WORKING PAPER n.05.03

June 2003

Evaluating Value-at-Risk measures in presence of long memory conditional volatility

M. Caporin[1]

[1] GRETA, Venice

Evaluating Value-at-Risk measures in presence of long memory conditional volatility

Massimiliano Caporin
Università "Ca' Foscari" di Venezia
Dipartimento di Scienze Economiche
Cannaregio
Fondamenta San Giobbe, 873
30121 Venezia
ITALY
and
GRETA Associati
San Marco 3870
30124 Venezia

This version August 2002

Abstract

This work analyze different approaches in the evaluation of Value-at-Risk measures when returns show long memory patterns in conditional variance. In a Montecarlo study we follow the approaches of Kupiec (1995), Christoffersen (1998), Christoffersen, Hahn and Inoue (2001) and Lopez (1998) using the suggested test and loss functions in choosing the best model among a group of alternatives (GARCH, IGARCH, the true FIGARCH DGP and the EWMA). Our Montecarlo analysis shows that the test of Kupiec and the loss function approach lead to the choice of a misspecified model, while the test of Christoffersen et al. (2001) correctly identify long memory. We apply then all the previous tests and measures in the comparison of different models for the Value-at-Risk of the returns of the FIB30, the future on the italian stock market index.

Keywords: Value-at-Risk, long memory, FIGARCH.

1 Introduction

In the last few years there has been a huge increment in analysis concerning Value-at-Risk (VaR), both from a theoretical point of view and from the empirical approach, in particular dealing with: the best methods to compute the risk exposure needed to satisfy regulators requirements, the choice of the best model for VaR computation, the evaluation of performances of different VaR models. The literature is still growing and with this work we will add some extensions showing how VaR is affected by model misspecification when variance follows a long memory conditional heteroskedastic process. This is related to the numerous findings of persistence in financial markets, coupled with the use of high frequency data for VaR computation, see among other Christoffersen and Diebold (2000) and Beltratti and Morana (1999). In many VaR papers the long memory behavior of the series has not been taken yet into account; even if Beltratti and Morana introduced a first empirical analysis, some problems arise, as pointed out by Christoffersen and Diebold (2000): what does really mean having a long range forecast with high frequency data? or in other words is it correct estimating 1-day VaR (or more) using intra-day observations? Beltratti and Morana (2000) solved that problem using the traditional $\sqrt{T}-$rule for computing s-step-ahead variance forecasts, but they concluded the analysis with a choice of a GARCH process for their foreign exchange data even if the observations showed a clear long memory behavior. They motivated the choice by the closeness of the results obtained by the long and short memory models, preferring then the simplest one, the GARCH. This is a particular effect, maybe due to the used data and it is not yet proved in a general context. Independently from these considerations the square root rule is not optimal as a scaling in a GARCH framework as Diebold et al. (1996) showed. Within this work and in a companion paper we will shed some light in this area considering the effects of misspecification on Value-at-Risk measures when the underlying generator is a long memory model and then we will compare Value-at-Risk estimates obtained from data with different frequencies (high frequency agaisnt lower frequency aggregated data). The first point will be the object of this study. In the next section after focusing our attention on a specific case, we will assume that the observed series we are analyzing follow a FIGARCH, we will present the forecasting equations for GARCH and FIGARCH specifications, precisely the forecasting equations for the mean square error of the mean predictor, when the residuals follow a conditional heteroskedastic model, extending in such a way the results of Baillie and Bollerslev (1992). In this part we will focus on point forecasts, not on density forecasts, for such an extension, which is straightforward, refers again to Baillie and Bollerslev (1992). In section 3 we will present a survey on the usual methods applied by banks and regulators to evaluate Value-at-Risk performances on their models, introducing a new loss function approach and applying very recent tests which will show, in section 4, the discrepancy between the best choice for the regulators and the best one for a bank, we will see that the regulators may push to the choice of a misspecified model; these results are obtained by a Montecarlo experiment with GARCH(1,1) and FIGARCH data generating processes, estimating then, on both DGP, GARCH, IGARCH and FIGARCH models. For all the considered models, even if incorrectly specified, we will compute VaR for 1-day horizon comparing the different results, in a backtesting framework, applying the evaluation techniques of section 3. Section 5 will conclude.

2 Prediction mean square errors with FIGARCH

In this section we will extend the approach of Baillie and Bollerslev (1992), who were considering prediction with dynamic models and conditional heteroskedasticity, to allow for long memory behavior. Assume that the series we are analyzing follow a generic process for the mean

$$y_t = \mu_t + \varepsilon_t$$

and that the residuals are such that $\varepsilon_t | I^{t-1} \sim iid\left(0, \sigma_t^2\right)$, where with I^{t-1} we identify the information set up to time t-1. Assuming that the mean term is always zero, we are in the framework of a GARCH-type process, where the forecast for the mean process is always zero and the Mean Square Error (MSE) depends on the s-step ahead prediction for the variance. The MSE will also depends nontrivially on the information set, an extensive discussion and numerous expressions can be found in the above cited paper. For the simple GARCH(1,1) the s-step ahead predictor for the variance (the MSE of the s-step ahead predictor for the mean) is:

$$E\left[\varepsilon_{t+s}^2|I^{t-1}\right] = E\left[\sigma_{t+s}^2|I^{t-1}\right] = \omega \sum_{i=1}^{s-1} (\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^{s-1} \sigma_{t+1}^2 \quad (1)$$

The FIGARCH process has been introduced by Baillie, Bollerslev and Mikkelsen (1996) as a generalization of the IGARCH, allowing for a non-integer integration coefficient. The volatility structure induced by a FIGARCH can be defined as follows (for details on the model properties and on the parameter estimation see Baillie et al. (1996), Bollerslev and Mikkelsen (1996) and Caporin (2002)):

$$\sigma_t^2 = \omega + \beta(L)\sigma_t^2 + \left[1 - \beta(L) - \Phi(L)(1-L)^d\right]\varepsilon_t^2$$

where $\beta(L) = \sum_{j=1}^{p} \beta_j L^j$, $\Phi(L) = \sum_{j=0}^{m} \phi_j L^j$ and $(1-L)^d$ is the fractional integration component. If the DGP is a FIGARCH process the predictor depends nontrivially on all past values therefor an expression like (1) cannot be obtained. In our analysis we will use the following representation, which is derived after some tedious algebra:

$$E\left[\sigma_{t+s}^2|I^{t-1}\right] = \theta_s\omega + \sum_{i=0}^{\infty} \psi_{i+1}\varepsilon_{t-i}^2 \quad (2)$$

$$\psi_k = \sum_{i=1}^{s} \phi_i \lambda_{k+s-i} \qquad \phi_1 = 1 \qquad \phi_i = \sum_{j=1}^{i-1} \lambda_j \phi_{i-j} \qquad \theta_s = \sum_{i=1}^{s} \phi_i$$

One question on the worthiness of the previous formula arises: why not using the following recursions formulas?

$$E\left[\sigma_{t+s}^2|I^{t-1}\right] = \theta_s\omega + \sum_{i=0}^{\infty} \lambda_{i+1} E\left[\varepsilon_{t+s-i}^2|I^{t-1}\right] \quad (3)$$

$$E\left[\varepsilon_{t+i}^2|I^{t-1}\right] = \varepsilon_{t+i}^2 \qquad \text{if } s \leq 0$$

$$E\left[\varepsilon_{t+i}^2|I^{t-1}\right] = E\left[\sigma_{t+i}^2|I^{t-1}\right] \qquad \text{if } s > 0$$

The main reasons is only based on computational advantages and rounding error that arise implementing the procedures with any software: in every point forecast of the conditional variance we use the past value of the observed series or residuals, given the infinite past dependence of any conditional variances with this simple recursion formula we induce a greater rounding error than the one induced by aggregating coefficients. By our formula we just induce one rounding error not the sum of s rounding errors.

