Assessing GARCH model’s predictive ability from Traders’ Point of View*
(PRELIMINARY VERSION)

Rocco Mosconi
Politecnico di Milano

October 16, 2006

Abstract
This paper shows that, even if volatility is accurately predicted by correctly specified GARCH models, however such predictions are not very useful for traders when the conditional volatility does not vary "enough" over time, being therefore quite close to the unconditional one. It is shown that a low $R^2$ in the Mincer-Zarnowitz regression implies flat (although correctly predicted) volatility, and therefore leads to extremely poor performances of GARCH based investment strategies. Conversely, a high $R^2$ is associated with good performances of GARCH based trading strategies.

KEY WORDS: GARCH, volatility forecasting.

JEL Classification: ...

1 Introduction and notation

It is acknowledged in the literature that good volatility forecasts are crucial for the implementation and evaluation of asset and derivative pricing theories as well as trading and hedging strategies. Meanwhile, it is a well-established fact, dating back to Mandelbrot (1963) and Fama (1965), that financial returns display pronounced volatility clustering. Only over the last decade have financial economists begun to seriously model these temporal dependencies. While the vast majority of the earlier studies relied on the Autoregressive Conditional Heteroskedastic (ARCH) framework pioneered by Engle (1982), there is now a large and diverse time-series literature on volatility modeling. Almost universally, reported results point towards a very high degree of intertemporal volatility persistence; see, e.g., Bollerslev, Chou and Kroner (1992), Bollerslev,

Yet, in spite of highly significant in-sample parameter estimates, numerous studies find that standard volatility models explain little of the variability in ex-post squared returns:\footnote{See, e.g., Cumby, Figlewski and Hasbrouck (1993), Figlewski (1997), and Jorion (1995, 1996)} This evidence is usually illustrated by reporting extremely low $R^2$, like 0.05, obtained in the so called "Mincer-Zarnowitz regression", where squared excess returns are regressed on a constant and the predicted conditional variance. This has led to the suggestion that these models may be of limited practical value. Andersen and Bollerslev (1997) strongly argue against this conclusion, showing that if GARCH models are correctly specified, they can provide excellent predictions of volatility, and this is perfectly compatible with poor predictions of realized squared returns.

I fully agree with the point made in Andersen-Bollerslev that volatility is very well predicted by GARCH models when they are correctly specified. However, I believe that if volatility does not vary "enough" over time, and therefore the conditional volatility is quite close to the unconditional one, then we should expect almost no gain in using GARCH models in trading strategies, with respect to considering the variance as constant. I will show in the following that a low $R^2$ in the Mincer-Zarnowitz regression implies flat (although correctly predicted) volatility, and therefore leads to extremely poor performances of GARCH based investment strategies. Notice that, with the sample size often encountered in modelling financial data, the volatility could be "almost constant" even in the presence of highly significant parameter estimates.

Through the paper we will use the following notation. Daily returns $r_t$ are defined as

$$r_t = p_t - p_{t-1}$$

where $p_t$ is the logaritmic price of a financial asset at the end of day $t$ ($t = 1, ..., T$). $r_t$ is assumed to be generated by a Gaussian GARCH(1,1) model with constant expected return:\footnote{It is my opinion that most of the conclusion reached in this paper can be extended to other discrete time volatility models: this is left for further research.} so that the conditional distribution of $r_t$ is given by

$$r_t \sim N \left( \mu, \sigma_t^2 \right)$$

$$\sigma_t^2 = \omega + \alpha (r_{t-1} - \mu)^2 + \beta \sigma_{t-1}^2$$

It is assumed that $\alpha + \beta \leq 1$: the IGARCH case is not discussed in this paper.

We denote by $\hat{\mu}$, $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\beta}$ the Maximum Likelihood estimates of $\mu$, $\omega$, $\alpha$ and $\beta$, and by $\hat{\sigma}_t^2$ the estimate of $\sigma_t^2$.

