Why does the GARCH(1,1) model fail to provide sensible longer-horizon volatility forecasts[1]

Stefano Herzel[2], Tomas Nord[3], and Cătălin Stărică[4]

First draft: July 2004

This version: April 2005

[1]This research has been supported by a grant from the Handelsbanken Research Foundation.

[2]Department of Economics, University of Perugia, Italy. Email: herzel@unipg.it

[3]Department of Applied Physics, Chalmers University of Technology, S–412 96, Gothenburg, Sweden. Email: nord@fy.chalmers.se.

[4]Department of Mathematical Statistics, Chalmers University of Technology, S–412 96, Gothenburg, Sweden. Email: starica@math.chalmers.se. Web page: www.math.chalmers.se/˜starica

Abstract

The paper investigates from an empirical perspective aspects related to the occurrence of the IGARCH effect and to its impact on volatility forecasting. It reports the results of a detailed analysis of twelve samples of returns on financial indexes from major economies (Australia, Austria, Belgium, France, Germany, Japan, Sweden, UK, and US).

The study is conducted in a novel, non-stationary modeling framework proposed in Stărică and Granger (2005). The analysis shows that samples characterized by more pronounced changes in the unconditional variance display stronger IGARCH effect and pronounced differences between estimated GARCH(1,1) unconditional variance and the sample variance. Moreover, we document particularly poor longer-horizon forecasting performance of the GARCH(1,1) model for samples characterized by strong discrepancy between the two measures of unconditional variance. The periods of poor forecasting behavior can be as long as four years. The forecasting behavior is evaluated through a direct comparison with a naive non-stationary approach and is based on mean square errors (MSE) as well as on an option replicating exercise.

JEL classification: C14, C16, C32.

Keywords and Phrases: stock returns, volatility forecasting, GARCH(1,1), IGARCH effect, hedging

1. Introduction

The GARCH conditional modeling framework often produces evidence that the conditional volatility process is highly persistent. In the case of the simple GARCH(1,1) process

$$r_t = z_t h_t^{1/2}, \quad h_t = \alpha_0 + \alpha_1 \, r_{t-1}^2 + \beta_1 \, h_{t-1},$$

(where z_t are iid, $Ez = 0$, $Ez^2 = 1$) this translates in the sum of the coefficients α_1 and β_1 being statistically equal to one, i.e. the so-called integrated GARCH (IGARCH) effect. As a consequence, this methodology suggests as a data generating process for returns a stationary model with an *infinite* second moment and in which shocks have a *permanent* effect on volatility. This last assumption has a serious impact on volatility forecasts: current information remains relevant when forecasting the conditional variance for all horizons.

This paper is motivated by growing empirical and theoretical evidence that the IGARCH effect might be an artifact due to structural changes in the unconditional variance process. The possible causal relation between non-stationarities and the IGARCH effect is a recurrent theme in the financial econometric literature (see Lamoureux and Lastrapes (1990), Hamilton and Susmel (1994), Cai (1994) among others) and can be traced back to Diebold (1986). Mikosch and Stărică (2004) show theoretically that, at least in the frame of the Whittle estimation, the IGARCH effect can be due to the behavior of the estimators under mis-specification. More concretely, they show that estimating a Garch(1,1) model on a sample displaying non-stationary changes of the unconditional volatility produces the IGARCH effect.

The aim of this paper is to investigate from an empirical perspective aspects related to the occurrence of the IGARCH effect and to its impact on volatility forecasting. The paper reports the results of a detailed analysis of twelve samples of returns on financial indexes from major economies (Australia, Austria, Belgium, France, Germany, Japan, Sweden, UK, and US) (see Table 1 for details).

The investigation is conducted in a novel, non-stationary modeling framework proposed in Stărică and Granger (2005) (see Section 2). There the authors argue that modeling the returns as non-stationary sequence of independent random variables with time-varying unconditional variance describes the dynamics of the S&P 500 log-returns better than GARCH-type or long memory-type models. As in Stărică and Granger (2005), we interpret the presence of significant autocorrelations in the absolute (square) values of returns as evidence of non-stationary changes in the unconditional second moment of the series of returns. Consistent with this interpretation, the success of the estimation is evaluated based on the removal of the long-memory look of the sample autocorrelation sample (SACF) in the absolute returns standardized with the estimated time-varying standard deviation.

The novel, non-stationary framework of our analysis is essential for the study of the impact of structural changes of the unconditional variance on the sum of the estimated GARCH(1,1) coefficients. Its use for the study of a large number of world's most important stock indexes is one of the original contributions of the paper. By successfully modeling twelve time series, our analysis brings further evidence that the framework developed Stărică and Granger (2005) is a viable set-up for the analysis of the dynamics of stock returns. Moreover, the analysis shows that samples characterized by more pronounced changes in the unconditional variance display stronger IGARCH effect.

The second goal of the paper is to investigate the impact of the IGARCH effect on volatility forecasting. The assumption of an integrated conditional volatility model has an heavy impact on volatility forecasts since an integrated data generating process implies that current information will be relevant when forecasting the conditional variance for all horizons. Stărică (2003) showed that the GARCH(1,1) model fails[5] to provide sensible longer-horizon[6] volatility forecasts on sub-samples of returns on the S&P500 index characterized by IGARCH effect. The author argues that the particularly poor forecast performance of the GARCH(1,1) model is due to the poor estimation of the unconditional volatility of the data caused by the IGARCH effect.

In this paper, we investigate volatility forecasting performance of the GARCH(1,1) model on different time series with and without the IGARCH effect. We confirm the findings in Stărică (2003) and show that the poor GARCH(1,1) forecasting performance reported there is widespread and, hence, not specific to the S&P 500 index. More specifically, we empirically identify periods of strong discrepancy between the estimated GARCH(1,1) unconditional volatility and the sample standard deviation in nine of the twelve series under scrutiny. For the samples characterized by the worse GARCH(1,1) mis-estimation of the variance (due to the IGARCH effect) we document particularly poor forecasting performance of the GARCH(1,1) model. On sub-samples

[5]The MSE error of the GARCH(1,1) model forecasts at 3 (6 respectively) month horizon were 2 (3 respectively) times bigger than those of the naive forecast that takes the past year's volatility as future volatility.

[6]The construction of long-horizon volatility forecasts are essential in many asset-pricing models.

not affected by the IGARCH effect, the longer-horizon volatility forecast performance of the GARCH(1,1) model is satisfactory.

The paper brings two other original methodological contributions. First concerns the statistical estimation of the time-dependent unconditional volatility in the non-stationary framework. The time-varying second moment of returns is estimated using the innovative non-parametric statistical methodology of Adaptive Weights Smoothing (AWS) proposed by Polzehl and Spokoiny (2003) (Section 3). Second contribution concerns the evaluation of the longer-horizon volatility forecasting performance. The GARCH(1,1) model is compared with a simple forecasting approach which assumes the volatility locally constant. The comparison cover horizons from one day to one business year and is done using two different measures. The first one is the classical mean square errors (MSE) of the variance forecasts.

The second, innovative approach compares the financial consequences of using the two volatility forecasts for pricing and hedging simple financial derivatives on indexes. This comparison is motivated by the observation that "a natural criteria for choosing between any pair of competing methods to forecast the variance of the rate of return on an asset would be the expected incremental profit from replacing the lesser forecast with the better one", as stated by Engle et al. (1993). The two volatility forecasts from the first comparison are employed to determine the initial prices of the replicating portfolios of at-the-money options as well as the dynamic strategies to be followed in hedging. Although motivated by the same idea, our approach differs in many ways from that in Engle et al. (1993) and (1997) (see Section 6 for details). More concretely, we focus on evaluating the ability of two competing modeling methodologies to help an investor to implement a dynamic strategy that replicates a given claim. The quality of the volatility forecasts of competing models is measured at the expiration. Our approach is based on the observation that more accurate volatility forecasts lead to smaller replication errors.

The paper is organized as follows. Section 2 introduces the modeling set-up of our non-stationary analysis of the dynamics of returns on stock indexes. Section 3 describes the non-parametric statistical methodology used in estimation of the time-varying second unconditional moment of returns. Section 4 presents the results of volatility estimation for the twelve series of indexes and assesses the goodness-of-fit of the non-stationary, unconditional approach. In Section 5 various sub-sample-specific measures of volatility and GARCH(1,1) modeling features are analyzed. The aim is to produce sub-sample-specific measures of the IGARCH effect as well as of the amount of change of the unconditional variance estimated in the previous section. Section 6 investigates the forecasting performance of the GARCH(1,1) model on specific sub-samples identified in the previous section while Section 7 concludes.

2. Non-stationary, unconditional modeling of index returns

In this section we introduce the modeling set-up of our in-depth analysis of the dynamics of stock indexes. Following the approach of Stărică and Granger (2005), the returns on a financial index (r_t) are described as

$$r_t = \sigma(t) z_t, \quad t = 0, 1, \ldots \tag{2.1}$$

where (z_t) is a sequence of *i.i.d.* random variables, $Ez = 0$, $Ez^2 = 1$ and $\sigma(t)$ a positive function of t. In the sequel, this function will be approximated by a step function, yielding a model with a piecewise constant variance. We are assuming the mean of the return to be zero. Working with de-meaned returns $r_t - \bar{r}$ does not change in any way the results of the analysis.

Equation (2.1) can be re-written as

$$r_t^2 = \sigma^2(t) + \widetilde{z}_t, \quad t = 0, 1, \ldots \tag{2.2}$$

where $\widetilde{z}_t = \sigma^2(t)(z_t^2 - 1)$, with $E\widetilde{z} = 0$, or like in Stărică and Granger (2005) as

$$\log |r_t| = \log \sigma(t) + \log |z_t|, \quad t = 0, 1, \ldots. \tag{2.3}$$

Note that both the equations (2.2) and (2.3) fit in the general non-parametric regression set-up

$$y_t = \mu(t) + s(t)\varepsilon_t, \quad t = 1, 2, ..., n, \tag{2.4}$$

where the time-varying trend μ and variance s^2 could be continuous or display jumps, the noise (ε_t) is assumed i.i.d. with zero mean and unit variance, not necessarily Gaussian. Hence the volatility function $\sigma^2(t)$ can be directly estimated using non-parametric smoothing techniques (to be discussed in the next section).

In words, the returns are modeled as *independent random variables* with a time-varying unconditional variance. They form a non-stationary sequence, free of any dependency[7] but with a

[7] Independent non-stationary sequences can display significant sample ACF. In particular, the long memory in volatility effect can occur for independent sequences with a time-varying unconditional variance. For more details on this issue, see Mikosch and Stărică (2004).

marginal distribution that evolves through time. Moreover, the only changing probabilistic feature of the marginal distribution is the unconditional variance. Consequently, the logarithm of the absolute returns are described as stochastic variations around a *time-varying trend* or expected level μ. An innovative statistical methodology, which we describe in Section 3, is used to produce a piecewise-constant approximation of the function μ (σ, respectively). In the analysis of the returns on the twelve indexes both equation (2.2) and (2.3) were used as basis for the estimation of time-varying second moment of the returns. The estimation results were identical. As we will see in Section 4, even the rough approximation of the variance dynamics by a step function is sufficient to explain most of the dependency structure present in the sample ACF of absolute return series, hence providing an explanation for the so called "long memory in volatility" effect.

The non-stationary framework allows for estimation of a time-varying second unconditional moment, an essential step for the study of the impact of structural changes in the volatility on the sum of the estimated GARCH(1,1) coefficients.

We continue with the description of the methodology used to estimate the function μ.

3. Non-parametric volatility estimation: Adaptive Weights Smoothing (AWS) methodology

In this section we describe the statistical methodology to be used in estimation of the volatility in the non-stationary modeling framework described in Sections 2. The main goal of our analysis is estimation of the time-varying volatility which appears in either of the two alternative forms, (2.2) or (2.3), of the model (2.1) as the trend function μ in (2.4).

