DOTTORATO IN
ECONOMIA E METODI QUANTITATIVI

CURRICULUM IN
METODI STATISTICI PER L'ECONOMIA E L'IMPRESA

XXIX CICLO

Tesi di dottorato

Three Essays on Structural Breaks in Financial Time Series

Dottorando: Davide De Gaetano[1]

Relatore: Prof. Marilena Barbieri

Coordinatore: Prof. Luca Salvatici

[1]E-mail: davide.degaetano@uniroma3.it - davide.degaetano89@gmail.com

I am a part of all that I have met.

(Ulysses, Lord A. Tennyson)

Acknowledgements

I would like to express my special appreciation and sincere gratitude to my advisor Professor Marilena Barbieri, for the continuous support of my Ph.D study and related research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis.
During the second year of my Ph.D., I had the opportunity to spend a period at School of Economics and Finance (Queen Mary, University of London). It was crucial for pushing this project forward and gave me the possibility to interact with many professors and Ph.D. students in a stimulating intellectual environment.
My sincere thanks also go to the referees, Professor Giovanni De Luca and Professor Antonio Parisi, for their insightful and stimulating comments and suggestions on the first draft which led to an improved final version.
I am extremely grateful to the Faculty and the administrative staff of the Department of Economics, University of Roma Tre, for their support and advice.
I would also like to thank all my friends and family for helping me survive all the stress from this year and not letting me give up.
Finally, I would like to express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Contents

List of Figures

List of Tables

Abstract

This dissertation consists of three related essays on structural breaks in financial time series. For independent reading purposes, some explanatory parts related to the conceptual framework of the three essays are repeated in aid of the reader.

In the first essay, *Forecast Combinations for Realized Volatility in Presence of Structural Breaks*, the problem of instability due to changes in the parameters of some Realized Volatility models has been addressed. The analysis is based on 5-minute RV of four U.S. stock market indices. Three different representations of the log Realized Volatility have been considered and rewritten as linear regression models.
In order to test the presence of structural breaks in Realized Volatility, we propose the use of standard tests for parameter instability, usually implemented in a regression context. In particular the attention has been focused on the recursive estimates test proposed by Ploberger *et al.* (1989). This choice is based on the fact that this test does not assume a particular pattern of deviation from the null hypothesis and it does not require the specification of the locations of the break points. Moreover, it is suitable for high dimensional data ensuring, at the same time, not high computational costs.
In order to account for potential structural breaks when generating forecasts, we have proposed the use of several forecast combinations. These tools, originally defined in a regression context, have been successfully adapted to the considered Realized Volatility models. All of the proposed forecast combinations are based on different estimation windows, with alternative weighting schemes, They do not take into account explicitly estimated break dates and, again, they are appropriate for high dimensional data.
In order to evaluated the advantages of the considered forecast combinations, for each Realized Volatility model, an out-of-sample forecasting exercise has been performed. The forecasting performances of the proposed approaches has been compared, in terms of two different loss functions, by using the Model Confidence Set procedure (Hansen *et al.*, 2011).
Our analysis has highlighted the importance of taking into account structural breaks in the Realized Volatility models. moreover it gives empirical evidence of the effectiveness of the proposed forecast combinations in this context.

In the second essay, *A Bootstrap Bias Correction of long run fourth order moment estimation in the CUSUM of squares test*, we have focused on the CUSUM of squares test proposed by Sansó *et al.* (2004) for the detection of structural breaks in financial data. It makes adjustments to the original proposal by Inclan and Tiao (1994) that allow the time series to obey a wide class of dependent processes, including GARCH and log-normal stochastic volatility processes, under the null. The test statistic is based on a consistent estimation of the the long run fourth order moment which can be obtained by using a kernel Heteroskedasticity Autocorrelation Consistent (HAC) estimator. Unfortunately, in the case of strong dependence and in relative small samples, the HAC estimation of the long run fourth order moment could be inefficient and, as a consequence, significant biases may arise leading to significant size distortion of the test (Rodriguez and Rubia, 2007).

In this essay we have proposed a bias correction of the estimation of the long run fourth order moment based on the stationary bootstrap proposed by Politis and Romano (1994). The choice of this resampling technique is justified by the stationarity and weakly dependence of the time series under the assumptions which ensure the existence of the limiting distribution of the test statistic, under the null hypothesis.

In order to evaluate the effect of the proposed bias correction we have implemented an extensive Monte Carlo experiment considering two alternative data generating processes, the GARCH(1,1) and the log-normal stochastic volatility. For both the specifications, we have considered a parametric space which includes empirical values typically observed in practice and different sample sizes.

The results give evidence that the bootstrap approach is better able to correctly identify the presence of structural breaks in the data. More specifically, in the GARCH(1,1) data generating process without breaks, the false rejection rates could be considered non statistically different from the nominal values when the persistence is not high and for large sample size even in the case of high persistence. These results are confirmed in the case of the log-normal stochastic volatility model. When the data are generated assuming the presence of a single break in the middle of the sample, the bias corrected estimation is able to correctly identify the break and its location, for all the sample sizes and in both the considered models.

The proposed procedure has been applied to analyse the presence of structural breaks in two real time series, IPC-Mexico and CNX Nifty-India. In both cases it seems to work quite well resulting more robust with respect to extreme observations.

In the last essay, *Forecasting with GARCH models under structural breaks: an approach based on combinations across estimation windows*, we have addressed the problem of forecasting in presence of structural breaks in the unconditional variance of a GARCH (1,1) model. The existence of structural breaks in the volatility can be the cause of estimation problems and forecast failures. If they are present in the data generating process but they are not considered in the specification of the model, the analysis could be biased toward a spurious persistence. The problem is not new in the econometric literature and generally it has been solved by selecting an arbitrary single estimation window in which only a fraction of the most recent observations is used to estimate the parameters and to generate the forecasts. In order to solve the problems arising with the choice of a single estimation window, it can be useful to consider forecast combinations generated by the same model but over different estimation windows.

In this essay, we have proposed some weighting schemes to average forecasts across different estimation windows to account for structural changes in the unconditional variance of a GARCH (1,1) model. Each combination is obtained by averaging forecasts generated by recursively increasing an initial estimation window of a fixed number of observations v. Three different choices of the combination weights have been proposed. In the first scheme, the forecast combination is obtained by using equal weights to average the individual forecasts; the second weighting method simply assigns heavier weights to forecasts that use more recent information; the third is a trimmed version of the forecast combination with equal weights where a fixed fraction of forecasts with the worst performance are discarded.

In order to evaluate the effects of the choice of the tuning parameter v and the effectiveness of the proposed procedures in accounting for structural breaks, an extensive Monte Carlo experiment has been implemented. The results give evidence that forecast combinations with high values of v seem to be effective in accounting for structural breaks. Moreover, they are also able to perform better with respect to alternative schemes proposed in the literature by Rapach and Strauss (2008) and Rapach *et al.* (2008). The proposed procedures have been evaluated on real data by considering the same data set used by Rapach and Strauss (2008). This empirical application has confirmed the results obtained in the simulations.

1 Forecast Combinations for Realized Volatility in Presence of Structural Breaks

Abstract

In this paper the problem of instability due to changes in the parameters of some Realized Volatility (RV) models has been addressed. The analysis is based on 5-minute RV of four U.S. stock market indices. Three different representations of the log-RV have been considered and, for each of them, the parameter instability has been detected by using the recursive estimates test. In order to analyse how instabilities in the parameters affect the forecasting performance, an out-of-sample forecasting exercise has been performed. In particular, several forecast combinations, designed to accommodate potential structural breaks, have been considered. All of them are based on different estimation windows, with alternative weighting schemes, and do not take into account explicitly estimated break dates. The model confidence set has been used to compare the forecasting performances of the proposed approaches. Our analysis gives empirical evidences of the effectiveness of the combinations which make adjustments for accounting the possible most recent break point.

Keywords: Forecast combinations, Structural breaks, Realized volatility.

JEL Code: C530, C580, G170.

1.1 Introduction

Modelling and forecasting the volatility is an important issue in empirical finance. Traditional approaches are based on the univariate GARCH class of models or stochastic volatility models. More recently, Realized Volatility (RV) has become very popular; it uses improved measures of ex post volatility constructed from high frequency data and provides an efficient estimate of the unobserved volatility of financial markets. In contrast with the GARCH approach, in which the volatility is treated as a latent variable, RV can be considered as an observable proxy and, as a consequence, it can be used in time series model to generate forecasts.

Many authors (see for example Liu and Maheau, 2008; Yang *et al.*, 2015) have highlighted the evidence for structural breaks in RV. Their presence in the data generating process can induce instability in the model parameters. Ignoring structural breaks and assuming wrongly that the structure of a model remains fixed over time, has clear adverse implications. The first finding is inconsistency of the parameter estimates. Moreover, structural changes are likely to be responsible for most major forecast failures of time invariant series models.

In this paper, three different model specifications of the log-RV have been considered. The first is the Heterogeneous Auto-Regressive model (HAR-RV) proposed by Corsi (2004); this model is able to capture many of the features of volatility including long memory, fat tails and self-similarity. The second is the Leverage Heterogeneous Auto-Regressive model (LHAR-RV) proposed by Corsi and Renó (2010); it is able to approximate both long-range dependence and the leverage effect. The last is the Asymmetric Heterogeneous Auto-Regressive model (AHAR-RV) which is a simplified version of the model proposed by Liu and Maheu (2008). In the spirit of the EGARCH model, it allows for asymmetric effects from positive and negative returns. These models, which have become very popular in the econometric literature on RV, have very parsimonious linear structures and, as a consequence, they are extremely easy to implement and to estimate. Moreover, they have good performances in approximating many features which characterize the dynamic of RV (Corsi, 2009).

The aim of this paper is to investigate empirically the relevance of structural breaks for forecasting RV of a financial time series. The presence of structural breaks in the considered RV-representations has been investigated and verified by resorting to a fluctuation test for parameter instability in a regression context. In particular the attention has been focused on the recursive estimates test proposed by Ploberger *et al.* (1989). This choice is motivated especially in cases where no particular pattern of the deviation from the null hypothesis of constant parameters is assumed. Furthermore the proposal does not require the specification of the locations of the break points.

In order to handle parameter instability, some specific forecast combinations have been introduced and discussed. They are based on different estimation windows with alternative weighting schemes. All of them do not incorporate explicitly the estimation of the break dates and, as shown by Pesaran and Timmermann (2007), they do not suffer from this estimation uncertainty.

The forecasting performances of the proposed forecast combinations for the three different specifications of RV models have been compared in terms of two loss functions, the MSE and QLIKE. They are the loss functions most widely used to compare volatility forecasting performances and, according to Patton (2011), they provide robust ranking of the models.

In order to assess statistically if the differences in the forecasting performances of the considered forecast combinations are relevant, the model confidence set, proposed by Hansen *et al.* (2011) has been used. This procedure is able to construct a set of models from a specified collection which consists of the best models in term of a loss function with a given level of confidence and it does not require the specification of a benchmark. Moreover, the model confidence set procedure is a stepwise method based on a sequence of significance tests; each test is based on a null hypothesis that the two models under comparison have the same forecasting ability against the alternative that they are not equivalent. The test stops when the first hypothesis is not rejected and, therefore, the procedure does not accumulate type I error (Hansen *et al.*, 2005). The critical values of the test as well as the estimation of the variance useful to construct the test statistic, is determined by using the block bootstrap. This re-sampling technique preserves the dependence structure of the series and it works reasonably well under very weak conditions on the dependency structure of the data.

The empirical analysis has been conducted on four U.S. stock market indices: S&P 500, Dow Jones Industrial Average, Russell 2000 and Nasdaq 100. For all the series the 5-minute RV has been considered; it is one of the most used proxy of the volatility and, as shown in Liu *et al.* (2015), no other realized measure outperforms it in terms of estimation accuracy of asset price variation.

The structure of this paper is as follows. Section 2 introduces the empirical models for RV and briefly illustrates the problem of structural breaks. In Section 3 some of the most used procedures to test parameters instability in regression framework have been reviewed. The attention has been focused on the class of fluctuation tests and, in particular, the Recursive estimates test has been discussed. Section 4 introduces the problem of forecasting in presence of structural breaks and discusses some forecast combinations able to take into account parameters instability. In section 5 the problem of comparing forecasting models is briefly illustrated and the model confidence set is discussed. In section 6 the empirical results on the four U.S. stock market indices are reported and discussed. Some final remarks close the paper.

1.2 Realized volatility models

In the econometric literature many approaches have been developed to model and forecast Realized Volatility with the aim of reproducing the main empirical features of financial time series such as long memory, fat tails and self-similarity. In this paper the attention has been focused on the classic Heterogeneous Auto-Regressive model of Realized Volatility (HAR-RV) and on some of its extensions.

The HAR-RV model, proposed by Corsi (2004), has a very simple and parsimonious structure and remarkably good forecasting performance (Corsi, 2009). Moreover, as highlighted by Vortelinos (2017), it is more accurate on realized volatility forecasting than some non-linear models, like Principal Components Combining, neural networks and GARCH.

In this model, lags of RV are used at daily, weekly and monthly aggregated periods. More precisely, let $v_t = log(RV_t)$ where RV_t is the Realized Volatility at time $t = 1, 2, \ldots, T$. The logarithmic version of the HAR-RV similar to that implemented by Andersen *et al.* (2007), is defined as:

$$v_t = \beta_0 + \beta_1 v_{t-1} + \beta_2 v_t^{(5)} + \beta_3 v_t^{(22)} + \epsilon_t \tag{1}$$

where $\epsilon_t \sim NID(0, \sigma^2)$ and $v_t^{(5)}$ and $v_t^{(22)}$ are defined, respectively, as:

$$v_t^{(5)} = \frac{v_{t-1} + v_{t-2} + \ldots + v_{t-5}}{5} \tag{2}$$

$$v_t^{(22)} = \frac{v_{t-1} + v_{t-2} + \ldots + v_{t-22}}{22} \tag{3}$$

The HAR-RV model is able to capture some well known features of financial returns such as long memory and fat tails (see Corsi, 2009).

The first extension of this model is the Leverage Heterogeneous Auto-Regressive model of Realized Volatility (LHAR-RV) proposed by Corsi and Renó (2010). This model is defined as:

$$\begin{aligned} v_t = \beta_0 + \beta_1 v_{t-1} + \beta_2 v_t^{(5)} + \beta_3 v_t^{(22)} + \beta_4 r_{t-1}^{-} + \beta_5 r_t^{(5)-} + \\ + \beta_6 r_t^{(22)-} + \beta_7 r_{t-1}^{+} + \beta_8 r_t^{(5)+} + \beta_9 r_t^{(22)+} + \epsilon_t \end{aligned} \tag{4}$$

where $\epsilon_t \sim NID(0, \sigma^2)$, r_t are the daily returns and:

$$r_t^{(1)-} = r_{t-1} I_{\{(r_{t-1}<0)\}} \tag{5}$$

$$r_t^{(1)+} = r_{t-1} I_{\{(r_{t-1}>0)\}} \tag{6}$$

$$r_t^{(5)-} = \frac{r_{t-1} + r_{t-2} + \ldots + r_{t-5}}{5} I_{\{(r_{t-1}+r_{t-2}+\ldots+r_{t-5})<0\}} \tag{7}$$

$$r_t^{(5)+} = \frac{r_{t-1} + r_{t-2} + \ldots + r_{t-5}}{5} I_{\{(r_{t-1}+r_{t-2}+\ldots+r_{t-5})>0\}} \tag{8}$$

$$r_t^{(22)-} = \frac{r_{t-1} + r_{t-2} + \ldots + r_{t-22}}{22} I_{\{(r_{t-1}+r_{t-2}+\ldots+r_{t-22})<0\}} \tag{9}$$

$$r_t^{(22)+} = \frac{r_{t-1} + r_{t-2} + \ldots + r_{t-22}}{22} I_{\{(r_{t-1}+r_{t-2}+\ldots+r_{t-22})>0\}} \tag{10}$$

where $I_{\{.\}}$ is the indicator function. The LHAR-RV model approximates both long-range dependence and the leverage effect. Some authors (Asai *et al.*, 2012) suggest to include only the nagative part of heterogeneous returns since the estimates of the coefficients of the positive ones are usually not significant.

The second extension is the Asymmetric Heterogeneous Auto-Regressive model of Realized Volatility (AHAR-RV) which is a simplified version of the model proposed by Liu and Maheu (2008). It is defined as:

$$v_t = \beta_0 + \beta_1 v_{t-1} + \beta_2 v_t^{(5)} + \beta_3 v_t^{(22)} + \beta_4 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} + \beta_5 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} I_{\{r_{t-1}<0\}} + \epsilon_t \tag{11}$$

The last two terms allow for asymmetric effects from positive and negative returns in the spirit of the EGARCH model.

All the considered models can be rewritten in a standard regression framework:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + \epsilon_t \tag{12}$$

where $y_t = v_t$, $\mathbf{x}_t$ is the $p \times 1$ vector of the regressors at time t and $\boldsymbol{\beta}$ is the $p \times 1$ vector of the corresponding coefficients. Of course the number p and the specification of the vector $\mathbf{x}_t$ are different for each model.

Many studies (see, for example, Liu and Maheau, 2008) agree on the existence of structural breaks in RV. If structural breaks are present in the data generating process, they could induce instability in the model parameters. Ignoring them in the specification of the model, could provide a wrong modelling and forecasting for the RV.

To dealing with structural breaks, the linear regression model (12) is assumed to have time varying coefficients and so it may be expressed as:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}_t + \epsilon_t \qquad\qquad t = 1, 2, \dots T \tag{13}$$

In many applications it is reasonable to assume that there are m breakpoints at the date $\tau_1, \tau_2, \cdots, \tau_m$ in which the coefficients shift from one stable regression relationship to a different one. Thus, there are $m + 1$ segments in which the regression coefficients are constant. Model (13) can be rewritten as

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}_{\tau_{j-1}+1:\tau_j} + \epsilon_t \qquad t = 1, 2, \dots T \qquad j = 1, 2, \dots m + 1 \tag{14}$$

and, by convention, $\tau_0 = 1$ and $\tau_{m+1} = T$.

1.3 Testing for structural breaks

The presence of structural breaks can be tested through the null hypothesis that the regression coefficients remain constant over time, that is:

$$H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta} \qquad t = 1, 2, \dots T$$

against the alternative that at least one coefficient varies over time.
In the statistical and econometric literature, testing for parameters instability in a regression framework has been treated using different approaches.
The classical test for structural change is the well known Chow Test (Chow, 1960). This testing procedure splits the sample into two sub-samples, estimates the parameters for each sub-sample, and then, using a classic F statistic, a test on the equality of the two sets of parameters is performed. For a review, which includes also some extensions in different contexts, see Andrews and Fair (1988). The principal issue of the Chow test is the assumption that the break-date must be known a priori. Generally this procedure is used by fixing an arbitrary candidate break-date or by selecting it on the base of some known feature of the data. However, the results can be highly sensitive to these arbitrary choices; in the first case, the Chow test may be uninformative and, in the second case, it can be misleading (Hansen 2001).
More recently, the literature has focused on a more realistic problem in which the number of break points and their location are supposed to be unknown (see Perron, 2006, for a survey). In this context, one of the major contribution is the strategy proposed by Bai and Perron (1998, 2003a, 2003b) who developed an iterative procedure which allows consistent estimation of the number and the location of the breaks points together with the unknown regression coefficients in each regime. In their procedure the breaks are considered deterministic parameters and so it is not required the specification of their underlying generating process. The number of breaks can be determined sequentially by testing for $q + 1$ against q or using a global approach of testing for q against no breaks. However the procedure needs the specification of some restrictions such as the minimum distance between breaks and their maximum number.
Another approach to change point testing is based on the generalized fluctuation tests (for a survey see Zeileis *et al.*, 2003). Such an approach has the advantage of not assuming a particular pattern of deviation from the null hypothesis. Moreover, although it is possible in principal to carry out the location of the break points, this method is commonly used only to verify their presence; with this aim the fluctuation tests will be used in this paper.

1.3.1 Fluctuation tests

Although the generalized fluctuation test framework includes formal significance tests, it is commonly used in a graphic way. The general idea is to fit a regression model to the data and derive the empirical process that captures the fluctuation in the residuals or in the parameter estimates. Under the null hypothesis of constant regression coefficients, fluctuations are governed by functional central limit theorems (Kuan and Hornik, 1995) and therefore boundaries can be found that are crossed by the corresponding limiting processes with fixed probability α. When the fluctuation of the empirical process increases, there is evidence of structural changes in the parameters. Moreover, its trajectory may also highlight the type of deviation from the null hypothesis as well as the dating of the structural breaks. As pointed out previously, the generalized fluctuation tests can be based on the residuals or on the parameter estimates of the regression model. The first class includes the classical CUSUM based on cumulative sums of recursive residuals (Brown *et al.*, 1975), the CUSUM test based on OLS residuals (Ploberger and Krämer, 1992) and the MOSUM tests (Chu *et al.*, 1995a). The second class includes the Recursive Estimates (RE) test (Ploberger *et al.*, 1989) and the Moving Estimates (ME) test (Chu *et al.*, 1995b). In both of them, the vector of unknown parameters is estimated recursively with a growing number of observations, in the RE test, or with a moving data window in the ME test and then compared to the estimates obtained by using the whole sample. Define:

$$\mathbf{y}'_{1:t} = (y_1, y_2, \ldots, y_t) \tag{15}$$

$$\mathbf{X}'_{1:t} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t)' \tag{16}$$

and let

$$\widehat{\boldsymbol{\beta}}_{1:t} = (\mathbf{X}'_{1:t}\mathbf{X}_{1:t})^{-1}\mathbf{X}'_{1:t}\mathbf{y}_{1:t} \qquad t = p, p+1, \ldots, T \tag{17}$$

be the ordinary least squares (OLS) estimate of the regression coefficients based on the observations up to t.

The basic idea is to reject the null hypothesis of parameter constancy if these estimates fluctuate too much. Formally the test statistic is defined as:

$$S^{(T)} = \sup_{0 \leq z \leq 1} \left\| B^{(T)}(z) \right\|_{\infty} \tag{18}$$

with

$$B^{(T)}(z) = \frac{\phi(z)}{\widehat{\sigma} T} (\mathbf{X}'_{1:T}\mathbf{X}_{1:T})^{1/2} (\widehat{\boldsymbol{\beta}}_{1:\phi(z)} - \widehat{\boldsymbol{\beta}}_{1:T}) \tag{19}$$

where

$$\widehat{\sigma} = \left[\frac{1}{T-p} \sum_{t=1}^{T} (y_t - \mathbf{x}_t' \widehat{\boldsymbol{\beta}}_{1:T}) \right]^{1/2} \tag{20}$$

and $\phi(z)$ is the largest integer less than or equal to $p + z(T-p)$ and $\|.\|_\infty$ is the maximum norm. As proved in Ploberger *et al.* (1989), $B^{(T)}(z)$ is a p-dimensional stochastic process such that

$$B^{(T)}(z) \xrightarrow{\mathcal{D}} B(z) \tag{21}$$

where $B(z)$ is a Brownian Bridge.

The distribution of $\sup_{0 \leq z \leq 1} \|B(z)\|_\infty$ is given Billingsley (1999):

$$P\left(\sup_{0 \leq z \leq 1} \|B(z)\|_\infty \leq x \right) = \left[1 + 2 \sum_{i=1}^{\infty} (-1)^i e^{-2i^2x^2} \right]^p \qquad x \geq 0 \tag{22}$$

This result allows to construct the asymptotic boundaries crossing probabilities.

1.4 Forecasting Methods in presence of structural breaks

Once the parameter instability due to the presence of structural breaks has been detected, the problem is how to account for it when generating forecasts. Indeed parameter instability could cause forecast failures in macroeconomic and financal time series (for a survey see Clements and Hendry, 2006).

When it is possible to identify the exact date of the last break, the standard solution is to use only observations over the post-break period. In practice, the dates of the break points are not known a priori and an estimation procedure has to be used. It could produce imprecise values which affect negatively the specification of the forecasting model and, as a consequence, poor performances of the forecasts. Furthermore, even if the last break date is correctly estimated, the forecasts generated by this scheme are likely to be unbiased and may not minimize the mean square forecast error (Pesaran and Timmermann 2007). Moreover, if the last break is detected close to the boundaries of the data sample, the parameters of the forecasting model are estimated with relatively short sample and the estimation uncertainty may be large.