3 Comparing Value-at-Risk estimates

The use of risk measures to determine the market risk implicit in any portfolio, investment or financial instrument is a need for all banks, investors and any firms that operate within financial markets. This need is particularly important for banks acting on both sides of the money market, investing with their funds and collecting savings, all banks have to fulfill requirements that are there to prevent a default that will be particularly burdensome for the collectivity. In this view most of the banks started in the last decades, given the increased sophistication in the financial markets, to measure the risk of their positions and balance sheets (the whole bank can be viewed as a portfolio of credits and debts, including by this way direct investments and other credit positions) with adequate and therefore complicated instruments. This leads to the diffusion of many "internal" models whose ultimate purpose was the same: monitoring the risk and the losses of all positions. In this situation the Basle Committee on Banking Supervision gave a regulated framework, with minimal requirements in term of model choice, to measure and compare the ability of internal models in meeting some very basilar qualifications, giving also an alternative valuation method, the "standardized approach". These rules were included in the accord of 1996, the well-known Amendment to the Capital accord to incorporate market risk (MRA). With this document the Basle Committee stated the formal rules that an internal model for market risk should meet, how it should compute the exposure to this kind of risk and how to define the minimal capital requirements needed to cover market risk. The MRA requires that each bank communicates daily the market exposure determined with any internal model or the standardized approach to the national regulator, this exposure has to be determined with a 99% one tail probability and with a holding period of 10 days. This measure of risk should represent the maximum loss with the 99% probability in the holding period, which is just the definition of the Value-at-Risk. Given these measures the regulator will verify if the internal model meets a minimal requirement, specifically, in the past year did this model give a 1% of failures or more? The verification is conducted with a technique described in the MRA accord, the backtesting approach, where the regulators verify the performances of the internal model in the last 250 days, and simply counts the exceptions (how many times the internal model fails). Given the number of exceptions the regulators classify the internal model with a grid 0-4, 5-9, more than 9, matched with a color green, yellow and red. The classification allows the regulators to impose some penalty, this because the MRA computes the correct VaR as the maximum between today's VaR and the average of last 60 VaR measures, multiplied by a scaling factor that depends on the previous classification. This methodology however may be inefficient for the banks, as it may lead to the application of a model that fulfills the requirements of the Basel accord but translates into a bigger cost: the minimal capital requirement can be viewed as an immobilization of resources, of liquidity, and given the operativity of the banks this represents an opportunity cost of investing resources.

The exposure measured by the VaR depends crucially on the underlying model employed for the return series of the financial instrument of interest. A group of questions arises: how can we judge if the underlying model is correct? how should the Value-at-Risk perform under different approaches? What are the consequences of a misspecification? In this section we will try to give an answer to some of these questions in a particular case: we will assume the true data generating process (DGP) follows a FIGARCH in the variance, and we will compare via tests and other approaches the true DGP with a group of misspecified models. The main works in this field are the ones of Kupiec (1995), Christoffersen (1998) and Lopez (1998) who proposed, respectively, a statistical based procedure and a loss function approach to test if the VaR estimates are correct and consistent with the data.

The reliability of VaR measures depends on the correct specification of the underlying models, this is necessary to provide an accurate measure of risk exposure. Considering the computation of Value-at-Risk using instruments (or portfolio) returns, indexing VaR estimates with time t, and model index m, assuming that the return follows a possibly time-dependent distribution f_t, the Value-at-Risk computed conditional on the information set on time t, for k-steps-ahead, is the α-quantile of the forecasted distribution f given for the model m. $\text{VaR}_{m,t}(\alpha, k)$ is the solution of the following equation

$$\int_{-\infty}^{VaR_{m,t}(\alpha,k)} f_{m,t+k}(x)\,dx = \alpha \tag{4}$$

Two different approaches are actually available to evaluate the VaR estimates: statistical based procedures, and loss functions approaches. The Proportion Failure test (or Unconditional coverage test), the Time Until First Failure test of Kupiec (1995) and the Conditional coverage test of Christoffersen (1998) and Lopez (1998) belong to the first group belong, while the approach of Lopez (1998) belong to the socond one. The main difference between the two is that with statistical procedure, inference analysis is available. The tests of Kupiec and Christoffersen are based on likelihood ratios, and on the assumption that VaR should exhibit a conditional or unconditional coverage equal to α.

The Unconditional Coverage test (UC) of Kupiec is based precisely on the first assumption: if VaR estimates are accurate, the exceptions x (the number of times return underperform VaR measures) can be modeled with a binomial distribution with probability of occurrence equal to α. In this case, by comparing the required unconditional coverage α (usually set to 0.05 or 0.01), with the measured coverage $\widehat{\alpha} = x/T$, it is possible to derive a likelihood ratio test under the null hypothesis $\alpha = \widehat{\alpha}$

$$LR_{UC} = 2\left[\ln\left(\widehat{\alpha}^{x}\left(1-\widehat{\alpha}^{T-x}\right)\right) - \ln\left(\alpha^{x}\left(1-\alpha^{T-x}\right)\right)\right] \tag{5}$$

Under the null hypothesis LR_{uc} is distributed as a $\chi^2(1)$. The UC test is also the statistical transposition of the procedure used by the regulator authority in judging if the internal model is accurate. As pointed out by Lopez (1998) this method does not show any power in distinguishing among different, but close alternatives.

This test, as pointed out by Christoffersen (1998), considers only exceptions over the sample size, however in presence of conditional heteroskedasticity, also the conditional coverage is important. Ignoring this issue, the volatility dynamics, we could have forecasts (VaR estimates with a GARCH-type model, include the forecast of the conditional variance as we will see) with correct unconditional coverage and incorrect conditional coverage, in this cases the UC test is of limited accuracy. Lopez adapted the general approach of Christoffersen formulating the following Conditional Coverage (CC) test. First a dummy variable is set to identify exceptions

$$D_{m,t+1} = \begin{cases} 1 \text{ if } \varepsilon_{t+1} < VaR_{m,t+1} \\ 0 \text{ if } \varepsilon_{t+1} \geq VaR_{m,t+1} \end{cases}$$

Under the null hypothesis that the VaR presents correct conditional and unconditional coverage, this indicator variable should be independent. Thus the CC test is computed as the sum of the UC test and of a test of independence on $D_{m,t+1}$, against a first-order Markov process. The independence test is constructed as follows: with $T_{i,j}$ we identify the number of observations in the sample T in state j after having been in state i, under the Markov process the likelihood function is

$$L_M = (1-\pi_{0,1})^{T_{0,0}} \pi_{0,1}^{T_{0,1}} (1-\pi_{1,1})^{T_{1,0}} \pi_{1,1}^{T_{1,1}} \tag{6}$$

where $\pi_{0,1} = T_{0,1}/(T_{0,0}+T_{0,1})$ and $\pi_{1,1} = T_{1,1}/(T_{1,0}+T_{1,1})$. Under serial independence the likelihood function is

$$L_I = (1-\pi)^{T_{0,0}+T_{1,0}} \pi^{T_{0,1}+T_{1,1}} \tag{7}$$

where $\pi = (T_{0,1}+T_{1,1})/T$. The test statistic is then

$$LR_{CC} = L_{UC} + 2[\ln(L_M) - \ln(L_I)] \tag{8}$$

and is distributed as a $\chi^2(2)$ under the null hypothesis of correct coverage (under the null hypothesis of independence the dependence test is a likelihood ratio test, whose limiting distribution is a $\chi^2(1)$).