The paper is organized as follows. In Section 2 we briefly summarize the debate on the predictive ability of GARCH models. Section 3 introduces a simple simulation exercise, aimed at showing under which circumstances the prediction from GARCH models are useful for setting up profitable trading rules. Section 4 describes the results of the simulation run under the assumption that
the investor knows the parameters of the GARCH model; some attention is
devoted, under this assumption, to the role of transaction costs. Section 5
illustrates .... Section 6 illustrates, using daily data from NYSE, one situation
where GARCH seems to provide useful information for traders and another
in which they do not. Section 7 summarizes and concludes. Two appendixes
illustrate some technical aspects of the paper.

2 The debate on predictive ability of GARCH
models

The majority of the volatility forecast evaluations reported in the literature rely
on some MSE criteria involving the ex-post squared or absolute returns over the
relevant forecast horizon. One particularly popular metric is obtained via the
ex-post squared return - volatility regression:

\[ (r_{t+1} - \hat{\mu})^2 = a + b\sigma_{t+1}^2 + u_{t+1} \]  

This regression equation provides an analogue to the commonly used procedure
for evaluating forecasts for the conditional mean, frequently referred to as a
Mincer-Zarnowitz regression (Mincer and Zarnowitz, 1969). If the model for the
conditional mean and the conditional variance is correctly specified, it follows
that, in population, \( a \) and \( b \) equals zero and unity, respectively. Of course, in
practice the values for \( \hat{\mu} \) and \( \hat{\sigma}_{t+1}^2 \) are subject to estimation error, resulting in
a standard errors-in-variables problem and a downward bias in the regression
estimate for \( b \). Nonetheless, the coefficient of multiple determination, or \( R^2 \),
from the regression in (1) provides a direct assessment of share of variability of
the ex-post squared excess returns that is explained by the particular estimates
of \( \hat{\sigma}_{t+1}^2 \). The \( R^2 \) is therefore often interpreted as the degree of predictability in
the volatility process, and hence of the potential economic significance of the
volatility forecasts.

Andersen and Bollerslev (1997) extensively report the findings in the litera-
ture making use of the Mincer-Zarnowitz regression, with minor differences
depending on the type of speculative return, sample period, data frequency, the
\( R^2 \) ranges from 0.001 to 0.106. The typical conclusion in the literature is that
these systematically low \( R^2 \)’s indicate that volatility models may be seriously
misspecified, provide poor volatility forecasts, and consequently are of limited,
if any, practical use. Andersen and Bollerslev (1997) strongly argue against this
conclusion. Building on Bollerslev (1986) they show that, if the data are gener-

3Day and Lewis (1992), Pagan and Schwert (1990), Jorion (1996), Cumby, Figlewski
and Hasbrouck (1993), West and Cho (1995), Akiray (1989), Boudoukh, Richardson and
Whitelaw (1997), Brailsford and Faff (1996), Canina and Figlewski (1993), Dimson and Marsh
(1995), Lamoureux and Lastrapes (1993), Schwert (1989, 1990a) and Schwert and Seguin
ated by a GARCH(1,1) model with finite fourth moment, then the population $R^2$ in (1) is given by

$$R^2 = \alpha^2 (1 - \beta^2 - 2\alpha \beta) \leq \kappa^{-1}$$

where $\kappa$ is the conditional kurtosis. Therefore with conditional Gaussian errors the $R^2$ from a correctly specified GARCH(1,1) model is bounded from above by $\frac{1}{4}$, while with conditional fat-tailed errors the upper bound is even lower. Moreover, with realistic parameter values for $\alpha$ and $\beta$, the population value for the $R^2$ statistic is significantly below this upper bound. For example, if the Data Generating Process is a GARCH(1,1) with Gaussian conditional distribution, $\alpha = 0.10$ and $\beta = 0.85$, the population $R^2$ would be 0.093. In other words, low $R^2$'s are not an anomaly, but rather a direct implication of standard volatility models.