However, as pointed out by Pesaran and Timmermann (2007), the pre-break observations could be informative for forecasting even after the break. More specifically, it is appropriate to choose a high fraction of the pre-break observations especially when the break size is small, the variance parameter increases at the break point and the number of post break observations is small. Furthermore the forecasting performance is sensitive to the choice of the observation window. A relative long estimation window reduces the forecast error variance but increases its bias; on the other hand, a short estimation window produces an increase in the forecast error variance although the bias decreases. Therefore, an optimal window size should balance the trade off between an accurate estimate of the parameters and the possibility that the data come from different regimes.

Although Pesaran and Timmermann (2007) have proposed an optimal estimation window[2], in practice, the selection of a single best estimation window is not an easy task and, in many empirical studies, it is arbitrarily determined.

Alternatively, in order to deal with the uncertainty over the size of the estimation window, it is possible to combine forecasts generated from the same model but over different estimation windows. This strategy is in the spirit of forecast combination obtained by estimating a number of alternative models over the same sample period (for a review see Timmermann, 2006). It has the advantage that it avoids the direct estimation of breakpoint parameters and it is applicable to general dynamic models and for different estimation methods (Pesaran and Timmermann, 2007). Moreover, Pesaran and Pick (2011), have developed theoretical results for random walks with breaks in the drift and volatility and for a linear regression model with a break in the slope parameter. In particular, they have shown that, when a small number of post-break data is likely to lead to imprecise estimates of the point and the size of the break, also a scheme based on forecast averaging over estimation window could lead to a lower bias and root mean square forecast error than forecasts based on a single estimation windows. Their results has been confirmed by an application to weekly returns on 20 equity index futures. This idea has been fruitfully applied in macroeconomic forecasting (see Assenmacher-Wesche and Pesaran, 2008 and Pesaran, *et al.*, 2009).

In view of the above considerations, in this paper the attention has been focused on the forecasts schemes generated from the same model but over different estimation windows. In particular for each of the considered Realized Volatility model, different forecast combina-

[2]The optimal estimation window leads to the lowest average out-of-sample loss over the cross-validation sample. More precisely, the cross-validation approach reserves some of the last observations for an out-of-sample estimation exercise and the estimation window is chosen in order to minimize the mean square forecast error on this sample.

tions have been considered focusing on those that do not incorporate explicitly the estimation of the break dates. Moreover, in the analysis one-step ahead forecasts have been considered and so it is assumed that no structural breaks occur in the forecast period (for forecasting with structural breaks over the forecast period see Pesaran *et al.*, 2006 and Maheu and Gordon, 2008).

Forecast combination with equal weights.

As pointed out previously, the forecast combination with equal weights is the simplest combination but it is robust to structural breaks of unknown break dates and sizes. Moreover it performs quite well especially when the break is of moderate magnitude and it is located close to the boundaries of the data sample (Pesaran and Pick, 2011).

Let ω be the minimum acceptable estimation window size[3]. The forecast combination with equal weights is defined by:

$$\widehat{y}_{T+1} = \frac{1}{T-\omega} \sum_{\tau=1}^{T-\omega} \left(\mathbf{x}'_{T+1} \widehat{\boldsymbol{\beta}}_{\tau+1:T} \right) \tag{23}$$

Many researches (see for example Pesaran and Timmermann, 2007) have highlighted the advantages of this scheme; it has good performances also when there is uncertainty about the presence in the data of structural breaks. This approach avoids also any estimation procedure for the weights.

Forecast combination with location weights.

By looking at equation (23) it is evident that the weights in the equally weighted combination can be converted into weights on the sample observations $\mathbf{x}_t$. As discussed in Tian and Anderson (2014), the ω most recent observations are used in all of the forecasts, whereas the older observations are used less. Furthermore the influence of each observation is inversely proportional to its distance from the forecasting origin: the most recent data are usually more relevant especially if the regression parameters have significant changes close to the end of the sample.

[3]Common to all the considered methods is that the estimation window should not be smaller than a minimum length ω that can be set equal to the number of regressors plus one. However, as pointed out by Pesaran and Timmerman (2007), to account for the very large effect of parameter estimation error, ω should be at least 3 times the number of unknown parameters.

A way to place heavier weights on the forecasts which are based on more recent data much more than under the equally weighted forecast combination is to use constant weights proportional to the location of τ in the sample.
More precisely, this combination, known as the forecast combination with location weights, is defined by:

$$\widehat{y}_{T+1} = \frac{1}{\sum_{\tau=1}^{T-\omega} \tau} \sum_{\tau=1}^{T-\omega} \tau \left(\mathbf{x}'_{T+1} \widehat{\boldsymbol{\beta}}_{\tau+1:T} \right) \tag{24}$$

Also in this case, no estimation of the weights is needed.

Forecast combination with MSFE weights

This approach, proposed by Pesaran and Timmerman (2007), is based on the idea that the weights of the forecasters obtained with different estimation windows should be proportional to the inverse of the associated out-of-sample mean square forecast error (MSFE) values. To this aim, a cross validation approach is used.
To better understand, let m be the generic start point of the estimation window and assume that $\tilde{\omega}$ is the number of the observations used in the cross validation set, that is the observations used to measure pseudo out-of-sample forecasting performance. The recursive pseudo out-of-sample MSFE value is computed as:

$$MSFE(m|T, \tilde{\omega}) = \tilde{\omega}^{-1} \sum_{\tau=T-\tilde{\omega}}^{T-1} \left(y_{\tau+1} - \mathbf{x}'_{\tau} \widehat{\boldsymbol{\beta}}_{m:\tau} \right)^2 \tag{25}$$

The forecast combination with MSFE weights is then defined as:

$$\widehat{y}_{T+1} = \frac{\sum_{m=1}^{T-\omega-\tilde{\omega}} \left(\mathbf{x}'_{T} \widehat{\boldsymbol{\beta}}_{m:T} \right) (MSFE(m|T, \tilde{\omega}))^{-1}}{\sum_{m=1}^{T-\omega-\tilde{\omega}} (MSFE(m|T, \tilde{\omega}))^{-1}} \tag{26}$$

Together with the parameter ω, the length of the minimal estimation window, this method requires also the choice of the parameter $\tilde{\omega}$, the length of the evaluation window. If this parameter is set too large, too much smoothing may result and, as a consequence, in the combination, the forecasting based on older data will be preferred. On the other hand, if $\tilde{\omega}$ is set too short, although a more precise estimation of the MSFE can be obtained, the ranking of the forecasting methods is more affected by noise. Of course the selection of this parameter depends on the problem at hand and on the length of the series.

Forecast combination with ROC weights

This approach has been proposed by Tian and Anderson (2014). It is based on by-products of the Reverse Ordered CUSUM (ROC) structural break test considered by Pesaran and Timmermann (2002).

It is a two stages forecasting strategy. In the first step a sequence of ROC test statistics, starting from the most recent observations and going backwards in time, is calculated. Each point in the sample is considered as a possible most recent break point.

This test is related to the classical CUSUM test but, in this case, the test sequence is made in reverse chronological order. In particular, the time series observations are placed in reverse order and the standard CUSUM test is performed on the rearranged data set.

In the paper by Pesaran and Timmermann (2002), the test statistics are used to perform a formal structural break test and to estimate the last breakpoint in the sample.

In the second step, the ROC statistics are used to weight the associated post break forecast, developing a forecast combination. Moreover, the weights do not depend on finding and dating a structural break but they are constructed in order to give more weights to observations subsequent to a potential structural break.

In the first step of the procedure, for $\tau = T - \omega + 1, T - \omega, \ldots, 2, 1$, let

$$\mathbf{y}'_{T:\tau} = (y_T, y_{T-1}, \ldots, y_{\tau+1}, y_\tau) \tag{27}$$

$$\mathbf{X}'_{T:\tau} = (\mathbf{x}_T, \mathbf{x}_{T-1}, \ldots, \mathbf{x}_{\tau+1}, \mathbf{x}_\tau) \tag{28}$$

be the observation matrices, and let

$$\widehat{\boldsymbol{\beta}}^{(R)}_{T:\tau} = (\mathbf{X}'_{T:\tau}\mathbf{X}_{T:\tau})^{-1}\mathbf{X}'_{T:\tau}\mathbf{y}_{T:\tau} \tag{29}$$

be a sequence of least squares estimates of $\boldsymbol{\beta}$ associated with the reverse-ordered data sets.

The ROC test statistics s_t are defined as:

$$s_\tau = \frac{\sum_{t=\tau}^{T-\omega} \xi_t^2}{\sum_{t=1}^{T-\omega} \xi_t^2} \qquad \text{for } \tau = T - \omega, T - \omega - 1, \ldots, 2, 1 \tag{30}$$

where ξ_t are the standardized one-step-ahead recursive residuals defined as:

$$\xi_t = \frac{y_t - \mathbf{x}_t'\widehat{\boldsymbol{\beta}}_{T:t+1}^{(R)}}{(1 + \mathbf{x}_t'(\mathbf{X}_{T:t+1}'\mathbf{X}_{T:t+1})^{-1}\mathbf{x}_t)^{1/2}} \tag{31}$$

In the second step of the procedure, all dates τ are considered as possible choices for the last breakpoint. The combination weight on each τ is constructed as:

$$cw_\tau = \frac{\left|s_\tau - \left(\frac{T-\omega-\tau+1}{T-\omega}\right)\right|}{\sum_{\tau=1}^{T-\omega}\left|s_t - \left(\frac{T-\omega-\tau+1}{T-\omega}\right)\right|} \qquad \tau = 1, 2, \ldots, T-\omega \tag{32}$$

Since, under the null hypothesis of no structural break in τ, it is:

$$E(s_\tau) = \frac{T-\omega-\tau+1}{T-\omega} \tag{33}$$

the combination weights vary according to the absolute distances between s_τ and its expected value. As a consequence, cw_τ is larger if this distance is large, that is if the evidence of a structural break is stronger. On the contrary, if in τ there is no evidence of substantial breakpoint, the associated weight is small.

Moreover, the weights do not depend on finding and dating a structural break. However, if the absolute values of the difference between the ROC statistics and their expectation, under the null, start to grow (giving evidence of a potential structural break), the weights cw_τ increase giving more weights to the observations on data, subsequent to τ.

The one-step ahead forecast based on ROC statistics is defined as:

$$\widehat{y}_{T+1} = \sum_{\tau=1}^{T-\omega}\left(cw_\tau(\mathbf{x}_{T+1}'\widehat{\boldsymbol{\beta}}_{\tau+1:T})\right) \tag{34}$$

Forecast combination with ROC location weights

In order to take into account a prior belief on the probability that a time τ could be the most recent break point, it is possible, in the definition of ROC weights, to incorporate an additional weight function l_τ.

Following Tian and Anderson (2014), the new weights are defined as:

$$cw_\tau = \frac{\left|s_\tau - \left(\frac{T-\omega-\tau+1}{T-\omega}\right)\right| l_\tau}{\sum_{\tau=1}^{T-\omega}\left|s_t - \left(\frac{T-\omega-\tau+1}{T-\omega}\right)\right| l_\tau} \qquad \tau = 1, 2, \ldots, T-\omega \tag{35}$$

For example, if a single break point seems to be equally likely at each time point, the natural choice is $l_\tau = 1$ for $\tau = 1, 2, \ldots, T - \omega$. In this case the weights depend only on the magnitude of the ROC statistics and the combination defined in (34) is obtained.

However, in a forecasting context, where the identification of the most recent break is essential, the prior weight l_τ could be chosen as an increasing function of the location of time τ in the full sample. In the spirit of forecast combination with location weights, the most natural choice is $l_\tau = \tau$. Of course different specifications are also allowed.

1.5 Comparing forecasting performances

In order to evaluate the advantages of the defined forecast combinations, for each RV-models, it is necessary to compare them in terms of predictive performances.

The natural way to compare the forecasting abilities of alternative methods is to perform out-of-sample validation in which the sample is divided in such a way that the first R observations are used for model estimation and the last P for forecasting evaluation.

The comparison is made by using a loss function. For univariate forecast, at time t, the loss function can be specified as:

$$L_{i,t} = L(y_t, \widehat{y}_t^{(i)}) \tag{36}$$

where y_t is the actual value at time t and $\widehat{y}_t^{(i)}$ is a forecast obtained using a particular method (i).

Loss functions are a crucial ingredient for assessing the accuracy of forecasts and for discriminating across competing forecasting methods. A generic loss function should satisfy these three proprieties: a) it is a non negative function; b) it is 0 if and only if the forecast equals the actual (no error and no loss); c) an increasing penalty should be applied to increasing forecast errors. Patton (2011) details the most used loss functions in a volatility context and gives evidence on the ability of the loss functions to identify superior forecasts.

The loss functions can be used to rank competing methods in term of forecasting accuracy. To this aim the average loss function for each method is usually calculated and the method with the lowest value is considered to be the ”best”.

Ranking the loss measures is the basic tool to discriminate among alternative methods. However, this procedure does not indicate whether the differences among the performances of the comparing methods are statistically significant.

The problem of assessing statistically if the differences in the performances are relevant, has received a growing interest in the econometric literature and, to this aim, a number of statistical tests has been proposed. Diebold and Mariano (1995) has reviewed some of these tests such as the F-test, the Morgan-Granger Newbold test and the Meese-Rogoff test. They have found that these tests can be negatively influenced by non-Gaussianity, contemporaneous and/or serial correlation of the forecast errors and small sample size. To solve these problems, the tests of equal predictive ability by Diebold and Mariano (1995) and by West (1996) have been proposed. Both of them test the hypothesis that two models have equal predictive ability; the difference between them is that the Diebold and Mariano test treats the forecast as given while the West test accounts for the use of estimated parameters in the forecasting model[4]. The advantages of these tests are that they do not require specific assumptions on the forecast errors and they are flexible about the specification of the loss function, in the sense that the loss function needs not be quadratic and symmetric.

However some problems arise when there are a large number of competing forecast methods. White (2000) has highlighted that in this case data snooping can arise, a problem which occurs when a given set of data is used more than once for inference or model selection. He has proposed a straightforward approach, the reality check procedure, for testing the null hypothesis that the best model has no predictive superiority over a given benchmark model. This permits data snooping to be undertaken with some degree of confidence.

Even if the problem of data snooping is solved, Hansen (2005) has pointed out that this test is affected by the inclusion of poor performing models. For this reason he has introduced the superior predictive ability test; this test improves the reality check test by using a studentized test statistic that reduces the influence of erratic forecasts and invoke a sample-dependent null distribution.

However the superior predictive ability test only provides evidence against a benchmark which, in many cases, could be difficult to specify. A solution to the limitation of the superior predictive ability test is the model confidence set approach, proposed by Hansen *et al.* (2003; 2011).

[4]Hansen (2005) has shown that the Diebold and Mariano test and the West test coincide if the forecasts are treated as given.

1.5.1 The Model Confidence Set

The goal of this procedure, which does not require the specification of a benchmark explicitly, is to determine which models, from an initial set M^0 of models indexed by $i = 1, \dots, M^0$, exhibit the same predictive ability in term of a loss function. Following Hansen *et al.* (2005), the term *model* is used loosely. It can refer to a model, a method, or a rule. In the following it is used to indicate a forecasting method associated to a RV-model.

The model confidence set is a stepwise procedure which starts by setting $M = M^0$. The algorithm is based on an equivalence test δ_M and an elimination rule e_M . The equivalence test assumes the value 0 if a properly defined null hypothesis $H_{0,M}$ is not rejected at significance level α and 1 otherwise. In the latter case the model that is to be removed from M is identified by the elimination rule e_M and the set M is updated. The procedure is repeated until the null hypothesis $H_{0,M}$ is not rejected. The collection of the optimal models at level $1 - \alpha$, denoted with $\widehat{M}^*_{1-\alpha}$, contains the surviving models.

Despite its sequential nature, the model confidence set procedure does not accumulate type I error. This is due to the fact that the test stops when the first hypothesis is not rejected.

To compare the relative performances of the models under comparison, for every $i = 1, \dots, M^0$ we specify $L_{i,t}$, a loss function associated with model i and time t, where $t = R+1, R+2, \dots, T$ while R is the length of the in-sample period and it will be further discussed in section 1.6.1. Let

$$d_{ij,t} = L_{i,t} - L_{j,t} \quad \forall i, j \in M \tag{37}$$

be the differential loss between model i and j. The hypothesis compared in each significance test are:

$$\mathrm{H}_{0,M} : \mathrm{E}(d_{ij,t}) = 0 \quad \text{for all } i, j \in M \tag{38}$$

$$\mathrm{H}_{A,M} : \mathrm{E}(d_{ij,t}) \neq 0 \quad \text{for some } i, j \in M$$

In order to test the hypothesis (38), let $\bar{L}$ be the loss of model i averaged over the forecasting period:

$$\bar{L}_i = \frac{1}{T-R} \sum_{t=R+1}^{T} L_{i,t} \tag{39}$$

$\overline{d}_{ij}$ the average loss of model i relative to model j:

$$\overline{d}_{ij} = \frac{1}{T-R} \sum_{t=R+1}^{T} d_{ij,t} = \bar{L}_i - \bar{L}_j \tag{40}$$

and $\overline{d}_{i.}$ the average loss of model i relative to all models:

$$\overline{d}_{i.} = \frac{1}{m} \sum_{j \in M} \overline{d}_{ij} = \bar{L}_i - \frac{1}{m} \sum_{j \in M} \bar{L}_j \tag{41}$$

where m is the total number of models in M.

The hypothesis (38) maps naturally into the test statistic:

$$T_{max,M} = \max_{i \in M} t_{i.} \tag{42}$$

where

$$t_{i.} = \frac{\overline{d}_{i.}}{\sqrt{\widehat{var}(\overline{d}_{i.})}} \tag{43}$$

and $\widehat{var}(\overline{d}_{i.})$ denotes the estimate of the variance of $\overline{d}_{i.}$.

When using this test statistic, it is necessary to define an elimination rule. The most reasonable choice associated to the $T_{max,M}$ statistic is:

$$e_{max,M} = \arg\max_{i \in M} t_i \tag{44}$$

Since the distribution, under the null hypothesis, of the test statistic is not known, the critical values of the test can not be determined analytically and re-sampling techniques have to be used. In the context of model confidence set, the block bootstrap is generally used. It also allows to obtain the estimate of the variances in (106) and the estimation of the p-value associated to the null hypothesis $H_{0,M}$.

The basic idea of the block bootstrap procedure is to resample randomly with replacement blocks of consecutive observations so that the original time series structure is preserved in each block. The blocks are assembled together in random order to obtain a simulated version of the original series.

The block bootstrap procedure works reasonably well under very weak conditions on the de-

pendency structure of the data. Moreover it does not require specific assumptions on the data generating process. As a consequence, it has been applied to a very broad range of applications especially in the context of time series analysis.
According to Hansen *et al.* (2011), the block bootstrap procedure can be adapted in the context of model confidence set as shown in algorithm 1 reported in Appendix 1.9.

1.6 Empirical application

The data is obtained from the Oxford-Man Institute's realised library. It consists of 5 minute Realized Volatility and daily returns of four U.S. stock market indices: S&P 500, Dow Jones Industrial Average, Russell 2000 and Nasdaq 100.
The 5 minute Realized Volatility is calculated by the sum of squared 5-minute returns within a day;

$$RV^{(5)} = \sum_{i=1}^{m} r^2_{t-1+5i,t-1+5(i-1)} \tag{45}$$

where

$$r^2_{t-1+5i,t-1+5(i-1)} = p_{t-1+5i} - p_{t-1+5(i-1)} \tag{46}$$

is the five-minute return computed as the log-difference of two consecutive index values over five-minutes, multiplied by 100; p_t denotes the logarithm of the index value; m is the number of five-minute returns per day calculated as $m = 390/5$ with 390 being the total minutes in a day.
The choice of 5-minute data is justified on the grounds of past empirical findings that show that at this frequency there is no evidence of micro-structure noise (Andreou *et al.*, 2012). Moreover, as shown in Liu *et al.* (2015), no other more sophisticated measures outperform 5-minute RV in terms of estimation accuracy.
The sample covers the period from January 1, 2012 to February 4, 2016. The plot of the RV series and the log-RV series are reported, respectively, in Figure 1 and 2 while in Table 1 some descriptive statistics of the log-RV series are shown.
In particular it is evident a strong serial correlation in accordance with the significant Ljung-Box statistics for all the considered log-RV series. In performing the test, the statistics have been based on 'lag' autocorrelation coefficients equal to the logarithm of the series length

which provides better power performance (Tsay, 2010).
In order to investigate the constancy of the regression coefficients in all the considered models (HAR-RV, LHAR-RV and AHAR-RV), an analysis based on the recursive estimates test has been employed for all the four series. The results are reported in Table 2. It is evident the rejection of the null hypothesis of constancy of the regression parameters for all the models and for all the series.
Moreover, in Figure 3, 4 and 5 the fluctuation process, defined in equation (18), is reported for each model specification and for each series, along with the boundaries obtained by its limiting process at level $\alpha = 0.05$. The paths of the empirical fluctuation process confirm the parameters instability: the boundaries are crossed for all the series and for all the models. The above analysis seems to confirm the effectiveness of parameter instability in all the three RV-models for all the considered indices. Moreover, it suggests the use of forecasting methods which take into account the presence of structural breaks.

1.6.1 Forecasting evaluation: out-of-sample analysis

In order to evaluate and to compare the forecasting performances of the proposed forecast combinations for each class of model, an initial sub-sample, composed by the data from $t = 1$ to $t = R$, is used to estimate the model and the 1-step ahead out-of-sample forecast is produced. The sample is then increased by one, the model is re-estimated using data from $t = 1$ to $t = R + 1$ and 1-step ahead forecast is produced. The procedure continues until the end of the available out-of-sample period. In the following examples R has been fixed such that the number of out of sample observations is 300. To generate the 1-step ahead forecasts the five competing forecast combinations, defined in section 1.4, have been considered together with a a benchmark method.
As a natural benchmark we refer to the Expanding window method, that ignores the presence of structural breaks. In fact it uses all the available observations. As pointed out in Pesaran and Timmermann (2007), this choice is optimal in situations with no breaks and it is appropriate for forecasting when the data is generated by a stable model. For each class of models, this method produces out-of-sample forecasts using a recursive expanding estimation window.
Common to all the considered forecast combinations is the specification of the minimum

acceptable estimation window size. As highlighted in section 1.4 it should be at least three times the number of unknown parameters; for simplicity this parameter has been held fixed at 40 for all the RV-model specifications.
Moreover, for the MSFE weighted average combination, the length of the evaluation window has been held fixed at 100. This value allows a good estimation of the MSFE and, at the same time, a not excessive loss of data at the end of the sample where a change point could be very influential for the forecasting.
In order to evaluate the quality of the volatility forecasts the MSE and the QLIKE loss functions have been considered. Their are defined as:

$$\mathrm{MSE}_t = (y_t - \widehat{y}_t)^2 \tag{47}$$

$$\mathrm{QLIKE}_t = \frac{y_t}{\widehat{y}_t} - \log\left(\frac{y_t}{\widehat{y}_t}\right) - 1 \tag{48}$$

where y_t is the actual value of the 5-minute RV at time t and $\widehat{y}_t$ is the corresponding RV forecast. They are the most widely used loss functions and they provide robust ranking of the models in the context of volatility forecasts (Patton, 2011).
The QLIKE loss is a a simple modification of the Gaussian log-likelihood in such a way that the minimum distance of zero is obtained when $y_t = \widehat{y}_{t+1}$. Moreover, according to Patton (2011), it is able to better discriminate among models and it is less affected by the most extreme observations in the sample .
These loss functions have been used to rank the six competing forecasting methods for each RV-model specification. To this aim, for every methods, the average loss (39) and the ratio between its value and the average loss of the benchmark method have been computed. Obviously, a value of the ratio below unit indicates that the forecasting method “beats” the benchmark according to the loss function metric.
Moreover, to assess statistically if the differences in the performances are relevant, the model confidence set procedure, introduced in the previous section, has been implemented in R-code using the package MCS (Catania and Bernardi, 2015) with $\alpha = 0.10$.
In the following the results obtained for the three different model specifications for the Realized Volatility of the four considered series are reported and discussed.
Table 3, 4 and 5 provide the results of the analysis for the HAR-RV, LHAR-RV and AHAR-RV model specification respectively for the four considered series. In particular they report

the average MSE and the average QLIKE for each forecasting method, as well as their ratio for an individual forecasting method to the benchmark expanding window method and the ranking of the considered method with respect to each loss function. Moreover it is also reported the value of the test statistic T_{MAX} of the model confidence set approach and the associate p-value.

