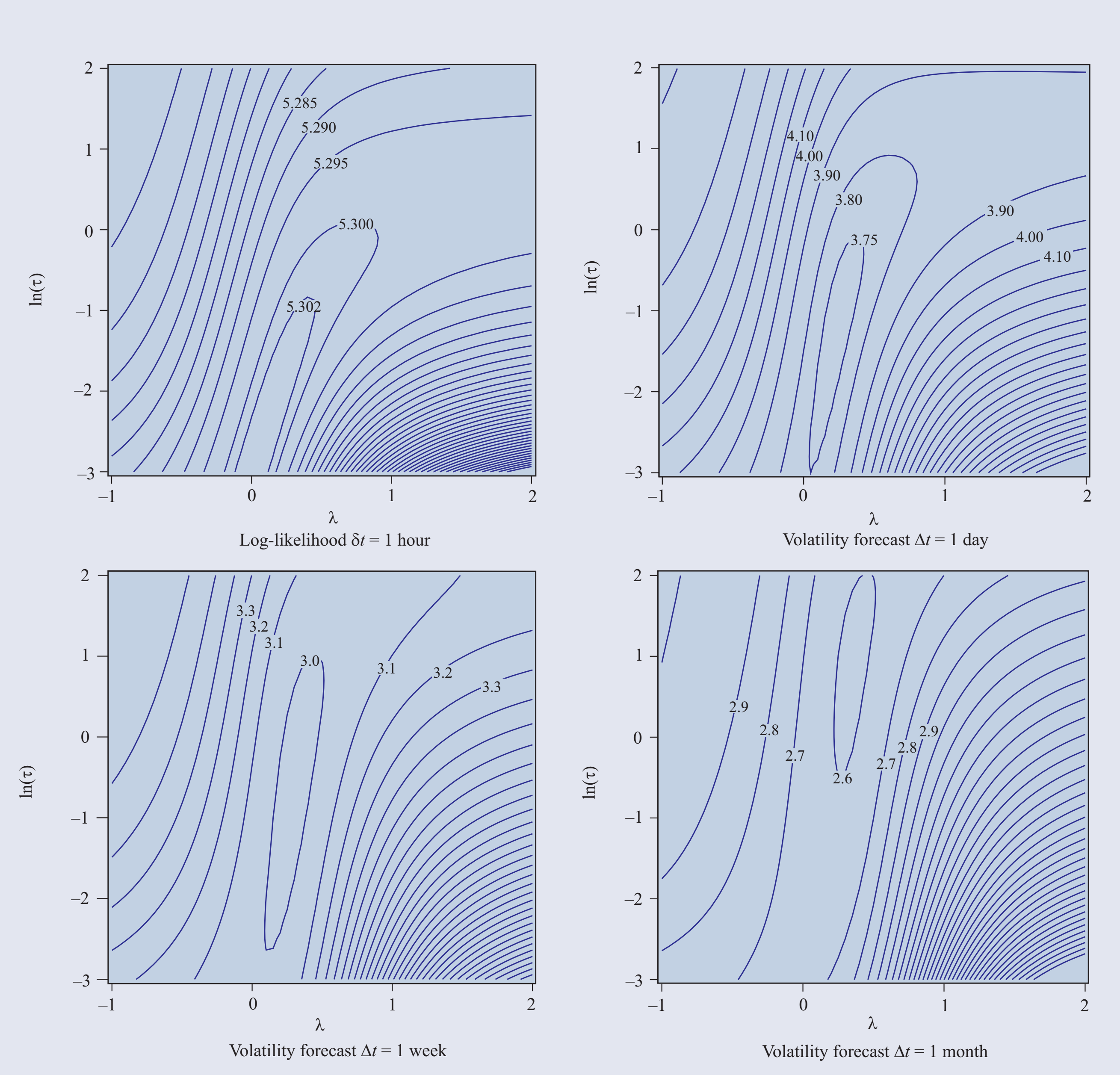

Figure 8. The cost function landscape for the LM-Mic-Lin-ARCH process.