The point in Andersen-Bollerslev (1997) is that correctly specified volatility model, despite giving very poor predictions of the realized squared excess return $(r_{t+1} - \mu)^2$, still can provide excellent forecasts of its ex-ante expectation, namely volatility. Their intuition is that, since

$$(r_t - \mu)^2 = \sigma_t^2 z_t^2,$$

in order to assess the ability of a model to predict $\sigma_t^2$ one should "clean up" $(r_t - \mu)^2$ from the idiosyncratic term $z_t^2$. In their paper, they suggest a procedure based on the availability of higher frequency data. Starting from continuous-time volatility modelling, they claim that a way less noisy measure of the realized volatility is given by the cumulative sum of squared higher frequency returns. Using DM-US$ exchange rate data for 260 weekday returns from October 1, 1992 to September 30, 1993, they show that when the squared daily excess return on the left hand side of (1) is replaced by the sum of 288 5-minutes squared excess returns in the same day, the $R^2$ rises from 0.047 to 0.479. Their conclusion is that ARCH models, contrary to the above contention, produce strikingly accurate interdaily forecasts for the latent volatility factor that is relevant for most financial applications.

We fully agree with the point made in Andersen-Bollerslev that volatility is very well predicted by GARCH models when they are correctly specified. However, we believe that indicators like the $R^2$ in the Mincer-Zarnowitz regression play some role, at least for trading strategies. The point here is extremely simple: a low $R^2$ implies that conditional volatility is almost constant over time (with respect to squared excess returns), and therefore quite close to the unconditional volatility: if this is the case we should expect almost no gain in trading strategies using GARCH models with respect to assuming constant variance.

Unconditional kurtosis is finite when

$$\kappa \alpha^2 + \beta^2 + 2\alpha \beta < 1$$

where $\kappa$ is the conditional kurtosis.

For example, squared excess return in day $t$ is replaced by the sum of 24 hourly squared excess returns in the same day.
3 An investment strategy based on GARCH

The predictive ability of GARCH models is assessed here through a simulation experiment. The setting is deliberately simple: an investor can choose among one risky asset and one risk free asset; he invests one monetary unit for $T$ time units (days); the portfolio can be modified any day, $\gamma_t$ being the weight of the risky asset in day $t$ ($0 \leq \gamma_t \leq 1$). The return from the risk free asset is $\mu_f$, while the return from the risky asset is assumed to follow a GARCH(1,1) process:

$$r_t \sim N(\mu, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha (r_{t-1} - \mu)^2 + \beta \sigma_{t-1}^2$$

When $\alpha + \beta < 1$, the unconditional variance is given by

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}$$

The first two moments of the portfolio are given by:

$$E(r_p^t) = (1 - \gamma_t) \mu_f + \gamma_t \mu$$

$$Var(r_p^t) = \gamma_t^2 Var(r_t)$$

Assume that the target portfolio standard deviation is given by $\delta \sigma$, where $0 \leq \delta \leq 1$. Two investment strategies are considered, depending on which notion of $Var(r_t)$ is used, unconditional or conditional:

$$\gamma_t^B = \delta$$

$$\gamma_t^G = \delta \sqrt{\frac{\sigma^2}{\omega + \alpha \sigma_{t-1}^2 + \beta \sigma_{t-1}^2}}$$

We denote by

$$r_t^B = (1 - \gamma_t^B) \mu_f + \gamma_t^B r_t$$

$$r_t^G = (1 - \gamma_t^G) \mu_f + \gamma_t^G r_t$$

the return of the base (B) and GARCH (G) strategies respectively. We want to measure the gain obtained by implementing the more sophisticated GARCH strategy with respect to the naive base strategy, under different plausible assumptions on the degree of persistence of volatility. If there is any gain, then we conclude that it is worth for investors to forecast volatility.