1.6.2 Results for the HAR model

For all the considered series and for both the loss functions there is a significant evidence that all the forecast combinations have better forecasting ability with respect the expanding window procedure; the ratio values are all less than one. Moreover, for both the loss functions, the best method is the forecast combination with ROC location weights for S&P500 and Dow Jones Industrial Average and with location weights for Russell 2000 and Nasdaq 100; this result confirms the importance of placing heavier weights of the forecast based on more recent data.

The model confidence set has the same structure for all the four series and for both the loss functions; it excludes only the forecast generated by the expanding window procedure and includes all the forecast combinations.

1.6.3 Results for the LHAR model

Focusing on the results of S&P500 and Dow Jones Industrial Average, which have very similar behaviour, the forecast combination with MSFE weights offers the best improvement in forecasting accuracy according to the MSE metric while the forecast combination with ROC weights according to the QLIKE metric. For Russell 2000 the forecast combination with ROC location weights beats all the competing models according to MSE loss function while, under the QLIKE, the best method is the forecast combination with location weights. This last method has better performance with respect to both of the loss function metrics for Nasdaq 100.

By looking at the MSE ratios for S&P500 and Dow Jones Industrial Average Index it is evident that the forecast combination with location weights is unable to beat the expanding window procedure; for all the others, the combinations are able to consistently outperform it. In the model confidence set, when the MSE loss function is used, the expanding window is eliminated from the model confidence set for all the series. However it is also eliminated the forecast combination with location weights for S&P500 and Dow Jones Industrial Average and that with MSFE weights for Russell 2000. A quite different situation arises when the loss function QLIKE is considered. In this case the only surviving models in the model confidence set are the two combinations based on ROC statistics for S&P500 and Dow Jones Industrial Average and those based on ROC location weights and on location weights for Russell 2000. For Nasdaq 100 all the combinations have the same forecasting accuracy. Excluding the last case, the QLIKE loss function, as pointed out previously, seems to better discriminate among forecasting methods.

1.6.4 Results for the AHAR model

In line with the previous results, the expanding window appears to offer the worst forecasting performance overall.
Moreover, for both MSE and QLIKE loss functions, the method which offers the major improvement in forecasting accuracy is the forecast combination with ROC location weights for S&P500 and Dow Jones Industrial Average Index and the the forecast combination with location weights for Russell 2000 and Nasdaq 100.
Focusing on the model confidence set, for the MSE loss function the expanding window is always eliminated for all the series together with the the forecast combination with MSFE weights for Russell 2000. With respect to the QLIKE loss function, for Dow Jones Industrial Average Index and Nasdaq 100 the only excluded method is the expanding window procedure. For the other series, the results confirm the better discriminate property of the QLIKE metric; the only surviving methods are forecast combination with ROC location weights and with location weights.

In conclusion, even if it is not clear which combination has the best forecasting performance, the forecast combination with ROC location weights and that with location weights seems to be always among the best methods. However the forecast combination with ROC location weights always outperform the expanding window method, it is always in the top position with respect the loss function ratio and it is never excluded by the model confidence set.

1.7 Concluding remarks

This paper has explored the relevance of taking into account the presence of structural breaks in forecasting Realized Volatility. The analysis has been based on 5-minute Realized Volatility of four U.S. stock market indices: S&P 500, Dow Jones Industrial Average, Russell 2000 and Nasdaq 100. Three different model specifications of the log-Realized Volatility have been considered. For all the considered market indices, the instability in the parameters of the RV-models has been verified through the recursive estimates test. In order to handle this problem, five forecast combinations, based on different estimation windows with alternative weighting schemes, have been introduced and compared with the expanding window method, a natural choice when the data are generated by a stable model. The forecasting performances have been evaluated, for each RV-model specification, through two of the most relevant loss functions, the MSE and the QLIKE. Moreover, to this aim the average loss function has been calculate and, to assess statistically if the differences in the performances are relevant, the model confidence set approach has been considered.

The analysis, repeated for each class of RV-models separately, has highlighted the importance of taking into account structural breaks; in fact the expanding window appears to offer the worst forecasting performances overall. In particular in almost all the considered cases the two combinations which make adjustments for accounting the possible most recent break point (the forecast combination with location weights and with ROC location weights) are placed in first position and, as a consequence, have better forecasting performances. Nevertheless the forecast combination with ROC location weights always outperforms the expanding window method, it is always in the top position with respect the loss function ratio and it is never excluded by the model confidence set.

1.8 Figures and Tables

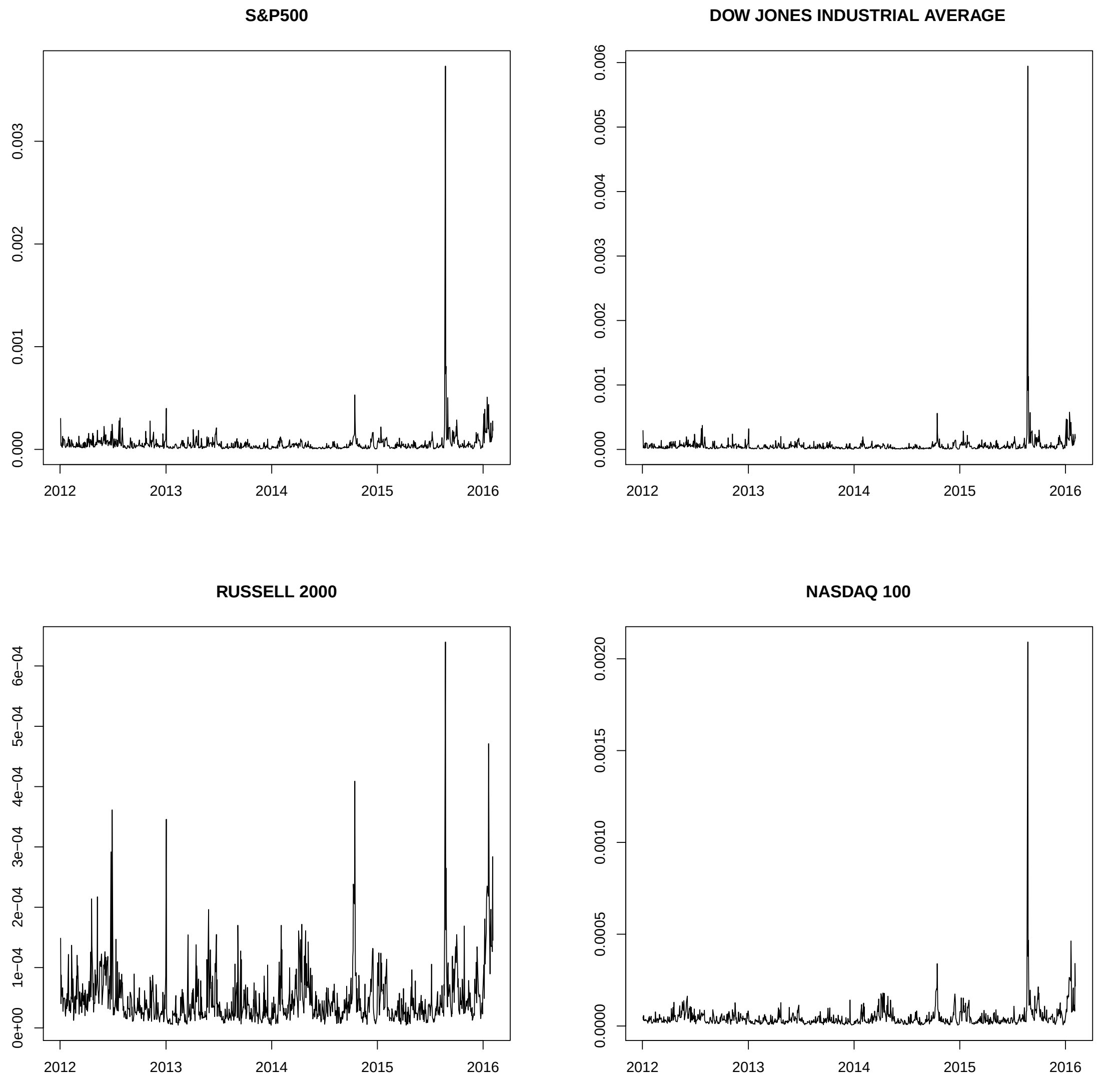


Figure 1: RV for S&P500, Dow Jones Industrial Average, Russell 2000 and Nasdaq 100.

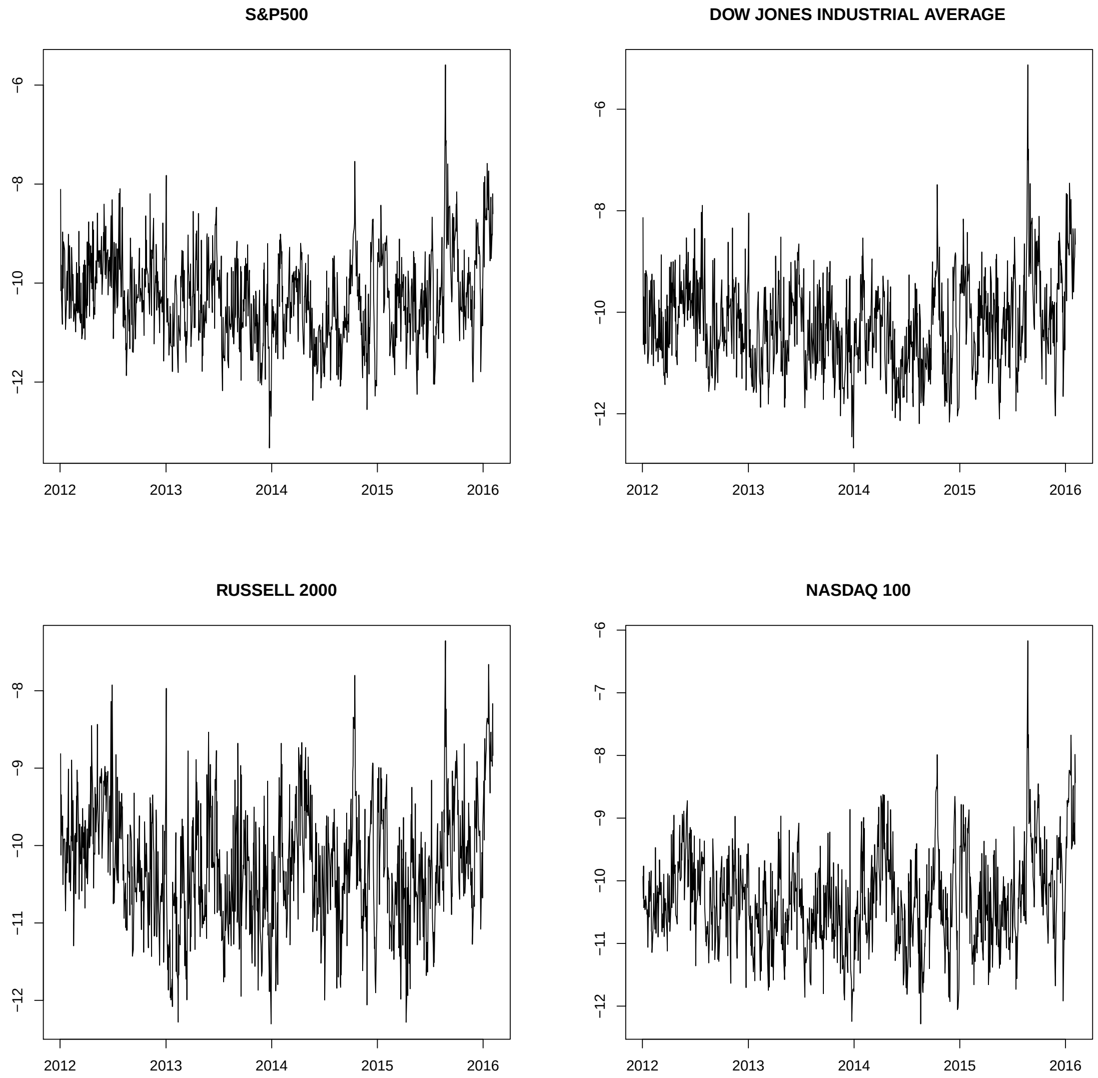


Figure 2: Log-RV for S&P500, Dow Jones Industrial Average, Russell 2000 and Nasdaq 100.

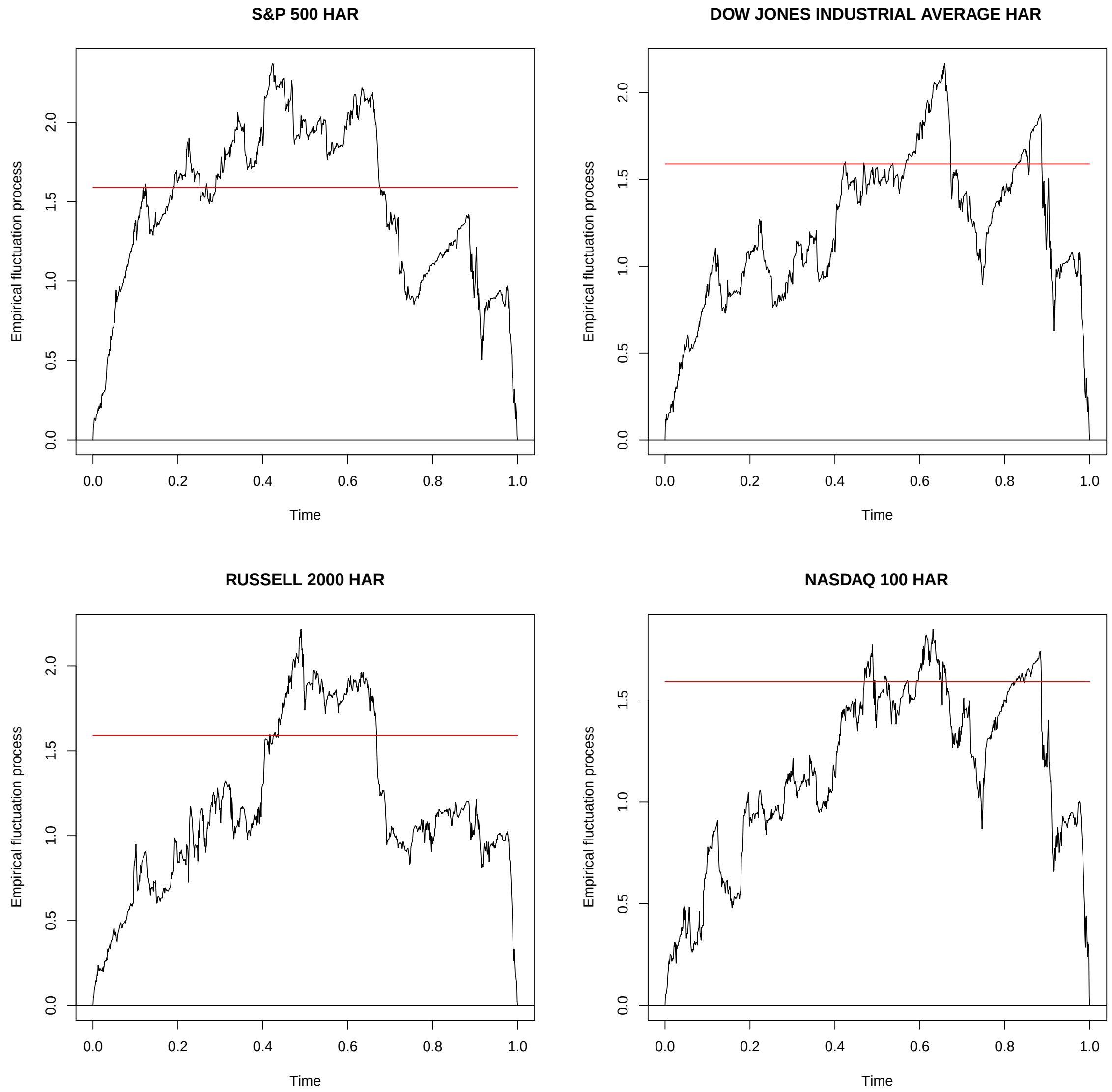


Figure 3: Empirical fluctuation process of the HAR-RV model parameters for S&P500, Dow Jones Industrial Average, Russell 2000 and Nasdaq 100. The red line refers to the boundary with $\alpha = 0.05$.

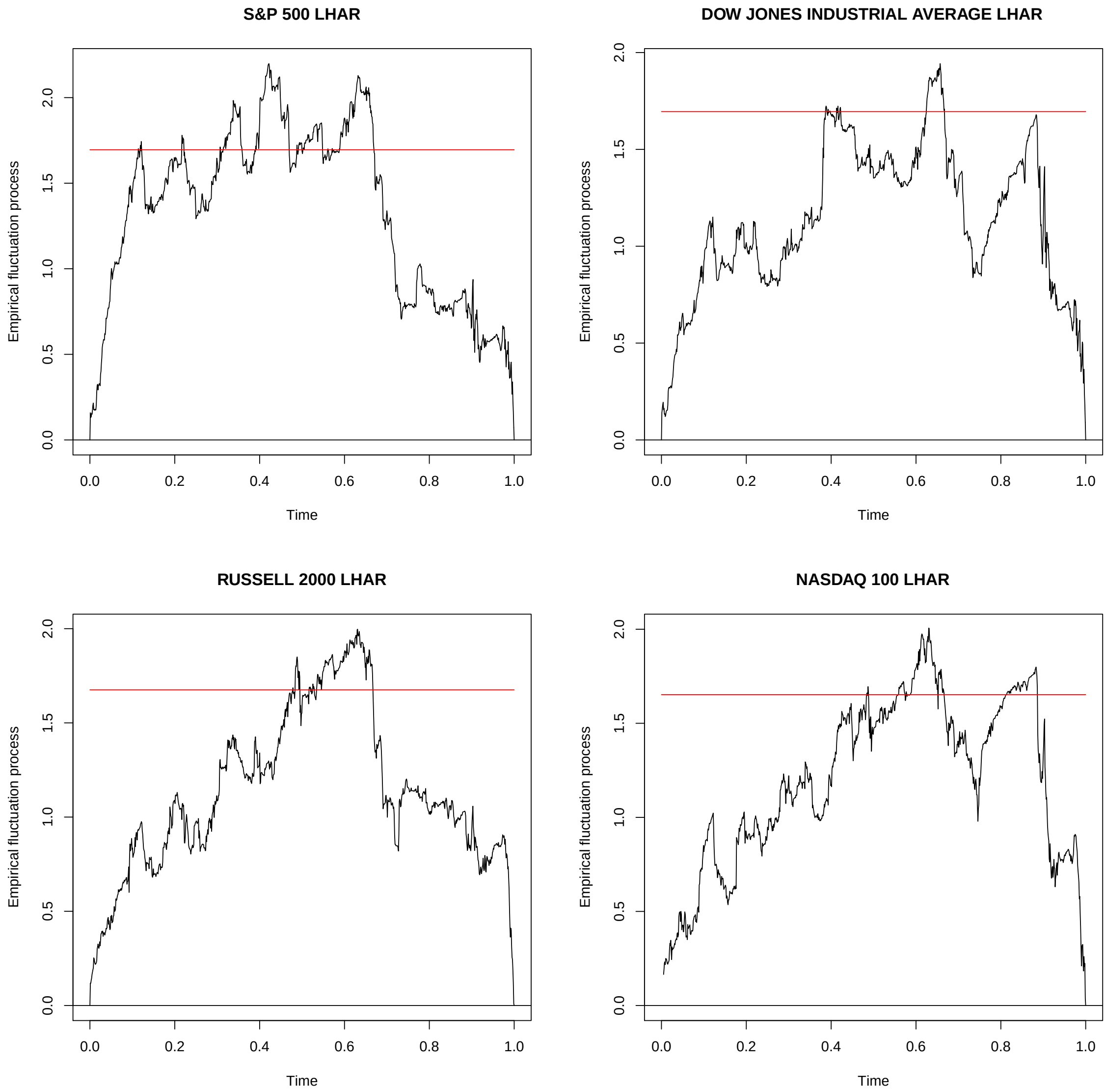


Figure 4: Empirical fluctuation process of the LHAR-RV model parameters for S&P500, Dow Jones Industrial Average, Russell 2000 and Nasdaq 100. The red line refers to the boundary with $\alpha = 0.05$.

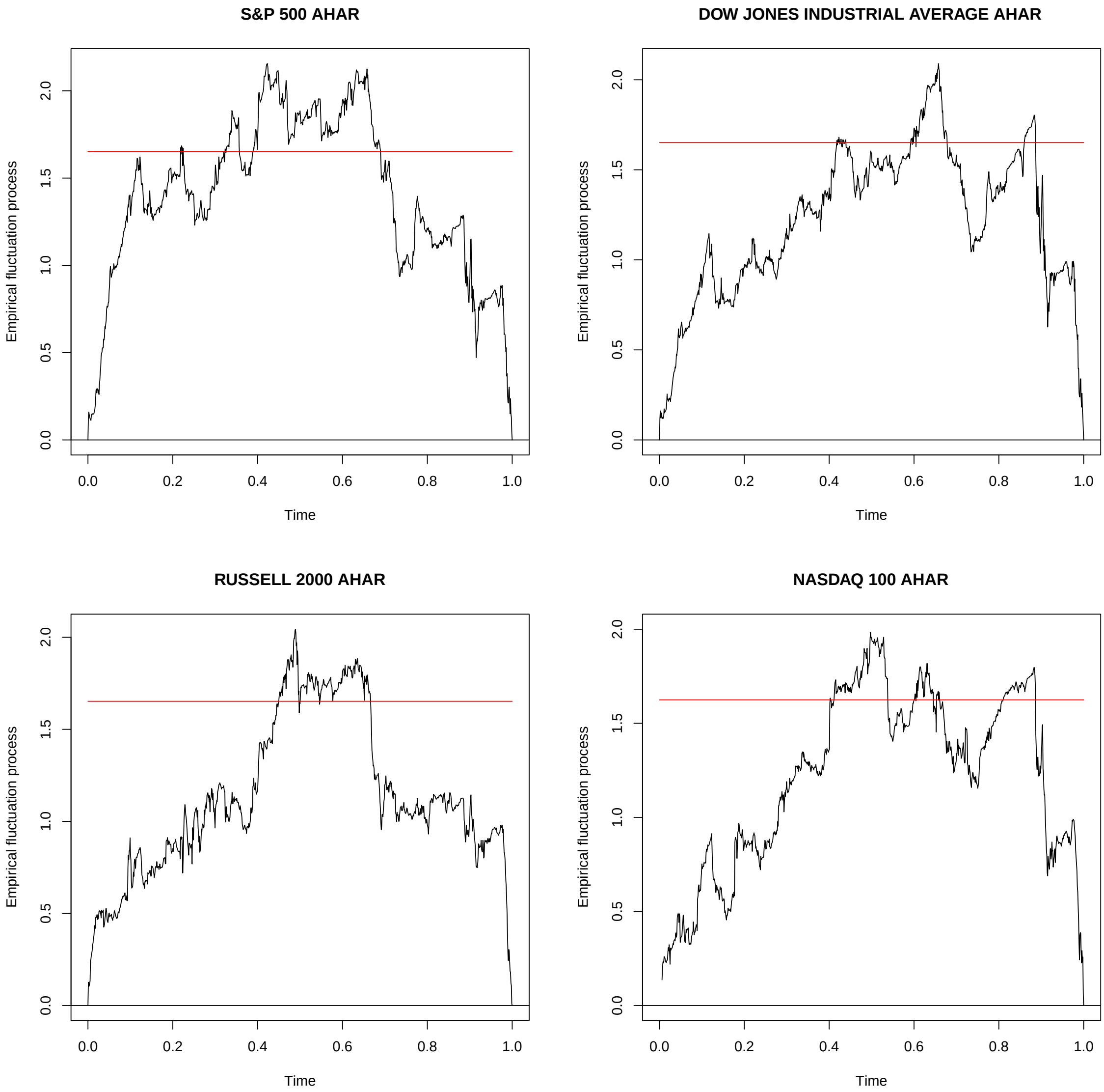


Figure 5: Empirical fluctuation process of the AHAR-RV model parameters for S&P500, Dow Jones Industrial Average, Russell 2000 and Nasdaq 100. The red line refers to the boundary with $\alpha = 0.05$.

Table 1: Summary statistics of log-RV for S&P500, Dow Jones Industrial Average, Russell 2000 and Nasdaq 100.