1 week). This originates in the rectangular shape of the memory kernel, a shape very different from the power law decay of the lagged correlation. Given its simplicity, the I-GARCH(1) makes a very honourable score. The I-GARCH(2) is consistently better than GARCH(1, 1). This result comes from the two exponentials of the memory kernel that are able to build a decent approximation of a power law. On the other hand, the GARCH(2, 2) provides for very minor improvements compared to GARCH(1, 1). This indicates that the GARCH(p, q) extension is not the right direction to generalize the GARCH(1, 1) process, as it does not incorporate in a parsimonious way the empirical properties of financial time series. The *-FIGARCH are (in general) better than GARCH(1, 1). Yet, the best forecasts are clearly provided by the LM processes, with no clear winner between the linear versus affine processes, or the aggregated versus microscopic processes. Nevertheless, the quantitative relative improvement of the forecast errors over GARCH(1, 1), in the range of 2–2.5%, is also a bit disappointing.

8.8. Robustness with respect to the data sample

The continuous estimation of the parameters on a moving 5 year window gives as a by-product the 'time evolution' of the parameters. Such a time trace is given for example in figure 9. We expect an efficient process to incorporate the most relevant market structures, and these structures should be fairly stable over time. These market structures should find their equivalent in the process equations, while the quantitative aspect of the market structure is incorporated in the process

Table 2. The relative log-likelihood value l, relative to GARCH(1, 1), in % (positive values indicate a better process). The process 'GARCH(1, 1)' has a log-likelihood of 5.303. For the LM process, the parameter 12 is the number of volatility components in σ_k, corresponding to an upper cut-off of $2^{12} = 4096$. For *-FIGARCH, the parameter 2048 gives the cut-off j_{max} in the expansion of the fractional derivative. The data set is described in section 8.1, with 1989 used for the build-up of the processes, and 1 January 1990 to 1 July 2000 for the in-sample computations. The Ljung–Box test is performed on the square residual over 6 h (6 points), one day (24 points) and one week (120 points). The * and ** symbols denote the significant values at the 95 and 99% level of confidence, indicating significant correlations of the residuals.

Process name	Relative log-likelihood	Ljung–Box test 6 h	1 day	1 week
ConstantVol	−0.92	940**	2675**	5249**
PastVolatility (1 day)	−0.37	53**	156**	232**
PastVolatility (1 week)	−0.19	263**	372**	550**
PastVolatility (1 month)	−0.38	884**	1828**	2247**
I-GARCH(1)	−0.062	94**	108**	196**
I-GARCH(2)	0.032	99*	39*	111
GARCH(1, 1)	0	48**	68**	157*
GARCH(2, 2)	0.016	22**	42**	132
LM-Mic-Lin-ARCH(12)	0.062	74	31	118
LM-Agg-Lin-ARCH(12)	0.066	35**	59**	123
LM-Mic-Aff-ARCH(12)	0.072	2.8	24	102
LM-Agg-Aff-ARCH(12)	0.076	21**	44**	116
Lin-FIGARCH(1, d, 0; 2048)	0.058	6.1	30	105
Aff-FIGARCH(1, d, 0; 2048)	0.047	13*	39*	159**

Table 3. The relative forecast error for the 1d realized volatility (annualized, in %). The negative values indicate better forecasts. The value for the GARCH(1, 1) process is the distance function d, in %.

Process name	Full sample	Continuous optimization
GARCH(1, 1) value	3.90	3.85
ConstantVol	21.7	26.4
PastVolatility (1 day)	18.8	15.0
PastVolatility (1 week)	5.9	6.1
PastVolatility (1 month)	7.8	9.7
I-GARCH(1)	1.7	1.7
I-GARCH(2)	−0.55	−1.46
GARCH(1, 1)	0	0
GARCH(2, 2)	−0.005	−0.005
LM-Mic-Lin-ARCH(12)	−1.11	−2.50
LM-Agg-Lin-ARCH(12)	−1.52	−2.24
LM-Mic-Aff-ARCH(12)	−1.15	−2.22
LM-Agg-Aff-ARCH(12)	−2.07	−2.32
Lin-FIGARCH(1, d, 0; 512)	0.40	−0.63
Lin-FIGARCH(1, d, 0; 2048)	−0.62	−1.91
Aff-FIGARCH(1, d, 0; 512)	−0.80	−1.43
Aff-FIGARCH(1, d, 0; 2048)	−1.09	