We will turn now to another approach, the one of loss functions. The main work in this area is the one of Lopez, based on computing a loss function distinguishing between exception and not-exception. In the general form he proposes the following formula

$$C_{m,t+1} = \begin{cases} f(\varepsilon_{t+1}, VaR_{m,t+1}) & \text{if } \varepsilon_{t+1} < VaR_{m,t+1} \\ g(\varepsilon_{t+1}, VaR_{m,t+1}) & \text{if } \varepsilon_{t+1} \geq VaR_{m,t+1} \end{cases} \tag{9}$$

where $f(x,y)$ and $g(x,y)$ are such that $f(x,y) \geq g(x,y)$. In this formulation higher values of the functions are associated with exceptions, thus summing $C_{m,t+1}$ over the backtesting sample used by regulators we obtain

$$C_m = \sum_{i=1}^{T} C_{m,t+i} \tag{10}$$

and the best model is the one that minimizes (10). The choice of the correct model can be done referring to a benchmark, once the functions have been specified. Lopez proposed different functions: one derived from the dummy for exception, another using weight as for the regulator choices, and then the following one, that takes into account the exception and the discrepancy between the realization and the VaR forecasted measure.

$$C_{m,t+1} = \begin{cases} 1 + (\varepsilon_{t+1} - VaR_{m,t+1})^2 \text{ if } \varepsilon_{t+1} < VaR_{m,t+1} \\ 0 \text{ if } \varepsilon_{t+1} \geq VaR_{m,t+1} \end{cases} \tag{11}$$

This function was suggested in order to take into account not only the risk but also the amount of the possible default in the position. This function was built mainly for regulatory purposes, helping the regulator in the evaluation of bank internal models. But there is an open point, with this function we may be tempted to reject a model only because, at parity of exceptions, it realizes a higher loss function. In this case we may reject a correct model, a correctly specified and identified model for the series of returns, choosing an incorrect model. Up to some extent this may be observed in the work of Beltratti and Morana (2000) on FX data, when they end up choosing a GARCH process for computing VaR even if the data show a clear long memory property, because the number of exceptions of the FIGARCH was lower, too conservative (this is a loss function based on the dummy). This can be clarified with an example: assume that two different models are fitted to a real series, a GARCH(1,1) and an IGARCH(1,1); the forecast from both models differs only in the wideness of 1-step-ahead prediction intervals for the mean, the one of the IGARCH is bigger; moreover assume that both models present exactly the same number of exceptions, then using the loss function suggested by Lopez we will choose the IGARCH model because its bands are wider and therefore the loss function is lower (the difference between VaR and the realization in the market is lower given that bands are wider); this will be translated in a higher cost for the bank, they will have to fulfill a higher capital margin to stick to the IGARCH bands, even if the exceptions of the two models are the same. To solve this point we suggest using different loss functions, dealing not only with the failure of the VaR measures but also taking into account the distance between the different forecasts and the past realizations. We suggest to check that the model fulfills the quantile requirement and that it also have to be stick to the realization of the underlying process. We propose three different distance measures, adopting the same terminology of Lopez:

$$
\begin{aligned}
{}^{1}f\left(\varepsilon_{t+1},VaR_{m,t+1}\right) &= \left|1-\left|\frac{\varepsilon_{t}}{VaR_{m,t+1}}\right|\right| \\
{}^{2}f\left(\varepsilon_{t+1},VaR_{m,t+1}\right) &= \frac{\left(\left|\varepsilon_{t}\right|-\left|VaR_{m,t+1}\right|\right)^{2}}{\left|VaR_{m,t+1}\right|} \\
{}^{3}f\left(\varepsilon_{t+1},VaR_{m,t+1}\right) &= \left|\varepsilon_{t}-VaR_{m,t+1}\right|
\end{aligned} \tag{12}
$$

In all three cases the best choice is the model that minimizes the loss function. Taking these as they are we can incur in the same problems when using the loss functions of Lopez: we may be not able to correctly choose the right model, preferring a solution with narrower bands. For this reason we suggest also to apply these loss functions not only to the exceptions but also to the whole sample:

$$
\begin{aligned}
{}^{1}f\left(\varepsilon_{t+1},VaR_{m,t+1}\right) &= {}^{1}g\left(\varepsilon_{t+1},VaR_{m,t+1}\right) \\
{}^{2}f\left(\varepsilon_{t+1},VaR_{m,t+1}\right) &= {}^{2}g\left(\varepsilon_{t+1},VaR_{m,t+1}\right) \\
{}^{3}f\left(\varepsilon_{t+1},VaR_{m,t+1}\right) &= {}^{3}g\left(\varepsilon_{t+1},VaR_{m,t+1}\right)
\end{aligned} \tag{13}
$$

The three functions suggested consider different approaches to testing the discrepancy between the identified model and the realizations: the first one considers the ratio between one step VaR and the realization, the second one is the squared error realized with the VaR, divided by the VaR itself to be standardized to the same quantity of the first function, this was done in order to be able to build a fourth criteria adding the first two measures, it maybe thought as a kind of first and second order loss; the third function takes into consideration only the difference between VaR measure and the realization in order to be summable with the Lopez loss function. The effect of such different approaches will be presented in the following chapter with a limited Montecarlo experiment (we deal with FIGARCH DGP, an extensive Montecarlo dealing with different generators will be object of future researches).

With these functions we can apply at a first stage the usual analysis of Kupiec and Christoffersen and then use the loss function approach to compare the cost of different admissible choices. Clearly from a regulatory point of view this choice may not be worthwhile because regulator's objective is to reduce the risk of default in case of extreme events, position represented by the loss function of Lopez, but the function we propose represents the best choice for bank purposes, choosing a model that fulfills regulatory requirements (compare with the Basel agreement...) and allows for a lower cost. Considering the system in a whole these functions may help in choosing a model that is closer to the data and in the meantime fulfills MRA requirements With this choice we ensure both conditional and unconditional coverage, instead of choices of misspecified models that may lead to incorrect conditional coverage.