3.1. **Local constant approximation of μ and s.** Our approach does not impose any global structural (parametric) assumption on the functions μ and s. Instead, we assume the following *local parametric structure*: for every time point t there exists a time interval around t in which the data can be well approximated by a simple parametric model. In this paper, both the variability s and the trend μ are locally approximated by constants, yielding step function approximations of the two functions of interest.

The statistical procedure we are about to describe focuses on constructing intervals where a parametric, stationary model provides a good approximation to the unknown true data generating process. On these intervals, called *intervals of homogeneity*, the parameters of the model can be consistently estimated. The size of these intervals is referred to as *degree of locality*.

One possible approach to building the homogeneity intervals consists in selecting a *bandwidth* h and in estimating the functions μ and s on the time window $[t-h, t+h]$ using the approximating model equation

$$y_u = \mu(t) + s(t)\varepsilon_u, \qquad u \in [t-h, t+h].$$

Such a simple parametric model with unknown coefficients $\mu(t)$ and $s(t)$ can be estimated using the standard least squares approach. The degree of locality is here determined by the bandwidth h. Its choice is crucial in applications. A small h means that only few data points are used for estimating the unknown parameters leading to insufficient noise reduction, while selection of a large bandwidth h may lead to a substantial bias due to the poor approximation to the true function μ that a constant model might provide on the long window $[t-h, t+h]$. The choice of the optimal bandwidth hence depends on the unknown shape of the function to estimate. A fixed bandwidth can be too restrictive for the analysis we aim to perform, hence we choose an approach with a bandwidth self-adapting to the data.

3.2. **Adaptive smoothing.** A more flexible approach, the Adaptive Weights Smoothing (AWS), was introduced in Polzehl and Spokoiny (2000) in the context of image de-noising and extended in Polzehl and Spokoiny (2002) to a large class of statistical models. The AWS method has a number of features which make it well-suited for the problem at hand. Firstly, it is completely data-driven and it adapts automatically to the unknown structure of the signal function μ in the model. In particular, it is very sensitive to structural changes and can identify the location of the break point with high precision. Secondly, it can be applied to a situation where the noise is heteroscedastic (as it is the case with reformulation (2.2) of the model (2.1)). In the case of a heteroscedastic regression, it can also be used to estimate the time-varying variance of the noise. Finally, in many special cases it provides nearly optimal noise reduction, see Polzehl and Spokoiny (2002), (2003).

To keep the exposition simple, let us first assume that we are in the case of a known variance $s^2(t)$. When that is not the case, we substitute it with an estimate $\hat{s}^2(t)$. The details on how to produce such estimate are given at the end of this section. The central idea of the approach is to construct, for every time point t, a set of non-negative weights $w_{t,u}$ satisfying $w_{t,u} \in [0, 1]$. These weights measure how relevant observation y_u is to the estimation of the function μ at moment t. The higher the weight $w_{t,u}$, the stronger the contribution of observation y_u to the estimation of the function μ at t. When $w_{t,u}$ is zero, observation y_s does not contribute to the estimation of μ at time t. Once constructed, the weights serve to produce the estimate $\mu(t)$ defined by the weighted least squares:

$$\hat{\mu}(t) = \operatorname*{argmin}_{\mu} \sum_u (y_u - \mu)^2 s^{-2}(u) w_{t,u}.$$

This is a quadratic optimization problem with the closed form solution

$$\hat{\mu}(t) = \left(\sum_s s^{-2}(u) w_{t,u} \right)^{-1} \sum_u y_u s^{-2}(u) w_{t,u}. \tag{3.1}$$

The weights $w_{t,u}$ are constructed from the data using the following iterative procedure. We start with the usual kernel weights $w_{t,u}^{(0)} = K(|t-u|^2/h_0^2)$ for some kernel K and a (small) bandwidth h_0. Let us denote by $w_{t,u}^{(k)}$ the weights after the k-th iteration, by h_k, the k-th iteration bandwidth, and by $\hat{\mu}^{(k)}(t)$ the local estimate given by (3.1) with weights $w_{t,u}^{(k)}$.

The weights $w_{t,s}^{(k+1)}$ are iteratively defined as

$$w_{t,u}^{(k+1)} = K(|t-u|^2/h_k^2)\ K(d_{t,u}^{(k)}),$$

where

$$d_{t,u}^{(k)} := |\hat{\mu}^{(k)}(t) - \hat{\mu}^{(k)}(u)|.$$

In words, the weights $w_{t,u}^{(k+1)}$ are a product of a *location penalty*, $K(|t-u|^2/h_k^2)$ with a *statistical penalty*, $K(d_{t,u}^{(k)})$. Note that, while in classical smoothing[8] only the information within a neighborhood of t, $[t-h, t+h]$ is pooled for estimation of $\mu(t)$, the AWS estimate at time t uses also observations that are chronologically remote from the moment t, *provided that the values of the estimated μ in those points are close to the estimate in t*. In other words, in estimating $\mu(t)$, the AWS methodology pools information from all episodes that are similar to what was going on at moment t. The weights $w_{t,u}^{(k)}$ and the estimates $\hat{\mu}^{(k)}(t)$ are recomputed at every step k as the bandwidth parameter h_k increases. The details of the procedure can be found in Polzehl and Spokoiny (2002), (2003).

For estimating the time-varying variance $s^2(t)$, we build the differences $\hat{\varepsilon}_t = (y_t - y_{t-1})/\sqrt{2}$. Since every $\hat{\varepsilon}_t^2$ has approximately the mean equal to s_t^2, we apply the AWS procedure for mean

[8]In classical smoothing the weights are defined as $w_{t,s} = K(|t-s|^2/h^2)$, where h is the optimal bandwidth.

estimation to $\hat{\varepsilon}_s$. This yields a locally constant approximation of the true time-varying variance function $s^2(t)$, see Polzehl and Spokoiny (2002). The expression (3.1) can be used to bound the standard deviation of the estimate $\hat{\mu}(t)$ and therefore, to construct the α-percent confidence intervals for this estimate.

4. Non-parametric volatility estimation: Empirical results

Our analysis is conducted on a set of daily returns on twelve stock market indexes (see the Appendix for more details on the indexes under study). The samples analyzed are described in Table 1.

	Index	Country	sub-sample	Full sample
1.	ASX	Australia	01/07/1995-05/06/2003	05/01/1985-26/05/2004
2.	ATX	Austria	07/01/1993-07/02/2001	07/01/1993-26/05/2004
3.	CAC 40	France	15/05/1995-23/04/2003	03/04/1990-15/04/2004
4.	FTSE 100	UK	21/04/1995-21/03/2003	06/05/1984-18/03/2004
5.	DAX	Germany	08/04/1995-21/03/2003	03/04/1990-17/03/2004
6.	OMX	Sweden	23/10/1994-25/10/2002	02/11/1986-18/04/2004
7.	Russell 3000	USA	16/09/1994-31/08/2002	07/01/1988-03/06/2004
8.	S&P/TSX	Canada	17/12/1994-01/12/2002	18/08/1984-18/03/2004
9.	BEL 20	Belgium	13/01/1995-22/03/2003	05/01/1985-26/05/2004
10.	NIKKEI 225	Japan	01/12/1985-21/01/1994	09/02/1984-18/03/2004
11.	FAZ	Germany	24/03/1995-21/03/2003	07/09/1984-19/03/2004
12.	DJI	USA	06/11/1994-16/10/2002	02/01/1988-18/04/2004

Table 1. Samples of index returns. The full sample is used in the analysis in Sections 4 and 5. The dates in the second column (sub-sample) correspond to 2000 observations used in evaluating volatility forecasting performance of the stationary, parametric, conditional GARCH(1,1) methodology in Section 6.

In the rest of the section we present the results of the non-stationary, non-parametric estimation of volatility based on the AWS methodology described in Section 3 and we evaluate the goodness-of-fit of the non-parametric estimation approach. In the analysis of the returns on the twelve indexes both equations (2.2) and (2.3) were used as basis for the estimation of time-varying second moment of the returns. The estimation results were identical.

Figure 4.1 displays the time-varying annualized standard deviation estimated using the methodology described in Section 3. The (annualized) absolute returns are plotted together with the volatility. The shaded area corresponds to the sub-samples specified in Table 1. This sub-samples are the object of a detailed analysis focusing on the forecasting performance of the GARCH(1,1) model in Section 6. The criteria for the selection of the sub-samples will be described in Section 5.

The graphs in Figure 4.1 show that the AWS approach identifies significant changes in the unconditional variance of the returns on the twelve indexes under scrutiny. The European and North-American series show a lower level of volatility in the middle of the 90's followed by an increase covering the second half of the decade (from 1996 on) and the beginning of the first decade of the new millennium. Most of them exhibit a lowering of the level of volatility in 2003. Note that the sub-samples, highlighted by the shaded areas, cover the periods animated by the most significant changes in the level of volatility.

Figure 4.2 displays the sample ACF of the absolute values of the returns and of the returns standardized with the estimated time-varying sd. All absolute returns display the so-called 'long-memory in volatility' effect, i.e. the presence of significant sample autocorrelations at large lags. A sample ACF that shows positive correlations at large lags (like those in Figure 4.2) could be a sign of non-stationarities in the second moment structure of the time series as well as a proof of a stationary, non-linear long-range dependence; see Mikosch and Stărică (2004). As in Stărică and Granger (2005), we interpret the presence of significant autocorrelations in the absolute (squared) values of returns as evidence of non-stationary changes in the unconditional second moment of the series of returns. Accordingly, the success of the estimation of time-varying unconditional volatilities is evaluated based on the removal of the long-memory aspect of the sample ACF for the absolute returns standardized with the estimated time-varying standard deviation. That is, the estimation procedure may be considered successful if the standardized returns look uncorrelated.

The graphs in Figure 4.2 show that even a rough approximation of the variance dynamics by a step function is sufficient to explain most of the dependency structure present in the sample ACF of absolute return series, thus providing an explanation for the so called "long memory in volatility" effect. In fact, the absolute returns standardized with the time-varying sd estimated by the AWS methodology, are practically uncorrelated.