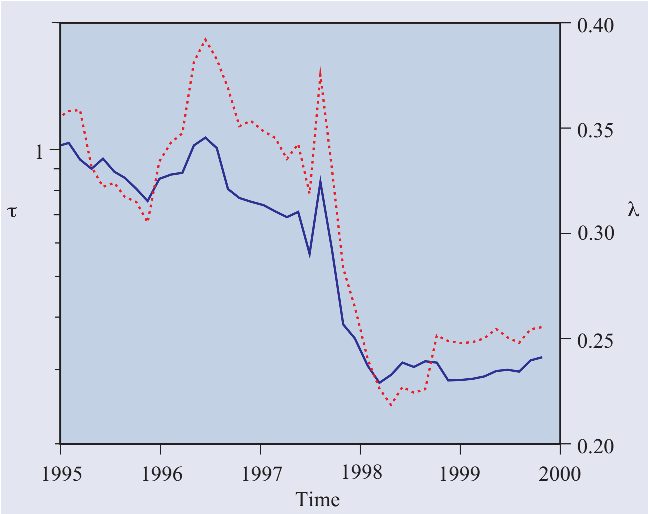

Figure 9. The parameters τ (full curve, left axis) and λ (dotted curve, right axis) for the LM-Mic-Lin-ARCH process estimated on a moving 5 year window. The time on the horizontal axis corresponds to the end of the 5 year estimation interval.

parameters. Therefore, for a 'good' process, we can expect that the estimated parameters are fairly stable over time. This begs for a measure of the fluctuation of the estimated parameters $\hat{\theta}$, for example computing the variance of $\hat{\theta}(t)$. The trouble with such a direct approach is that it depends on a particular coordinate system in the parameter space. For example, in figure 9, we can use μ or $\ln \tau$ instead of τ. Therefore, a direct measure of the parameter fluctuations cannot be compared across processes as they are coordinate dependent. In order to measure the parameter fluctuations, the key idea is to use a function that is invariant through reparametrizations, and the volatility and volatility forecasts are precisely such functions. Therefore, we define a measure of robustness with respect to changing data sets by

$$Q^2 = \sum_t (\bar{\mathcal{F}}[1d, 5y; \sigma^2_{\text{eff}}](t) - \bar{\mathcal{F}}[1d, 10y; \sigma^2_{\text{eff}}](t))^2 \quad (45)$$

where $\bar{\mathcal{F}}[1d, 5y; \sigma^2_{\text{eff}}]$ is a one day forecast for the realized volatility, with the parameters estimated on the subsample from $t - 5y$ to t. The term $\bar{\mathcal{F}}[1d, 10y; \sigma^2_{\text{eff}}]$ is the forecast with the parameters estimated on the full 10 year sample (in-sample forecast). Essentially, this measures the difference in the forecasts when estimating the parameters in a short sample

Table 4. The measure of robustness Q, in %, for the volatility forecast with respect to changes in the data set. The 'PastVolatility' for all time horizons has no parameter, and therefore $Q = 0$.

Process name	Q
ConstantVol	0.72
PastVolatility(*)	0
I-GARCH(1)	0.10
I-GARCH(2)	0.57
GARCH(1, 1)	0.26
GARCH(2, 2)	0.26
LM-Mic-Lin-ARCH(12)	0.13
LM-Agg-Lin-ARCH(12)	0.24
LM-Mic-Aff-ARCH(12)	0.17
LM-Agg-Aff-ARCH(12)	0.49
Lin-FIGARCH(1, d, 0; 512)	0.11
Lin-FIGARCH(1, d, 0; 2048)	0.07
Aff-FIGARCH(1, d, 0; 512)	0.23

(5 years) compared to a long sample (10 years), and with the forecasts computed, respectively, out-of-sample and in-sample. The results are given in table 4. These numbers can be directly interpreted as a kind of standard deviation for the annualized volatility, in per cent. The relevant numbers to compare with are the mean volatility $E[\sigma] \simeq 10\%$ and the forecast error $d \simeq 4\%$. Overall, the forecasts are fairly robust, with variations of the order of 0.3% for a mean forecast of 10%. Interestingly, the worst result is for the constant volatility, which is equivalent to a 5 and 10 year historical average. On the other side, the I-GARCH(1) is one of the most robust, and this is where the process's simplicity pays. Similarly, the linear processes are systematically more stable. These tell us that estimating a mean volatility parameter is not very robust against changing the data sample, even at time horizons of several years

9. Conclusions

There already exists in the literature a very large number of modifications and extensions of the basic univariate GARCH(1, 1) process. The LM processes presented in this work are interesting because they open a new direction in the space of processes, and because they incorporate in a parsimonious way the most salient empirical fact about the volatility memory, namely the power law decay of the correlations. Yet, as with all other extensions of GARCH(1, 1), the quantitative improvements are not spectacular, regardless of the chosen criteria (log-likelihood or forecast error). This can be understood intuitively from figure 10, which displays the one day realized volatility and its forecast. Improving the process equations leads to a shorter response to abrupt volatility increases, and to a slower relaxation to the mean, whereas GARCH(1, 1) needs to compromise between sharp increases and slow relaxation. Yet, the forecast error is dominated by sudden volatility increases, and the associated large values are weakly dependent on the detailed dynamic of a particular process. Choosing a L_1 distance for the forecast error makes the quantitative differences between processes larger (as less weight is given to large and unpredictable