A GMM-based testing approach Recently Christoffersen, Hahn and Inoue (2001) introduced a new approach in the evaluation of Value-at-risk measures. In a general approach we can define the VaR via a quantile regression:

$$VaR_{m,t}\left(\alpha,\beta\right)=\beta_{1,m}\left(\alpha\right)+\beta_{2,m}\left(\alpha\right)\sigma_{t,m} \tag{14}$$

where the conditional volatility depends on the model we are using and parameters depend both on the model and on the significance level (coverage probability). Then we can state the following

Definition 1 *(CHI 2001 definition 1) The VaR is efficient with respect to the information set Ψ^{t-1} when*

$$E\left[I\left(\varepsilon_t<VaR_{m,t}\left(\alpha\right)\right)-p|\Psi^{t-1}\right]=0$$

where I(.) is the indicator function

Using then this efficient condition we can test if VaR measures satisfy it, but also we can compare different VaR even if misspecified. The methodology of the analysis requires conditioning on some information set, and the choice among different models. The first point is achieved considering as the information set at time t, as the measure of volatility in time t-1 obtained with the different models we are comparing and with a constant

$$E\left[\left(I\left(\varepsilon_t < VaR_{m,t}\left(\alpha,\beta\right)\right) - p\right) \times k\left(1,\sigma_{t-1,m1},\sigma_{t-1,m2},\sigma_{t-1,m3}...\right)\right] = 0 = E\left[f\left(\varepsilon_t,\beta\right)\right] \tag{15}$$

Specification testing is achieved using the test suggested by Kitamura and Stutzer (1997), the information theoretic alternative to a general method of moments (GMM) based test. Define the following quantity

$$M_T\left(\beta,\gamma\right) = \frac{1}{T}\sum_{i=1}^{T}\exp\left(\gamma' f\left(\varepsilon_t,\beta\right)\right)$$

and maximizing over the two parameter sets

$$\hat{M}_T\left(\hat{\beta}_T,\hat{\gamma}_T\right) = \max_{\beta}\min_{\gamma}\frac{1}{T}\sum_{i=1}^{T}\exp\left(\gamma' f\left(\varepsilon_t,\beta\right)\right)$$

then the Kitamura Stutzer test has the following equation

$$\kappa_T = -2T\log\left(\hat{M}_T\right) \rightarrow \chi^2\left(r-k\right) \tag{16}$$

where r is the number of conditioning information variables (constant included) and k is the dimension of the estimated parameters vector (β) in the quantile regression. The null hypothesis of this test is that the VaR measures satisfy the efficiency condition, therefore accepting the null will mean that the VaR model is correctly specified. In this approach we have, however, a challenge: the function $f\left(\varepsilon_t,\beta\right)$ is non-differentiable due to the presence of the indicator function. This problem will cause the traditional optimization techniques to burn down, requiring simulation based methods to estimate parameters or to employ generalized algorithm such as the simplex method or simulated annealing. This problem can be easily avoided in our case: considering that we focus on GARCH-type models, the VaR measure depends only on the evaluated conditional variance and on the coverage probability

$$VaR_{m,t}(\alpha,\beta) = \Phi^{-1}(\alpha)\,\sigma_{t,m} \tag{17}$$

using the cumulative standard normal inverse and excluding the effect of a constant. By this way we exclude the optimization over the parameters in the quantile regression and the traditional optimization routines can be used without problems.

Christoffersen et al. (2001) introduced another testing approach that allows to compare directly two different VaR measures. This test is based on the difference between two KLIC distances. If we consider two different VaR measures $m1$ and $m2$, and we define the KLIC respectively as

$$\hat{M}_{T,m1}\left(\hat{\beta}_T, \hat{\gamma}_T\right) \text{ and } \hat{M}_{T,m2}\left(\hat{\beta}_T, \hat{\gamma}_T\right)$$

Christoffersen et al. (2001) generalizing a result of Kitamura (1997) states the following:

Theorem 2 *(CHI theorem 1) Let*

$$\begin{aligned} M_{m1,T}(\beta_1^*, \gamma_1^*) &= \max_{\beta_1}\min_{\gamma_1} M_{m1,T}(\beta_1, \gamma_1) \\ M_{m2,T}(\beta_2^*, \gamma_2^*) &= \max_{\beta_2}\min_{\gamma_2} M_{m2,T}(\beta_2, \gamma_2) \end{aligned}$$

Under the null that $M_{m1}(\beta_1^*, \gamma_1^*) = M_{m2}(\beta_2^*, \gamma_2^*)$ *we have*

$$\sqrt{T}\left(\hat{M}_{T,m1}\left(\hat{\beta}_T, \hat{\gamma}_T\right) - \hat{M}_{T,m2}\left(\hat{\beta}_T, \hat{\gamma}_T\right)\right) \to N\left(0, \sigma_\infty^2\right)$$

where $\sigma_\infty^2 = \lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\exp\left(\gamma_1^{*\prime} f(\varepsilon_t, \beta_1^*)\right) - \exp\left(\gamma_2^{*\prime} f(\varepsilon_t, \beta_2^*)\right)\right)\right)$ *and the* T *subscript denote quantities computed with* T *observations instead of the infinite past.*

In this case the rejection of the null hypothesis will imply that the two measures do not match equally well the efficiency condition in favor of the model 2. When the null is accepted a positive measure implies the preference of model 1, a negative result the preference of model 2.

4 VaR and Long memory GARCH

We analyse the performances of tests and loss functions for the identification and choice the best model for VaR computation. We run a Montecarlo experiment dealing with a group of simulating DGP, eight FIGARCH with different orders and parameter values and a GARCH(1,1) used as a comparative test for evaluating the ability of tests and measures on Value-at-Risk when the DGP is a short memory one. The DGPs are described in the following table.

DGP	μ	ω	d	β	ϕ
FIGARCH(1,d,1)	0	0.01	0.8	0.5	0.3
FIGARCH(1,d,1)	0	0.01	0.8	0.5	0.05
FIGARCH(1,d,0)	0	0.01	0.8	0.5	0
FIGARCH(0,d,0)	0	0.01	0.8	0	0
FIGARCH(1,d,1)	0	0.01	0.1	0.4	0.5
FIGARCH(1,d,1)	0	0.01	0.4	0.3	0.2
FIGARCH(1,d,0)	0	0.01	0.4	0.3	0
FIGARCH(0,d,0)	0	0.01	0.4	0	0
GARCH(1,1)	0	0.01	0	0.65	0.3 (α)