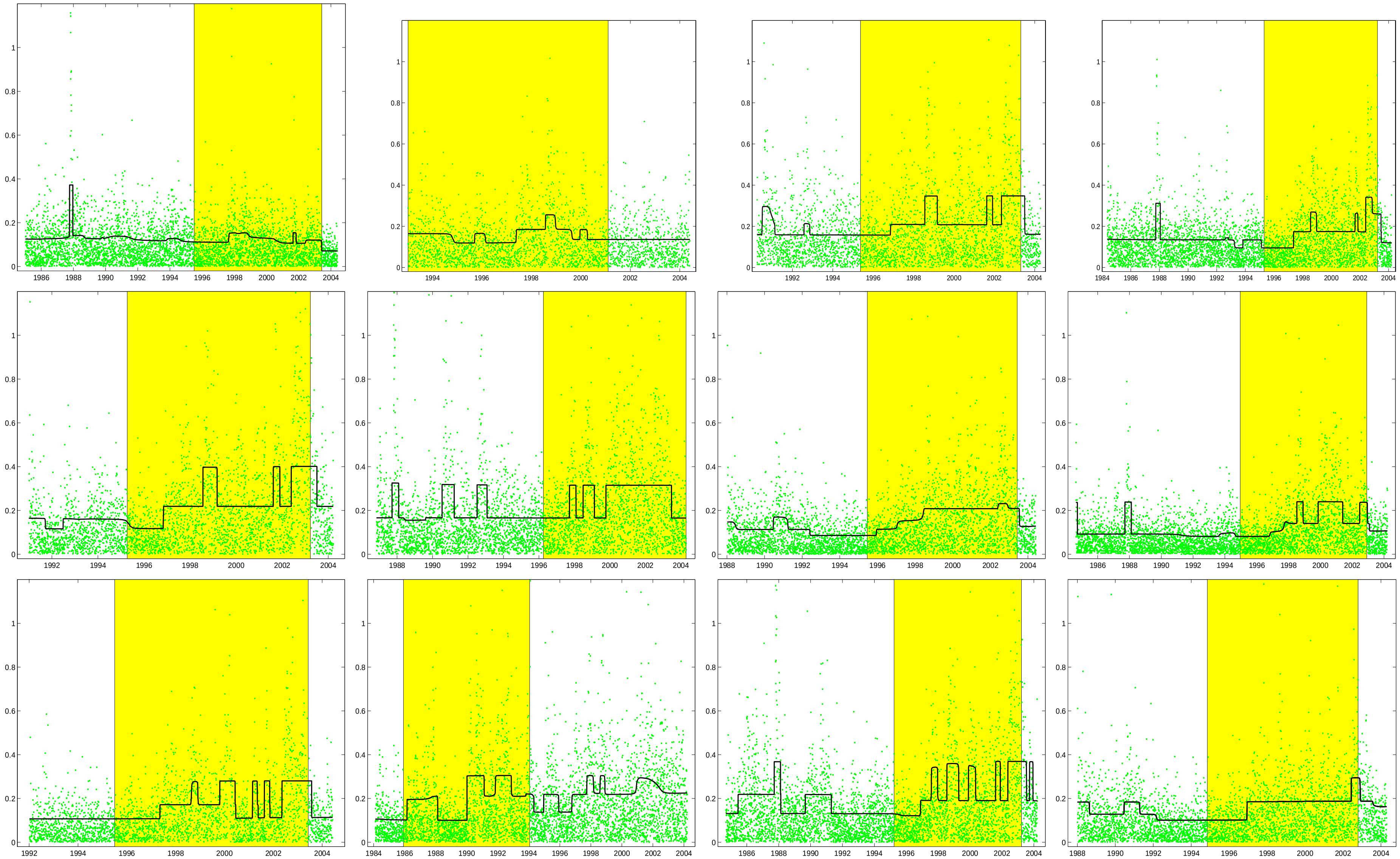

Figure 4.1. The time-varying unconditional annualized volatility estimated using the AWS approach in Section 3. The order from the top-left to bottom-right corresponds to the order in Table 1). The (annualized) absolute returns are plotted together with the volatility. The shaded area corresponds to the sub-samples specified in third column of Table 1.

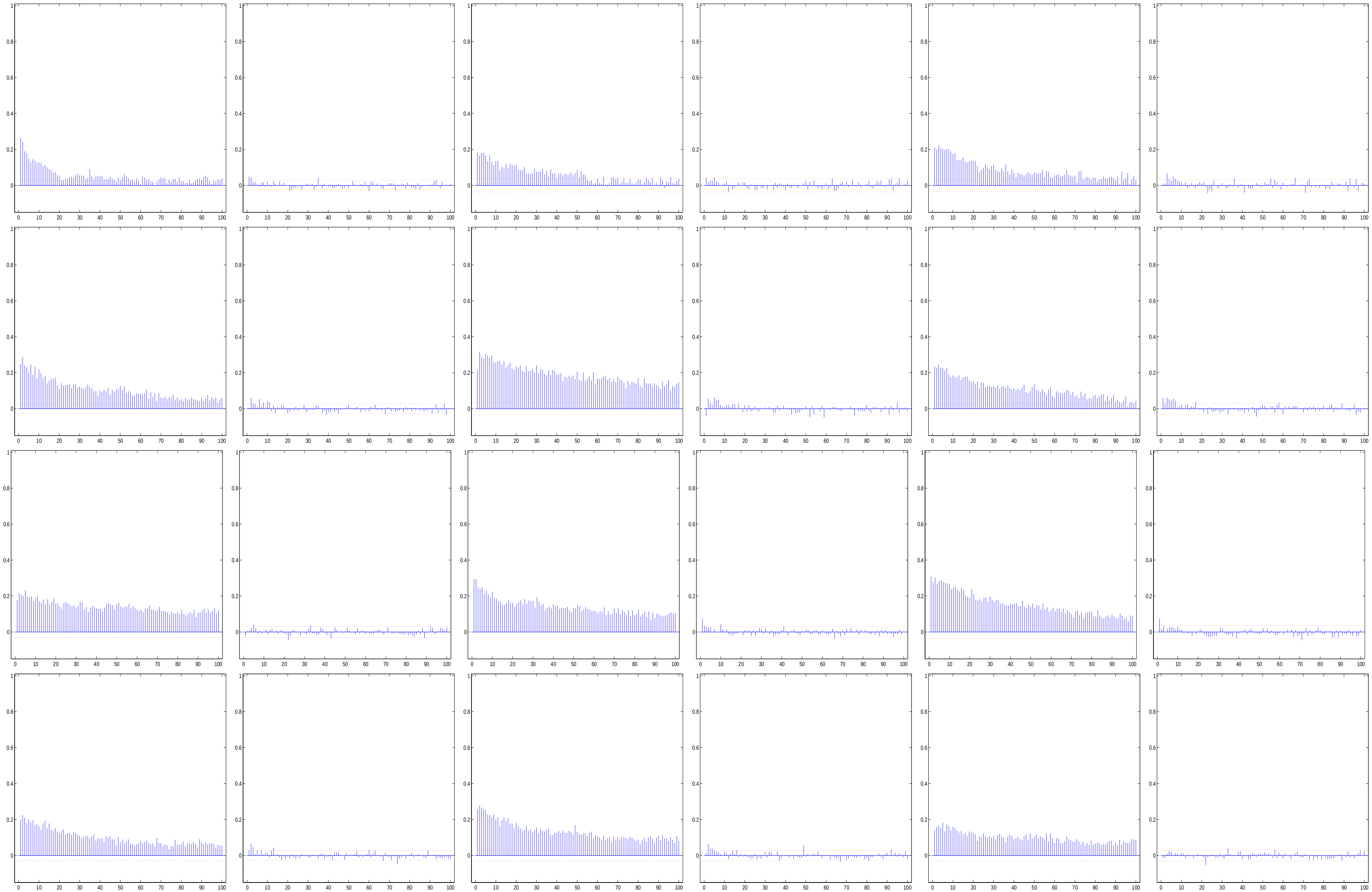

Figure 4.2. Sample ACF of absolute returns and absolute standardized returns. The order from the top-left to bottom-right corresponds to the order in Table 1). Two (consecutive) graphs are displayed for each index.

5. Changes in the unconditional volatility and the IGARCH effect

In this section we investigate the dynamics of the IGARCH effect for the twelve indexes under scrutiny. When modeling the returns on an index in the stationary, parametric, conditional ARCH framework, the *working assumption* is often that the data generating process is the stationary GARCH(1,1) model

$$r_t = z_t h_t^{1/2}, \quad h_t = \alpha_0 + \alpha_1 \, r_{t-1}^2 + \beta_1 \, h_{t-1}, \tag{5.1}$$

where (z_t) are iid, $Ez = 0$, $Ez^2 = 1$. Condition $\alpha_1 + \beta_1 < 1$ is necessary and sufficient for the process to be weakly stationary[9]

The IGARCH effect consists in the sum $\alpha_1 + \beta_1$ being (slightly smaller and) close to one. Under the assumption that the returns have finite second moment, the unconditional variance of the GARCH(1,1) model (5.1) is given by

$$\sigma^2_{GARCH(1,1)} := \alpha_0/(1 - \alpha_1 - \beta_1). \tag{5.2}$$

Replacing the GARCH(1,1) coefficients in (5.2) with estimated values yields the estimated *GARCH-(1,1) unconditional variance*, $\widehat{\sigma}^2_{GARCH(1,1)}$. Note that (5.2) implies that the stronger the IGARCH effect, i.e. the closer $\widehat{\alpha}_1 + \widehat{\beta}_1$ is to one, the larger the estimated GARCH(1,1) unconditional volatility becomes.

In the recent financial econometric literature, many authors (some of which were cited in the Introduction) have argued that there is a causal connection between the IGARCH effect and structural changes in the unconditional variance of returns. That is, estimating a Garch(1,1) model on a sample displaying non-stationary changes of the unconditional volatility, may induce a spurious IGARCH effect.

The non-stationary paradigm of modeling and estimating the unconditional variance of returns described in the previous section offers a consistent set-up for an empirical investigation of such connection. The investigation is carried through sub-sample-specific measures of volatility and GARCH(1,1) modeling features. The sub-sample-specific measures quantify and compare the strength of the IGARCH effect and the amount of change of the unconditional sd (as estimated in Section 4) in a window moving through the data.

---

[9]If this condition is not fulfilled, the GARCH(1,1) process, if (strongly) stationary, has infinite variance.

Let us now define precisely the two mentioned sub-sample-specific measures. To measure the intensity of the IGARCH effect in the sample $[t-a,t]$, a GARCH(1,1) model is estimated using the quasi-ML estimation method. A sample size of $a = 2000$ is commonly assumed to be sufficient for a precise estimation of a GARCH(1,1) model. Sample sizes that are significantly smaller yield unacceptably large standard deviations for the estimated parameters. This is the sample size that we use in the sequel analysis.[10] Besides the statistical motivation, the choice of a window of length 2000 incorporates the belief, common in the econometric community, that return time series can be safely modeled by stationary models, i.e. the stochastic features of the data are relativelty stable in time. Denote by $\widehat{\sigma}_{GARCH(1,1)}\ (t)$ the estimated GARCH(1,1) unconditional sd of the sample $[t-a,t]$ and by $\widehat{\sigma}(t)$ that sample's sd $\widehat{\sigma}(t) := (\sum_{i=t-a}^{t} r_i^2)/a$. The strength of the IGARCH effect in the sample $[t-a,t]$ is measured by *its impact on the estimation of the unconditional variance of that sample*. More concretely, the ratio

$$\nu(t) := \frac{\widehat{\sigma}_{GARCH(1,1)}\ (t)}{\widehat{\sigma}(t)}, \tag{5.3}$$

will be used in the sequel as a quantitative measure of the intensity of the IGARCH effect. A particularly strong IGARCH effect in the sample $[t-a,t]$ will produce an estimated $\widehat{\sigma}_{GARCH(1,1)}\ (t)$ much greater than the sample sd, and hence a ratio $\nu(t)$ much greater than one. A ratio $\nu(t)$ close to one identifies the sub-samples on which the GARCH(1,1) estimated variance matches the sample variance (due to the absence of the IGARCH effect). We interpret a strong discrepancy between the two estimates of the standard deviation of the data as a clear indication that GARCH(1,1) fails to model the dynamics of the returns.

To measure the amount of change in the unconditional volatility on the interval $[t-a,t]$, first the mean $\bar{\sigma}(t)$ of the AWS unconditional volatility estimate $\widehat{\sigma}_{AWS}(u)$, $u \in [t-a,t]$ is computed

$$\overline{\sigma}(t) = \frac{1}{a}\sum_{i=0}^{a} \widehat{\sigma}_{AWS}(t-a+i).$$

[10] The analysis was run on smaller sample sizes of $a = 1750$ and $a = 1500$ observations. While the details change, the overall qualitative results do not. A sample size of 1500 is the absolute minimum in terms of statistical precision of the estimated coefficients. See Straumann (2005).

Then the following relative measure of the variation of the unconditional volatility to the mean unconditional volatility is built

$$l(t) := \frac{\sum_{i=0}^{a} |\sigma_{AWS}(t-a+i) - \overline{\sigma}(t)|/a}{\overline{\sigma}(t)}. \tag{5.4}$$

In words, $l(t)$ measures how much the unconditional volatility has changed in the window of a observations ending at t, relative to the unconditional volatility of the window $[t-a,t]$.