In the simulations, we will maintain the following

$$\mu_f = 0.0001$$

$$\mu = 0.0003$$

$$\sigma = 0.01$$

$$\delta = 0.5$$

$$T = 5,000,000$$

---

$^6$IGARCH is not addressed in this paper, but we will discuss explicitly GARCH with no finite fourth moment.
The values for $\mu_f$ and $\mu$ correspond on yearly basis to a return of roughly 0.025 and 0.075 for risk free and risky asset, while the value for $\sigma$ corresponds on early basis to a standard deviation of roughly 0.16; notice that fixing $\sigma = 0.01$ means that $\omega$ is determined as a function of $\alpha$ and $\beta$ as

$$\omega = \sigma^2 (1 - \alpha - \beta).$$

The values chosen for $\alpha$ and $\beta$ will be discussed in the next Section. It is important to remark that, to avoid unnecessary randomness, the standard Gaussian error driving the process will be the same for any value of $\alpha$ and $\beta$.

4 Benefit from GARCH when the parameters are known

In this Section, two different exercises are carried on; in both of them the hypothesis of perfect knowledge of the parameters is maintained (it will be removed in the next Section). In the first exercise transaction costs are assumed out: this hypothesis is clearly the most favourable for the GARCH strategy, which implies that the portfolio is rebalanced everyday. The results show that it is always preferable to adopt the GARCH strategy, since it gives an higher expected return with the same volatility. However, the benefit may be very close to zero for some values of $\alpha$ and $\beta$, even very far away from zero. In the second exercise we make an assumption on the transaction costs, inspired by the current market conditions in Italy. It is show that unless a huge amount of money is invested, transaction costs may cancel the benefit coming from adopting the GARCH strategy. This obviously holds even more true for those values of $\alpha$ and $\beta$ where the benefit of the GARCH strategy is small.

4.1 No transaction costs

The values for $\alpha$ and $\beta$ used in the simulation are given in Table 1, along with some of the results. Some explanation is needed to read the Table.

- The column headed $E(r^4)$ gives the result of the formula $\kappa \alpha^2 + \beta^2 + 2\alpha\beta$, where $\kappa = 3$ is the conditional kurtosis. This figure must be less than 1 as a condition of existence of finite fourth moments. Notice that in cases 5, 10, 15 and 20 there are no finite fourth moments.

- The column $R^2_{\text{theo}}$ gives the result of the formula $\frac{\alpha^2}{1 - \beta^2 - 2\alpha\beta}$; when finite fourth moments exist, this figure represent the population $R^2$ in the Mincer-Zarnowitz regression (1), otherwise it does not have any precise meaning. Notice that for each of the four different values of $\alpha$ (0.05, 0.10, 0.20 and 0.22), five increasing values of $\beta$ are chosen such that $\frac{\alpha^2}{1 - \beta^2 - 2\alpha\beta}$ takes on the values 0.05, 0.10, 0.20, 0.30 and 0.60. Actually, in the latter case the finite fourth moment condition is violated, and therefore 0.60
Table 1: Values of alfa and beta

cannot be interpreted as population $R^2$, since the $R^2$ is undefined. The values of $\beta$ are given with 6 decimals to minimize the rounding error in $R^2$.

- The column $R^2_{emp}$ gives the empirical $R^2$ when regression (1) is run on the $T = 5,000,000$ simulated data, with $\hat{\sigma}_t^2$ replaced by $\sigma_t^2$ since the parameters are known. Notice that when $R^2$ approaches the upper bound $\frac{1}{3}$, $R^2_{emp}$ becomes significantly smaller than $R^2_{theo}$. The distance becomes extremely large when fourth moments do not exist.

- The columns $E(\tau^B_i)$ and $E(\tau^G_i)$ give the average return on the $T = 5,000,000$ returns from the base and GARCH strategies respectively; actually, to improve readability the averages have been multiplied by 250 to transform them into annual returns. Not surprisingly, $E(\tau^B_i)$ is always approximately 0.05: in fact, the risk free asset’s return is 0.0001*250=0.025, the risky asset’s expected return is 0.0003*250=0.075, and $\delta = 0.5$.