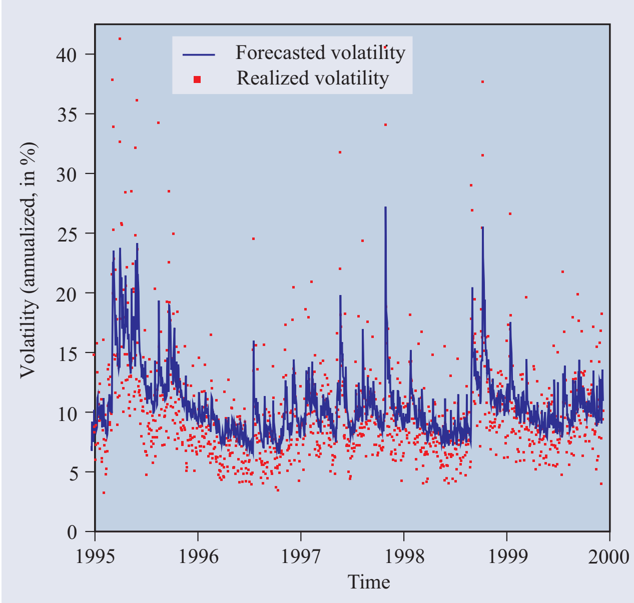

Figure 10. The daily realized volatility with its forecast computed with the LM-Mic-Lin-ARCH process.

volatility increases), but the conclusions are similar. In our view, the major advantage of the LM processes lies not in quantitative improvements, but in their capacity to describe the empirical data well, from 1 h to 1 month, with *the same parameters*. It is mainly in this sense that the LM processes are better than GARCH(1, 1), as they capture more accurately the properties of the empirical data. A second advantage of the LM processes is that their mathematical structures are intuitively related to the structure of financial markets. This make them easier to modify in order to incorporate other features present in the empirical data.

For the purpose of forecasting, the linear LM processes are particularly interesting as they achieve very good forecast performances with only two parameters. They are consistently better than GARCH(1, 1), but with one parameter less. Moreover, the parameters do not include the mean volatility, and therefore they may only weakly depend on the particular time series. In short, the LM-Mic-Lin-ARCH process is a very good candidate to produce general forecasts, and to replace the widely used I-GARCH(1) or GARCH(1, 1) processes. In particular, this process is a good candidate to extend the RiskMetric methodology toward high frequency data, which should be rewarding for horizons up to a few weeks. We have also shown that measures of the historical volatility provide for bad forecasts. As data generating processes (for example for Monte Carlo simulations), the LM-Mic-Aff-ARCH or LM-Agg-Aff-ARCH processes are good candidates. Yet, for time horizons up to the upper cut-off, the linear processes can also be used, for example in stress testing, with the advantage of having only two parameters.

For the structure of the equations, with respect to the microscopic (Mic) versus aggregated (Agg) processes, there is no clear winner, at least according to the measures studied in this paper. The aggregation seems not to be a determinant factor for the quality of a process, in contradiction to the argument developed before equation (10) that the market participants active over short (long) time horizons are using high (low) frequency data. This also goes against one of the main motivations for the HARCH processes as introduced in Müller (1995) and Dacorogna *et al* (1998), namely an asymmetry in the information flow between fine and coarse volatilities. Although this asymmetry is a clearly stylized fact, its quantitative effect seems marginal. With the present quantitative results, simplicity suggests using microscopic processes as they are easier to implement. Yet, further statistical studies may lead to a sharper distinction.

Regarding the GARCH(2, 2) process, it leads to a very marginal improvement in the results, at the cost of two extra parameters. In the author's view, the GARCH(p, q) extension, originating in the 'traditional' time series analysis and the expansions in the lag operator, is not the right direction to go, because it lacks a sparse modelling of the volatility memory.