We emphasize that although the ratios $\nu(t)$ and $l(t)$ are indexed by t, they measure features of *the data in the window of* a *observations ending at time* t. One needs to keep this in mind when interpreting the graphs in Figures 5.1, 5.2, and 5.3.

The two measures are re-calculated every 50 days on a window of past a observations (in the case of the ratio ν, the GARCH(1,1) model is re-estimated on the new sub-sample). The results are displayed in Figures 5.1, 5.2, and 5.3. Besides the two measures $\nu(t)$ and $l(t)$, we also display the sum $\widehat{\alpha}_1 + \widehat{\beta}_1$ with the upper one-sided 95% confidence intervals and the estimated GARCH(1,1) unconditional sd together with the sample sd.

Sub-samples with a particularly pronounced IGARCH effect are identified in most of the twelve time series. It is particularly significant the fact that sub-samples with a more pronounced IGARCH effect as measured by $\nu(t)$ are also characterized by higher measures of the amount of unconditional volatility change as measured by $l(t)$. Hence, our analysis seems to give evidence in favor of the hypothesis of a connection between non-stationarities in the second moment structure and the IGARCH effect.

The series can be divided into three groups, according to the level attained by the measure $\nu(t)$ of the IGARCH effect.

5.1. **No IGARCH effect.** The first three time series from Table 1, i.e. ASX, ATX, CAC 40 indexes, are characterized by the absence of the IGARCH effect (see Figure 5.1). The upper 95% confidence bound is strictly smaller than 1. Hence the point estimate $\widehat{\alpha}_1 + \widehat{\beta}_1$ is significantly different from 1. Moreover, the ratio $\nu(t)$ remains smaller than or equal to one (and always close to it), showing a good match between the estimated GARCH(1,1) unconditional variance and the sample variance. It is worth noticing that $l(t)$ is bounded by 30%, indicating moderate changes of the sd.

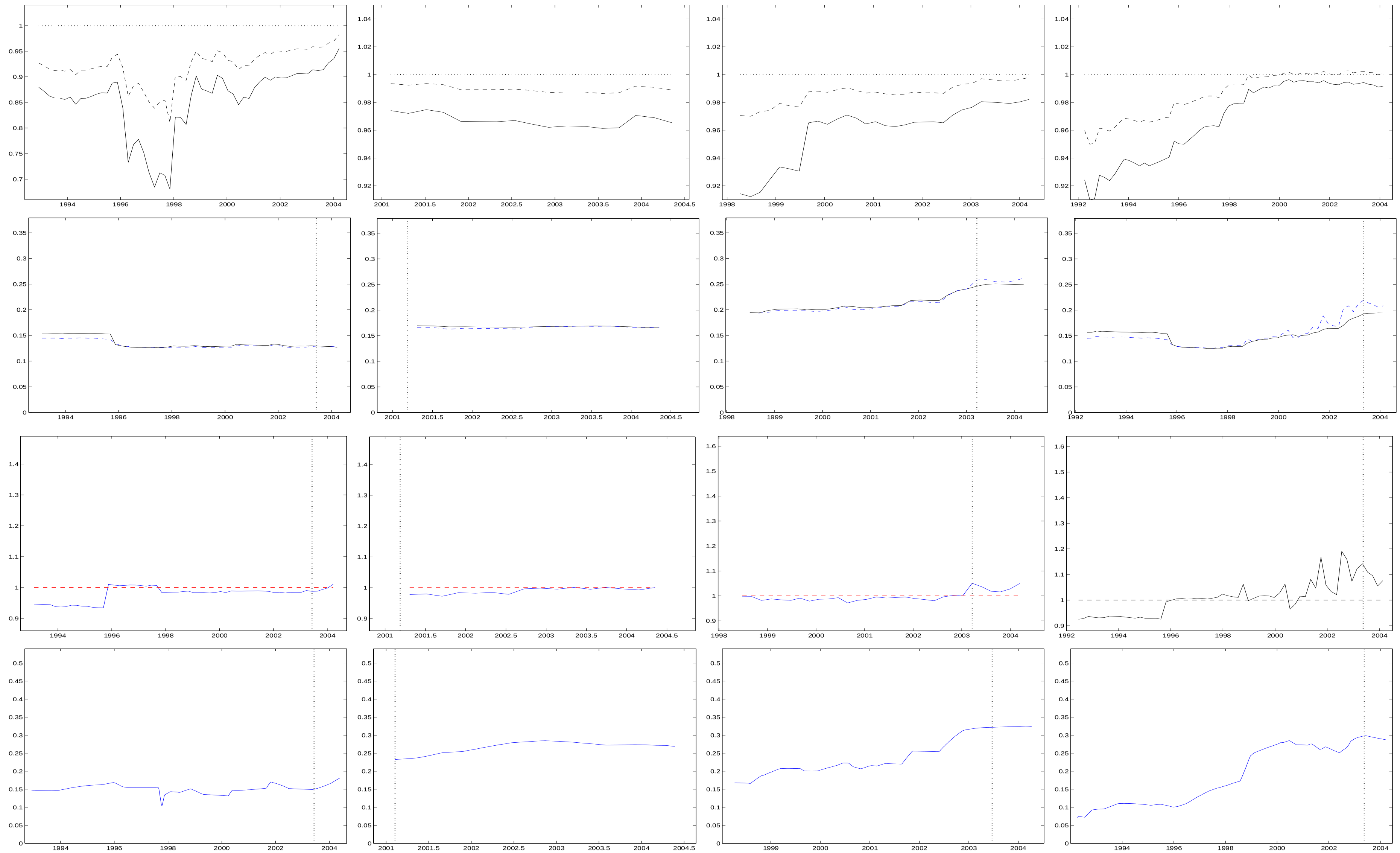

Figure 5.1. Changes in the unconditional volatility and the IGARCH effect for ASX, ATX, CAC 40, and FTSE 100 indexes. *Top:* The sum $\widehat{\alpha}_1 + \widehat{\beta}_1$ (full line) with the upper one-sided 95% confidence interval (dotted). *Second row:* The sample sd (full line) and the GARCH(1,1) estimated sd (dotted). *Third row:* The ν measure defined in (5.3). *Bottom:* The l measure defined in (5.4). The vertical dotted lines mark the sub-sample used for the detailed forecasting comparison in Section 6.

5.2. **Mild IGARCH effect.** For the next three samples (series 4 to 7 from Table 1, i.e., FTSE 100, DAX, and OMX indexes), the IGARCH effect is rather light but noticeable (see Figures 5.1 and 5.2). While one is on the boundary or slightly inside the asymmetric confidence interval, the point estimates mostly remain away from it. We note that the IGARCH effect appears towards the end of the samples examined. The initial estimates on these samples are characterized by a sum of GARCH(1,1) coefficients $\widehat{\alpha}_1 + \widehat{\beta}_1$ that is significantly away from one and by values of $l(t)$, the measure of the amount of changes in the unconditional volatility, smaller than in the end of the samples. The end of the samples is animated by more pronounced changes in the unconditional second moment (higher values of $l(t)$). The IGARCH effect takes hold, while the two measures of sd, the estimated GARCH(1,1) unconditional sd and the sample sd drift apart. The maximum of the ratio $\nu(t)$ is 1.2-1.4 while the measure of the amount of variation of the unconditional volatility is bounded by 35%. Note that higher values of $\nu(t)$ usually corresponds to higher values of $l(t)$.

5.3. **Strong IGARCH effect.** For the remaining six time series (series 8 to 12 from Table 1, i.e. Russell 3000, S&P/TSX, BEL 20, NIKKEI 225, FAZ, and DJI indexes), the IGARCH effect is pronounced (see Figures 5.2 and 5.3). The value one is well inside the one-sided confidence interval while the point estimates also come close to one (in some cases being practically equal to it). The maximum of the ratio $\nu(t)$ is bigger than 1.4 while the maximum of the relative measure $l(t)$ gets close and sometimes trespasses the threshold of 40%. As before, periods animated by significant changes in the unconditional variance of the returns are also characterized by strong IGARCH effect.

5.4. **The choice of the sub-samples in Table 1.** Since for the model (5.1) the volatility forecast at longer horizons is, practically, the unconditional variance (see equation (6.1)), poor point estimates for this last quantity will, most likely, have a strong impact on the longer horizon volatility forecasting performance of the model. To substantiate this conjecture in the next section we analyze the forecasting performance of the Garch(1,1) model on sub-samples that are characterized by a strong IGARCH effect. The sub-samples were chosen to cover the periods when the level of the measure $\nu(t)$ is at its peak. As Figures 5.1, 5.2, and 5.3 show, the measure $l(t)$ quantifying the amount of variation of the unconditional volatility of these sub-samples is, often, also at its highest, i.e. the sub-samples we chose to analyze in a forecasting set-up are often those that show the largest (or close to the largest) relative amount of changes in the unconditional variance.

It is worth noticing that eleven of the twelve sub-samples analyzed cover a eight year period between 1995 and 2004 with only one other, i.e. the NIKKEI 250 covering the period 1985-1994 interval. The choice of the periods, i.e. full samples, within which the sub-samples to be analyzed in detail were selected, is due to the limited availability of data. We believe that the fact that the selected sub-samples coincide with the known intervals of stock market upheaval (the end of the 90's for the Western stock markets and the end of the 80's and the beginning of the 90's for the Japanese stock market) is not a coincidence. It is in fact precisely during these turbulent intervals, characterized by relevant changes in the unconditional variance, that the Garch(1,1) model performs poorly.

6. Forecasts of future volatility

In this section we evaluate the performance of the GARCH(1,1) in volatility forecasting on the sub-samples of length 2000 days, reported in Table 1. The dotted vertical bars in Figures 5.1, 5.2, 5.3 mark the end of the sub-samples for each index.

Under the assumption of a GARCH(1,1) data generating process (5.1) that satisfies $\alpha_1+\beta_1 < 1$, the minimum Mean Square Error (MSE) forecast at time t for r^2_{t+p} is

$$\sigma^{2,\,GARCH}_{t+p} := E_t r^2_{t+p} = \sigma^2_{GARCH(1,1)} + (\alpha_1+\beta_1)^{p-1}(h_t - \sigma^2_{GARCH(1,1)}), \tag{6.1}$$

where $\sigma^2_{GARCH(1,1)}$ is the unconditional variance defined in (5.2). Consequently, the minimum MSE forecast for the variance of the cumulative return over the next p days, is given by

$$\overline{\sigma}^{2,\,GARCH}_{t,p} \quad := \quad E_t(r_{t+1}+\ldots+r_{t+p})^2 = \sigma^{2,\,GARCH}_{t+1} + \cdots + \sigma^{2,\,GARCH}_{t+p}.$$

From Equation (6.1) it follows that, for large p, the forecast $\sigma^{2,\,GARCH}_{t+p}$ is close to the unconditional variance, $\sigma^2_{GARCH(1,1)}$. Therefore, failing to produce accurate point estimates for this last quantity will, most likely, produce poor longer horizon volatility forecasts. Stărică (2003) showed that for sub-samples of returns on the S&P500 index characterized by IGARCH effect, the GARCH(1,1) model fails to provide sensible longer-horizon volatility forecasts. In the sequel we bring further empirical evidence supporting this finding. We also document the fact that for sub-samples on which the two measures of volatility are in good match, the forecasting behavior of the GARCH-(1,1) model is satisfactory.

Our evaluation includes a direct comparison with a simple forecasting approach which assumes that the volatility is locally constant (this choice of an alternative is motivated by the finding in Section 4). A second approach consists in a portfolio option replication exercise where the two volatility forecasts from the first comparison are used to set the price of at-the-money options with various maturities through a dynamic strategy that replicates the instrument to be priced.