	S&P 500	**DJIA**	**RUSSELL 2000**	**NASDAQ 100**
Mean	-10.260	-10.230	-10.280	-10.270
Minimum	-13.330	-12.670	-12.300	-12.280
1st. Quart	-10.880	-10.850	-10.810	-10.780
Median	-10.310	-10.280	-10.330	-10.330
3rd. Quart	-9.644	-9.624	-9.746	-9.811
Maximum	-5.591	-5.125	-7.354	-6.170
Std.Dev.	0.897	0.894	0.799	0.757
Skewness	0.346	0.498	0.272	0.453
Kurtosis	0.652	0.941	0.038	0.825
LjungBox.stat	1399.597	1188.962	1111.333	1506.633
LjungBox.pval	0.000	0.000	0.000	0.000

Table 2: Recursive Estimates Test

S&P 500			**DJIA**		
Model	statistic	p-value	Model	statistic	p-value
HAR-RV	2.369	0.0001	HAR-RV	2.167	0.0006
LHAR-RV	2.198	0.0010	LHAR-RV	1.943	0.0084
AHAR-RV	2.156	0.0011	AHAR-RV	2.089	0.0019
RUSSELL 2000			**NASDAQ 100**		
Model	statistic	p-value	Model	statistic	p-value
HAR-RV	2.216	0.0004	HAR-RV	1.848	0.0086
LHAR-RV	1.997	0.0048	LHAR-RV	1.659	0.0104
AHAR-RV	2.044	0.0028	AHAR-RV	1.782	0.0093

Table 3: Out of sample forecasting result for the HAR-RV model

S&P 500

	MSE	Ratio	rk	MCS	**QLIKE**	Ratio	rk	MCS
Expanding Window	0.5166	1.0000	6	2.154 (0.00)	0.4082	1.0000	6	2.035 (0.00)
MSFE weights	0.5016	0.9710	5	1.010 (1.00)	0.3879	0.9500	4	0.610 (1.00)
ROC weights	0.4986	0.9653	2	-0.956 (1.00)	0.3825	0.9370	2	-0.753 (1.00)
ROC location weigh.	0.4980	**0.9639**	1	-1.410 (1.00)	0.3794	**0.9294**	1	-1.534 (1.00)
Equal Weights	0.5015	0.9708	4	0.917 (1.00)	0.3901	0.9557	5	1.191 (0.99)
Location weights	0.5007	0.9694	3	0.437 (1.00)	0.3874	0.9489	3	0.492 (1.00)

DJIA

	MSE	Ratio	rk	MCS	**QLIKE**	Ratio	rk	MCS
Expanding Window	0.6182	1.0000	6	2.112 (0.00)	0.5610	1.0000	6	2.007 (0.00)
MSFE weights	0.6088	0.9849	5	1.344 (0.97)	0.5410	0.9643	4	0.441 (1.00)
ROC weights	0.6066	0.9813	3	-0.213 (1.00)	0.5388	0.9603	2	-0.102 (1.00)
ROC location weigh.	0.6046	**0.9781**	1	-1.599 (1.00)	0.5319	**0.9480**	1	-1.789 (1.00)
Equal Weights	0.6079	0.9834	4	0.690 (1.00)	0.5441	0.9699	5	1.208 (1.00)
Location weights	0.6066	0.9813	2	-0.220 (1.00)	0.5402	0.9629	3	0.252 (1.00)

RUSSELL 2000

	MSE	Ratio	rk	MCS	**QLIKE**	Ratio	rk	MCS
Expanding Window	0.3491	1.0000	6	1.904 (0.00)	0.2066	1.0000	6	1.649 (0.00)
MSFE weights	0.3450	0.9881	5	1.235 (0,98)	0.2039	0.9868	4	1.018 (1.00)
ROC weights	0.3445	0.9868	4	0.925 (1.00)	0.2040	0.9872	5	1.071 (0.65)
ROC location weigh.	0.3424	0.9808	2	-0.548 (1.00)	0.2011	0.9733	2	-0.702 (1.00)
Equal Weights	0.3430	0.9826	3	-0.106 (1.00)	0.2025	0.9798	3	0.123 (1.00)
Location weights	0.3410	**0.9768**	1	-1.500 (1.00)	0.1998	**0.9670**	1	-1.507 (1.00)

NASDAQ 100

	MSE	Ratio	rk	MCS	**QLIKE**	Ratio	rk	MCS
Expanding Window	0.3509	1.0000	6	2.061 (0.00)	0.2397	1.0000	6	1.897 (0.00)
MSFE weights	0.3448	0.9825	4	1.007 (1.00)	0.2307	0.9697	5	1.106 (1.00)
ROC weights	0.3449	0.9829	5	1.130 (1.00)	0.2319	0.9676	4	0.907 (1.00)
ROC location weigh.	0.3433	0.9784	3	-0.229 (1.00)	0.2285	0.9535	2	-0.484 (1.00)
Equal Weights	0.3432	0.9781	2	-0.319 (1.00)	0.2300	0.9595	3	0.108 (1.00)
Location weights	0.3417	**0.9738**	1	-1.578(1.00)	0.2257	**0.9417**	1	-1.631 (1.00)

Note: For both the loss functions (MSE and QLIKE) the entries are: the values of the average loss; the ratio of the average loss to that of the expanding window method; the rank (rk) according to the average loss function; the T_{MAX} statistic and the p-value of the Model Confidence Set (MCS) procedure with $\alpha = 0.10$. A bold entry denotes the value of the smallest average loss

Table 4: Out of sample forecasting result for the LHAR-RV model.

S&P 500

	MSE	Ratio	rk	MCS	**QLIKE**	Ratio	rk	MCS
Expanding Window	0.4247	1.0000	5	1.776 (0.00)	0.2858	1.0000	6	2.143 (0.00)
MSFE weights	0.4176	**0.9834**	1	-1.258 (1.00)	0.2666	0.9328	3	1.316 (0.01)
ROC weights	0.4183	0.9851	2	-0.617 (1.00)	0.2657	**0.9298**	1	-0.632 (1.00)
ROC location weigh.	0.4204	0.9899	4	1.275 (0.13)	0.2658	0.9301	2	0.618 (0.55)
Equal Weights	0.4197	0.9882	3	0.599 (1.00)	0.2678	0.9370	4	1.557 (0.00)
Location weights	0.4260	1.0031	6	1.536 (0.00)	0.2695	0.9430	5	1.673 (0.00)

DJIA

	MSE	Ratio	rk	MCS	**QLIKE**	Ratio	rk	MCS
Expanding Window	0.5291	1.0000	5	1.191 (0.09)	0.4236	1.0000	6	1.998 (0.00)
MSFE weights	0.5255	**0.9932**	1	-1.181 (1.00)	0.4012	0.9471	3	1.368 (0.00)
ROC weights	0.5260	0.9942	2	-0.844 (1.00)	0.3996	**0.9433**	1	-0.913 (1.00)
ROC location weigh.	0.5290	0.9998	4	1.092 (0.57)	0.3999	0.9440	2	0.912 (0.44)
Equal Weights	0.5268	0.9957	3	-0.311 (1.00)	0.4063	0.9592	4	1.633 (0.00)
Location weights	0.5325	1.0063	6	1.783 (0.00)	0.4091	0.9658	5	1.522 (0.00)

RUSSELL 2000

	MSE	Ratio	rk	MCS	**QLIKE**	Ratio	rk	MCS
Expanding Window	0.3096	1.0000	6	2.082 (0.00)	0.1828	1.0000	6	1.564 (0.00)
MSFE weights	0.3058	0.9877	5	1.783 (0.00)	0.1821	0.9964	5	1.464 (0.00)
ROC weights	0.3044	0.9833	4	0.777 (1.00)	0.1809	0.9895	4	1.334 (0.00)
ROC location weigh.	0.3035	**0.9805**	1	-1.676 (1.00)	0.1794	0.9812	2	-0.989 (1.00)
Equal Weights	0.3044	0.9832	3	0.686 (1.00)	0.1804	0.9867	3	1,268 (0.00)
Location weights	0.3042	0.9827	2	0.210 (1.00)	0.1789	**0.9786**	1	-0.993 (1.00)

NASDAQ 100

	MSE	Ratio	rk	MCS	**QLIKE**	Ratio	rk	MCS
Expanding Window	0.3180	1.0000	6	2.083 (0.00)	0.2070	1.0000	6	1.984 (0.00)
MSFE weights	0.3112	0.9787	4	1.058 (0.99)	0.1979	0.9562	5	1.066 (1.00)
ROC weights	0.3113	0.9788	5	1.086 (0.99)	0.1977	0.9551	4	0.970 (1.00)
ROC location weigh.	0.3099	0.9747	3	-0.034 (1.00)	0.1947	0.9406	2	-0.315 (1.00)
Equal Weights	0.3093	0.9726	2	-0.586 (1.00)	0.1953	0.9436	3	-0.047 (1.00)
Location weights	0.3082	**0.9691**	1	-1.516 (1.00)	0.1915	**0.9252**	1	-1.665 (1.00)

Note: For both the loss functions (MSE and QLIKE) the entries are: the values of the average loss; the ratio of the average loss to that of the expanding window method; the rank (rk) according to the average loss function; the T_{MAX} statistic and the p-value of the Model Confidence Set (MCS) procedure with $\alpha = 0.10$. A bold entry denotes the value of the smallest average loss

Table 5: Out of sample forecasting result for the AHAR-RV model.

S&P 500

	MSE	Ratio	rk	MCS	**QLIKE**	Ratio	rk	MCS
Expanding Window	0.4576	1.0000	6	2.186 (0.00)	0.3315	1.0000	6	2.151 (0.00)
MSFE weights	0.4397	0.9608	5	1.230 (0.94)	0.3072	0.9267	5	1.268 (0.00)
ROC weights	0.4371	0.9552	2	-0.860 (1.00)	0.3043	0.9179	3	1.393 (0.00)
ROC location weigh.	0.4364	**0.9537**	1	-1.416 (1.00)	0.3012	**0.9085**	1	-0.973 (1.00)
Equal Weights	0.4391	0.9596	4	0.781 (1.00)	0.3063	0.9240	4	1.413 (0.00)
Location weights	0.4385	0.9582	3	0.252 (1.00)	0.3016	0.9097	2	-0.977 (0.38)

DJIA

	MSE	Ratio	rk	MCS	**QLIKE**	Ratio	rk	MCS
Expanding Window	0.5669	1.0000	6	2.189 (0.00)	0.4718	1.0000	6	2.091 (0.00)
MSFE weights	0.5534	0.9762	5	1.015 (0.99)	0.4418	0.9365	5	0.986 (1.00)
ROC weights	0.5529	0.9754	3	0.291 (1.00)	0.4404	0.9335	3	0.688 (1.00)
ROC location weigh.	0.5516	**0.9731**	1	-1.744 (1.00)	0.4308	**0.9132**	1	-1.337 (1.00)
Equal Weights	0.5532	0.9758	4	0.679 (0.99)	0.4407	0.9341	4	0.749 (1.00)
Location weights	0.5526	0.9748	2	-0.219 (1.00)	0.4320	0.9158	2	-1.083 (1.00)

RUSSELL 2000

	MSE	Ratio	rk	MCS	**QLIKE**	Ratio	rk	MCS
Expanding Window	0.3193	1.0000	6	1.964 (0.00)	0.1898	1.0000	6	1.706 (0.00)
MSFE weights	0.3132	0.9811	5	1.498 (0.00)	0.1868	0.9842	5	1.348 (0.00)
ROC weights	0.3113	0.9750	4	1.183 (0.19)	0.1854	0.9769	4	1.385 (0.00)
ROC location weigh.	0.3086	0.9665	2	-0.735 (1.00)	0.1824	0.9610	2	0.995 (0.20)
Equal Weights	0.3107	0.9731	4	0.760 (1.00)	0.1849	0.9741	3	1.385 (0.00)
Location weights	0.3079	**0.9644**	1	-1.204 (1.00)	0.1818	**0.9579**	1	-0.995 (1.00)

NASDAQ 100

	MSE	Ratio	rk	MCS	**QLIKE**	Ratio	rk	MCS
Expanding Window	0.3299	1.0000	6	2.078 (0.00)	0.2223	1.0000	6	1.992 (0.00)
MSFE weights	0.3207	0.9719	4	0.825 (1.00)	0.2122	0.9545	4	0.936 (1.00)
ROC weights	0.3213	0.9738	5	1.212 (1.00)	0.2125	0.9560	5	1.073 (1.00)
ROC location weigh.	0.3192	0.9675	3	-0.092 (1.00)	0.2088	0.9394	2	-0.429 (1.00)
Equal Weights	0.3189	0.9665	2	-0.298 (1.00)	0.2101	0.9451	3	0.085 (1.00)
Location weights	0.3167	**0.9599**	1	-1.643 (1.00)	0.2058	**0.9257**	1	-1.651 (1.00)

Note: For both the loss functions (MSE and QLIKE) the entries are: the values of the average loss; the ratio of the average loss to that of the expanding window method; the rank (rk) according to the average loss function; the T_{MAX} statistic and the p-value of the Model Confidence Set (MCS) procedure with $\alpha = 0.10$. A bold entry denotes the value of the smallest average loss

1.9 Appendix. Block bootstrap procedure for Model Confidence Set

Algorithm 1 Block bootstrap procedure for model confidence set

Choose the block-length bootstrap parameter l as the max number of significant parameters obtained by fitting an AR(p) model on all the d_{ij} terms.

Let $g = [T/l]$ where $[x]$ denote the smaller integer greater or equal to x.

Generate B bootstrap resamples of $\{1, \dots, T\}$

- Let $(\psi_1^b, \dots, \psi_g^b)$ *i.i.d.* uniform on $\{1, \dots, T\}$ for all $b = \{1, \dots, B\}$
- Define $\tau_k^b = (\psi_k^b, \psi_k^b + 1, \dots, \psi_k^b + l - 1)$ for all $k = \{1, \dots, g\}$. For convention $T + i = T$ for $i \geq 1$ and, if T is not a multiple of l, only $T + l - gl$ observations from the last block are used.
- The b bootstrap rasample is $\tau^b = (\tau_1^b, \dots, \tau_g^b)$.

Compute $L^*_{b,i,t} = L_{i,\tau^b,t}$ and $\bar{L}^*_{b,i} = T^{-1} \sum_{t=1}^{T} L^*_{b,i,t}$, the corresponding bootstrap of $L_{i,t}$ and $\bar{L}_i$ respectively, for each model i=(1, …,m) and for each t=(1, …,T)

Calculate the variables $\xi^*_{b,i} = \bar{L}^*_{b,i} - \bar{L}_i$

Initialize by setting $M = M_0$

repeat

  Let $\xi^*_{b,.} = m^{-1} \sum_{i=1}^{m} \xi^*_{b,i}$

  Determine $t^*_{b,i} = (\xi^*_{b,i} - \xi^*_{b,.}) / \sqrt{var^*(\bar{d}_{i,.})}$
  where $var^*(\bar{d}_{i,.}) = B^{-1} \sum_{b=1}^{B} (\xi^*_{b,i} - \xi^*_{b,.})^2$ is the bootstrap variance.

  Define $T^*_{b,max} = \max_i t^*_{b,i}$ for $b = \{1, \dots, B\}$.

  Calculate the p-value of $H_{0,M}$ given by:
  $P_{H_{0,M}} = B^{-1} \sum_{b=1}^{B} I_{(T_{max} > T^*_{b,max})}$
  where I is the indicator function

  if $P_{H_{0,M}} < \alpha$ **then**

    $H_{0,M}$ is rejected and the model with $e_M = \arg\max_i t_i$ is eliminated from M

  end if

until $P_{H_{0,M}} \geq \alpha$ for the first time.

The model confidence set is the set $\widehat{M}^*_{1-\alpha}$

2 A Bootstrap Bias Correction of long run fourth order moment estimation in the CUSUM of squares test

Abstract

The aim of this paper is to propose a bias correction of the estimation of the long run fourth order moment in the CUSUM of squares test proposed by Sansó *et al.* (2004) for the detection of structural breaks in financial data. The correction is made by using the stationary bootstrap proposed by Politis and Romano (1994). The choice of this resampling technique is justified by the stationarity and weak dependence of the time series under the assumptions which ensure the existence of the limiting distribution of the test statistic, under the null hypothesis.

Some Monte Carlo experiments have been implemented in order to evaluate the effect of the proposed bias correction considering two particular data generating processes, the GARCH(1,1) and the log-normal stochastic volatility. The effectiveness of the bias correction has been evaluated also on real data sets.

Keywords: CUSUM of squares test, Structural breaks, Bias correction.

JEL Code: C120, C580, G170.

2.1 Introduction

Many studies (see for example Andreou and Ghysels, 2009 and the references therein) agree on the possible existence of structural breaks in the variance of financial time series and on the importance of their detection.

The literature on testing for structural breaks is extensive and a lot of procedures and tests have been proposed (for a review see Perron, 2006). An important role in analysing variance constancy, both in methodological advances and in applications, is played by statistical methods based on the CUSUM of squares test proposed by Brown *et al.* (1975).

The widespread use of these tests is mainly due to their simplicity and to the lack of the need to specify a particular pattern of variation. These tests are non parametric and, as a consequence, they ensure asymptotic invariance against a fairly general class of generating processes. Moreover they do not require any previous knowledge of the timing and the number of the shifts.

These properties make these tests very popular in many empirical applications especially in a financial context, where different conditional heteroskedasticity patterns can be assumed.

The original procedure for detecting breaks in the variance has been developed, for independent processes, by Inclan and Tiao (1994). More recently it has been extended to some types of dependent processes; in particular Kokoszka and Leipus (2000) have considered the case of β-mixing processes assuming an ARCH (∞) process.

In a more general framework, starting from a different set of assumptions, Sansó *et al.* (2004) have proposed a test which is similar to Kokoszka and Leipus test. It makes adjustments to the original proposal that allow the time series to obey a wide class of dependent processes, including GARCH and log-normal stochastic volatility processes, under the null.

This test has been widely used in the econometric literature; recently, it has been employed to identify structural breaks in US GDP and oil price growth rates (Gadea *et al.*, 2016) , in daily returns of U.S. and BRICS stock markets (Mensi *et al.*, 2016) and in stock return data on two Chinese stock return market indexes (Ni *et al.*, 2016).

Optimal properties of the test statistic are obtained by means of a consistent estimator of the long-run fourth order moment of the considered series, which provides the right standardization to obtain a statistic free of nuisance parameters. A popular strategy, used in the econometric literature and in the original paper, is based on a non parametric kernel method often referred to as the kernel Heteroskedasticity Autocorrelation Consistent (HAC) estimator.

When using this estimation in the context of long run variance estimation, it is quite common to consider the Bartlett Kernel. However the kernel function depends on an additional parameter, the bandwidth, which controls the degree of complexity of the estimation and, consequently, plays a fundamental role in order to obtain accurate properties of the resultant HAC estimator.

In the econometric literature it is quite common the use of data dependent rules in which this parameter is fixed as a function of the correlation structure of the series. In this context the proposal by Newey and West (1994) is a classical choice. Although other methods, which lead to inconsistent estimation, have been proposed (Kiefer and Vogelsang, 2002; 2005), the Newey and West procedure remains the most used and it will be considered in this paper. Moreover, when using this approach, the test is asymptotically correct in presence of short-run dynamics or conditional heteroskedasticity. On the contrary, in the case of strong dependence and in relative small samples, the HAC estimation of the long run fourth order moment could be inefficient and, as a consequence, significant biases may arise leading to significant size distortion of the test (Rodriguez and Rubia, 2007).

The aim of this paper is to propose a bias correction of the estimation of the long run fourth order moment in the CUSUM of squares test proposed by Sansó *et al.* (2004). The correction is made by using a resampling technique, namely the stationary bootstrap, proposed by Politis and Romano (1994). This choice is justified by the stationarity and weakly dependence of the time series which ensure the existence of the limiting distribution of the test statistic under the null hypothesis. Following Politis and White (2004), the optimal expected block size for the stationary bootstrap has to be fixed to $c * T^{1/3}$ where T is the time series length and c is a constant that has been fixed in a suitable way depending on the specific properties of the data generation process.

The structure of the paper is as follows. Section 2 provides an overview of testing procedures for single and multiple breaks in the unconditional variance. Starting from the first proposal by Inclan and Tiao (1994), in the case of *i.i.d.* Normal data, a correction to the cumulative sum of squares test, proposed by Sansó *et al.* (2004) is discussed and some theoretical results on its limiting distribution are reported. In section 3 the problem of the estimation of the long run fourth order moment is addressed reviewing the most popular approaches and focusing on the data driven rule. Section 4 introduces the bootstrap approach in the context of dependent data and discusses the bootstrap bias correction of the long run fourth order moment. In particular the stationary bootstrap is illustrated and a procedure to select the optimal expected block size is proposed. In section 5 Monte Carlo experiments are implemented in order to evaluate the effect of the bias correction on the considered CUSUM of squares test. The attention is focused on two particular data generating processes, the GARCH (1,1) and the log-normal stochastic volatility which are the most used models in financial applications. In section 6 the effectiveness of the bias correction is evaluated on a real data set consisting of two financial time series; the IPC Index - Mexico and CNX Nifty Index - India. Some final remarks close the paper.

2.2 Tests for structural breaks

Let $\{a_t\}, t = 1, \cdots, T$ denotes the time series of interest and assume, for simplicity, that the unconditional and the conditional mean is 0. The problem of testing the hypothesis that the unconditional variance of a_t is constant could be formulated as:

$$H_0 : \sigma_t^2 = \sigma_a^2 \quad \text{for all } t = 1, \cdots, T \tag{49}$$

against the alternative hypothesis of $m + 1$ regimes

$$H_A : \sigma_t^2 = \begin{cases} \sigma_1^2 & \text{for } t = 1, \cdots, \tau_1 \\ \sigma_2^2 & \text{for } t = \tau_1 + 1, \cdots, \tau_2 \\ \vdots & \quad\ \vdots \quad \vdots \\ \sigma_{m+1}^2 & \text{for } t = \tau_m + 1, \cdots, T \end{cases} \tag{50}$$

The number of breakpoints m and their locations $\tau_1, \cdots, \tau_m$, are assumed to be unknown. The problem will be analysed by considering at first the case of a single change-point and then the extension to the more general case of multiple change-points.

2.2.1 Testing for a single structural break

The most popular test for a change point is based on the IT statistic proposed by Inclan and Tiao (1994). It is based on cumulative sum of squares statistic and it defined as:

$$IT = \sup_k |(T/2)^{1/2} D_k| \tag{51}$$

where $D_k = (C_k/C_T) - (k/T)$ and $C_k = \sum_{t=1}^{k} a_t^2$ for $k = 1, \ldots, T$. The value of k corresponding to the maximum of $|(T/2)^{1/2} D_k|$ is the estimate of the break date.

Proposition 1 (Inclan and Tiao, 1994) *Under the hypothesis that $\{a_t\}$ are i.i.d. with distribution $N(0, \sigma_a^2)$, the asymptotic distribution of the IT statistic is given by:*

$$IT \overset{A}{\to} \sup_r |W^*(r)| \tag{52}$$

where $W^*(r) \equiv W(r) - rW(1)$ is a Brownian Bridge and $W(r)$ is a standard Brownian motion[5].