In this experiment we act as the simulated series were daily series, simulating 2250 observations, using the first 2000 to estimate the model and the last 250 to assess the validity of Value at risk measures in a backtesting framework. In order to simulate the 2250 observations needed the recursion formulas we used generates an additional 2000 observations to avoid any dependence from initial values. We estimate on all simulated series 4 different models: the true DGP (for details on the simulation and estimation procedures and on the identification problem see Caporin 2002), a GARCH(1,1), an IGARCH(1,1) and an exponentially weighted moving average (EWMA, the well known RiskMetrics model), with smoothing parameter set to 0.97. The different specifications (apart the EWMA) are estimated with quasi maximum likelihood, moreover to induce a faster convergence of the estimate of the memory parameter d, in the FIGARCH specification, an additional presample of 2000 observations is used, this presample is set equal to the sample variance of the estimated series, following the approach of Teyssière (1996). Given the estimated parameters and conditional variances we compute VaR and then we test the correctness of these risk measures. We use the tests and loss functions described in the previous section. For all DGP we ran 1000 replications. The results are summarized in a large set of tables, here we present only a single DGP, all other tables are available from the author upon request. The tables are grouped with respect to the DGP and contain in the order (inside each group): the average number of exceptions across the 1000 replications, for each of the four fitted models, the standard deviation and the average percentage of exceptions; the frequency of less exception, that is, we count how many times each model is the one that gives a lower number of exceptions, note that the cumulate frequency can be above one since different models can lead to the same number of exceptions; the frequency of accepting the null hypothesis for the test of Unconditional Coverage (UC), Independence (I) and Conditional Coverage (CC); the frequency of model selection using Lopez loss function, that is we count how many times each model minimizes the loss function; the frequency of model selection with the alternative loss functions previously suggested, and their combinations, computed only on exceptions (E) or on the full sample (T); the results of the model comparison test of Christoffersen et al. (2000), we consider 4 different VaR p-levels (1%, 5%, 10% and 25% to compare results with the cited paper), and we report the frequencies of having a significant test statistics and the frequency of choice of the first or of the second model, all at confidence levels of 1%, 5% and 10%; finally the results of the model specification test of Christoffersen et al. (2000), again computed at the previous 4 VaR p-values and confidence levels. The tests developed by Christoffersen et al. (2000) presume a comparison of non-nested models; as we point out in Caporin (2002a) FIGARCH and GARCH (or IGARCH) most of the time are non-nested models, this let us compute the tests and perform the analysis. However GARCH, IGARCH and EWMA are nested models, therefore we expect significant results comparing long and short memory models, while we will have to take with care results among short memory specification. The following conclusions arise from the results:

Average exceptions and MRA. In most of the cases (excluding only the FIGARCH(0,d,0) with d=0.8) at 1% Value-at-Risk p-level, the RiskMetrics model is too conservative, leading to an average number (and percentage) of exceptions strictly below 2.5 (correspondent to 1%). This effect is present, even if with less evidence, also at 5% level and is influenced by the memory property of the generator: with higher memory (lower d) the RiskMetrics is much more conservative. This is probably due to the different structure of the two processes: in the FIGARCH case a bigger weight is given to past innovations, so there is a greater sensitivity to market movements, this implies a variance forecast with abrupt changes without signals of convergence of variances to an unconditional level, while in the RiskMetrics, a particular IGARCH model, the parameter configuration gives much more importance to shifts in the variances (the β parameter is 0.97) leading to gradual movements and slower convergence to the unconditional variance level. This effect remains also in GARCH and IGARCH specifications, since no constraints are imposed (except the one for positivity of variances) on the parameters, and this leads to an estimated β much smaller than 0.9. Comparing then FIGARCH, GARCH and IGARCH results we can see that they are very close showing that even a misspecified model can be good enough to fulfill MRA requirements, however we must remark that the forecasts obtained with misspecified models lead to incorrect conditional coverage. In all cases, on average, all models strongly satisfy the requirements of the amendment to Basle accord for market risks, leading to Green zone positioning (exceptions lower than 4). Considering now the frequencies of model selection in particular just the number of exceptions, the best choice is most of the time the EWMA, but this result is strictly related to the fact that this is the most conservative of the models, and is therefore of limited significance.

Tests of Conditional and Unconditional Coverage. As in the previous work of Lopez (1999), we find that these tests show no power in distinguishing among different models. All null hypothesis of correct unconditional or conditional coverage and of independence among exception are accepted with a percentage ranging from 75% to 100% at the 1% level of the test and for both 1% and 5% VaR. Results do not depend on parameter values. For the test at 5% significance level the null hypothesis is rejected with higher probability, especially for the Independence test, however this is true for all the 4 models, again we cannot infer on the best solution for our purposes.

Loss functions. We can observe that the Lopez loss function, given its formulation, depends crucially on the number of exceptions, this influences its value and therefore the model selection frequencies based on it. In all cases considered (again apart the FIGARCH(0,d,0) with d set to 0.8) the Lopez approach leads to the choice of the RiskMetrics as the best model for Value-at-Risk computation. This is in the sense that the best model is the one that minimizes the cost of an exception, it is a choice based on the risk of default, a choice driven by regulators objectives. However this does not imply that the best model is the true generator or the one that minimizes the cost for a private bank: as we can observe from figure (4) and (8) the EWMA has a smaller number of exceptions, since its VaR bands are much wider compared to the bands of the true generator, this can be interpreted as a higher cost for the bank, in fact the VaR level represents a minimal capital requirement that banks must hold on to cover market risks. Immobilizing this capital translates into an opportunity cost of liquidity resources, and reduces the operativity for the bank. A VaR based on the true generator meets the Basle MRA requirements and gives a correct conditional coverage for market risks, with narrower VaR bands. In spite of that none of the loss functions lead to a correct choice of the generator as the best model. All the functions considered, if applied only on the exceptions, select most of the times the EWMA, with percentage ranging fro 40% to 60%, second best choice switch between GARCH and IGARCH, in none of the cases the FIGARCH is chosen. Considering the whole sample the FIGARCH does not appear as the best model, even if its frequencies of selections increase. In this case the best choice switches between GARCH and EWMA, leading again a possible choice of a mis-specified model. Now this solution can be considered from a different point of view: should we prefer a model that minimizes the number of exceptions but imposes a greater opportunity cost, or would it be better to choose a model that is closer to the true generator, that satisfies in the meantime regulators requirements and allows for narrower VaR bands? The answer depends on the subject whom is posed: a regulator will surely prefer the first solution, while private banks will chose the second one. A consideration on the GARCH generator case: the model is correctly chosen with our alternative loss functions, but only if we consider the whole sample, not if IGARCH or EWMA is preferred.

Model comparison test. Now choices change. A first group of observation on the tables: the test is labeled as "not significant" when the two models equally well match the efficient moment condition, therefore the label "significant" is given to the rejection of the null hypothesis; we can observe that the null is accepted with high percentage when we compare very close models, that is the case of GARCH and IGARCH, when the GARCH parameters are close to the constraint $\alpha + \beta < 1$; the frequencies of selection of the first or of the second model are computed as percentages on the "significant" tests, they always sum to one, moreover I can always choose between the two models, provided I rejected the null, depending on the sign of the test statistics. All tables show a similar behavior, the EWMA model is never preferred to the DGP with a percentage greater than 40%, and most of the time this is true also for GARCH and IGARCH. This can be interpreted as a result of our observations on the correct conditional coverage given by the true generator, a condition that is extracted form the information set (here this is represented by the forecasts obtained with the four models in the past) by the estimation procedure. Moreover the true FIGARCH generator is preferred also to the GARCH and IGARCH with frequencies always above 50% in all cases considered.

Model specification test. In this case the test shows dependence on the VaR p-level, leading to very poor results, none of the models are correctly specified for the simulated series, for the 1% case, while for the remaining the percentage of accepting the null (the model is correctly specified) increases with p, with a jump from 1% to 5%. This may be due to the very limited number of exceptions in the 1% case, not sufficient to extract an indication on the ability of the model in matching the efficient moment condition. This result will probably change extending the backtesting period, however we will not pursue this point since we focus on the selection process of a model that should be analyzed by a regulator who uses 250 period for backtesting (see MRA).