With respect to the fractional processes, the Lin-FIGARCH and Aff-FIGARCH processes need to be considered because of the subtleties in the practical implementation of the fractional difference operator with a finite cut-off. The memory structure of these processes is very similar to the LM*ARCH processes, with the ARCH(∞) representation having the same asymptotic decay for the coefficients. Quantitatively, the fractional processes are slightly inferior to the LM processes. But their main drawback is to impose a very high computational burden, as the computational time increases as the maximum cut-off of the fractional difference operator. Moreover, the cut-off needs to be chosen large enough. This is a clear problem, particularly when using high frequency data. Another potential pitfall with the *-FIGARCH processes is that the cost function landscape presents two minima. The fractional processes lack an intuitive connection with our understanding of the market structure, and it is difficult to modify and extend the *-FIGARCH processes.

Finally, another advantage of the LM processes is that they can be easily modified or extended along many directions, while keeping the simple structure leading to the power law memory. For example, non-quadratic processes can be written by changing the exponent used for computing the volatility or, in the case of stocks, the asymmetry between positive and negative returns can be easily included. Clearly, in view of this wide open space, fine statistical analysis of the empirical data should guide our way in the relevant direction. A first step in such a direction is taken in Zumbach and Lynch (2001) and Lynch and Zumbach (2003), where a finer modellization of the market components is explored while using the correlation between the volatility derivative and realized volatility as a guideline.

Acknowledgments

The bulk of this research was conducted when the author was working at Olsen & Associates, Research Institute for Applied Economics, Seefeldstrasse 233, 8008 Zürich, Switzerland. Moreover, I want to thank Höskuldur A Hauksson for his suggestion regarding the robustness with respect to the data sample.

References

Baillie R T, Bollerslev T and Mikkelsen H-O 1996 Fractionally integrated generalized autoregressive conditional heteroskedasticity *J. Econometrics* **74** 3–30

Bera A K and Higgins M L 1993 Arch models: properties, estimation and testing *J. Economic Surveys* **7** 305–62

Bollerslev T 1986 Generalized autoregressive conditional heteroskedasticity *J. Econometrics* **31** 307–27

Bollerslev T, Chou R Y and Kroner K F 1992 ARCH modeling in finance *J. Econometrics* **52** 5–59

Breidt F, Crato N and de Lima P 1998 The detection and estimation of long memory in stochastic volatility *J. Econometrics* **83** 325–48

Breymann W, Zumbach G, Dacorogna M M and Müller U A 2000 Dynamical deseasonalization in OTC and localized exchange-traded markets *Internal Document* WAB.2000-01-31, Olsen & Associates, Seefeldstrasse 233, 8008 Zürich, Switzerland

Chung C 1999 Estimating the fractionally integrated GARCH model *Unpublished Working Paper* National Taiwan University, Taipei, TW

Corsi F, Zumbach G, Müller U A and Dacorogna M M 2001 Consistent high-precision volatility from high-frequency data *Econ. Notes—Rev. Banking, Finance Monetary Economics* **30** 183–204

Dacorogna M M, Müller U A, Olsen R B and Pictet O V 1998 Modelling short-term volatility with GARCH and HARCH models *Nonlinear Modelling of High Frequency Financial Time Series* ed C Dunis and B Zhou (Chichester: Wiley) pp 161–76

Davidson J 2002 Moment and memory properties of linear conditional heteroscedasticity models *Preprint* Cardiff University

Ding Z and Granger C 1996 Modeling volatility persistence of speculative returns: a new approach *J. Econometrics* **73** 185–215

Engle R F and Bollerslev T 1986 Modelling the persistence of conditional variances *Econometric Rev.* **5** 1–50

Giratis L, Kokoszka P and Leipus R 2000 Stationary arch models: dependence structure and central limit theorem *Econometric Theory* **16** 3–22

Granger C and Ding Z 1996 Varieties of long memory models *J. Econometrics* **73** 61–77

Lynch P and Zumbach G 2003 Volatility causal structure and market heterogeneities *Quant. Finance* **3** 320–3

Müller U A 1995 HARCH—a new model to analyse and predict volatility *FX and MM—J. Foreign Exchange Money Market, London* **2** 59–63

Nelson D 1990 Stationarity and persistence in the GARCH(1, 1) model *Econometric Theory* **6** 318–4

Teyssière G 1996 Double long-memory financial time series *QMW Working Paper* 348 University of London, UK

Zaffaroni P 2000 Stationarity and memory of arch models(∞) *Preprint* London School of Economics and Political Science

Zumbach G 2000 The pitfalls in fitting GARCH processes *Advances in Quantitative Asset Management* ed C Dunis (Dordrecht: Kluwer)

Zumbach G and Lynch P 2001 Heterogeneous volatility cascade in financial markets *Physica* A **298** 521–9