Proof The proof follows most of the steps used by Inclan and Tiao (1994).
Let $a_1, a_2, \ldots$ be a sequence of independent, identically distributed Normal$(0, \sigma_a^2)$ and $\xi_t = a_t^2 - \sigma_a^2$ with $\mathrm{E}[\xi_t] = 0$ and Var$(\xi_t) = 2\sigma_a^4$. Let:

$$X_T(r) = \frac{1}{\sqrt{T}\sqrt{2\sigma_a^4}} S_{\lfloor rT \rfloor} + (rT - \lfloor rT \rfloor) \frac{1}{\sqrt{T}\sqrt{2\sigma_a^4}} \xi_{\lfloor rT \rfloor + 1} \tag{53}$$

where $S_T = \xi_1 + \ldots \xi_T$, $r \in [0, 1]$ and $\lfloor . \rfloor$ is the floor function.
By Donsker's theorem (Billingsley 1999, thm. 8.2) $X_T(r) \overset{\mathcal{D}}{\to} W(r)$ and so $X_T(r) - rX_T(1) \overset{\mathcal{D}}{\to} W^*(r)$.
Let $rT = k$, $k = 1, \ldots T$. By simple algebra, it is:

$$\begin{aligned} X_T(r) - rX_T(1) &= \frac{1}{\sqrt{2}\sigma_a^2\sqrt{T}} \left[S_k - \frac{k}{T} S_T \right] + \frac{(rT - \lfloor rT \rfloor)}{\sqrt{2}\sigma_a^2\sqrt{T}} \xi_{\lfloor rT \rfloor + 1} \\ &= \frac{1}{\sqrt{2}\sigma_a^2\sqrt{T}} \left[\sum_{t=1}^{k} (a_t^2 - \sigma_a^2) - \frac{k}{T} \sum_{t=1}^{T} (a_t^2 - \sigma_a^2) \right] \frac{(rT - \lfloor rT \rfloor)}{\sqrt{2}\sigma_a^2\sqrt{T}} \xi_{\lfloor rT \rfloor + 1} \end{aligned}$$

[5] see appendix 2.9.1 for definition of Brownian Bridge and Brownian Motion

$$= \frac{1}{\sqrt{2}\sigma_a^2\sqrt{T}}\left[\sum_{t=1}^{T} a_t^2\right] D_k + \frac{(rT - \lfloor rT \rfloor)}{\sqrt{2}\sigma_a^2\sqrt{T}}\xi_{\lfloor rT \rfloor + 1}$$

It follows that:

$$\sqrt{\frac{T}{2}} D_k \left[\frac{\frac{1}{T}\sum_{i=1}^{T} a_t^2}{\sigma_a^2}\right] = X_T(r) - rX_T(1) - \frac{(rT - \lfloor rT \rfloor)}{\sqrt{2}\sigma_a^2\sqrt{T}}\xi_{\lfloor rT \rfloor + 1}$$

As $T \to \infty$

$$\frac{\frac{1}{T}\sum_{i=1}^{T} a_t^2}{\sigma_a^2} \to 1 \quad \text{by the law of large number}$$

and

$$\sup_r \left| \frac{(rT - \lfloor rT \rfloor)}{\sqrt{2}\sigma_a^2\sqrt{T}}\xi_{\lfloor rT \rfloor + 1} \right| \xrightarrow{\mathcal{P}} 0$$

it is

$$\sqrt{T/2} D_k \xrightarrow{\mathcal{D}} W^*(r) \tag{54}$$

Applying the Continuous Mapping Theorem the result is proven. □

The distribution of $\sup_r |W^*(r)|$ is given in Billingsley (1999); finite-sample critical values for the test can be determined by simulation.

The most serious weakness of the IT statistic is that its asymptotic distribution is strongly dependent on the assumption that the random variables $\{a_t\}$ have a Normal independent and identical distribution. The following proposition generalized the asymptotic distribution defined previously in the case of $\{a_t\} \sim i.i.d(0, \sigma^2)$

Proposition 2 (Sansó *et al.*, 2004) *If* $\{a_t\} \sim i.i.d(0, \sigma_a^2)$ *and* $\mathrm{E}(a_t^4) = \eta_4 < \infty$*, then:*

$$IT \xrightarrow{A} \sqrt{\frac{\eta_4 - \sigma_a^4}{2\sigma_a^4}} \sup_r |W^*(r)| \tag{55}$$

Proof The proof follows most of the steps used in the proposition 1. Note that, for not mesokurtic random variable $\mathrm{Var}(\xi_t) = \eta_4 - \sigma_a^4$. It is:

$$\sqrt{\frac{2\sigma_a^4}{\eta_4 - \sigma_a^4}}\sqrt{\frac{T}{2}} D_k = \sqrt{\frac{2\sigma_a^4}{\eta_4 - \sigma_a^4}}\sqrt{\frac{T}{2}}\left(\frac{C_k}{C_T} - \frac{k}{T}\right)$$

$$
\begin{aligned}
&= \frac{\sqrt{T}\sigma_a^2}{\sqrt{\eta_4 - \sigma_a^4}} \left(\frac{C_k - \frac{k}{T} C_T}{C_T} \right) \\
&= \frac{1}{\sqrt{T}\sqrt{\eta_4 - \sigma_a^4}} \left(C_k - \frac{k}{T} C_T \right) \left(\frac{\sigma_a^2}{\frac{1}{T} C_T} \right) \\
&= \frac{1}{\sqrt{T}\sqrt{\eta_4 - \sigma_a^4}} \left(\sum_{t=1}^{k} a_t^2 - \frac{k}{T} \sum_{t=1}^{T} a_t^2 \right) \left(\frac{\sigma_a^2}{\frac{1}{T} \sum_{t=1}^{T} a_t^2} \right) \\
&= \frac{1}{\sqrt{T}\sqrt{\eta_4 - \sigma_a^4}} \left(\sum_{t=1}^{k} (a_t^2 - \sigma_a^2) - \frac{k}{T} \sum_{t=1}^{T} (a_t^2 - \sigma_a^2) \right) \left(\frac{\sigma_a^2}{\frac{1}{T} \sum_{t=1}^{T} a_t^2} \right) \\
&= \frac{1}{\sqrt{T}\sqrt{\eta_4 - \sigma_a^4}} \left(\sum_{t=1}^{k} \xi_t - \frac{k}{T} \sum_{t=1}^{T} \xi_t \right) \left(\frac{\sigma_a^2}{\frac{1}{T} \sum_{t=1}^{T} a_t^2} \right)
\end{aligned}
$$

As $T \to \infty$, by the law of large number:

$$\frac{\sigma_a^2}{\frac{1}{T} \sum_{t=1}^{T} a_t^2} \to 1$$

and, by Donsker's theorem (Billingsley 1999, thm. 8.2), it is:

$$\frac{1}{\sqrt{T}\sqrt{\eta_4 - \sigma_a^4}} \sum_{t=1}^{\lfloor rT \rfloor} \xi_t \xrightarrow{\mathcal{D}} W(r) \tag{56}$$

From the previous results, it follows that:

$$
\begin{aligned}
\sqrt{\frac{2\sigma_a^4}{\eta_4 - \sigma_a^4}} \sqrt{\frac{T}{2}} D_k &\xrightarrow{\mathcal{D}} W(r) - rW(1) \\
&\equiv W^*(r)
\end{aligned}
$$

and so:

$$\sqrt{T/2} D_K \xrightarrow{\mathcal{D}} \sqrt{\frac{\eta_4 - \sigma_a^4}{2\sigma_a^4}} W^*(r) \tag{57}$$

Applying the Continuous Mapping Theorem the result is proven. □

Remark 1 *If $\{a_t\}$ is a Gaussian process, $\eta_4 = 3\sigma_a^4$ and the result in proposition 1 holds. When $\eta_4 < 3\sigma_a^4$ the distribution is platykurtic and the test could be too conservative. On the contrary, if $\eta_4 > 3\sigma_a^4$ the distribution is leptokurtic and, as a consequence, the effective size of the IT test could be greater than the nominal one*

Hence, the asymptotic distribution is not free of nuisance parameters and, when using the critical values of the supremum of a Brownian Bridge, size distortion should be expected.

Despite these problems, the IT test has been used for detecting changes in the volatility of financial time series (see Aggarwal *et al.*, 1999; Huang and Yang, 2001). However its reliance on the Gaussian distribution is problematic since financial data are known to be non Gaussian and can exhibit heavy tailed behaviour. However, as it will become apparent later, adjustments can be applied to IT statistic that allow $\{a_t\}$ to obey a wide class of dependent processes, including GARCH processes, under the null. In this direction, two different tests, K_1 and K_2, which consider the fourth moment properties of the disturbances and the conditional heteroskedasticity, have been proposed. The K_1 test corrects for non-mesokurtosis. It is defined as:

$$K_1 = \sup_r |T^{-1/2} B_k| \tag{58}$$

where:

$$B_k = \frac{C_k - (k/T)C_T}{\sqrt{\widehat{\eta}_4 - \widehat{\sigma}_a^4}} \tag{59}$$

with $\widehat{\eta}_4 = T^{-1}\sum_{t=1}^{T} a_t^4$ and $\widehat{\sigma}_a^2 = T^{-1}C_T$

Proposition 3 (Sansó *et al.*, 2004) *Under the hypothesis that* $\{a_t\} \sim i.i.d(0, \sigma_a^2)$ *and* $\mathrm{E}(a_t^4) < \infty$ *the asymptotic distribution of the* K_1 *statistic is given by*

$$K_1 \overset{A}{\to} \sup_r |W^*(r)| \tag{60}$$

Proof The proof follows immediately from Proposition 2. Indeed, from (57) it is:

$$\sqrt{\frac{2\sigma_a^4}{\eta_4 - \sigma_a^4}}\sqrt{\frac{T}{2}} D_K \overset{\mathcal{D}}{\to} W^*(r) \tag{61}$$

By simple algebra and by using $\widehat{\sigma}_a^2 = T^{-1}C_T$, it is easy to show that:

$$\begin{aligned}\sqrt{\frac{2\sigma_a^4}{\eta_4 - \sigma_a^4}}\sqrt{\frac{T}{2}} D_K &= \frac{1}{\sqrt{T}\sqrt{\eta_4 - \sigma_a^4}}\left[C_k - \frac{k}{T}C_T\right] \\ &\equiv T^{-1/2}B_k \end{aligned} \tag{62}$$

Applying the Continuous Mapping Theorem the result is proven. □

The statistics IT and K_1 are both based on the assumption of independence of the sequence of random variables and, as a consequence, they are not suitable for financial time series. To overcome this problem, Sansó *et al.* (2004) have proposed to correct the cumulative sum of squares with an estimation of the persistence. However some conditions on $\{a_t\}$ are required.

Assumption 1 $\mathrm{E}(a_t) = 0$ *and* $\mathrm{E}(a_t^2) = \sigma_a^2 < \infty \; \forall t \geq 1$

Assumption 2 $\sup_t \mathrm{E}(|a_t|^{\psi+\epsilon}) < \infty$ *for some* $\psi \geq 4$ *and* $\epsilon > 0$

Assumption 3 '$\omega_4 = lim_{T\to\infty}\mathrm{E}(T^{-1}(\sum_{t=1}^{T}(a_t^2 - \sigma_a^2))^2)$

Assumption 4 $\{a_t\}$ *is* α*-mixing with coefficients* α_j *satisfying:*

$$\sum_{j=1}^{\infty} \alpha_j^{(1-2/\psi)} < \infty$$

Assumption 1 imposes that the unconditional mean of the process is 0 and the unconditional variance exists and it is constant. The latest is the hypothesis which has to be tested.
Assumption 2 states the existence of moments of order greater than four although it does not impose them a constant value and, as a consequence, some sort of non stationarity is allowed.
Assumption 3 establishes a value for ω_4 which can be interpreted as the long run fourth order moment of $\{a_t\}$.
Assumption 4 allows for temporal dependence in the process $\{a_t\}$ (see appendix B for definition of mixing processes) so that, although there may be substantial dependence among recent event, events which are separated by long intervals of time are almost independent.
From the previous assumptions, it is evident that the existence of the fourth order moment and the finiteness of the long-run fourth order moment are both necessary conditions to derive the asymptotic distribution of the test statistic. Therefore, classes of processes which do not satisfy these conditions cannot be considered. For example, conditions 2 and 3 are not fulfilled if $\{a_t\}$ is a sequence of random variables independent and identically distributed as a t-Student with three degrees of freedom.
Under the previous assumptions, Sansó *et al.* (2004) have proposed the following statistic:

$$K_2 = \sup_k |T^{-1/2} G_k| \tag{63}$$

where $G_k = \widehat{\omega}_4^{-1/2}[C_k - (k/T)C_T]$ and $\widehat{\omega}_4$ is a consistent estimator of ω_4.
The asymptotic distribution of the three statistics IT, K_1 and K_2 under the variance-persistence processes are illustrated in the following proposition.

Proposition 4 (Sansó *et al.*, 2004) *Under the assumptions 1-4, the asymptotic distribution of the statistics are given by*

$$\begin{aligned}
a) &\quad IT \xrightarrow{A} \sqrt{\frac{\omega_4}{2\sigma_a^4}} \sup_r |W^*(r)| \\
b) &\quad K_1 \xrightarrow{A} \sqrt{\frac{\omega_4}{\eta_4 - \sigma_a^4}} \sup_r |W^*(r)| \\
c) &\quad K_2 \xrightarrow{A} \sup_r |W^*(r)|
\end{aligned}$$

Proof The proof follows most of the step used by Sansó *et al.* (2004). For the assumption 4, $\{a_t\}$ are α-mixing and then they are also ξ_t. Thus it is possible to use the Herrndorf's theorem (Herrndorf, 1984) being assumptions 1-4 a restrictive case of the conditions of the theorem.
Let us consider:

$$\begin{aligned}
\frac{1}{\sqrt{T}\sqrt{\omega_4}}\left(C_k - \frac{k}{T}C_T\right) &= \frac{1}{\sqrt{T}\sqrt{\omega_4}}\left(\sum_{t=1}^{k}\xi_t - \frac{k}{T}\sum_{t=1}^{T}\xi_t\right) \\
&\xrightarrow{\mathcal{D}} W(r) - rW(1) \\
&\equiv W^*(r)
\end{aligned}$$

Suppose that $\widehat{\omega_4}$ is a consistent estimator of ω_4. For the previous result:

$$\begin{aligned}
\frac{1}{\sqrt{T}}G_K &= \frac{1}{\sqrt{T}}(\widehat{\omega_4})^{-1/2}\left(C_k - \frac{k}{T}C_T\right) \\
&\xrightarrow{\mathcal{D}} W^*(r)
\end{aligned}$$

Applying the Continuous Mapping Theorem the result c) is proven.
Given that

$$\frac{1}{\sqrt{T}}\left(C_k - \frac{k}{T}C_T\right) \xrightarrow{\mathcal{D}} (\widehat{\omega_4})^{1/2}W^*(r) \tag{64}$$

by simple algebra:

$$\begin{aligned}\sqrt{\frac{T}{2}}D_k &= \sqrt{\frac{T}{2}}\left(\frac{C_k}{C_T} - \frac{k}{T}\right)\\ &= \frac{1}{\sqrt{T}}\left(C_k - \frac{k}{T}C_T\right)\frac{1}{\sqrt{2}}\frac{T}{C_t}\\ &\xrightarrow{\mathcal{D}} \sqrt{\frac{\omega_4}{2\sigma_a^4}}W^*(r)\end{aligned}$$

Applying the Continuous Mapping Theorem the result a) is proven.
Moreover it is:

$$\begin{aligned}\frac{1}{\sqrt{T}}B_k &= \frac{1}{\sqrt{T}}\frac{1}{\sqrt{\widehat{\eta}_4 - \widehat{\sigma}_a^4}}\left(C_k - \frac{k}{T}C_T\right)\\ &\xrightarrow{\mathcal{D}} \sqrt{\frac{\omega_4}{\eta_4 - \sigma_a^4}}W^*(r)\end{aligned}$$

Applying the Continuous Mapping Theorem the result b) is proven. □

2.2.2 Testing for multiple structural breaks

Many of the procedures for analysing single break point can be adapted to the analysis of multiple break points by recursively applying a single change point method. Different approaches have been proposed in the statistical and the econometric literature, such as the ICCS algorithm (Inclan and Tiao, 1994), the Binary Segmentation (Edwards and Cavalli-Sforza, 1965), the Segmented Neighborhood search (Braun *et al.*, 2000) and the Minimum description method (Davis *et al.*, 2006). Here, the attention has been focused on the binary segmentation which is based on successive application of the test to sub-series obtained consecutively after a change-point is found. The procedure starts by applying the detection method to the whole series. If no change-point is found the procedure is stopped; otherwise the data are split into two segments and the detection method is applied to each of them. The procedure is repeated until no further change-points are detected.
The choice of binary segmentation algorithm is justified by its simplicity and efficiency; it is

very fast and it could be implemented with a low computational cost.
However, the procedure could produce spurious break points because of the presence of extreme observations which can be erroneously interpreted as being change points (Ross, 2013). To partially solve this problem and to better identify the break points location, it is necessary to implement a pruning procedure. In this paper a pruning procedure, in the spirit of the ICCS algorithm (Inclan and Tiao, 1994), has been suggested.
In the case that m breaks have been detected at times $\tau_1, \tau_2, \ldots, \tau_m$, with $\tau_0 = 1$ and $\tau_{m+1} = T$, the pruning procedure can be implemented as follows.

- The detection method is applied to the segment $[\tau_{i-1}, \tau_{i+1}]$ for $i = 1, \ldots, m$
- If no change-point is found in the segment $[\tau_{i-1}, \tau_{i+1}]$, the break at τ_i is not considered as change point. If a new change point is detected, it replaces the old one at τ_i.
- The procedure is repeated until the number of change points does not change and the points found in each new step are "close" to those on the previous step.

The main problem, when a detection method is applied with a searching algorithm, is that the use of the same critical value for any segments may distort the performance of the iterative procedure. To overcome this problem, it is possible to use the response surfaces methodology (MacKinnon, 1994). This approach allows to approximate the asymptotic distribution of a test and to obtain critical values for the statistic that converges to this distribution. This method is based on the determination of the following regression:

$$q_T^\alpha = \sum_{j=1}^{d} \theta_{p_j}^\alpha T^{p_j} + v_T \tag{65}$$

where q_T^α is the $\alpha-$quantile of the test statistic for a sample size T, θ_{p_j} are the parameters to be estimated and the regressors are powers of T. The values of q_T^α could be obtained by using Monte Carlo experiments for different values of sample size T. These empirical quantiles are than used in the response surface regression to calculate the estimated parameters. For the K_2 test statistics. Sansó *et al.* (2004) reported the final estimates of the response surface with $\alpha = 0.05$.

$$
\begin{aligned}
\hat{q}_T^{0.05} &= 1.405828 - 3.317278 * T^{-0.5} + 31.22133 * T^{-1} - 1672.206 * T^{-2} + \\
&+ 52870.53 * T^{-3} - 411015 * T^{-4}
\end{aligned} \tag{66}
$$

The parameter estimates are all significant, the fit can be considered quite good, being $R^2 = 0.999$, and the residual standard deviation is very small ($\hat{\sigma}_v = 0.013$).
The equation (66) can be used to compute the critical values of the K_2 test for different sample sizes T.

2.3 Long-run fourth order moment estimation

When using the statistic K_2, it is necessary to obtain a consistent estimator of ω_4 the long-run fourth order moment of a_t. Since it provides the right standardization to obtain a statistic free of nuisance parameters, the determination of this estimator plays a crucial role in implementing the CUSUM of squares test.
The long-run fourth order moment of a_t can be rewritten in terms of the long-run variance of the zero mean random variable $\xi_t = a_t^2 - \sigma_a^2$ and, as a consequence, ω_4 can be estimated by using the spectral density of ξ_t at frequency zero. One possibility is to use a parametric approach based on the Akaike estimator of the spectrum. However, a more popular strategy is based on a non parametric kernel method often referred to as the kernel Heteroskedasticity Autocorrelation Consistent (HAC) estimator. It is defined as:

$$
\widehat{\omega}_4 = \widehat{\gamma}_0 + 2\sum_{l=1}^{m}[w(l,m)]\widehat{\gamma}_l \tag{67}
$$

where

$$
\widehat{\gamma}_l = T^{-1}\sum_{t=l+1}^{T}(a_t^2 - \widehat{\sigma}_a{}^2)(a_{t-l}^2 - \widehat{\sigma}_a{}^2) \tag{68}
$$

$w(.)$ is a suitable kernel weighting function and m is a truncation lag or bandwidth parameter. The kernel function, which is assumed to verify the standard conditions of continuity, squared integrability and symmetry, ensures non-negative estimates. In the context of long run variance estimation, it is quite common, to consider the Bartlett Kernel. It is defined as:

$$w(l,m) = 1 - \frac{l}{m+1} \tag{69}$$

The choice of the bandwidth parameter m plays a fundamental role in order to obtain accurate properties of the resultant HAC estimator. It captures the covariance structure of the underlying series; if the temporal dependence increases, a larger value of m is necessary to account for a larger number of non-zero covariances.

From an asymptotic point of view, this parameter is a function of the series length T, $m = m_T$ and it should verify $m_T \to \infty$ as $T \to \infty$ in such a way that $m_T/T \to 0$, *i.e.*, $m_T = o(T)$. Under very general assumptions on the data generating process, this approach provides consistent estimates.

In empirical applications, different procedures to choose m_T have been proposed based on deterministic rules or on data dependent methods.

In a first proposal, the bandwidth depends only on the sample size T; it can be fixed by setting $m_T = [bT^a]$ where a and b are suitable constants such that $b > 0$ and $0 < a < 1$. A common choice is $m_T = \left[4(T/100)^{2/9}\right]$ (see Rodrigues and Rubia, 2007).

The second approach assumes that the bandwidth can be chosen as a function of the correlation structure of the series and thus it results data dependent (see Andrews, 1991; Newey and West, 1994). In the case of Bartlett kernel, the optimal bandwidth is defined as:

$$m_T = min\left\{T, \lfloor \theta T^{\frac{1}{3}} \rfloor\right\} \tag{70}$$

where

$$\theta = 1.447(S_1^2/S_0^2)^{\frac{1}{3}} \tag{71}$$

$$S_j = \widehat{\gamma}_0 + 2\sum_{i=1}^{m_T^*} i^j \widehat{\gamma}_j \qquad j = 0, 1 \tag{72}$$

and m_T^* is a given deterministic pre-bandwidth parameter.

More recently, in a different context, inconsistent long run variance estimators have been proposed (Kiefer and Vogelsang, 2002; 2005). They are based on a bandwidth verifying the condition $m_T/T \to b$, where b is a fixed constant such that $0 < b \leq 1$. This is equivalent to fix the bandwidth as $m_T = bT$ where b represents the proportion of sample autocovariances included in the estimation. Under this approach $m_T = O_P(T^2)$ and, as a consequence, the

HAC estimator is not consistent.
A common choice is to fix $b = 1$ (Kiefer and Vogelsang, 2002); in this case $m_T = T$ and so the estimator uses the entire sample. Although such estimators are inconsistent, when they are used in test statistics, the resulting asymptotic approximation seems to have some advantages in improving the size properties compared to the conventional approach. However, as pointed out by Cai an Shintani (2006) for unit-toot testing, the power of the test could decrease significantly.
In the CUSUM framework, this problem has been addressed by Rodriguez and Rubia (2007). They have found that when the data generating process is a GARCH (1,1) model, the CUSUM test seems to over-reject the null hypothesis when a consistent estimator of ω_4 is used and to under-reject in the case of inconsistent estimator. Moreover they have found a significant distortion of the test resulting from the long run variance estimation in small sample and, even in large samples, when the empirical process exhibits persistent volatility.
All the previous results confirm the important role of the estimation of the long run fourth order moment to perform the CUSUM test. Unfortunately, even if a consistent estimator is considered and although inference based on the resulting test statistic is asymptotically valid, it is well known that in finite sample HAC estimator could have a significant bias which is evident not only in samples of small dimension but also in large samples, especially when the data generating process exhibits persistent volatility.

In this paper we propose a bias correction of the estimation of ω_4 based on the use of a resampling technique, namely the stationary bootstrap. With this approach we will be able to reduce the bias of the estimator of the long run fourth order moment and, as a consequence, the size distortion of the CUSUM test.

2.4 The bootstrap bias correction

The bootstrap is an approach to approximate the distribution of a statistic of interest and to assign measures of accuracy to statistical estimates. This technique involves resampling the original data to form many pseudo-time series and re-estimating the statistic, or the quantity of interest, on each bootstrap sample.
Starting from the pioneeristic Efron's paper (Efron, 1979) in the context of independent and

identically distributed data, many bootstrap procedures have been proposed in the literature, in order to solve more complex statistical problems.

In the context of dependent data, a natural approach is to model explicitly the dependence in the data through a parametric model; the bootstrap sample is then drawn from the fitted model. Such model-based approach is inconsistent if the model used for resampling is mis-specified. Alternatively, nonparametric model free bootstrap schemes can be used.

In particular the block bootstrap scheme is one of the most widely used bootstrap methods in time series analysis. In this scheme, blocks of consecutive observations are resampled randomly with replacement, from the original series and assembled by joining the blocks together in random order in order to obtain a simulated version of the original series. Therefore this resampling technique maintains the dependence structure of the data within a pseudo sample.

There are basically two different ways of proceeding, depending on whether the blocks are overlapping or non-overlapping. Both approaches have been introduced by Hall (1985) in the context of spatial data. For univariate time series, Carlstein (1986) has proposed non-overlapping blocks, whereas Künsch (1989) has suggested overlapping blocks.

The block bootstrap method does not require the selection of a model but it is sensitive to the block length specification, which is the crucial point of the procedure. If the blocks are long, a poor estimate of the distribution of the statistic could be obtained whereas too short blocks could not preserve, in the resampled series, the original dependence in the data.