6.1. **Direct comparison of volatility forecasts.** This subsection describes the set-up for direct evaluation of short- and longer-horizon volatility forecasting performance of a GARCH(1,1) model.

The benchmark model (BM) for volatility forecasting is the simple non-stationary model (2.1). Since no dynamics is specified for the variance, future observations $r_{t+1}, r_{t+2}, \ldots$ are modeled as iid with constant variance $\widehat{\sigma}^2_{250}(t)$, an estimate of $\sigma^2(t)$. In the sequel, we use the sample variance of the previous year of returns as the estimate for $\sigma^2(t)$. The forecast is then given by

$$\sigma^{2,\,BM}_{t+p} := \widehat{\sigma}^2_{250}(t) = \frac{1}{250}\sum_{i=1}^{250} r^2_{t-i+1}. \tag{6.2}$$

The forecast for the variance of the next p aggregated returns is then, simply,

$$\overline{\sigma}^{2,\,BM}_{t,p} := p\,\widehat{\sigma}^2_{250}(t). \tag{6.3}$$

To measure of the realized volatility in the interval [t+1, t+p] we define

$$\overline{r}^{\,2}_{t,p} := \sum_{i=1}^{p} r^2_{t+i}, \tag{6.4}$$

moreover, we compare the following MSE on n forecasts performed,

$$MSE^*(p) := \sum_{t=1}^{n} (\overline{r}^{\,2}_{t,p} - \overline{\sigma}^{2,\,*}_{t,p})^2 \tag{6.5}$$

where "$*$", here and in the sequel, stands for "BM" or "GARCH". The MSE (6.5) is preferred to the simpler MSE

$$\sum_{t=1}^{n} (r^2_{t+p} - \sigma^{2,\,*}_{t+p})^2$$

since this last one uses a poor measure of the realized return volatility[11]. Through averaging some of the idiosyncratic noise in the daily squared return data is canceled yielding (6.4), a better measure against which to check the quality of the two forecasts.

The direct comparison of short- and longer-horizon volatility forecasts was performed on the twelve sub-samples of length 2000 reported in Table 1. The GARCH(1,1) model is estimated initially on the first 1000 data points from every sample. Consistent with the assumption of stationarity, fundamental to the ARCH methodology, the model is re-estimated every week (i.e. every 5 days) using the observations from the beginning of the sample up to the moment of re-estimation. At the same time, $\hat{\sigma}^2_{250}(t)$ is also estimated. After every re-estimation, volatility forecasts are made for the next year ($p = 1, \ldots, 250$) using (6.2) and (6.3). Following the out-of-sample forecasting paradigm, the quantities $MSE^{GARCH}(p)$ and $MSE^{BM}(p)$ defined in (6.5) are calculated based on the observations from the year that followed. The graphs in Figure 6.1 display the ratio $MSE^{BM}(p)/MSE^{GARCH}(p)$. A ratio smaller than one at horizon p indicates that the volatility forecast of the GARCH(1,1) parametric, conditional methodology for the interval of next p days is poorer than that based on the simple approach that assumes that the history of the past year will repeat. Figure 6.1 demonstrates strong variation in the quality of the GARCH(1,1) forecast. The first two graphs demonstrate an overall good performance at all forecasting horizons. The third and the fourth show only good shorter-horizon performance, with a deterioration of the quality of forecast at horizons beyond three or four months. For the rest of the sub-samples (from five to twelve) and for periods as long as four business years, the GARCH(1,1) model provides poor shorter- and longer- volatility forecasts (sometimes with exceptions of forecasts of at most ten days ahead).

6.2. **Volatility forecasts for option replication.** In this subsection we perform an indirect evaluation of short- and long-horizon volatility forecasting performance of a Garch(1,1) model. This evaluation consists in an option replication exercise. The goal is to get a financially sound measure of the accuracy of different variance forecasts.

The use of option prices to measure the forecasting capabilities of a model was considered (among others) by Engle et al. (1993), with the motivation that "the pricing of the options

[11]It is well known (see Andersen and Bollerslev [1]) that the realized square returns are poor estimates of the day-by-day movements in volatility, as the idiosyncratic component of daily returns is large.

provides the appropriate test of forecasts of asset volatility". Engle et al. (1997) compare the behavior of different specifications of GARCH when applied to pricing options on the NYSE index, by checking if the option price based on a given model is a good forecast of the final payoff of the option.

In what follows we use an alternative approach based on option replicating strategies implied by each model. That is, we focus on evaluating the ability of two competing modeling methodologies to help an investor to implement a dynamic strategy that replicates a given claim. Our approach is based on the assumption that more accurate volatility forecasts lead to smaller replication errors.

We compare the performances of Garch(1,1) model and of the simple model BM (2.1). The option replication exercise goes as follows. At time t_k we start a self-financing strategy, involving the underlying asset and a bank account. We consider the usual hypotheses of "perfect market" with no transaction costs and zero interest rate. The goal of the strategy is to replicate the payoff at time $t_k + T$ of an at-the-money straddle, i.e. a portfolio consisting of a European call option and a European put option, both at-the-money and with the same maturity. Independently of the model imposed on the underlying, the same hedging strategy, namely Black-Scholes Delta-hedging, is used to define the composition of the replicating portfolio. This implies that differences will be mostly due to model-specific estimates of the volatility of the underlying. One could object that, coherently with the assumption of a Garch(1,1) model, one should use a Garch(1,1) option pricing methodology, as proposed in Duan (1995) for example. There are a few reasons for which we do not follow this approach. First, Garch(1,1) pricing does not yield closed-form expressions neither for pricing nor for hedging. Hence one would have to relay on a time-consuming Monte Carlo methodology. Second, according to our experience (and also shown by Choi (2005)) the differences between Black-Scholes and Garch(1,1) prices are rather small, especially when options are near moneyness. The third reason is that the market practice is to consistently use the Black-Scholes formula, even when the hypothesis under which this formula holds might be violated. This last argument we find particularly compelling since our goal is to evaluate the relevance of various modeling approaches to the *practice* of pricing. Note that our article of reference in this part of the comparison, i.e. Engle et al. (1997), follows the same approach.

Let us now describe the construction of the replicating strategy. Let $r_t, t = 1, \ldots, 2000$, be the series of log-returns in the sub-sample. A new strategy is started every week, at times t_k, (where $t_1 = 1000$ and $t_{k+1} - t_k = 5$) and the composition of the replicating portfolio is adjusted every day until maturity. The goal of each strategy is to replicate, at maturity, the payoff of an at-the-money straddle, i.e. of a portfolio consisting of a call and a put, both at-the-money. At time t, the number of shares of the underlying in the replicating portfolio is given by the Black-Scholes hedge ratio formula for a straddle with strike K and time to maturity τ, that is

$$\Delta(S_t, \overline{\sigma}^*_{t,\tau}) = 2\Phi\left(\frac{\log S_t/K + \overline{\sigma}^{2,*}_{t,\tau}/2}{\overline{\sigma}^*_{t,\tau}}\right) - 1,$$

where $\Phi(\cdot)$ is the standard normal cumulative density function, S_t is the price of the underlying at time t, $\sigma^*_{t,\tau}$ is the volatility forecasted at time t by model $*$ for the period from t to $t+\tau$. While for the BM model the only sensible choice of such forecast is represented by (6.3), for the GARCH model one could use either the stationary variance (as it is often done, see Duan (1995)) or the conditional forecast as defined in (6.2) and employed for instance by Engle et al. (1997). Choi (2005) proved that in this way one obtains a good approximation of the exact GARCH pricing formula. The value S_t is obtained from the historical time series of log-returns by setting $S_{t_k} = 1$, for each starting time t_k. Since the options to be replicated are at-the-money, we set the strike price K to one. The replicating portfolio is daily re-adjusted according to the new values of $\Delta(\cdot)$. In order to make the strategy self-financing, the money involved in buying or selling of the shares is withdrawn from (or deposited into) a bank account. We assume that the interest rate is zero. The initial cost of the strategy, under the hypothesis that the underlying follows the model $*$, is $C^*_{t_k} = BS(\overline{\sigma}^*_{t_k,\tau})$, where $BS(\cdot)$ is the Black-Scholes pricing formula for the straddle. That is, if the model $*$ is correct, investing $C^*_{t_k}$ at time t_k and following the appropriate strategy, one should get at time $t_k + T$, with probability one, the payoff of the straddle.[12] We denote by $e^*_{H,t_k}(T)$ the difference between the final value of the replicating portfolio, based on model $*$ and started at time t_k, and the payoff of the straddle at maturity T. The smaller the absolute value of the error is, the more accurate are the volatility forecasts of model $*$.

We applied the replicating procedure on the sub-samples of Table 1. Both models are re-estimated weekly, that is at each time t_k. Consistent with the hypothesis of stationarity, the Garch(1,1) estimation uses all returns available from the beginning of the sample up to time t_k. The BM estimation uses only the previous 250 returns (roughly one business year of data).

[12] We are neglecting the discretization error because it would affect both the models and, the trading interval being rather fine, it should be small with respect to model error.

The first replication exercise compares the BM strategy to the GARCH strategy when the stationary variance is employed. Figure 6.2 displays the mean values of the hedging errors for the 12 series for maturities of five and twenty days. From this figure it is apparent how the problems affecting the GARCH estimates of series 7 to 13 lead to significant errors in replication exercise, much greater indeed than those produced in the simple BM set-up. The performance of the approach using GARCH stationary variance gets worse for longer maturities.

In the second replication exercise the Garch(1,1) volatility, $\overline{\sigma}_{t,\tau}^{GARCH}$ is computed every day t according to formula (6.2), using the most recent parameter estimates available. We performed the strategy for four maturities: $T = 60, 120, 180, 250$ days. For each maturity we performed as many replication exercises as allowed by the length of the sample, that is $188, 176, 164$, and 150 respectively. We then computed the mean of the hedging errors of each series for any given maturity. The results are reported in Figures 6.3. Although the use of the conditional variance brings an obvious improvement, we see that the Garch strategies tends to produce a great error for those series that display a particularly large ratio MSE_{GARCH}/MSE_{BM}. In fact, while for series 1 to 6 the overall behavior of the two approaches is similar for all maturities, for series 7 to 12, the Garch(1,1) model has an average error significantly greater than that produced by the naive BM approach. Note that the difference between errors increases with maturity becoming very relevant for $T = 250$.

To evaluate the significance of the errors produced by the two approaches and represented by the graphs in Figure 6.3 we used two statistical tests. The tests are made under the assumption of stationarity of the return series.[13] If $D_t(T) := e_{H,t}^{GARCH}(T) - e_{H,t}^{BM}(T)$, we test the hypothesis

$$H_0 : ED_t = 0 \quad \text{against} \quad H_1 : ED_t > 0, \tag{6.6}$$

(an one-sided test). Note that, since the hedging errors are computed on overlapping returns, they are obviously dependent.

Both tests use the sample mean $\overline{D}$ as the test statistics and differ in the way the asymptotic distribution of the test statistic is obtained. The first test uses the overlapping-block bootstrap (Künsch (1998)) with the block-length selection proposed by Politis and White (2003). The second one computes the asymptotic distribution from the Central Limit Theorem for stationary

[13] It is hard to test the significance of the differences in Figure 6.3 under the working hypothesis of non-stationary returns.

sequences with sumable covariances (note that Diebold and Mariano (1995) and by Harvey et. al (1997) use a similar test).