We conclude this section with a word of advise on the results we obtained, compared to the ones of Christoffersen et al. (2001): we developed this Montecarlo on a backtesting approach in order to verify the power of the VaR specification test and VaR comparison test in the framework used by regulators following the MRA, that is on 250 observations. In this setup the number of exception is very limited and the size and power of the two test is affected: the tests are built on an efficiency condition that depend on an indicator function selecting exceptions, lower the number of exceptions lower will be the number of significant points used in (15)and in the tests. Moreover we want to stress that once the number of exceptions are the same in two or more models, the VaR specification test will lead to the very same result and the VaR comparison test will show clustered results including one or more groups of zeros. We tried, in a limited Montecarlo, to compute tests on the whole sample, results seem not to differ from the ones here presented, however an additional analysis in this direction will be necessary and left for future researches, but we stress that it must be developed as a suggestion for an alternative framework that will allow regulators to test the reliability of internal models, otherwise, with the current MRA, the results of this work apply.

5 A case study

In this section we will apply the tests presented in this work to the variance of a real series. Data were provided by the Italian Stock Exchange (Borsa Italiana S.p.A.). We used a one-year database of transaction data on the FIB30 market segment, ranging from 20 march 2000 to 15 march 2001. The FIB30 is the Future on the stock market index, the MIB30, that collect the first 30 firms for capitalization quoted in the Italian Stock Exchange. We extracted from the provided database the series of 5-minute log-returns and filtered them from a periodic daily component. A detailed discussion of the database and of the extraction and filtration process of the 5-minute series can be directly requested to the author. Here we focus our attention on a subset of filtered data, consisting in 2200 observations, recorded across february and march 2001. In Figure 1 we report a graph of the returns, while in Figure 2 the autocorrelations of returns and absolute returns (proxy for the volatility) are represented.

[Insert here Graph 1 and Graph 2]

As we can observe data clearly show a long memory behaviour in the variances, while in the mean there is limited evidence of an ARMA structure. On this set of data we fitted four different models, a FIGARCH(1,d,1), a GARCH(1,1) an IGARCH(1,1) and the EWMA with smoothing paramter 0.97. Estimation results are reported in table 8. All the models are estimated on the first 2000 observations, while the last 200 are used to compute the tests in a backtesting approach. Tests result are reported in table 9, where we consider Value-at-Risk at 1% and 5% coverage level. At first we consider the tests of unconditional coverage and conditional coverage: in all cases the null hypothesis of of correct coverage is accepted with the only exception (at 95%) of the CC test for the GARCH model and with VaR covergae level at 5%. If we consider the number of exceptions the EWMA seems to be the better choice, as confirmed by most of the loss functions. Alternatively the IGARCH(1,1) represents a good model (note that the α parameter for the EWMA is set to 0.7, very close to the estimated parameter for the IGARCH model). As noted in the Montecarlo study loss functions may lead to the choice of a misspecified model, clearly GARCH and IGARCH formulations do not take into account long memory. If we consider the VaR model specification and comparison tests results are different: all models are rejected when VaR coverage is at 1% and are accepted whit VaR coverage at 5%; when VaR p-level is 1% the model comparison test is significant only if we consider FIGARCH compared to alternatives and the preferred model is the long memory one in all cases; for the VaR at 5% coverage the model comparison test is significant in most of the cases and the preference is again for the FIGARCH.

6 Conclusions

In this paper we considered the computation of Value-at-Risk when the conditional variance follows a FIGARCH process. After presenting the recursion formula to evaluate the mean square error with a FIGARCH in the variance we reviewed the current techniques used to evaluate and compare Value-at-Risk measures, extending the loss function approach of Lopez (1998) with new loss functions and considering the whole sample and not only the exceptions. Our Montecarlo experiment shows that the regulators scheme, the tests of Kupiec and the loss function approach of Lopez (even with our extensions) lead to the choice of a misspecified model, switching between the GARCH and the RiskMetrics. This indicate that a simpler model is chosen but we will have incorrect conditional coverage and wider Value-at-Risk bands, which imply higher opportunity costs for banks. Different results are obtained with the tests of Christoffersen et al. (2001): the test of model comparison correctly chooses the long memory GARCH in all cases and interestingly the RiskMetrics is never considered as the best model even if compared with GARCH and IGARCH. Unfortunately the model specification test performs very poorly. Finally we applied the test and loss function to a real case, analysing the returns on the future of the italian stock market index, a series that show evident long memory behaviour in the variance.

References

[1] Baillie R.T and T. Bollerslev, 1992, Prediction in dynamic models with time-dependent conditional variances, *Journal of Econometrics*, 52, 91-113

[2] Baillie, R.T, Bollerslev, T., and Mikkelsen, H.O., (1996), "Fractionally integrated generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics,* 74, 3-30

[3] Bollerslev, T and Mikkelsen, H.O. (1996), "Modeling and pricing long memory in stock market volatility", *Journal of Econometrics,* 73, 151-184

[4] Basle Committee on Banking Supervision, 1996, Amendment to the capital market accord to incorporate market risks, Bank for International Settlement

[5] Basle Committee on Banking Supervision, 1996, Supervisory framework for the use of "Baktesting" in conjunction with the internal models approach to market risk capital requirements, Bank for International Settlement

[6] Caporin M., 2002a, FIGARCH processes: stationarity, estimation and the identification problem, GRETA working paper series 0202, www.greta.it

[7] Christoffersen P.F., 1998, Evaluating interval forecasts, *International economic review*, 39, 841-862

[8] Christoffersen P.F. and F.X. Diebold 2000, How relevant is volatility forecasting for financial risk management, *Review of Economics and Statistics*, 82-1, 12-22

[9] Christoffersen P.E., J. Hahn and A. Inoue, 2001, Testing and comparing value-at-risk measures, CIRANO working paper 2001s-03

[10] Beltratti A. and C. Morana, 1999, Computing Value at Risk with high frequency data, *Journal of empirical finance*, 6 (5), 431-455

[11] Diebold F.X., A. Hickman, A. Inoue and T. Schuermann, 1996, Converting 1-Day volatility to h-Day volatility: scaling by $\sqrt{h}$ is worse that you think, University of Pennsilvania working paper

[12] Drost F.C. and T.E. Nijman, 1993, Temporal aggregation of GARCH processes, *Econometrica*, 4, 909-927

[13] J.P. Morgan, 1996, *RiskMetrics Technical Document*, 4th edition, New York

[14] Kitamura Y., 1997, Comparing misspecified dynamic econometric models using nonparametric likelihood, University of Wisconsin, Madison, Department of Economics

[15] Kitamura Y., 1997, Information theoretic alternative to generalized method of moments estimation, *Econometrica*, 65, 861-874

[16] Kupiec P., 1995, Techniques for verifying the accuracy of value-at-risk measurement models, *Journal of Derivatives*, 3, 73-84

[17] Lopez A.J., 1999, Methods for evaluating value-at-risk estimates, *Federal Reserve Bank of San Francisco Economic Review*, 2

[18] Teyssière, G, 1996, *Double long-memory financial time series*, GREQAM working paper

Table 1: DGP FIGARCH(1,d,0) $d = 0.4$ $\beta = 0.3$

Average number of exceptions, standard deviation, *average percentage of exception* - 1000 replications – 250 forecasts				
	Fitted models			
	Figarch(1,d,0)	Garch(1,1)	Igarch(1,1)	EWMA(0.97)
1% VaR	2,606 1,641 *1,042*	2,704 1,848 *1,082*	2,270 1,516 *0,908*	1,152 1,086 *0,461*
5% VaR	12,771 3,548 *5,108*	12,867 4,191 *5,147*	11,671 3,221 *4,668*	11,449 2,967 *4,580*

The percentage beside VaR is the coverage rate.