- Finally, $\Delta = E(\tau^G_i) - E(\tau^B_i)$ is the gain from adopting the GARCH strategy. It ranges from about 0.001 to about 0.018. The primary goal of the analysis is to discuss what $\Delta$ depends on. It seems rather clear that $\Delta$ grows as a function of $R^2$, almost linearly up to $R^2 = \frac{1}{3}$, and then even faster in the area where the fourth moments are undefined.
When $R^2 = 0.05$ - a figure very close to those reported by Andersen and Bollerslev (1997) - irrespective of the value of $\alpha + \beta$, $\Delta$ is very small. For example, in case 1 $\alpha + \beta = 0.976$, a figure that might suggest strong persistence, which indeed is there: however, the low $R^2$ implies that excess squared returns have an high variance with respect to the variance of $\sigma_t^2$, and therefore are highly unpredictable, leading to a poor performance of the investment strategy. On the opposite side, case 19 has a smaller $\alpha + \beta$, but a higher $R^2$, which delivers a much better performance.

Table 2 gives more insight on the advantages of the GARCH strategy. Both the base and the GARCH strategy seem to be able to control volatility in the portfolio, reaching the target $\sigma_p = \delta \sigma$. Therefore, the gain in the expected return is obtained at no cost in terms of volatility. Moreover, the Table illustrate the considerable advantages of the GARCH strategy in terms of regularity of the return distribution. When the GARCH strategy is adopted, skewness and kurtosis of the portfolio returns are perfectly Gaussian up to the fourth moment. Conversely, the returns from the base strategy suffer from some skewness and excess kurtosis. Notice that the non normality problems in the base strategy

$$\delta \sigma \sqrt{250} = 0.079057.$$
seems also strongly dependent on \( \frac{\alpha^2}{1-\beta^2-2\alpha\beta} \), being negligible when this is 0.05, and severe when it is 0.6.

4.2 The Impact of transaction costs

Let us now discuss the role of transaction costs. Clearly, when the parameters of the system are known, the base strategy implies to keep the portfolio fixed for all the duration of the investment, and therefore does not generate any transaction cost. On the contrary, the GARCH strategy requires to change the portfolio every day (not necessarily major changes). We make an assumption on the transaction costs, inspired by the current market conditions in Italy, and discuss the results.

STILL TO BE WRITTEN, BUT THE RESULTS ARE THERE.

It is show that unless a huge amount of money is invested, transaction costs may cancel the benefit coming from adopting the GARCH strategy. This obviously holds even more true for those values of \( \alpha \) and \( \beta \) where the benefit of the GARCH strategy is smal.

5 Estimated parameters

Maximum Likelihood estimates of the parameters of the GARCH model have been proved to be consistent and asymptotically normal (see Lumsdaine, 1991). However, the properties in finite samples have not been fully studied. Simulation experiments easily show that the estimates of the parameters in the GARCH(1,1) model are very inefficient and significantly biased even in relatively large samples (say \( T = 250 \) or \( T = 500 \)), the bias being upwards for \( \omega \) and \( \alpha \), and downwards for \( \beta \). Moreover, due to the large variance, the estimated sum \( \hat{\alpha} + \hat{\beta} \) may well be below zero or above one. We want to discuss in this Section if the poor properties of finite sample estimates might reduce the advantage of using GARCH based trading strategies with respect to simple strategies based on unconditional variances.

THE SIMULATION HAS TO BE COMPLETED AND THE RESULTS HAVE TO BE WRITTEN.

6 Empirical illustration

In this Section I will illustrate, using daily data from NYSE, one situation where GARCH seems to provide useful information for traders and another in which it does not.

STILL TO BE WRITTEN

7 Summary and conclusions
References


A  Finite sample bias in GARCH models