For the first test, the distribution of the test statistic is obtained as follows. Define the overlapping blocks of size b

$$\mathbf{D}_1 = \{D_1, \ldots, D_b\}, \ldots, \mathbf{D}_{n-b+1} = \{D_{n-b+1}, \ldots, D_n\}.$$

New samples $\mathbf{D}_1^*, \ldots, \mathbf{D}_m^*$ are drawn with replacement from the set $\mathbf{D}_1, \ldots, \mathbf{D}_{n-b+1}$. The simulated samples preserve most of the original dependence structure since the dependency inside any block remains untouched. The samples are used to calculate sample means $\overline{D}_1^*, \ldots, \overline{D}_m^*$. Since m can be made as big as one wants, the procedure yields a good approximation of the sampling distribution of $\overline{D}^*$. The results in Künsch (1998) show that, under general conditions, the sampling distribution of $\overline{D}^*$ approximates that of $\overline{D}$ (for details see Künsch (1998)).

For the second test, the asymptotic distribution of the test statistic is given by

$$\overline{d} \sim N(0, v^2), \quad v^2 = \gamma(0) + 2\sum_{h=1}^{\infty} \gamma(h), \tag{6.7}$$

where $\gamma(\cdot)$ is the autocovariance function that is assumed sumable.[14] For statistical purposes, the theoretical autocovariance function is replaced with the sample version.

Figure 6.4 displays the results of the first test. It shows relevant information on the bootstrapped sampling distribution of the test statistic (obtained with the block bootstrap method discussed above). More concretely, the 5%th (downwards-pointing triangle), 10%th (cross), 90%th (cross) and 95%th (upwards-pointing triangle) quantiles of the bootstrapped sampling distribution of the test statistic $\overline{D}$, together with the mean of the distribution (square) are displayed. One notices that, while for $T = 60$ the support of most of the sampling distributions contains zero, i.e. the errors of the two models are not significantly different, for longer horizons, the support of several sampling distributions does not include the zero value any more, indicating that the GARCH(1,1) model produces *statistically significant bigger errors.* In particular, for the

[14]The statistical analysis of the series $d_t(T)$ show that the autocovariances are never significant beyond the first 40 lags. Most of the time the number of significant lags is smaller or equal to 6 and when significant autocorrelations are present at larger lags they are barely significant. These facts confirm the appropriateness of the hypothesis of sumability of the autocovariance function.

one year horizon, the hedging errors of the GARCH(1,1) model for the series 2 and 6 to 12 are statistically bigger than those of the simple BM at a 95% confidence level.

Series	Horizon			
	60 days	120 days	180 days	250 days
1.	**0.01**	0.06	0.10	0.11
2.	**0.00**	**0.00**	**0.00**	**0.00**
3.	0.43	0.37	0.24	0.10
4.	0.75	0.72	0.71	0.46
5.	0.82	0.74	0.83	0.63
6.	0.36	0.15	0.08	0.08
7.	0.48	0.08	**0.00**	**0.00**
8.	0.19	0.06	**0.01**	**0.00**
9.	**0.01**	**0.00**	**0.00**	**0.00**
10.	0.13	**0.01**	**0.00**	**0.00**
11.	0.45	0.28	0.12	**0.02**
12.	**0.00**	**0.00**	**0.00**	**0.00**

TABLE 2. p-values for the test (6.6) based on the CLT asymptotic distribution applied to D(T)=$e_H^{GARCH}(T) - e_H^{BM}(T)$, the differences between hedging errors at different maturities $T = 60, 120, 180, 250$. The null hypothesis is $ED = 0$. The samples are those of Table 1. The series and the periods where the null hypothesis is rejected are in bold face. The bold face values indicate that the Garch model produces *statistically significant greater* hedging errors than the BM model.

Table 2 displays the p-values of the test based on the CLT, i.e. the probability under the null hypothesis that the test statistic would take values larger or equal to its actual sample value. Rejections of the null, indicating the series and the maturities where the mean of the hedging errors based on a Garch(1,1) modeling is significantly bigger that that of the errors incurred using the BM approach at 95% confidence level, are in bold. We see that, as the horizon increases (notably for $T = 180$ and $T = 250$), the Garch(1,1) model produces significantly larger errors on average on the series that display a particularly large ratio MSE_{GARCH}/MSE_{BM}, i.e. the series 7 to 14 (with the exception of series 12 which for $T = 180$ displays a significant p-value, while for $T = 250$ the value is barely significant at 95% confidence level). For the first six series (with the exception of series 2), for $T = 180$ and $T = 250$, the two approaches are not significantly different. The statistical testing confirms the overall picture given by the graphs in Figure 6.3.

Series	p-values for Garch(1,1) hedging				p-values for BM hedging			
	60 days	120 days	180 days	250 day	60 days	120 days	180 days	250 days
1.	**0.00**	**0.02**	0.05	**0.00**	0.08	0.09	0.10	0.05
2.	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.06	0.05
3.	0.21	0.24	0.38	0.36	0.51	0.50	0.68	0.87
4.	0.32	0.35	0.32	0.30	0.28	0.16	0.06	0.52
5.	0.18	0.26	0.39	**0.02**	0.18	0.09	**0.00**	**0.01**
6.	0.27	0.11	0.13	0.14	0.30	0.34	0.34	0.80
7.	**0.00**	**0.00**	**0.00**	**0.00**	**0.03**	**0.01**	**0.00**	**0.00**
8.	**0.00**	**0.00**	**0.00**	**0.00**	0.35	0.29	0.36	0.45
9.	**0.01**	**0.03**	0.11	**0.00**	**0.03**	0.29	0.34	0.17
10.	**0.00**	**0.00**	**0.00**	**0.00**	0.16	0.21	0.17	0.23
11.	**0.03**	**0.04**	**0.01**	**0.00**	0.11	0.09	0.08	0.12
12.	**0.00**	**0.00**	**0.00**	**0.00**	0.44	0.42	0.29	0.35

TABLE 3. p-values for test (6.6) applied to $e_H^{Garch}(T)$ (the left side of the table) and $e_H^{BM}(T)$ (the right side of the table), the hedging errors at different maturities $T = 60, 120, 180, 250$. The null hypothesis is $E\ e_H^* = 0$. The samples are those of Table 1. The series and the periods where the null hypothesis is rejected are in bold face. The higher number of bold face values on the left of the table indicate that *the Garch model produces more often hedging errors that are statistically different from 0* on average.

A similar test can be applied to the individual means of the hedging errors. At maturity, a correct hedging strategy leads, with probability one, to no hedging error. Of course, in practice, the error will never be exactly 0. However, applying the correct hedging strategy should produce, on average, no hedging error. Hence, testing the hypothesis $Ee_{H,t}^*(T) = 0$ allows us to evaluate statistically the quality of the hedging strategy based on model $*$ (as usual $*$ stands for GARCH or BM).

Table 3 contains the p-values of the test (6.6) based on the sample mean (6.7) with $D_t := e_{H,t}^*(T)$. The sampling distribution of the test statistic is obtained by overlapping-block bootstrap with the block length selection criteria of Politis and White (2003). The results in Table 3 confirms the previous findings. The Garch strategy produces statistically significant errors on the series with a particularly non-favorable (to Garch) ratio MSE_{BM}/MSE_{GARCH} while the BM approach seems to produce, with few exceptions, a correct replicating strategy, i.e. a mean zero average error.

To summarize, it appears that, for the time series considered, the Garch(1,1) model does not outperform the simple BM model. Moreover, for most of the twelve series, the BM model produces better results, i.e. smaller replication errors. The results of the exercise seems to indicate that, for the series under scrutiny, the Garch(1,1) dynamics provides a poorer description of the longer horizon evolution of the price process than a naive modeling strategy that simply takes the past for the future.

7. Conclusions

We investigated the relationship between non-stationarities and the IGARCH effect in a novel modeling framework proposed in Stărică and Granger (2005) that treats the returns as independent observations with a time-varying unconditional second moment. By successfully modeling twelve series of index returns, we brought further evidence that this non-stationary framework is a viable set-up for the analysis of the dynamics of stock returns. The novel modeling set-up was complemented with an innovative estimation approach of the unconditional time-varying volatility based on the Adaptive Weights Smoothing approach of Polzehl and Spokoiny (2003).

As a corollary, our analysis gave empirical evidence of the possible causal relationship between shifts in the unconditional volatility and the IGARCH effect as emphasized by Diebold (1986), Lamoureux and Lastrapes (1990), Hamilton and Susmel (1994), Cai (1994), Mikosch and Stărică (2004) among others. It indicated that periods of relative small changes of the unconditional variance are characterized by the absence of the IGARCH effect. On the other hand, we found that the periods displaying significant changes of the unconditional variance of returns were often characterized by the presence of the IGARCH effect.

We showed that GARCH(1,1) process often fails to model the dynamics of index returns producing estimates of the unconditional variance that are significantly bigger than the variance of the sample. We evaluated the forecasting performance of the GARCH(1,1) model on such samples strongly affected by the IGARCH effect. We showed that the GARCH(1,1) model often fails to produce reasonable longer horizon forecasts and that poor forecasting episodes can last at least four years (sometimes much longer than that). We found that the poor forecasting behavior affects negatively the quality of GARCH(1,1) replicating strategies of simple claim. More concretely, the GARCH(1,1) produces replication errors that are significantly greater than those of a naive modeling approach that takes the past as the future.

8. Appendix.

All Ordinaries (All Ords ASX) index is made up of the weighted share prices of about 500 of the largest Australian companies. The Austrian Traded Index (ATX) is a capitalization-weighted index of the most heavily traded stocks on the Vienna Stock Exchange. The CAC-40 Index is a narrow-based, modified capitalization-weighted index of 40 companies listed on the Paris Bourse. The German Stock Index (DAX) is a total return index of 30 selected German blue chip stocks traded on the Frankfurt Stock Exchange. The FTSE 100 Index is a capitalization-weighted index of the 100 most highly capitalized companies traded on the London Stock Exchange. The Stockholm Options Market Index (OMX) is a capitalization-weighted index of the 30 stocks that have the largest volume of the trading on the Stockholm Stock Exchange. The Russell 3000 Index is a total return index of 3000 companies representing approximately 98% of the U.S. market. The S&P/Toronto Stock Exchange Composite Index (S&P/TSX) is a capitalization-weighted index designed to measure market activity of stocks listed on the TSX. The BEL 20 Index is a modified capitalization-weighted index of the 20 most capitalized and liquid Belgian stocks that are traded on the Brussels Stock Exchange. The Nikkei-225 Stock Average is a price-weighted average of 225 top-rated Japanese companies listed in the First Section of the Tokyo Stock Exchange. FAZ is a stock index produced by the Frankfurter Allgemeine Zeitung, representing 500 German stocks. FAZ index is share price index and only reflect price trends. (By contrast the DAX, a performance index, also take dividends and rights issues into account.) The Dow Jones Industrial Average is a price-weighted average of 30 blue-chip stocks that are generally the leaders in their industry.

References

[1] Andersen, T, Bollerslev, T. (1998) Answering the critics: Yes, ARCH models do provide good volatility forecasts. *International Economic Review* **39**, 885–905.

[2] Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994) GARCH models. In: Engle, R.F. and McFadden, D.L. (Eds.) *Handbook of Econometrics.* Vol. **4**, pp. 2961–3038. Elsevier, Amsterdam.

[3] Cai, J. (1994) A Markov model of unconditional variance in ARCH. *J. Business and Economic Statistics* **12**, 309–316.

[4] Choi, Y. (2005) An analytical approximation to the option formula for the GARCH model *International Review of Financial Analysis* **14**, 149–164.

[5] Diebold, F.X. (1986) Modeling the persistence of the conditional variances: a comment. *Econometric Reviews* **5**, 51–56.

[6] Engle, F., Hong, T., Kane, A., and Noh, J. (1993) Arbitrage valuation of variance forecasts. *Advanced Futures and Options Research* **6**, 393–415.

[7] Engle, F., Kane, A., and Noh, J. (1997) Index-Option Pricing with Stochastic Volatility and the Value of Accurate Variance Forecasts. *Review of Derivatives Research* **1**.