Table 2: DGP FIGARCH(1,d,0) $d = 0.4$ $\beta = 0.3$

Frequency of less exceptions – 1000 replications – 250 forecasts				
	Fitted models			
	Figarch(1,d,0)	Garch(1,1)	Igarch(1,1)	EWMA(0.97)
1% VaR	0,243	0,278	0,319	0,946
5% VaR	0,352	0,229	0,274	0,634

The percentage beside VaR is the coverage rate. The table reports the frequency of selection of the models based on the number of exceptions: the sum by rows is greater than one since different models can lead to the same number of exceptions.

Table 3: DGP FIGARCH(1,d,0) $d = 0.4$ $\beta = 0.3$

Tests - Frequency of accepting H_0 – 1000 replications – 250 forecasts					
	α	Fitted models			
		Figarch(1,d,0)	Garch(1,1)	Igarch(1,1)	EWMA(0.97)
Test of Unconditional Coverage of Kupiec					
1% VaR	1%	0,995	0,989	0,997	1,000
	5%	0,900	0,869	0,892	0,684
5% VaR	1%	0,987	0,968	0,990	0,994
	5%	0,935	0,881	0,942	0,961
Test of Independence of Christoffersen-Lopez					
1% VaR	1%	0,779	0,780	0,744	0,623
	5%	0,313	0,370	0,286	0,336
5% VaR	1%	0,973	0,976	0,981	0,951
	5%	0,909	0,924	0,923	0,852
Test of Conditional Coverage of Christoffersen-Lopez					
1% VaR	1%	0,964	0,968	0,980	0,987
	5%	0,756	0,737	0,734	0,621
5% VaR	1%	0,970	0,960	0,977	0,966
	5%	0,895	0,859	0,897	0,857

The percentage beside VaR is the coverage rate. The table reports the frequency of acceptance of the null hypothesis of the different test considered.

Table 4: DGP FIGARCH(1,d,0) $d = 0.4$ $\beta = 0.3$

Lopez loss function – frequency of model selection 1000 replications – 250 forecasts				
	Fitted models			
	Figarch(1,d,0)	Garch(1,1)	Igarch(1,1)	EWMA(0.97)
1% VaR	0,094	0,123	0,115	0,923
5% VaR	0,094	0,065	0,038	0,803

The percentage beside VaR is the coverage rate. The table reports the frequency of model selection based on the loss function suggested by Lopez. The sum by rows is above one since GARCH, IGARCH and EWMA may lead to the very same 1-step-ahead forecast of conditional variance, this happen when the GARCH collapse on an IGARCH and/or when IGARCH α parameter is equal to 0.7

Table 5: DGP FIGARCH(1,d,0) $d = 0.4$ $\beta = 0.3$

Loss - Frequency of model selection – 1000 replications – 250 forecasts					
		Fitted models			
		Figarch(1,d,0)	Garch(1,1)	Igarch(1,1)	EWMA(0.97)
Loss Function 1: absolute value of return VaR measure ratio					
1% VaR	E	0,092	0,125	0,115	0,923
	T	0,052	0,193	0,472	0,283
5% VaR	E	0,043	0,170	0,189	0,598
	T	0,052	0,193	0,472	0,283
Loss Function 2: square return-VaR normalized by absolute VaR measure					
1% VaR	E	0,090	0,118	0,118	0,929
	T	0,066	0,339	0,000	0,595
5% VaR	E	0,009	0,081	0,021	0,889
	T	0,034	0,234	0,000	0,732
Loss Function 3: absolute of return-VaR					
1% VaR	E	0,089	0,119	0,115	0,932
	T	0,094	0,375	0,000	0,531
5% VaR	E	0,010	0,105	0,068	0,817
	T	0,073	0,331	0,000	0,596
Loss Function 1 + Loss Function 2					
1% VaR	E	0,093	0,124	0,115	0,923
	T	0,031	0,208	0,000	0,761
5% VaR	E	0,036	0,163	0,164	0,637
	T	0,004	0,036	0,000	0,960
Loss Function 1 + Loss Function 3					
1% VaR	E	0,092	0,125	0,115	0,923
	T	0,059	0,307	0,000	0,634
5% VaR	E	0,034	0,163	0,167	0,636
	T	0,021	0,096	0,000	0,883
Loss Function 2 + Loss Function 3					
1% VaR	E	0,089	0,118	0,114	0,934
	T	0,082	0,361	0,000	0,557
5% VaR	E	0,009	0,093	0,049	0,849
	T	0,051	0,302	0,000	0,647
Loss Function 1 + Loss Function 2 + Loss Function 3					
1% VaR	E	0,091	0,124	0,115	0,925
	T	0,065	0,324	0,000	0,611
5% VaR	E	0,030	0,154	0,141	0,675
	T	0,020	0,166	0,000	0,814

See footnote to table 4

Table 6: DGP FIGARCH(1,d,0) $d = 0.4$ $\beta = 0.3$

Test of model comparison – 1000 replications – 250 forecasts							
Frequencies of	α	Model comparison					
		1-2	1-3	1-4	2-3	2-4	3-4
VaR(1%)							
Test is significant	1%	0,552	0,606	0,785	0,510	0,778	0,724
	5%	0,553	0,607	0,787	0,510	0,778	0,725
	10%	0,554	0,608	0,787	0,510	0,778	0,726
Prefer 1st model	1%	0,478	0,711	0,925	0,747	0,919	0,870
	5%	0,479	0,712	0,925	0,747	0,919	0,870
	10%	0,478	0,712	0,925	0,747	0,919	0,869
Prefer 2nd model	1%	0,522	0,289	0,075	0,253	0,081	0,130
	5%	0,521	0,288	0,075	0,253	0,081	0,130
	10%	0,522	0,288	0,075	0,253	0,081	0,131
VaR(5%)							
Test is significant	1%	0,848	0,877	0,955	0,778	0,948	0,914
	5%	0,851	0,882	0,958	0,782	0,954	0,917
	10%	0,852	0,888	0,958	0,785	0,956	0,920
Prefer 1st model	1%	0,560	0,673	0,711	0,622	0,679	0,592
	5%	0,559	0,670	0,709	0,620	0,676	0,592
	10%	0,560	0,668	0,709	0,619	0,677	0,591
Prefer 2nd model	1%	0,440	0,327	0,289	0,378	0,321	0,408
	5%	0,441	0,330	0,291	0,380	0,324	0,408
	10%	0,440	0,332	0,291	0,381	0,323	0,409
VaR(10%)							
Test is significant	1%	0,910	0,939	0,983	0,860	0,977	0,970
	5%	0,916	0,944	0,985	0,864	0,980	0,973
	10%	0,919	0,948	0,986	0,864	0,980	0,974
Prefer 1st model	1%	0,575	0,649	0,635	0,610	0,598	0,513
	5%	0,575	0,648	0,636	0,611	0,598	0,513
	10%	0,575	0,649	0,636	0,611	0,598	0,513
Prefer 2nd model	1%	0,425	0,351	0,365	0,390	0,402	0,487
	5%	0,425	0,352	0,364	0,389	0,402	0,487
	10%	0,425	0,351	0,364	0,389	0,402	0,487
VaR(25%)							
Test is significant	1%	0,893	0,953	0,979	0,854	0,987	0,983
	5%	0,894	0,953	0,983	0,855	0,988	0,985
	10%	0,897	0,953	0,984	0,855	0,989	0,987
Prefer 1st model	1%	0,560	0,594	0,612	0,546	0,579	0,534
	5%	0,559	0,594	0,610	0,546	0,578	0,534
	10%	0,561	0,594	0,610	0,546	0,577	0,534
Prefer 2nd model	1%	0,440	0,406	0,388	0,454	0,421	0,466
	5%	0,441	0,406	0,390	0,454	0,422	0,466
	10%	0,439	0,406	0,390	0,454	0,423	0,466