Several rules have been suggested to select the optimal length of the blocks; they are based on different criteria but often they are useful only as rough guides. However, unless the length of the series is long enough to allow long and numerous blocks, the dependence structure may be difficult to preserve, especially when the data have a complex and long range dependence structure. In such cases, the block bootstrap scheme tends to generate resampled series that are less dependent than the original ones. Moreover by construction, in the block bootstrap method, the bootstrapped time series has a nonstationary (conditional) distribution. Politis and Romano (1994) have introduced a variant of the block bootstrap, namely the stationary bootstrap, that is generally applicable for stationary weakly dependent time series.

Unlike the block bootstrap method, where the block length is fixed, in the stationary bootstrap method the length of each block of observations is generated from a geometric distribution with parameter Q. As a consequence, the pseudo time series generated by this procedure is actually stationary and so, the stationarity property of the original series is replicated in the resample time series (Politis and Romano, 1994).

In this paper this resampling technique has been used to obtain a bias correction of the estimator of the long run fourth order moment (defined in equation 67) in the case of a strictly stationary and weakly dependent process. In this context, given the real number $0 < Q < 1$, the procedure runs as follows:

Step 1. Randomly select an observation, say a_t, from the data as the first bootstrapped observation a_1^*.

Step 2. With probability Q, a_2^* is set to a_{t+1}, the observation following the previously sampled observation, and with probability $1 - Q$, the second bootstrapped observation a_2^* is randomly selected from the original data.

Step 3. Repeat recursively step 2 to form $(a_1^*, \ldots, a_T^*)$, the bootstrap series with T observations.

In this scheme, the block size is the (random) number of bootstrapped observations that are drawn consecutively. The block size follows a geometric distribution with parameter Q, and so the bootstrapped results depend on the choice of this parameter. The average length of a block is $1/Q$, where Q is the parameter of the geometric distribution. Thus, $1/Q$ should play the same role as the block length parameter in the block bootstrap. Of course, when Q approaches zero, the resulting stationary bootstrap is like the *i.i.d.* bootstrap. However, as pointed out by Politis and Romano (1994), the stationary bootstrap is less sensitive to the choice of Q than the block bootstrap is to the choice of the block length. Politis and White (2004) have suggested that, under very general conditions, the optimal expected block size for the stationary bootstrap is $c * T^{1/3}$ where c is a constant depending on the specific properties of the data generation process and on the problem at hand. In our case, the constant c has been fixed equal to θ defined in equation (71).
For the long run fourth order moment, the bootstrap bias correction could be obtained by a Monte Carlo procedure which has been implemented as follows.

Step 4. Generate B different bootstrap series $(a^*_{1,b}, \ldots, a^*_{T,b})$ for $b = 1, , \ldots, B$. For each generated series calculate the long run fourth order moment $(\widehat{\omega}^*_{4,1}, \ldots, \widehat{\omega}^*_{4,B})$.

Step 5. Estimate the bias:

$$\widehat{bias} = \frac{1}{B}\sum_{b=1}^{B} \widehat{\omega}^*_{4,b} - \widehat{\omega_4} \tag{73}$$

Step 6. The bias corrected estimator of the long run fourth order moment is:

$$\widehat{\omega_4}^{bc} = \widehat{\omega_4} - \widehat{bias} = 2\widehat{\omega_4} - \frac{1}{B}\sum_{b=1}^{B} \widehat{\omega}^*_{4,b} \tag{74}$$

2.5 Monte Carlo Experiment

In this section results from a set of Monte Carlo simulation experiments have been reported and discussed. The aim is to evaluate the performance of the K_2 test in which the estimation of the long run fourth order moment has been bias corrected using the stationary bootstrap. In particular, we compare our proposal with that obtained by using a classical HAC estimation. We considered two alternative data generating processes, namely GARCH(1,1) with Normal error and log-Normal stochastic volatility. In the next two subsections we review the conditions necessary to guarantee that the processes match the assumptions 1-4 under which the test has been developed.

2.5.1 The data generating process: GARCH (1,1)

The GARCH(1,1) model, proposed by Bollerslev (1986), is the most commonly applied parameterization in empirical econometrics; it is a very parsimonious model and usually it is adequate to obtain good performances in terms of fitting and forecasting (Hansen and Lunde, 2005).

The canonical GARCH(1,1) process $(a_t)_{t\in\mathbb{Z}}$ with volatility process $(\sigma_t)_{t\in\mathbb{Z}}$ is defined as:

$$a_t = \sigma_t \epsilon_t \quad t \in \mathbb{Z} \tag{75}$$

with

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \qquad \alpha_0 > 0 \quad \alpha_1, \beta_1 \geq 0 \tag{76}$$

where $\{\epsilon_t\}$ is a sequence of *i.i.d.* random variables with mean zero and unit variance. The conditions on α_0, α_1 and β_1 ensure that the conditional variance σ_t^2 is positive.

The probabilistic proprieties of the GARCH (1,1) process are well known (for a review see Lindner, 2009). In the following the questions of stationarity [6], moment conditions and strong mixing proprieties will be reviewed.

Nelson (1990) has proved that the GARCH(1,1) process with $\alpha_0, \alpha_1, \beta_1 > 0$ is strictly stationary if and only if

$$-\infty < \mathrm{E}(\ln(\beta_1 + \alpha_1 \epsilon_t^2)) < 0 \tag{77}$$

By the Jensen inequality and under the assumption that $\{\epsilon_t\}$ has unit variance, it is easy to show that $\alpha_1 + \beta_1 < 1$ is a sufficient condition for (77). Moreover Bollerslev (1986) has stated that the GARCH(1,1) process is weakly stationary if and only if $\alpha_1 + \beta_1 < 1$. Bougerol and Picard (1992) have also shown that every strictly stationary GARCH(1,1) process is ergodic. Carrasco and Chen (2002) have proved that, under very general conditions, if the process is strictly stationary then a_t and σ_t^2 are β-mixing[7] with geometric rate. Because of the β-mixing implies α-mixing, the processes a_t and σ_t^2 are also α-mixing with geometric rate.

In conclusion, the condition $\alpha_1 + \beta_1 < 1$ is necessary and sufficient for weakly stationarity, sufficient for strictly stationarity and ergodicity and it ensures that the process is β-mixing with geometric rate.

Once the stationarity and mixing conditions are defined, it is important to define conditions for the existence and the finiteness of the moments of order higher than two. For the GARCH(1,1) model a necessary and sufficient condition has been given by Bollerslev (1986) for Normal innovation and by He and Teräsvirta (1999a; 1999b) for general noise sequence. In particular, if ϵ_t is standard Normally distributed, the condition for the existence of the fourth order moment is

$$3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1 \tag{78}$$

[6] When considering a stationary GARCH(1,1) process, the stationarity is referred to the bivariate process $(a_t, \sigma_t)_{t \in \mathbb{N}_0}$.

[7] Mixing conditions describe some types of asymptotic independence which allow the use of suitable limit theorems. The common types of mixing conditions have been introduced in appendix 2.9.2; for an extensive treatment see Doukhan (1994).

This condition is clearly more restrictive then $\alpha_1 + \beta_1 < 1$ and so it ensures all the previous proprieties.

2.5.2 The data generating process: Log-Normal Stochastic Volatility

The canonical log-normal Stochastic Volatility (SV) process $(a_t)_{t\in\mathbb{Z}}$ with volatility process $(\sigma_t)_{t\in\mathbb{Z}}$ is defined (Taylor, 2007) as:

$$a_t = \sigma_t \epsilon_t \quad t \in \mathbb{Z} \tag{79}$$

$$\ln(\sigma_t^2) = \phi_0 + \phi_1 \ln(\sigma_{t-1}^2) + v_t \qquad \phi_0, \phi_1 > 0 \tag{80}$$

where $\{\epsilon_t\}$ is a sequence of *i.i.d.* random Normal variables with mean zero and unit variance, $\{v_t\}$ is a sequence of *i.i.d.* random Normal variables with mean zero and variance σ_v^2, $\{\epsilon_t\}$ and $\{v_t\}$ are mutually independent.
Denoting $h_t = \ln(\sigma_t^2)$ the models (79) and (80) can be rewritten as:

$$a_t = \exp\left(\frac{1}{2}h_t\right)\epsilon_t \tag{81}$$

$$h_t = \phi_0 + \phi_1 h_{t-1} + v_t \tag{82}$$

This model is the natural discrete-time approximation to the continuous-time Orstein-Uhlenbeck process used in finance theory.
As pointed out by Harvey *et al.* (1994) if $|\phi_1| < 1$, the process h_t is ergodic and stationary both in the strict and weak sense. It is trivial to show that its mean is $\mu_h = \phi_0/(1 - \phi_1)$ and its variance is $\sigma_h^2 = \sigma_v/(1 - \phi_1^2)$.
The condition $|\phi_1| < 1$ ensures also the stationarity of a_t being the product of two stationary processes. Moreover the condition is sufficient by itself to ensure the existence of the moments. The odd moments of a_t are all zero because ϵ_t is symmetric; the even moments can be obtained by using standard results, being $\exp(h_t)$ lognormal. In particular, it is easy to show that the variance of a_t is equal to $\exp(\mu_h + \frac{1}{2}\sigma_h^2)$ and, unlike a GARCH model, the fourth order moment always exists when h_t is stationary.
Finally, Carrasco and Chen (2002) have shown that if a_t is a stationary process it is also β-mixing and so the previous results apply.

2.5.3 Simulation design and results

In order to analyse the false rejection rates, data generating processes without structural breaks have been considered. In particular for the GARCH(1,1) process, we have considered a parametric space which includes empirical values typically observed in practice. In particular, we have fixed $\alpha_0 = 0.1$ and we have set $\alpha_1 \in \{0.05, 0.075, 0.01\}$ and β_1 in such a way that $\alpha_1 + \beta_1 \in \{0.7, 0.8, 0.9, 0.95\}$. For the SV model we have set $\phi_1 \in \{0.7, 0.8, 0.9, 0.95\}$, $\sigma_v \in \{0.1, 0.2, 0.3\}$ and $\phi_0 = -\frac{1}{2}\frac{\sigma_v}{1+\phi_1}$, so that the unconditional variance of a_t is equals to 1.

To evaluate small samples as well as asymptotic behaviour of the analysed test, four different sample sizes $T \in \{500, 1000, 2000, 4000\}$ have been considered.

The simulations are based on 1000 Monte Carlo runs and, in the bootstrap implementation, the number of replications has been fixed equal to 999. The nominal level of the test is 0.05.

As pointed out previously, the expected block size in the stationary bootstrap scheme should be fixed to $c * T^{1/3}$ where c is a suitable constant. In this experiment, we have proposed to fix c according to the optimal bandwidth parameter of the Bartlett kernel used in the HAC estimation of the long run fourth order moment. More precisely, c is set to θ defined in equation (71). The value of the expected block size are reported in table 6 for the GARCH (1,1) specification and in table 7 for the log-normal stochastic volatility process. As expected, the values increase with the sample size and with the persistence.

The empirical false rejection rates for the different specifications of the GARCH (1,1) process are reported in table 8. The entries in bold indicate the values inside the asymptotic acceptance interval at level 0.95. [8]

It is quite evident that for all specifications of the parameters the false rejection rates of the test when a bootstrap bias correction is used are always less than the corresponding value obtained with the HAC estimator.

The test with HAC estimation has high values of the false rejection rate for small sample size ($T = 500$ and $T = 1000$); for large sample sizes reasonable rates are achieved only in the case of time-series with not extremely persistent conditional heteroskedasticity ($\alpha_1 + \beta_1 \leq 0.8$).

When the proposed bootstrap bias correction is used, for all sample sizes the false rejec-

[8]The asymptotic acceptance interval at 95% is defined as: $\widehat{p} \pm 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{T}}$ where $\widehat{p}$ is fixed at nominal value 0.05.

tion rates could be considered not statistically different from the nominal value when the persistence is not high ($\alpha_1 + \beta_1 \leq 0.8$). For large sample size, also for high persistence ($\alpha_1 + \beta_1 = 0.9$) the values are inside the acceptance interval.

The previous results are confirmed in the case of log-normal stochastic volatility model (table 9); the bootstrap bias correction always reduces the false rejection rates for all the parameter specifications and for all the sample sizes. However the test with HAC estimation generates false rejection rates not significantly different from the nominal value for $\phi_1 = 0.7$, in the case of small and moderate sample size ($T = 500$, $T = 1000$ and $T = 2000$) and for $\phi_1 \leq 0.8$ in the case of high sample size ($T = 4000$). Again, the false rejection rates achieve the nominal value, in the case of the test with bootstrap bias correction, for $\phi_1 \leq 0.8$, when the sample size is small and moderate ($T = 500$, $T = 1000$ and $T = 2000$) and for $\phi_1 \leq 0.9$ in the case of high sample size ($T = 4000$).

In order to analyse if the test is able to identify a break present in the data generating process, we have generated time series with a single structural break located in the middle of the sample. In this case three different sample sizes $T \in \{1000, 2000, 4000\}$ have been fixed. The variance in the first of the two regimes, identified by the structural breaks, has been fixed to 1 while the variance in the second to $\{2, 2.5, 3\}$. For the GARCH (1,1) process the parameters are $\alpha_1 = 0.1$ and $\beta_1 \in \{0.7, 0.8, 0.9, 0.95\}$; for the log-normal stochastic volatility model we have set $\phi_1 \in \{0.7, 0.8, 0.9, 0.95\}$ and $\sigma_v = 0.3$. With respect to the previous simulation design, the parameters α_1 and σ_v have been fixed since the performances of the test with HAC estimation and with the bootstrap bias correction seem to be not sensitive to the choice of these parameters.

For both models the constant has been fixed in such a way that the variances in the regimes assume the specified values. The simulations are based on 1000 Monte Carlo runs; the bootstrap parameters have been fixed as in the previous analysis. For each sample size, the relative frequency of zero break, a single break and more than one break have been computed for both the long run fourth order moment estimations. In this simulation we have also examined if the procedure is able to well identify the location of the break. To this aim it is assumed that a break is correctly located if it is identified in an interval of length 40 around the exact location. When the procedure identifies more than 1 break, it has been also computed the relative frequency of exact location of ”at least one” of the identified breaks.

In table 10 the results for the GARCH (1,1) process are reported. In general, it is quite evident

that the bootstrap approach is able to better identify the presence of one structural break in the data for all the specifications of the parameters and for all the sample sizes. Moreover, when one break is identifies, the test with the bias correction produces a hight relative frequency of exact location. The differences between the two approaches seem to be more clear when the persistence and the sample size increase. By looking at the case in which more than one breaks are identified, it is evident that the bootstrap procedure is more conservative, in the sense that the relative frequency is lower in all the considered simulation schemes. Furthermore, the zero break relative frequency tends to 0 as the sample size increases highlighting not significant differences between the two long run fourth order estimations.
The previous results are confirmed in the case of log-normal stochastic volatility model (table 11). In particular, the two approaches seem to have equal performances for $\phi_1 = 0.70$; for the other values of the parameter ϕ_1, the test with the bootstrap bias correction seems to significantly outperform that obtained with the HAC estimator, especially when the sample size increases.

2.6 An example on real data

In order to evaluate the effectiveness of the proposed procedure, an example on real data has been implemented.
The series are the IPC index (Índice de Precios y Cotizaciones) which is the broadest indicator of the overall performance of the Mexican Stock Exchange and the CNX Nifty index which is the National Stock Exchange of India's benchmark stock market index for Indian equity market.
The data are obtained from the Oxford-Man Institute's realised library and the sample covers the period from January 3, 2000 to May 20, 2016 for the IPC Index and from January 1, 2003 to May 20, 2016 for the CNX Nifty index.
In order to investigate the presence of break proints in the unconditional variance of both the series, an analysis based on the K_2 test with the HAC estimator of the long run variance and the proposed bootstrap bias correction has been employed. In both cases, a binary segmentation with the pruning method described in section 2.2 has been used.
For both the series, the identified break points with the two procedures are reported in table 12 and 13 respectively for the IPC-Mexico and CNX Nifty-India; in figure 6 and 7 the two-

standard-deviation bands for each of the regimes defined by the structural breaks are shown. For the IPC index, the procedure based on HAC estimation identifies seven break points. Most of them could be considered as spurious false positive due to the presence of extreme observations which correspond to very small period of high variability. When applying the bootstrap bias correction to the estimation of the long run fourth order moment, the number of the break points has reduced. In particular, the proposed procedure identifies a single break point located at July 2009. The reduction of the volatility after this break point could be due to the effects of the end of the global financial crisis. Moreover the location of the break is consistent with that identified by the test based on HAC estimation; the difference between the two break dates is not relevant.

For the CNX Nifty Index both procedures identify three break points located in the same dates. The first break occurs at November 2007 which can be considered as the starting point of the global financial crisis. The second break point is located, as in the previous case, in July 2009 that could be assumed as the end of the financial crisis. The last break point is located at May 2012 and it is characterized by a reduction of volatility. In this case both procedures seem to work quite well and to be robust to extreme observations. The results of this toy example confirm that the test based on bootstrap estimation seems to be less sensitive to the extreme values and, as a consequence, more conservative with respect to the classical HAC estimation.

2.7 Concluding remarks

In this paper we have proposed a bias correction of the estimation of the long run fourth order moment in the CUSUM of squares test proposed by Sansó *et al.* (2004). It is a development of the Inclan and Tiao test (1994) for the detection of changes in unconditional variance suitable for financial time series analysis.

The correction has been made by using the stationary bootstrap which has been justified by the stationarity and the weak dependence of the time series under the assumptions which ensure the existence of the limiting distribution of the test statistic under the null. After a brief introduction and discussion on the CUSUM of squares test and on the bootstrap approach in the context of dependent data, a procedure to select the optimal expected block size has been proposed.

In order to evaluate the effects of the proposed bias correction, Monte Carlo experiments have been implemented focusing on two particular data generating processes, the GARCH(1,1) and the log-normal stochastic volatility models. The results give evidence that the bootstrap approach is better able to correctly identify the presence of structural breaks in the data. More specifically, in the GARCH(1,1) data generating process without breaks, the false rejection rates could be considered non statistically different from the nominal values when the persistence is not high and for large sample size even in the case of high persistence. These results are confirmed in the case of the log-normal stochastic volatility model.
When the data are generated assuming the presence of a single break in the middle of the sample, the bias corrected estimation is able to correctly identify the break and its location, for all the sample sizes and in both the considered models.
The proposed procedure has been applied to analyse the presence of structural breaks in two real time series, IPC-Mexico and CNX Nifty-India. In both cases it seems to work quite well resulting more robust with respect extreme observations.

2.8 Figures and Tables

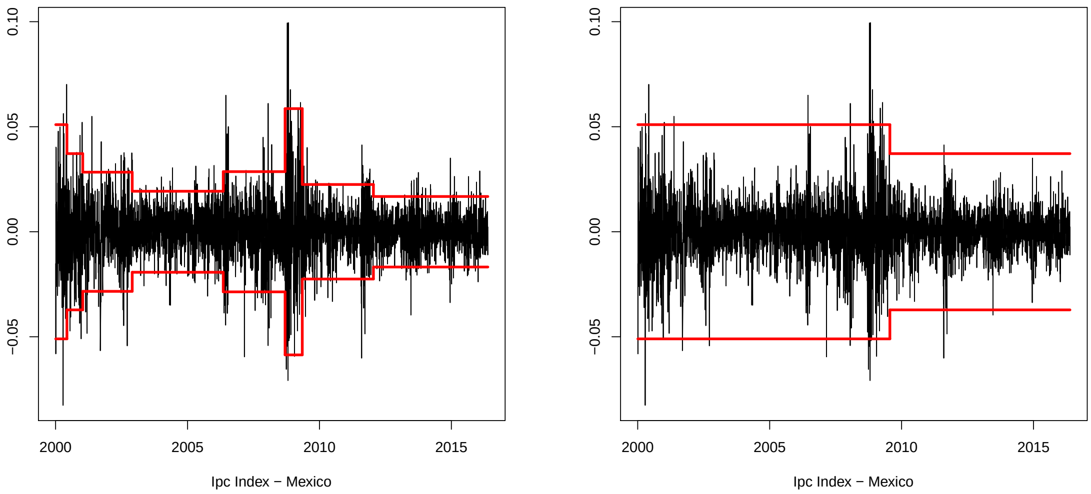


Figure 6: IPC Index-Mexico.

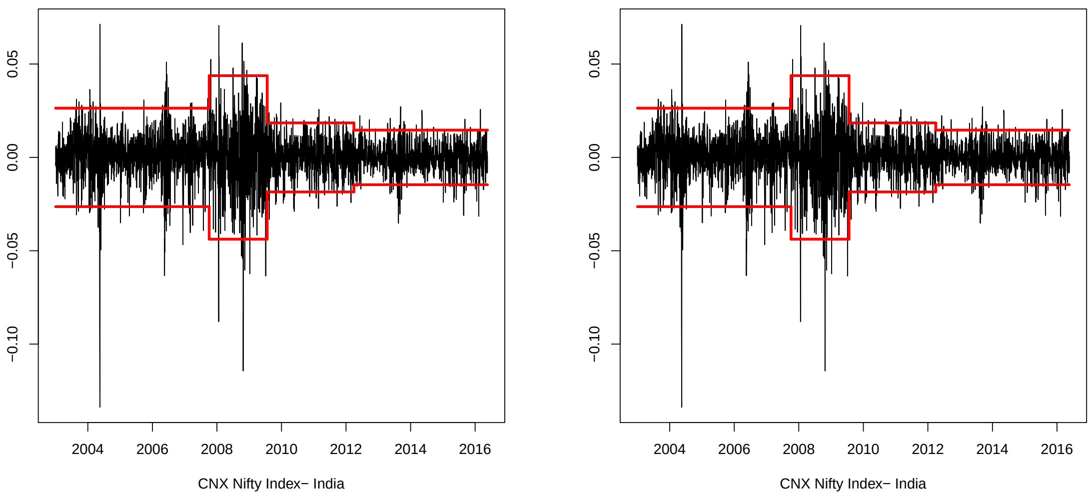


Figure 7: CNX Nifty Index-India.

Note: The red lines are two-standard-deviation bands for the regimes defined by the structural breaks identified by the binary segmentation and the pruning procedure. On the left the breaks have been identified by the K_2 test with the HAC estimator of the long run variance; on the right with the proposed bootstrap bias correction.

Table 6: Expected block size of the stationary bootstrap for the GARCH (1,1) data generating process.

α_1	$\alpha_1 + \beta_1$	**T=500**	**T=1000**	**T=2000**	**T=4000**
	0.70	6.51	8.26	10.52	14.17
0.05	0.80	7.11	9.69	13.34	19.26
	0.90	8.09	12.02	17.47	26.80
	0.95	8.89	13.90	20.48	32.15
	0.70	7.20	9.65	12.83	17.46
0.075	0.80	8.18	11.72	16.44	23.63
	0.90	9.63	14.64	21.06	31.78
	0.95	10.77	16.70	24.23	37.29
	0.70	7.92	10.93	14.75	20.04
0.10	0.80	9.16	13.35	18.63	26.65
	0.90	10.91	16.41	23.37	34.97
	0.95	12.12	18.44	26.46	40.24

Table 7: Expected block size of the stationary bootstrap for the log-normal stochastic volatility data generating process.

σ_v	ϕ_1	**T=500**	**T=1000**	**T=2000**	**T=4000**
	0.70	6.09	7.11	7.96	9.89
0.10	0.80	6.21	7.23	8.00	10.81
	0.90	6.44	8.13	10.63	16.64
	0.95	7.44	11.13	16.82	26.73
	0.70	6.17	7.24	8.73	11.33
0.20	0.80	6.46	8.42	11.25	16.27
	0.90	8.58	13.18	19.21	28.92
	0.95	11.51	17.43	25.13	38.06
	0.70	6.22	8.06	10.39	13.77
0.30	0.80	7.85	11.04	15.69	22.16
	0.90	11.09	16.43	23.06	34.34
	0.95	13.21	19.33	27.30	40.92

Table 8: GARCH(1,1) data generating process: False Rejection Rate for the K_2 test with the HAC estimator of the long run variance and with the bootstrap bias correction.