[8] Duan, J.-C. (1995) The Garch option pricing model. *Mathematical Finance* **5**.

[9] Hamilton, J. and Susmel, R. (1994) Autoregressive conditional heteroskedasticity and changes in regime. *J. Econometrics* **64**, 307–333.

[10] Kallsen, J. and Taqqu, M. S. (1998) Option pricing in ARCH-type models. *Mathematical Finance* **8**, 13–26.

[11] Lamoureux, C.G. and Lastrapes, W.D. (1990) Persistence in variance, structural change and the GARCH model. *J. of Business and Economic Statistics* **8**, 225–234.

[12] Mikosch, T. and Stărică, C. (2004) Non-stationarities in financial time series, the long-range dependence and IGARCH effects. *Rev. of Economics and Statistics* **86**, 378–390.

[13] Polzehl, J. and Spokoiny, V. (2000). Adaptive weights smoothing with applications to image segmentation. *J. of Royal Stat. Soc., Series B*, **62**, 335–354.

[14] Polzehl, J. and Spokoiny, V. (2002). Local likelihood modeling by adaptive weights smoothing. WIAS Preprint 787. `http://www.wias-berlin.de/publications/preprints/787`.

[15] Stărică , C. (2003) Is GARCH(1,1) as good a model as the accolades of the Nobel prize would imply? Working paper. Available under www.math.chalmers.se/∼starica.

[16] Stărică , C. and Granger, C. (2005) Non-stationarities in stock returns. Forthcoming in *Rev. of Economics and Statistics*, **87**.

[17] Straumann, D. (2005) Estimation in Conditionally Heteroscedastic Time Series Models *Lecture Notes in Statistics*, **Vol. 181**, Springer Verlag, ISBN: 3-540-21135-7

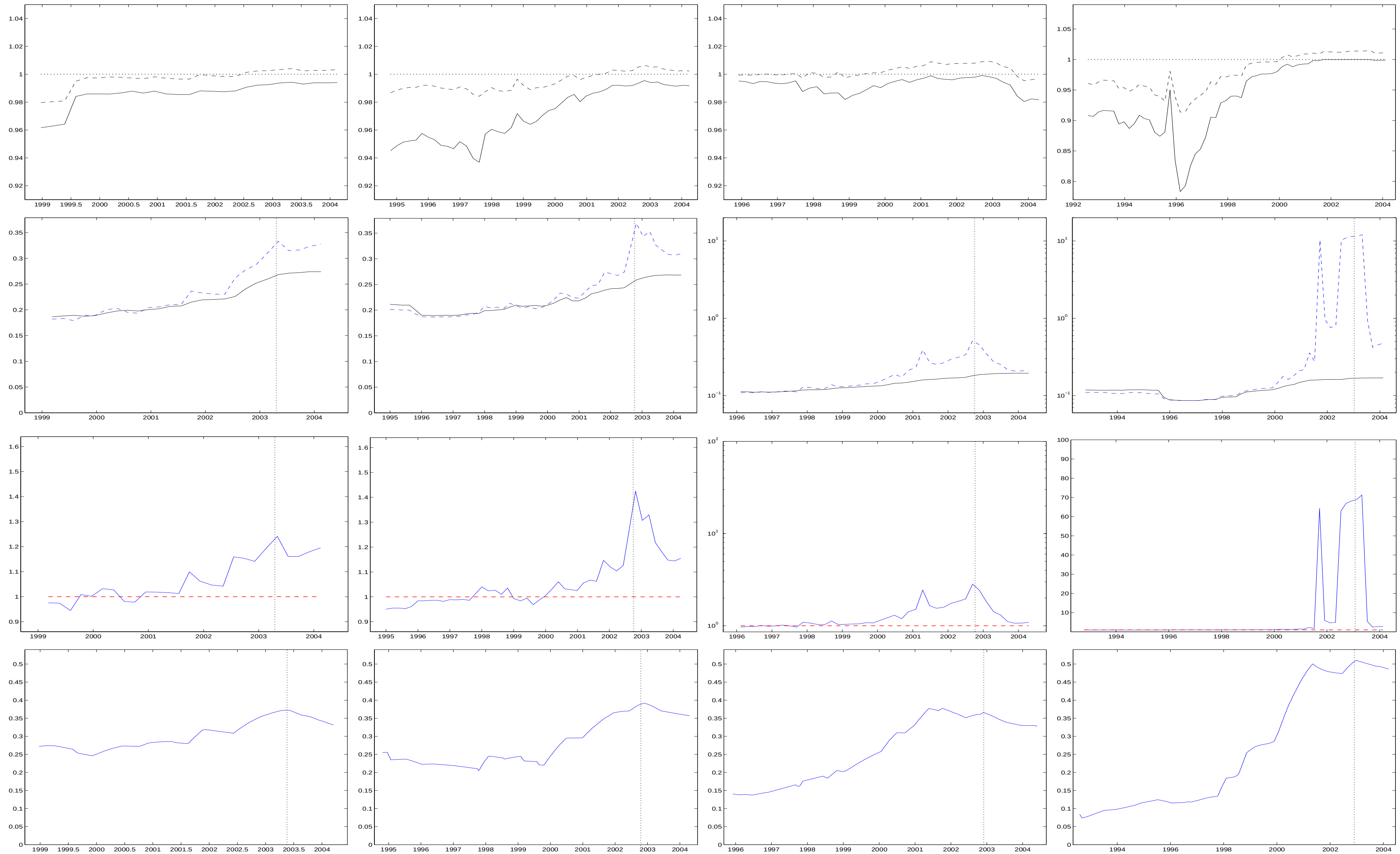

Figure 5.2. Changes in the unconditional volatility and the IGARCH effect for DAX, OMX, Russell 3000, and S&P/TSX indexes. *Top:* The sum $\widehat{\alpha}_1 + \widehat{\beta}_1$ (full line) with the upper one-sided 95% confidence interval (dotted). *Second row:* The sample sd (full line) and the GARCH(1,1) estimated sd (dotted). *Third row:* The ν measure defined in (5.3). *Bottom:* The l measure defined in (5.4). The vertical dotted lines mark the sub-sample used for the detailed forecasting comparison in Section 6.

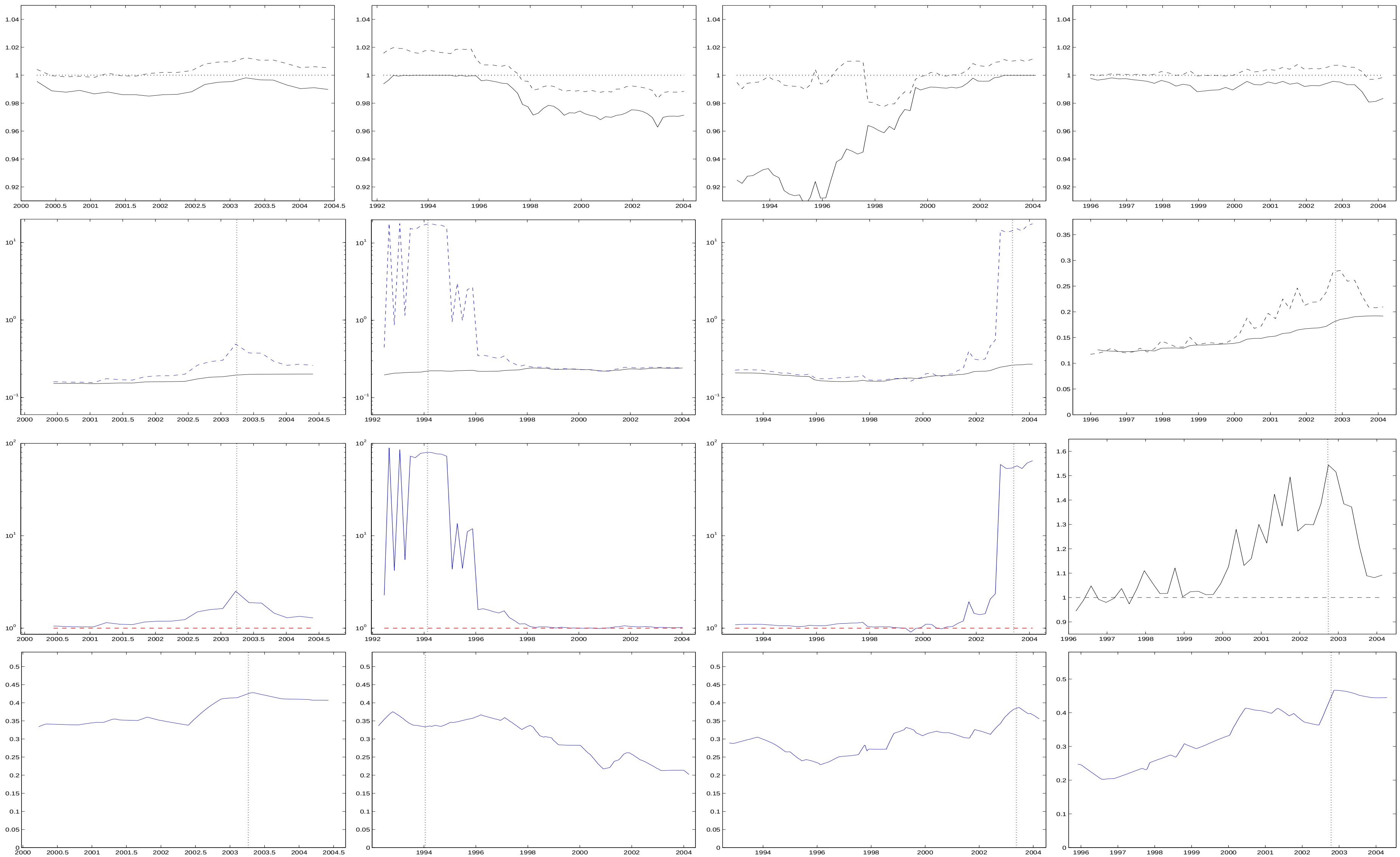

Figure 5.3. Changes in the unconditional volatility and the IGARCH effect for BEL 20, NIKKEI 225, FAZ, and DJI indexes. *Top:* The sum $\widehat{\alpha}_1 + \widehat{\beta}_1$ (full line) with the upper one-sided 95% confidence interval (dotted). *Second row:* The sample sd (full line) and the GARCH(1,1) estimated sd (dotted). *Third row:* The s measure defined in (5.3). *Bottom:* The l measure defined in (5.4). The vertical dotted lines mark the sub-sample used for the detailed forecasting comparison in Section 6.