The percentage beside VaR is the coverage rate. The table reports for each Value-at-Risk level the frequency of having a significant test (the two compared models differently satisfy the efficiency condition) and the frequency of preferring the first or the second model in percentage within the significant tests. The α represents the test confidence level.

Table 7: DGP FIGARCH(1,d,0) $d = 0.4$ $\beta = 0.3$

Test of VaR model specification (null: VaR(p) is correctly specified) Frequency of accepting H_0 – 1000 replications – 250 forecasts					
VaR p-value	Test α-value	Fitted models			
		Figarch(1,d,0)	Garch(1,1)	Igarch(1,1)	EWMA(0.97)
1%	1%	0,030	0,035	0,015	0,002
	5%	0,022	0,018	0,009	0,001
	10%	0,020	0,017	0,008	0,001
5%	1%	0,456	0,415	0,312	0,279
	5%	0,326	0,300	0,235	0,180
	10%	0,268	0,227	0,184	0,137
10&	1%	0,588	0,565	0,487	0,490
	5%	0,446	0,410	0,340	0,331
	10%	0,347	0,329	0,271	0,258
25%	1%	0,750	0,722	0,690	0,698
	5%	0,581	0,543	0,525	0,528
	10%	0,463	0,440	0,423	0,412

The table reports the frequency of accepting the null hypothesis.

Table 8: Fitted models on the FIB30 series

Model	Parameter	Estimate	Standard error	T-statistics
FIGARCH(1,d,1)	μ	-0.00255	0.00256	-0.99673
	ω	0.00062	0.00022	2.80426
	d	0.45070	0.08457	5.32932
	β	0.64779	0.06491	9.98031
	ψ	0.22977	0.04930	4.66105
GARCH(1,1)	μ	-0.00246	0.00259	-0.94998
	ω	0.00039	0.00014	2.83954
	α	0.06482	0.01200	5.40055
	β	0.91323	0.01697	53.80210
IGARCH(1,1)	μ	-0.00243	0.00257	-0.94412
	ω	0.00013	0.00005	2.86684
	α	0.06954	0.01281	5.42835

Table 9: Tests and loss functions computed on VaR bounds

	FIGARCH(1,d,1)	GARCH(1,1)	IGARCH(1,1)	EWMA(0.97)
Exceptions 1%	5	4	4	4
Exceptions 5%	13	18	12	11
CC 1%	3.209	1.565	1.565	1.565
CC 5%	0.869	5.502	0.397	0.102
UC 1%	5.245	4.313	4.313	4.313
UC 5%	4.227	8.093	4.529	5.125
Lopez 1%	5.103	4.134	4.089	4.098
Lopez 5%	13.367	18.417	12.311	11.333
F1 1% E	6.381	5.696	5.195	5.314
F1 1% T	66.210	71.602	63.930	64.825
F1 5% E	18.257	24.340	16.307	15.591
F1 5% T	84.410	84.985	81.458	83.608
F2 1% E	0.247	0.349	0.210	0.235
F2 1% T	122.272	117.557	126.890	125.554
F2 5% E	1.189	1.443	0.997	1.080
F2 5% T	99.882	97.274	102.556	101.775
F3 1% E	0.584	0.658	0.504	0.552
F3 1% T	102.836	97.015	108.407	106.819
F3 5% E	1.688	1.924	1.404	1.478
F3 5% T	73.722	69.790	77.434	76.351
VaR sig. 1%	55.170	87.066	87.066	87.066
VaR sig. 5%	6.658	10.410	6.681	9.557

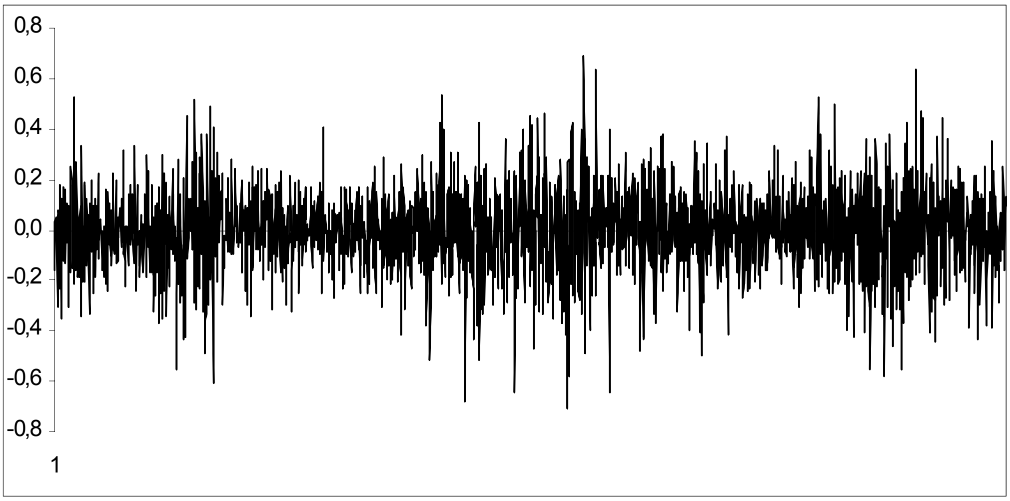
The percentage beside the descriptions indicates the VaR coverage level. CC stands for conditional coverage test while UC stands for unconditional coverage test. Lopez idicates the Lopez loss function, while with F1, F2 and F3 we indicate the loss functions suggested in this work. E stands for the loss functions computed only with the exceptions, while T stands for the whole sample. The last two rows reports the test statistics for the VaR model specification test. The χ^2 distribution has the following 1% (5%) critical values: k=1 (degrees of freedom) 6.635 (3.841); k=2 9.210 (5.991); k=5 15.086 (13.388).

Table 10: model comparison tests

Model comparison	VaR 1%	VaR 5%
FIGARCH vs GARCH	5.733	507.639
FIGARCH vs IGARCH	5.733	1.432
FIGARCH vs EWMA	5.733	155.604
GARCH vs IGARCH	0.000	-357.434
GARCH vs EWMA	0.000	-63.761
IGARCH vs EWMA	0.000	132.738

The table reports the test statistics for the VaR model comparison test. The test is distributed as a standardized normal and the null hypothesis is that the two model equally match the VaR efficiency condition. A positive sign indicatea preference for the first model.

Graph 1: FIB30 filtered logreturns - 2200 data points



Graph 2: Autocorrelations of the FIB30 filtered logreturns