α_1	$\alpha_1+\beta_1$	**T=500**		**T=1000**		**T=2000**		**T=4000**	
		HAC	BOOT	HAC	BOOT	HAC	BOOT	HAC	BOOT
0.05	0.70	0.073	**0.057**	0.066	**0.055**	**0.056**	**0.045**	**0.057**	**0.047**
	0.80	0.083	**0.062**	0.078	**0.063**	0.064	**0.049**	**0.057**	**0.047**
	0.90	0.149	0.097	0.126	0.087	0.100	0.071	0.066	**0.054**
	0.95	0.268	0.199	0.235	0.166	0.205	0.141	0.140	0.090
0.075	0.70	0.072	**0.053**	0.068	**0.057**	**0.056**	**0.045**	**0.055**	**0.047**
	0.80	0.085	**0.062**	0.076	**0.055**	0.065	**0.046**	**0.051**	**0.046**
	0.90	0.156	0.103	0.133	0.083	0.100	0.068	0.069	**0.047**
	0.95	0.284	0.206	0.254	0.155	0.197	0.126	0.133	0.077
0.10	0.70	0.076	**0.050**	0.065	**0.047**	**0.055**	**0.044**	**0.050**	**0.045**
	0.80	0.088	**0.059**	0.075	**0.051**	0.066	**0.044**	**0.050**	**0.045**
	0.90	0.154	0.093	0.127	0.073	0.096	**0.058**	0.072	**0.044**
	0.95	0.281	0.186	0.242	0.136	0.179	0.113	0.122	0.070

Note: The nominal level of the test is 0.05. In bold the values inside the asymptotic acceptance interval at 95%.

Table 9: Log-normal stochastic volatility data generating process: False Rejection Rate for the K_2 test with the HAC estimator of the long run variance and with the proposed bootstrap bias correction.

σ_v	ϕ_1	**T=500**		**T=1000**		**T=2000**		**T=4000**	
		HAC	BOOT	HAC	BOOT	HAC	BOOT	HAC	BOOT
0.10	0.70	**0.057**	**0.052**	**0.057**	**0.053**	**0.053**	**0.052**	**0.054**	**0.052**
	0.80	0.071	**0.062**	**0.059**	**0.056**	**0.055**	**0.051**	**0.054**	**0.052**
	0.90	0.096	0.085	0.081	0.066	0.084	0.070	0.068	**0.056**
	0.95	0.205	0.165	0.186	0.155	0.176	0.142	0.133	0.096
0.20	0.70	**0.053**	**0.051**	**0.058**	**0.056**	**0.056**	**0.053**	**0.053**	**0.049**
	0.80	**0.068**	**0.062**	0.069	**0.056**	0.065	**0.051**	0.059	**0.052**
	0.90	0.135	0.103	0.116	0.084	0.117	0.083	0.077	**0.054**
	0.95	0.290	0.192	0.240	0.155	0.198	0.127	0.127	0.077
0.30	0.70	**0.054**	**0.041**	**0.055**	**0.044**	**0.056**	**0.049**	**0.054**	**0.047**
	0.80	0.086	**0.057**	0.068	**0.049**	0.073	**0.051**	**0.050**	**0.047**
	0.90	0.154	0.093	0.127	0.073	0.096	**0.058**	0.061	**0.048**
	0.95	0.260	0.145	0.188	0.110	0.139	0.076	0.118	0.060

Note: The nominal level of the test is 0.05. In bold the values inside the asymptotic acceptance interval at 95%.

Table 10: Relative frequency of zero break, a single break an more than one break for the GARCH(1,1) model generated with a break in the middle on the sample size. The values σ_1^2 and σ_2^2 indicate the unconditional variance in the two regimes identified by the break. The entries in brackets are the relative frequencies of exact location (inside an interval of length 40) of "at least one" of the identified breaks.

GARCH(1,1)	T=1000						T=2000						T=4000					
	HAC			**BOOT**			**HAC**			**BOOT**			**HAC**			**BOOT**		
	0	1	>1	0	1	>1	0	1	>1	0	1	>1	0	1	>1	0	1	>1
$\alpha_1 = 0.1$ $\beta_1 = 0.60$																		
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.0$	0.00	0.89	0.11	0.00	0.92	0.08	0.00	0.89	0.11	0.00	0.91	0.09	0.00	0.90	0.10	0.00	0.93	0.07
		[0.86]	[0.10]		[0.88]	[0.08]		[0.86]	[0.10]		[0.88]	[0.08]		[0.87]	[0.09]		[0.89]	[0.06]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.5$	0.00	0.87	0.13	0.00	0.90	0.10	0.00	0.88	0.12	0.00	0.91	0.09	0.00	0.89	0.11	0.00	0.92	0.08
		[0.86]	[0.12]		[0.89]	[0.09]		[0.87]	[0.11]		[0.90]	[0.09]		[0.88]	[0.10]		[0.91]	[0.08]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 3.0$	0.00	0.86	0.14	0.00	0.91	0.09	0.00	0.87	0.13	0.00	0.91	0.09	0.00	0.90	0.10	0.00	0.93	0.07
		[0.86]	[0.14]		[0.91]	[0.09]		[0.87]	[0.12]		[0.90]	[0.09]		[0.90]	[0.09]		[0.92]	[0.07]
$\alpha_1 = 0.1$ $\beta_1 = 0.70$																		
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.0$	0.01	0.86	0.13	0.01	0.90	0.09	0.00	0.88	0.12	0.00	0.91	0.09	0.00	0.89	0.11	0.00	0.92	0.08
		[0.81]	[0.12]		[0.84]	[0.09]		[0.81]	[0.11]		[0.84]	[0.08]		[0.84]	[0.9]		[0.87]	[0.07]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.5$	0.00	0.85	0.15	0.00	0.89	0.11	0.00	0.86	0.14	0.00	0.90	0.10	0.00	0.88	0.12	0.00	0.91	0.09
		[0.83]	[0.15]		[0.87]	[0.11]		[0.85]	[0.13]		[0.88]	[0.10]		[0.86]	[0.12]		[0.89]	[0.09]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 3.0$	0.00	0.84	0.16	0.00	0.89	0.11	0.00	0.86	0.14	0.00	0.91	0.09	0.00	0.88	0.12	0.00	0.92	0.08
		[0.83]	[0.15]		[0.88]	[0.11]		[0.85]	[0.13]		[0.90]	[0.09]		[0.88]	[0.11]		[0.92]	[0.07]
$\alpha_1 = 0.1$ $\beta_1 = 0.80$																		
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.0$	0.03	0.77	0.20	0.06	0.82	0.12	0.01	0.79	0.20	0.01	0.88	0.11	0.00	0.82	0.18	0.00	0.90	0.10
		[0.65]	[0.17]		[0.70]	[0.11]		[0.66]	[0.16]		[0.72]	[0.10]		[0.70]	[0.15]		[0.76]	[0.09]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.5$	0.00	0.76	0.24	0.01	0.85	0.14	0.00	0.79	0.21	0.00	0.87	0.13	0.00	0.83	0.17	0.00	0.90	0.10
		[0.71]	[0.21]		[0.78]	[0.13]		[0.73]	[0.19]		[0.80]	[0.12]		[0.78]	[0.15]		[0.83]	[0.09]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 3.0$	0.00	0.74	0.26	0.01	0.83	0.17	0.00	0.78	0.21	0.00	0.87	0.13	0.00	0.83	0.17	0.00	0.89	0.11
		[0.72]	[0.24]		[0.80]	[0.15]		[0.76]	[0.20]		[0.84]	[0.11]		[0.79]	[0.16]		[0.85]	[0.10]
$\alpha_1 = 0.1$ $\beta_1 = 0.85$																		
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.0$	0.13	0.54	0.33	0.19	0.62	0.19	0.05	0.61	0.34	0.06	0.75	0.19	0.01	0.70	0.29	0.01	0.80	0.19
		[0.41]	[0.28]		[0.45]	[0.16]		[0.42]	[0.24]		[0.51]	[0.13]		[0.47]	[0.18]		[0.52]	[0.12]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.5$	0.04	0.58	0.38	0.07	0.68	0.25	0.01	0.64	0.35	0.01	0.76	0.23	0.00	0.71	0.29	0.00	0.81	0.19
		[0.48]	[0.33]		[0.55]	[0.21]		[0.50]	[0.28]		[0.59]	[0.17]		[0.56]	[0.23]		[0.64]	[0.15]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 3.0$	0.01	0.59	0.40	0.03	0.70	0.27	0.01	0.61	0.38	0.01	0.75	0.24	0.00	0.70	0.30	0.00	0.80	0.20
		[0.52]	[0.36]		[0.61]	[0.25]		[0.52]	[0.31]		[0.64]	[0.20]		[0.60]	[0.25]		[0.69]	[0.16]

Table 11: Relative frequency of zero break, a single break an more than one break for the log-normal stochastic volatility model generated with a break in the middle on the sample size. The values σ_1^2 and σ_2^2 indicate the unconditional variance in the two regimes identified by the break. The entries in brackets are the relative frequencies of exact location (inside an interval of length 40) of "at least one" of the identified breaks.

Log-Normal Stochastic Volatility	T=1000						T=2000						T=4000					
	HAC			BOOT			HAC			BOOT			HAC			BOOT		
	0	1	>1	0	1	>1	0	1	>1	0	1	>1	0	1	>1	0	1	>1
$\phi_1 = 0.7$ $\sigma_v = 0.3$																		
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.0$	0.00	0.91 [0.91]	0.09 [0.09]	0.00	0.92 [0.92]	0.08 [0.08]	0.00	0.90 [0.89]	0.10 [0.09]	0.00	0.91 [0.91]	0.09 [0.08]	0.00	0.91 [0.90]	0.09 [0.09]	0.00	0.91 [0.91]	0.09 [0.08]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.5$	0.00	0.91 [0.90]	0.09 [0.09]	0.00	0.92 [0.91]	0.08 [0.08]	0.00	0.90 [0.89]	0.10 [0.09]	0.00	0.92 [0.91]	0.08 [0.08]	0.00	0.91 [0.90]	0.09 [0.08]	0.00	0.92 [0.91]	0.08 [0.07]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 3.0$	0.00	0.91 [0.91]	0.09 [0.08]	0.00	0.92 [0.91]	0.08 [0.08]	0.00	0.89 [0.89]	0.11 [0.10]	0.00	0.90 [0.90]	0.10 [0.09]	0.00	0.90 [0.90]	0.10 [0.09]	0.00	0.91 [0.91]	0.09 [0.09]
$\phi_1 = 0.8$ $\sigma_v = 0.3$																		
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.0$	0.00	0.85 [0.84]	0.15 [0.14]	0.00	0.89 [0.88]	0.11 [0.11]	0.00	0.86 [0.85]	0.14 [0.13]	0.00	0.90 [0.88]	0.10 [0.10]	0.00	0.89 [0.88]	0.11 [0.09]	0.00	0.91 [0.89]	0.09 [0.08]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.5$	0.00	0.87 [0.85]	0.13 [0.12]	0.00	0.91 [0.89]	0.09 [0.08]	0.00	0.87 [0.86]	0.13 [0.12]	0.00	0.90 [0.88]	0.10 [0.10]	0.00	0.87 [0.84]	0.13 [0.12]	0.00	0.89 [0.86]	0.11 [0.10]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 3.0$	0.00	0.84 [0.84]	0.16 [0.15]	0.00	0.89 [0.89]	0.11 [0.10]	0.00	0.85 [0.84]	0.15 [0.14]	0.00	0.89 [0.88]	0.11 [0.10]	0.00	0.89 [0.87]	0.11 [0.09]	0.00	0.92 [0.89]	0.08 [0.07]
$\phi_1 = 0.9$ $\sigma_v = 0.3$																		
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.0$	0.13	0.70 [0.55]	0.17 [0.15]	0.21	0.70 [0.55]	0.10 [0.08]	0.02	0.83 [0.63]	0.15 [0.11]	0.03	0.87 [0.67]	0.10 [0.07]	0.00	0.87 [0.64]	0.13 [0.09]	0.00	0.92 [0.67]	0.08 [0.06]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.5$	0.03	0.77 [0.67]	0.20 [0.18]	0.06	0.81 [0.71]	0.13 [0.12]	0.00	0.83 [0.70]	0.17 [0.14]	0.00	0.88 [0.75]	0.12 [0.10]	0.00	0.83 [0.68]	0.17 [0.13]	0.00	0.89 [0.73]	0.11 [0.08]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 3.0$	0.01	0.77 [0.70]	0.22 [0.20]	0.02	0.82 [0.75]	0.16 [0.14]	0.00	0.80 [0.71]	0.20 [0.18]	0.00	0.87 [0.78]	0.13 [0.11]	0.00	0.84 [0.71]	0.16 [0.12]	0.00	0.89 [0.76]	0.11 [0.07]
$\phi_1 = 0.95$ $\sigma_v = 0.3$																		
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.0$	0.18	0.56 [0.38]	0.26 [0.20]	0.30	0.56 [0.38]	0.14 [0.11]	0.06	0.65 [0.39]	0.29 [0.20]	0.10	0.73 [0.44]	0.17 [0.12]	0.01	0.75 [0.45]	0.24 [0.14]	0.01	0.85 [0.50]	0.14 [0.08]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 2.5$	0.27	0.48 [0.31]	0.25 [0.18]	0.40	0.49 [0.32]	0.11 [0.08]	0.15	0.61 [0.34]	0.24 [0.15]	0.23	0.62 [0.35]	0.15 [0.09]	0.03	0.75 [0.39]	0.22 [0.11]	0.05	0.83 [0.43]	0.12 [0.06]
$\sigma_1^2 = 1.0$ - $\sigma_2^2 = 3.0$	0.18	0.54 [0.38]	0.28 [0.20]	0.29	0.56 [0.39]	0.15 [0.12]	0.07	0.63 [0.39]	0.30 [0.20]	0.11	0.72 [0.4]	0.17 [0.10]	0.01	0.75 [0.44]	0.24 [0.14]	0.01	0.84 [0.48]	0.15 [0.09]

Table 12: Volatility breaks dates for IPC Index - Mexico.

IPC Index - Mexico

HAC	BOOT
2000-01-03	2000-01-03
2000-06-06	-
2001-01-11	-
2002-11-27	-
2006-05-10	-
2008-09-12	-
2009-05-08	2009-07-23
2012-01-18	-
2016-05-20	2016-05-20

Note: On the left the breaks have been identified by the K_2 test with the HAC estimator of the long run variance; on the right with the proposed bootstrap bias correction. In both cases, the binary segmentation and the proposed pruning procedure have been used.

Table 13: Volatility breaks dates for CNX Nifty Index - India.

CNX Nifty Index - India

HAC	BOOT
2003-01-01	2003-01-01
2007-10-05	2007-10-05
2009-07-23	2009-07-23
2012-03-30	2012-03-30
2016-05-20	2016-05-20

Note: On the left the breaks have been identified by the K_2 test with the HAC estimator of the long run variance; on the right with the proposed bootstrap bias correction. In both cases, the binary segmentation and the proposed pruning procedure have been used.

2.9 Appendix

2.9.1 Brownian Motion and Brownian Bridge

Definition (Brownian motion) A stochastic process $\{W(t)\}_{t>0}$ is a Brownian Motion (or Wiener Process) if the following properties hold:

1. $W(0) = 0$.
2. For $t_1 < t_2$, $W(t_2) - W(t_1) \sim N(0, t_2 - t_1)$.
3. For any $t_1 < t_2 < \cdots < t_K$, the random variables $W(t_k) - W(t_{k-1})$ for $k = 2, 3 \cdots K$ are independent.
4. $W(t)$ is continuous in t with probability 1.

Remark 2 *For any $t_1, \cdots, t_K$, the random vector $(W(t_1), \cdots, W(t_K))$ has a multivariate Normal distribution with 0 mean and $Cov(W(t), W(s)) = min(t, s)$.*

Definition (Brownian Bridge) A stochastic process $\{W^*(t)\}_{t\in[0,1]}$ is a Brownian Bridge if the following properties hold:

1. $W^*(0) = W^*(1) = 0$.
2. For any $t_1, \cdots, t_K$, the random vector $(W^*(t_1), \cdots, W^*(t_K))$ has a multivariate Normal distribution with 0 mean.
3. $Cov(W^*(t), W^*(s)) = min(t, s) - st$.
4. $W^*(t)$ is continuous in t with probability 1.

Remark 3 *If $\{W(t)\}_{t>0}$ is a Brownian Motion, then $W^*(r) = W(r) - rW(t)$ is a Brownian Bridge. This property can be proven by showing that $W^*(r)$ satisfies the definition of Brownian Motion.*

2.9.2 Stationarity and mixing

Definition (Strict Stationarity) The stochastic process $\{Y_t\}_{t\in\mathbb{Z}}$ is **strictly stationary**, if for all $k \in \mathbb{N}$, $h \in \mathbb{Z}$ and $(t_1, \ldots, t_k) \in \mathbb{Z}^k$

$$(Y_{t-1}, Y_{t_2}, \ldots, Y_{t_k}) \stackrel{d}{=} (Y_{t_1+h}, Y_{t_2+h}, \ldots, Y_{t_k+h}) \tag{83}$$

where $\stackrel{d}{=}$ denotes equality in distribution.

Definition (Weak Stationarity) The stochastic process $\{Y_t\}_{t\in\mathbb{Z}}$ is **weakly stationary** or **covariance stationary** if, for alll $h, t \in \mathbb{Z}$:

$$E(Y_t) = \mu \tag{84}$$

$$Cov(Y_t, Y_{t+h}) = \gamma(h) \tag{85}$$

with $\gamma(0) < \infty$

Remark 4 *If The process $\{Y_t\}_{t\in\mathbb{Z}}$ is strictly stationary then it is also weakly stationary if and only if $Var(Y_t) < \infty$*

Definition (Mixing) Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a strictly stationary process defined on a probability space $(\Omega, \mathcal{F}, P)$. Denote by $\mathcal{F}^0_{-\infty}$ the $\sigma-$algebra generated by $(Y_s : s \leq 0)$ and by $\mathcal{F}^\infty_t$ the $\sigma-$algebra generated by $(Y_s : s \geq 0)$ and for $k \in \mathbb{N}$ let:

$$\alpha_k = \sup_{C\in\mathcal{F}^0_{-\infty}, D\in\mathcal{F}^\infty_t} |P(C\cap D) - P(C)P(D)| \tag{86}$$

$$\beta_k = \frac{1}{2} \sup \sum_{i=1}^{I} \sum_{j=1}^{J} |P(C_i \cap D_j) - P(C_i)P(D_j)| \tag{87}$$

where in the definition of β_k the supremum is taken over all pairs of finite partitions $\{C_1, \ldots, C_I\}$ and $\{D_1, \ldots, D_I\}$ of Ω such that $C_i \in \mathcal{F}^0_{-\infty}$ and $D_j \in \mathcal{F}^\infty_t$ for each i and j. The coefficients α_k and β_k are the $\alpha-$mixing coefficients and $\beta-$mixing coefficients.
$(Y_t)_{t\in\mathbb{Z}}$ is called **$\boldsymbol{\alpha}$-mixing** or (**strongly mixing**) if $\lim_{k\to\infty} \alpha_k = 0$.
$(Y_t)_{t\in\mathbb{Z}}$ is called **$\boldsymbol{\beta}$-mixing** or (**absolutely regular**) if $\lim_{k\to\infty} \beta_k = 0$.

$(Y_t)_{t\in\mathbb{Z}}$ is called **α-mixing with geometric rate** if α_k decay at exponential rate i.e. there are constants $\lambda \in (0,1)$ and c such that $\alpha_k \leq c\lambda^k$ for every k.
$(Y_t)_{t\in\mathbb{Z}}$ is called **β-mixing with geometric rate** if β_k decay at exponential rate i.e. there are constants $\lambda \in (0,1)$ and c such that $\beta_k \leq c\lambda^k$ for every k.

Remark 5 *Since $\alpha_k \leq \frac{1}{2}\beta_k$, β-mixing implies α-mixing*

3 Forecasting with GARCH models under structural breaks: an approach based on combinations across estimation windows

Abstract

This paper proposes some weighting schemes to average forecasts across different estimation windows to account for structural changes in the unconditional variance of a GARCH (1,1) model. Each combination is obtained by averaging forecasts generated by recursively increasing an initial estimation window of a fixed number of observations v. Three different choices of the combination weights are proposed. In the first scheme, the forecast combination is obtained by using equal weights to average the individual forecasts; the second weighting method simply assigns heavier weights to forecasts that use more recent information; the third is a trimmed version of the forecast combination with equal weights where a fixed fraction of forecasts with the worst performance are discarded. Simulation results show that forecast combinations with high values of v seem to be effective in accounting for structural breaks. Moreover, they are also able to perform better with respect to alternative schemes proposed in the literature. An application to real data confirms the simulation results.

Keywords: Forecast combinations, Structural breaks, GARCH models.

JEL Code: C530, C580, G170

3.1 Introduction

The Generalized Autoregressive Conditional Heteroskedastic model GARCH (p,q) has been found to be a particularly useful parametrization in modelling and forecasting financial time series. Numerous applications of GARCH models have already appeared in the statistical and econometric literature, since the seminal papers by Engle and Bollerslev (1986).
Many studies (see for example Andreou and Ghysels, 2009) agree that the existence of structural breaks in the volatility can be the cause of estimation problems and forecast failures. If structural breaks are present in the data generating process but they are not considered in the specification of the model, the analysis could be biased toward a spurious persistence.
The hypotheses that structural changes may misinterpret the persistence estimation in GARCH models has been first highlighted by Diebold (1986). Lamoureux and Lastrapes (1990) have confirmed the Diebold's idea by allowing for changing states of the constant term of the conditional variance of a GARCH(1,1) model.
Mikosch and Stărică (2004) have explained how changes in the unconditional variance can be the cause of two stylized facts observed in long log-return series: the long-range dependence in volatility and the integrated GARCH (IGARCH) effect. In the first case the sample autocorrelation functions of the absolute values of the log returns series and their squares are all positive, decay relatively fast at the first lags and tend to stabilize around a positive value for larger lags. Concomitantly, the periodograms blow up at frequencies near zero. In the second case, whereas for shorter samples the estimated parameters of the GARCH(1,1) model sum to values significantly different from one, for longer samples their sum becomes close to one. This motivated the introduction of the integrated GARCH(1,1) model, IGARCH(1,1) by Engle and Bollerslev (1986) as a possible generating process for returns. As a consequence, the fitted GARCH model often appears to be very close to an IGARCH model only because structural breaks in unconditional volatility are ignored.
Hillebrand (2005) has shown that if structural changes in the conditional volatility process are not taken into account in the model specification, they could cause a substantial overestimation of the autoregressive parameters of the conditional variance.
The presence of structural breaks is even more fateful in a forecasting context and it constitutes one of the primary reasons for forecast failures in practice.
In particular, West and Cho (1995) have shown that a GARCH (1,1) that allows for structural

breaks in the unconditional variance, could have better forecasting performances with respect to methods which do not take into account their presence. This result has been confirmed by Stărică *et al.* (2005) in the context of long-horizon forecasts of stock return volatility and by Rapach and Strauss (2008) for eight daily U.S. dollar exchange rate return series.

A common strategy to handle parameter instability, in the context of GARCH models, is to select a single estimation window in which only a fraction of the most recent observations is used to estimate the parameters and to generate the forecasts. However, while in the regression framework solutions have been proposed to derive an optimal estimation window size (Pesaran and Timmerman, 2007; Pesaran *et al.*, 2013; Giraitis *et al.*, 2013; Inoue *et al.*, 2017), in a GARCH context this is not an easy task. One possible approach is to identify the last structural break in the series and to use only observations over the post-break period. This procedure could be not reliable in many applications because it depends on the correct identification of the last break, which is unknown in terms of location and magnitude.

Another approach is to use an estimation window whose size is proportional to the number of observations in the sample, independently of the last break location (Rapach and Strauss, 2008). However the forecasting performance is sensitive to the choice of the observation window due to the bias-variance trade off. More precisely a relative long estimation window reduces the forecast error variance but increases its bias; on the other hand, a short estimation window produces an increase in the forecast error variance although the bias decreases.

In order to solve the problems arising with the choice of a single estimation window, it can be useful to consider forecast combinations generated by the same model but over different estimation windows. As highlighted by Pesaran and Pick (2011) in the case of random walks and for a linear regression model, this strategy could be superior to forecasts generated by a single estimation window. However, while this procedure is suitable in a regression context and has been widely used in the econometric literature, in the context of GARCH models some problems arise. They are related to the very high length of the involved series, to the more complex generation of forecasts and to the possible large estimation uncertainty.