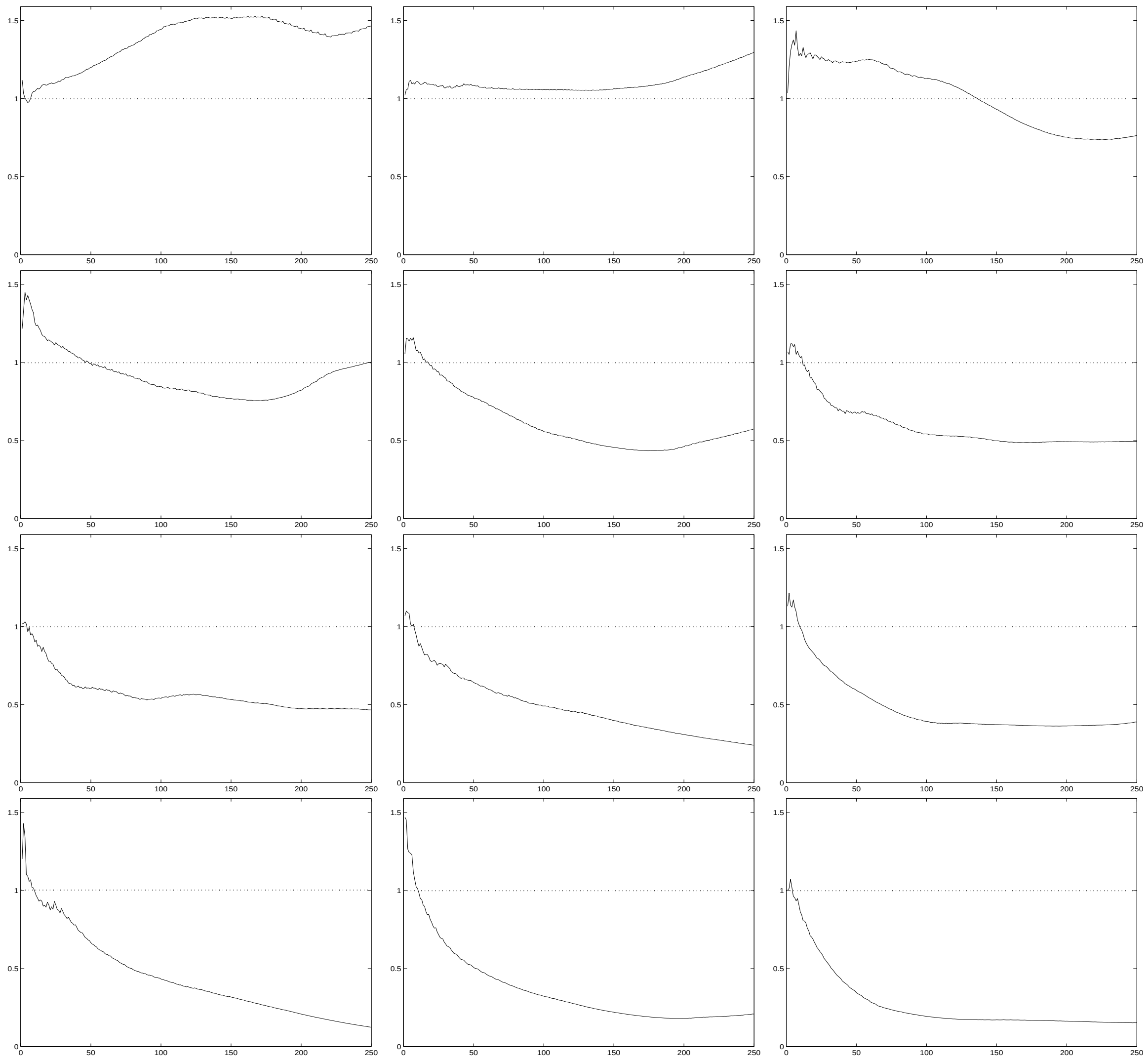

Figure 6.1. The ratio $MSE^{BM}(p)/MSE^{GARCH}(p)$ defined in (6.5) for the sub-samples in Table 1. The order from top-left to bottom-right corresponds to that in the table. A ratio smaller than 1 at horizon p indicates that Garch(1,1) volatility forecast for the next interval of p days is poorer than that based on the simple BM approach.

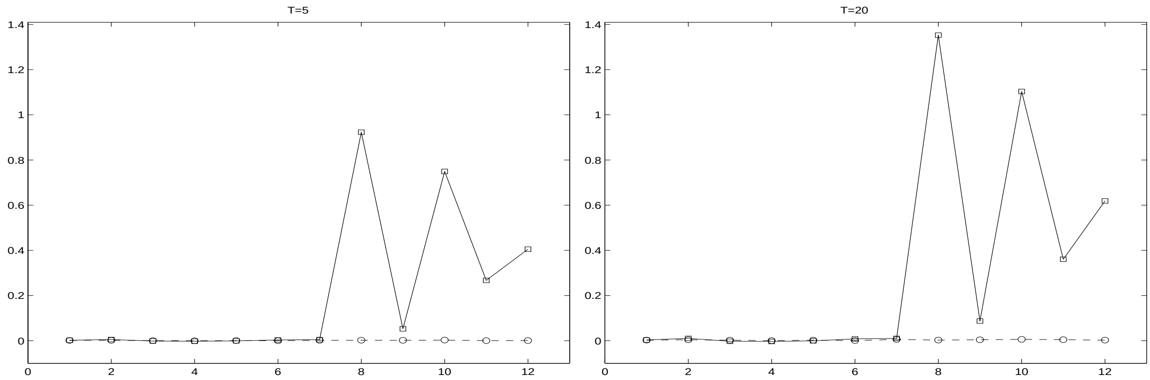

Figure 6.2. Means of the hedging errors of the replicating strategies for straddles with maturities $T = 5$ (left) and $T = 20$ days (right) for the twelve series of Table 1, when the stationary variance of GARCH is employed.

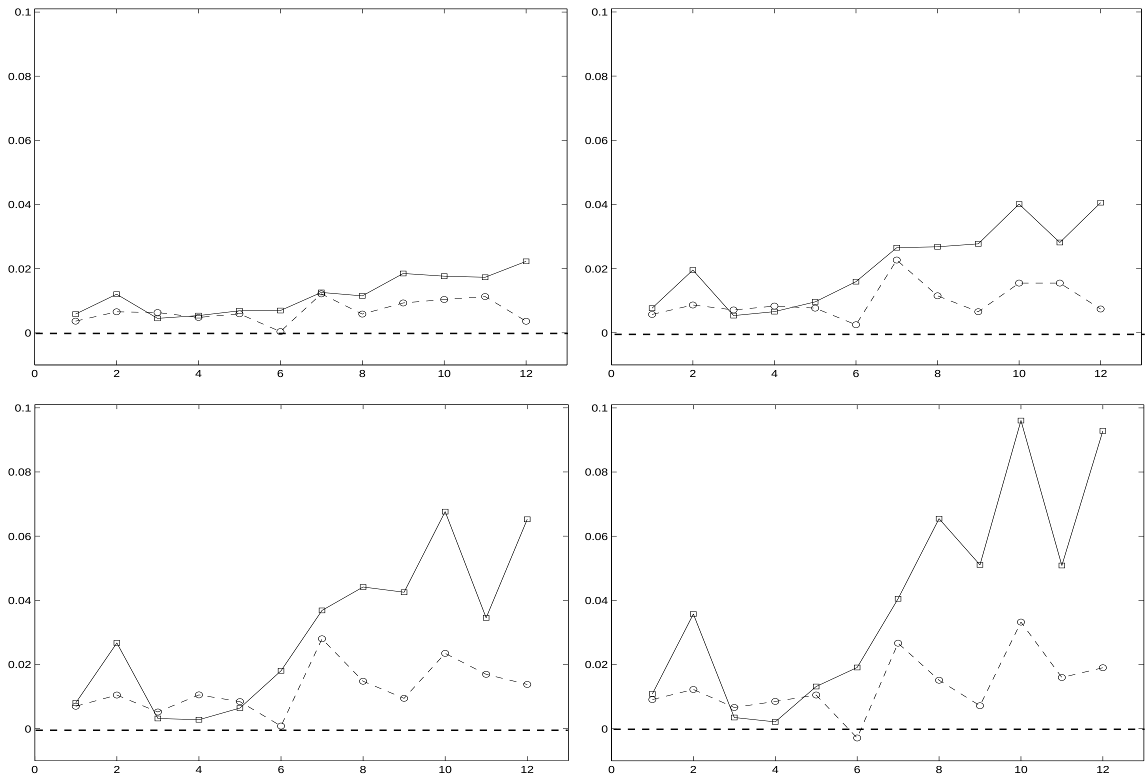

Figure 6.3. Means of the hedging errors of the replicating strategies for straddles with maturities $T = 60, 120, 180, 250$ for the twelve series of Table 1.

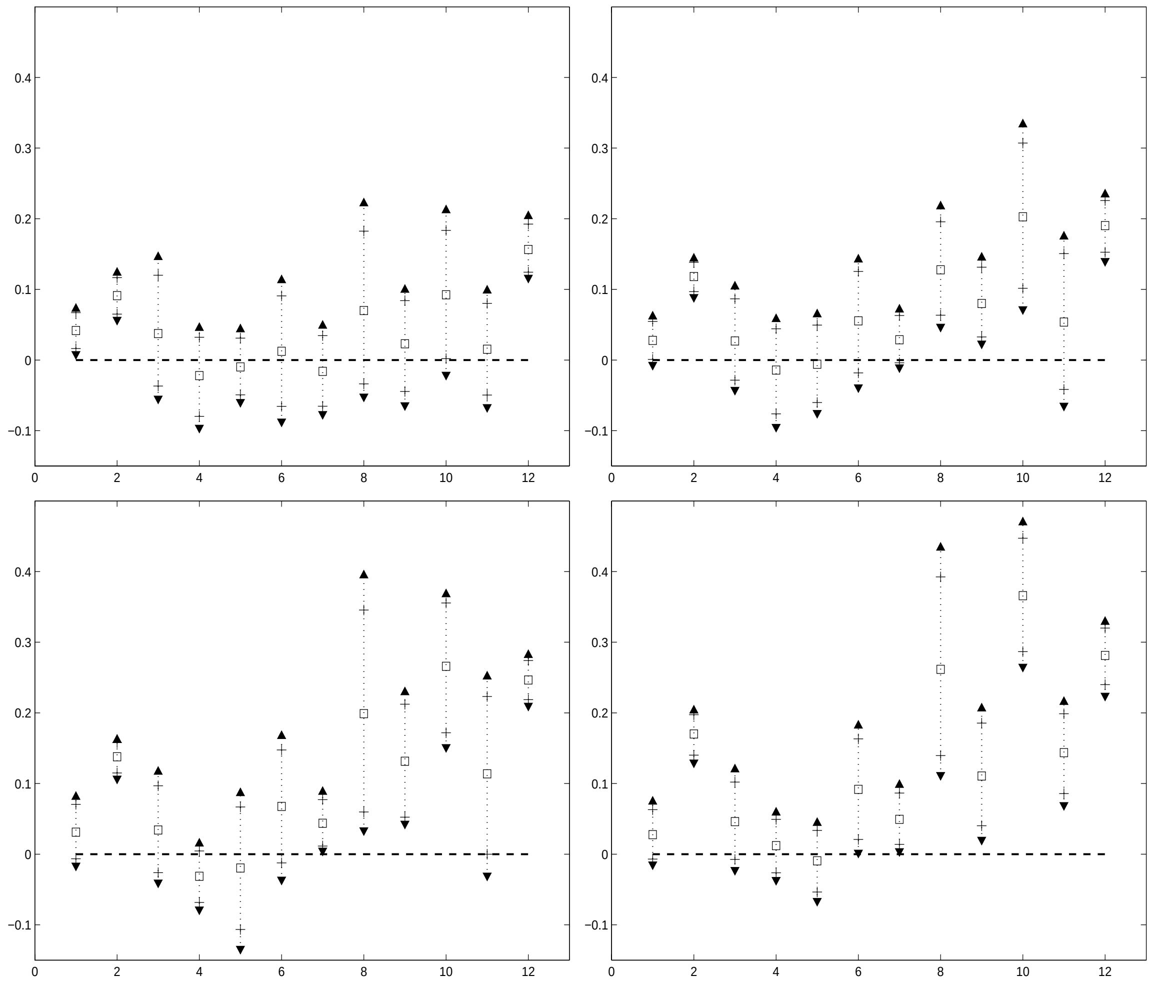

Figure 6.4. The 5%th (downward-pointing triangle), 10%th (cross), 90%th (cross), 95%th (upward-pointing triangle) quantiles and the mean (square) of the bootstrapped sampling distribution of the test statistic $\overline{D}$, corresponding to the differences of hedging errors of replicating strategies for straddles with maturities $T = 60$ (upper-left), 120 (upper-right), 180 (lower-left), 250 (lower-right) for the twelve series of Table 1.