The aim of this paper is to propose some alternative weighting schemes that, in the spirit of Pesaran and Timmerman (2007), average forecasts across estimation windows, in order to account for structural breaks in the unconditional variance of a GARCH (1,1) model. Common to all the proposed forecast combinations is that the individual forecasts are obtained by expanding the length of an initial estimation window backwards of a fixed number of ob-

servations. The length of the initial window is introduced to allow a convergent estimation of a GARCH model whereas the length of the expanding windows controls the number of individual forecasts entering into the combination. The proposed combinations are usually effective to account for structural breaks even in the common situation when their location and magnitude are not known a priori. At the same time, they ensure low computation costs, avoiding the estimation of thousands of GARCH models that could be obtained when all possible window sizes are used. The proposed schemes differ for the specification of the combination weights: in the first scheme, the forecast combination is obtained by using equal weights to average the individual forecasts; the second weighting method simply assigns heavier weights to forecasts that use more recent information; the third is a trimmed version of the forecast combination with equal weights where a fixed fraction of forecasts with the worst performances are discarded.
The paper is organized as follows. Section 2 introduces the problem of forecasting under structural breaks and briefly overviews the most used methods. Starting from the single estimation window approach, the use of forecasting combinations is addressed, highlighting their advantages. In Section 3, the proposed forecast combinations across estimation windows are illustrated and discussed. In section 4 a Monte Carlo experiment has been implemented. The aim is to evaluate how the performance of the proposed forecast combinations, with different values of the tuning parameter and different choices of the weights, are influenced by the location and the size of the break. Moreover, through the implemented experiment, the forecast combinations have been also compared to some alternative schemes proposed in the literature. The forecasting performances of all the considered forecast combinations have been compared in terms of a loss functions, through the model confidence set procedure, proposed by Hansen *et al.* (2011). In section 5 the effectiveness of the proposed forecast combinations is evaluated on the same data set used by Rapach and Strauss (2008). It consists of daily returns of the U.S. dollar against the currencies of Canada, Denmark, Germany, Japan, Norway, Switzerland and the daily returns corresponding to the U.S. trade-weighted exchange rate. Some final remarks close the paper.

3.2 Forecast methods

In the paper we will refer to a GARCH (1,1) model since it is a very parsimonious model and usually it is adequate to obtain good performances in terms of fitting and forecasting (Hansen and Lunde, 2005).

The canonical GARCH (1,1) model for the zero-mean series a_t can be expressed in the form:

$$a_t = h_t^{0.5}\epsilon_t \tag{88}$$

with

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 h_{t-1} \tag{89}$$

where $\{\epsilon_t\}$ is a sequence of *i.i.d.* random variables with zero mean and unit variance. Conditions on α_0, α_1 and β_1 need to be imposed for the previous equation to be well defined. In particular, $\alpha_0 > 0$ and $\alpha_1, \beta_1 \geq 0$ are imposed to ensure that the conditional variance h_t is positive. Moreover $\alpha_1 + \beta_1 < 1$ ensures that the process is stationary. For a GARCH (1,1) process, the unconditional variance is defined as $\sigma^2 = \alpha_0/(1 - \alpha_1 - \beta_1)$.

The parameters are estimated by using the Quasi Maximum Likelihood Estimation in which the likelihood corresponding to the assumed distribution of ϵ_t, is maximized under the previous assumptions.

In the case of stable GARCH(1,1) model, the one-step-ahead forecast at time $T+1$, obtained by using the observations from 1 to T is defined as

$$\widehat{h}_{T+1} = \widehat{\alpha_0} + \widehat{\alpha_1} a_T^2 + \widehat{\beta}_1 h_T \tag{90}$$

As it is well known, the presence of structural breaks can have serious consequences on the performance of the forecasts. To deal with this problem, many procedures have been suggested in the econometric literature.

One of the most popular approach, in the context of GARCH models, is to select a single estimation window to generate a single forecast. In this perspective, a possible approach is to identify the presence of potential breaks in the data, by means of the numerous tests proposed in the econometric literature, and to predict with a GARCH model estimated using only the data from the last break (Rapach and Strauss, 2008). However, this approach is not free from criticism. First of all, the detection of breaks could be imprecise especially in the context of financial time series in which the classical tests proposed in the literature could not distinguish between break points and extreme observations (Ross, 2013). Secondly, as pointed out by Pesaran and Timmerman (2007) in the regression framework, the forecasts generated by this scheme is likely to be biased and may not minimize the mean square forecast error even when the last break date is correctly detected. This suggests that the pre-break observations could be useful for forecasting even after the break. Moreover, when the last break is small and the variance parameter increases at the break point it could be advisable to use also a fraction of the pre-break observations. Finally, if the last detected break point is very close to the last observation, the sample used for the estimation of the parameters of the forecast model is relatively short causing a very large estimation uncertainty. This is a strong weakness in GARCH model in which the parameters estimations need quite a high number of observations to converge.

Alternatively it is possible to consider forecasting methods which implicitly take into account the presence of breaks. In these cases, it is not necessary to specify the number and the locations of the breaks but they are taken into account in the generation of forecasts by using a GARCH model estimated with a fixed sample size. The forecasting performance of these methods is obviously sensitive to the choice of the observation window which should be chosen in a way to balance the trade off between bias and variability. More specifically, using data from different data generating processes, may lead to biased parameter estimates and forecasts; such bias can accumulate leading to large mean square forecast errors. This bias could be reduced using only that data relevant to the present. On the contrary, in order to reduce the heterogeneity, it could be appropriate to reduce the sample but this could increase the variance of the parameter estimates. This increase in variance maps into the forecast errors and, again, it causes an increment of the mean square forecast error. Hence, a plausible estimation window should guarantee an accurate estimate of the GARCH parameters and a not extensively use of observations from different regimes (Clark and McCracken, 2009).

However, in this approach, the identification of the optimal estimation window size is not an easy task. In the context of regression, Pesaran and Timmermann (2007) propose some methods to select the window size in the case of multiple discrete breaks when the errors of the model are serially uncorrelated and the regressors are strictly exogenous; Pesaran *et al.* (2013) derive optimal weights under continuous and discrete breaks in the case of independent errors and exogenous regressors; Giraitis *et al.* (2013) propose to select a tuning parameter to downweight older data by using a cross-validation-based method in the case of models without regressors; Inoue *et al.* (2017) suggest to choose the optimal window size that minimizes the conditional mean square forecast error (MSFE).

For the GARCH models the selection of this single estimation window remains an open issue and in many empirical studies, it is arbitrarily determined. For example, Rapach and Strauss (2008) propose different specifications for this parameter; in particular, in their analysis, the estimation window size is fixed to one-half and one-quarter of the length of the in sample period.

To overcome the problem of the selection of a single estimation window, an alternative strategy can be used. It is based on combinations of forecasts generated from the same model but over different estimation windows.

3.2.1 Forecast combinations

Starting from the pioneeristic paper of Bates and Granger (1969), there has been a growing interest, in the econometric literature, for the combinations of forecasts obtained by estimating a number of alternative models over the same sample period. Clemen (1989) has provided a review highlighting that forecast accuracy can be substantially improved through the combination of multiple individual forecasts. Makridakis and Hibon (2000), by analysing 3003 time series, have concluded that the accuracy of forecast combinations have better performances on average with respect to each specific method which is present in the combination. The same conclusions have been confirmed in the papers by Stock and Watson (2001,2004) and Marcellino (2004) in the context of economic and financial variables. More recently, Timmerman (2006) has theoretically analysed the factors that determine the advantages in combing forecasts focusing, in particular, on model misspecification, instability and estimation error.

In a regression framework, Pesaran and Timmerman (2007) have proposed forecast combinations formed by averaging across forecasts generated by using all possible window size subject to a minimum length requirement. Based on the same idea, more complex forecasting schemes have been proposed (see, for example, Tian and Anderson, 2014 and Pesaran *et al.*, 2013).

The idea of forecast averaging over estimation windows has been fruitfully applied also

in macroeconomic forecasting, in particular in the context of vector autoregressive models with weakly exogenous regressors (Assenmacher-Wesche and Pesaran, 2008; Pesaran *et al.*, 2009), and in the context of GDP growth on the yield curve (Schrimpf and Wang, 2010).
Pesaran and Pick (2011) have discussed the theoretical advantages of using such combinations considering random walks with breaks in the drift and volatility and a linear regression model with a break in the slope parameter. They have shown that averaging forecasts over different estimation windows leads to a lower bias and root mean square forecast error than forecasts based on a single estimation window for all but the smallest breaks. Similar results are reported in Clark and McCracken (2009); they have highlighted that, in presence of structural breaks, averaging forecasts obtained by using all the observations in the sample and forecasts obtained by using a window can be useful for forecasting[9].
All the discussed approaches are feasible for linear regression models and moderate sample size. However, when dealing with GARCH models, they became not suitable because of the estimation of thousands models just to form a single combination forecast.
In the light of this result, Rapach *et al.* (2008), in the case of stock return volatility, have proposed the following forecast combinations which they have found to be especially effective for forecasting in presence of breaks.

Rapach and Strauss forecast combination.
This scheme, labelled as RS combined, is a combination which averages individual forecasts obtained with four different estimation windows. The first two have fixed sizes proportional to the number of observations in the sample; in particular they are equal to one-half and one-quarter of the size of the sample period; the third window is formed by observations from the final break point to the end of sample. In this case, the last break has been identified by using the test proposed by Sansó *et al.* (2004). The last estimation window uses all the observations in the sample.

Trimmed Rapach and Strauss forecast combination.
This scheme, denoted as Trimmed-RS combined, is the average of the individual forecasts after excluding the highest and lowest from the previous scheme.

[9] In this case forecasts from only two different windows are combined and so this procedure can be seen as a limited version of that proposed by Pesaran and Timmermann (2007)

Clark and McCracken forecast combination.

The last scheme, denoted as CM combined, is the average of two individual forecasts: the first uses all the observations in the sample and the second only the observations from 0.75T to T. This is in the spirit of Clark and McCracken (2009).

3.3 Forecast combinations across estimation windows in GARCH(1, 1) models

In this section we propose some alternative forecast combinations which use different estimation windows to account for structural breaks, in the context of GARCH models.

Let T be the number of observations in the sample and ω the minimum acceptable estimation window size. The forecast combination at time $T+1$, denoted with $\widehat{h}_{T+1}$, is obtained by averaging across forecasts generated by increasing recursively, of a fixed number v, the minimum estimation window ω. More precisely:

$$\widehat{h}_{T+1} = \sum_{\tau=0}^{k-1} c_\tau \widehat{h}_{T+1}^{[T-w-\tau v:T]} \tag{91}$$

where $\widehat{h}_{T+1}^{[T-w-\tau v:T]}$ is the one-step-ahead forecast obtained by using the observations from $(T-w-\tau v)$ to T; c_τ are combination weights and

$$k = \left\lceil \frac{T-\omega}{v} \right\rceil \tag{92}$$

being $\lceil x \rceil$ the smallest integer greater than or equal to x.

In the equation (91), the individual forecasts are generated by expanding the length of the estimation window backwards of v observations after reserving the most recent ω observations. Therefore, in the combination, the last ω observations are used in all the forecasts, whereas the observations at the beginning of the sample are used less.

The proposed forecast combination scheme depends on the parameters ω and v which have to be fixed and to the weights c_τ.

The estimation windows should not be smaller than a minimum length ω. This parameter should be set in such a way that allows the parameter estimation of the GARCH (1,1) model to converge. Hwang and Valls Pereira (2006) have shown that the estimates of the popular GARCH(1,1) model are significantly negatively biased in small samples and that in many cases converged estimates are not possible with Bollerslev's non-negativity conditions. Considering the size of biases and convergence errors, they have proposed at least 500 observations for GARCH(1,1) models.

The parameter v controls the number of observations which are added to the minimum estimation window ω and, as a consequence, to the number of individual forecasts which enters in the combination. The lower the value of v, the more individual forecasts enter in the combination. Moreover if the location of the last break is near to the end of the sample, the higher the value of v, the less the number of windows containing many pre-break observations are in the combination scheme. This could be advantageous since the forecasts generated by using many pre-break observations could be biased especially when the size of the breaks is high. Note that the logic behind this approach is similar to that proposed by Pesaran and Timmermann (2007) in a regression context. In this latter case, the parameter v is set equal to 1 since the number of observations is generally not so high and the linearity of the models ensures the feasibility, in term of computational costs, of the generation of the many individual forecasts involved. As previously pointed out, in the case of GARCH models, the choice of $v = 1$ is unrealistic; in this context, a value for v should guarantee the effectiveness of the forecast combination in accounting for possible structural breaks in the series and, at the same time, it should ensure not too high computational costs[10].

The selection of an optimal value of v is an open issue. However, it is possible to select it by simulations, searching, for example, the value which minimizes a fixed loss function. This is the approach which will be used in this paper.

As far as the vector of the weights $c_0, c_1, \ldots, c_{k-1}$ is concerned, as it is usual in the literature of forecast combination, we will impose that they satisfy the constrains:

$$c_\tau \geq 0 \quad \tau = 0, 1, \ldots, k-1 \qquad \text{and} \qquad \sum_{\tau=0}^{k-1} c_\tau = 1 \tag{93}$$

These conditions ensure that the equation (91) is a convex linear combination of the individual

[10]For example, in the case of a GARCH(1,1) model, consider a time series with $T = 4000$ observations and fix $\omega = 800$. The one-step haed forecast at time $T + 1$ obtained by using a value of $v = 50$, is generated by averaging 64 individual forecasts.

forecasts. Moreover we assume that the weights are such that

$$c_0 \geq c_1 \geq \ldots \geq c_{k-1} \tag{94}$$

being c_τ the weight associated with the forecast obtained by using the observations from $(T - w - \tau v)$ to T. This assumption ensures that the forecasts obtained by the most recent observations have higher weights than those further away from the forecasting origin. Although all the sequences satisfying the previous assumptions can be used in the equation (91), some trivial choices have been proposed.

Mean window forecast combination with equal weights.
In this scheme, the forecast combination is obtained by using equal weights to average the individual forecasts. More precisely, the weights are defined as:

$$c_\tau = \frac{1}{k} \qquad \tau = 0, 1, \ldots, k-1 \tag{95}$$

and, as a consequence, the proposed forecast combination (91) is:

$$\widehat{h}_{T+1} = \frac{1}{k} \sum_{\tau=0}^{k-1} \widehat{h}_{T+1}^{[T-w-\tau v:T]} \tag{96}$$

Many researches (see for example Pesaran and Timmermann, 2007) have highlighted the advantages of the equally weighted forecast combination; this scheme is easy to compute and often has performances as good as more complicated schemes, also when there is uncertainty about the presence of structural breaks in the data.

Mean window forecast combination with location weights.
In this scheme, heavier weights are assigned to forecasts that use more recent information. As in Tian and Anderson (2014), we have proposed the following linear function of the location of time τ in the full sample:

$$c_\tau = \frac{k-\tau}{\sum_{\tau=0}^{k-1}(k-\tau)} = \frac{2(k-\tau)}{k(k+1)} \qquad \tau = 0, 1, \ldots, k-1 \tag{97}$$

As a consequence, the proposed forecast combination (91) is:

$$\widehat{h}_{T+1} = \frac{1}{\sum_{\tau=0}^{k-1}(k-\tau)} \sum_{\tau=0}^{k-1}(k-\tau)\widehat{h}_{T+1}^{[T-w-\tau v:T]} \tag{98}$$

The use of these weights implies that heavier weights are placed on the forecasts which are based on more recent parts of the sample.

Trimmed mean window forecast combination.

In this scheme, rather than combining the full set of forecasts, a fraction of those with the worst performance is discarded. In particular a fraction of the highest and lowest individual forecasts is excluded from the analysis. The remaining individual forecasts are then combined by using an equal weighting scheme. This approach is in the spirit of Armstrong (1989) and has been used by Rapach *et al.* (2008) in the context of forecasting stock return volatility. This weighting scheme could be useful since it is less sensitive to possible implausible forecasts.

3.4 Monte Carlo experiment

The aim of this simulation experiment is twofold. First our objective is to evaluate the effectiveness of the proposed procedures in accounting for structural breaks and the effects of the choice of the tuning parameter v. The results of this check will be used to identify an empirical value v^* of v such that the corresponding forecast combinations have better forecasting performances compared to other alternative ones. The second aim is to compare the proposed forecast combinations in correspondence of v^* with those proposed by Rapach and Strauss (2008) and Rapach *et al.* (2008).

As data generating process we have used a GARCH (1,1) with a single break point and, as a consequence, two different regimes are present:

$$a_t = h_t^{0.5}\epsilon_t \tag{99}$$

$$\begin{aligned} h_t = & \left[\alpha_{01} + \alpha_{11}a_{t-1}^2 + \beta_{11}h_{t-1}\right] I_{(t\leq\tau)}(t) \\ & + \left[\alpha_{02} + \alpha_{12}a_{t-1}^2 + \beta_{12}h_{t-1}\right] I_{(t>\tau)}(t) \end{aligned} \tag{100}$$

where $(\alpha_{01}, \alpha_{11}, \beta_{11})$ and $(\alpha_{02}, \alpha_{12}, \beta_{12})$ are the parameters of the GARCH(1,1) model respectively in the first and in the second regime, τ is the location of the break point and $I_A(x)$ is the indicator function defined as:

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases} \tag{101}$$

Three different parameter specifications of the model (100) have been considered:

$$\text{M1}: \begin{cases} (\alpha_{11}, \beta_{11}) = (0.10, 0.80) & \alpha_{01} \text{ such that } \sigma_1^2 = 2 \\ (\alpha_{12}, \beta_{12}) = (0.05, 0.90) & \alpha_{02} \text{ such that } \sigma_2^2 = 3 \end{cases}$$

$$\text{M2}: \begin{cases} (\alpha_{11}, \beta_{11}) = (0.15, 0.70) & \alpha_{01} \text{ such that } \sigma_1^2 = 1 \\ (\alpha_{12}, \beta_{12}) = (0.10, 0.80) & \alpha_{02} \text{ such that } \sigma_2^2 = 2 \end{cases}$$

$$\text{M3}: \begin{cases} (\alpha_{11}, \beta_{11}) = (0.15, 0.70) & \alpha_{01} \text{ such that } \sigma_1^2 = 1 \\ (\alpha_{12}, \beta_{12}) = (0.05, 0.90) & \alpha_{02} \text{ such that } \sigma_2^2 = 3 \end{cases}$$

where σ_1^2 and σ_2^2 are, respectively, the unconditional variances in the first and in the second regime.

Concerning the break, five different locations τ have been considered; they have been fixed as function of the time series length T. In particular:

$$\tau = \psi T \quad \text{with } \psi \in \{0.50, 0.60, 0.70, 0.80, 0.90\}$$

Therefore, the simulation experiment consists of fifteen different specifications of the model (100): three arising from the parameter specifications and five from the different locations of the break. Moreover two different series lengths $T \in \{3000, 4000\}$ have been considered. For each model specification and for both the sample sizes, 500 time series have been generated.

In order to evaluate the forecasting performances of the proposed forecast combinations for each model specification, an initial sub-sample, composed by the data from $t = 1$ to $t = R$, is used to estimate the model and the 1-step ahead out-of-sample forecast is produced. The sample is increased by one, the model is re-estimated using data from $t = 1$ to $t = R + 1$ and 1-step ahead forecast is produced. The procedure continues until the end of the available out-of-sample period. In the following, R has been fixed so that the number of out of sample observations is 300.

Common to all the considered forecast combinations is the specification of the minimum acceptable estimation window size ω. This parameter has been fixed equal to 800 in order to guarantee a convergent estimation of the GARCH(1,1) parameters.

For all the fifteen model specifications, for both sample sizes and for each of the 500 simulated series, the forecast combinations have been compared in terms of their predictive performances by using the model confidence set (Hansen *et al.*, 2011).

As pointed out in the Appendix 3.8, this procedure is able to construct a set of combinations which exhibit the same predictive ability, at a given level of confidence, in terms of a predefined loss function. The choice of this function is arbitrary and depends on the nature of the competitors. However, as detailed by Patton (2011), when comparing conditional variance forecasts, the use of imperfect, but conditionally unbiased, volatility proxy can lead to objectionable results. In this context, where squared returns are used as proxy of the realized volatility, the most used loss function is the QLIKE function defined as:

$$\text{QLIKE}_t = \frac{\tilde{h}_t}{\widehat{h}_t} - \log\left(\frac{\tilde{h}_t}{\widehat{h}_t}\right) - 1 \tag{102}$$

where $\tilde{h}_t$ is some realized volatility measure and $\hat{h}_t$ is the punctual volatility forecast. It yields a ranking of volatility forecasts that is robust to noise in the proxy; moreover it is able to better discriminate among models and it is less affected by the most extreme observations in the sample. Once the analysis has been performed for all the 500 simulated series, the relative frequencies a given forecast combination enters in the Model Confidence Set (MCS) have been determined.

3.4.1 Evaluation of the proposed forecast combinations

In order to evaluate the effect of different values of the tuning parameter v on the proposed forecast combinations, the following different values of v

$$v \in \{50, 100, 200, 300, 400, 500, 600, 700, 800, 900\}$$

have been considered. This range of values allows to evaluate how the proposed forecast combinations vary with this parameter. For each value of v, the proposed forecast combinations with equal weights, defined in (96) and with location weights defined in (98) have been considered. Moreover, when it is reasonable, a trimmed version, with a fraction of trimmed forecasts equal to 0.2, has been also included in the analysis.

Moreover, to assess the effectiveness of the proposed combinations, a benchmark forecasting method has been introduced in the analysis. As a natural benchmark, a GARCH (1,1) model estimated by using an expanding window method has been considered. This method uses all the available observations and, as a consequence, it ignores the presence of structural breaks. This choice is optimal in situations with no breaks and it is appropriate for forecasting when the data is generated by a stable model.

As previously pointed out, the model confidence set has been employed and it has been determined the relative frequency a given forecast combination enters in the set of superior model. Two values of the confidence levels, $\alpha = (0.10, 0.25)$, for both the statistics used in the MCS procedure (see Appendix 3.8 for details) have been considered.

Tables 14, 15 and 16 report the relative frequencies for the parameter specification M1, M2 and M3, respectively, for $T = 3000$; the same parameter specifications for $T = 4000$ are reported in Tables 17, 18 and 19.

From Table 14, it is evident that when the size of break is not so high (specification M1) the proposed combinations have, in general, good predictive performances for all the values of the tuning parameter v.

The relative frequencies of the proposed forecast combinations are higher than those of the expanding method, which does not take into account the presence of the break. This behaviour is more evident in correspondence of a confidence level equal to 0.25 and when the location of the break approaches the end of the sample. With reference to the three specifications of the combination weights, the combinations based on location weights seem to have better performances with respect to the other two weighting schemes, independently from the break locations. The tuning parameter v seems not so relevant when the break is far from the end of the sample, $\tau = (0.5T, 0.6T)$; in the other cases, the combinations with high values v and with the same weighting schemes, have relative frequencies significantly higher than those with small values of v. In particular the combination with location weights and $v = 800$ seems to have the best performances in all the cases.

In the case of specification M2 reported in Table 15, with respect to the previous case, the expanding method has performances even worse than the proposed combinations. Moreover, except for the case in which the break is located in the middle of the sample, the relative frequencies it enters in the set of superior model are very low and they decrease as the break location approaches the end of the sample. With respect to the three weighting schemes, it is more evident that the forecast combinations with location weights in general outperform those with the equal weights and the trimmed ones. Finally, regarding the tuning parameter v, the Table shows that, again, high values of v improve the forecasting performance of the considered combinations.

When considering the M3 specification, reported in Table 16, it is also more evident that not accounting for the break can have dramatic consequences in forecasting. Indeed, the behaviour of the expanding window method is very similar to that observed in the previous Table but the values of the relative frequencies are significantly reduced, especially when the confidence level is $\alpha = 0.25$. Moreover, it is confirmed that the forecast combinations with location weights and high values of the tuning parameter, especially $v = 800$, have better performances with respect all the other schemes.

In Tables 17, 18 and 19 the relative frequencies for the parameter specification M1, M2 and M3, respectively, for $T = 4000$ are reported. For all the model specifications, the results support the effectiveness of the proposed combinations with respect to the expanding procedure. Indeed, the expanding method enters in the set of superior model with very low relative frequencies if the comparison is made with the corresponding results obtained with $T = 3000$. The general behaviour of the proposed forecast combinations is confirmed: also in this case, for all the model specifications, the location schemes associated with a high values of the tuning parameter v seem to have better performances, especially when the break is located at the end of the sample size. The relative frequencies corresponding to these combinations are all very large especially when the confidence level is $\alpha = 0.25$.

3.4.2 A comparison with alternative forecast combinations

The proposed forecast combinations have been compared with some alternative combination schemes proposed in the econometric literature.

In particular the forecast combinations with equal weights and location weights associated with a tuning parameter equal to 800 (labelled, respectively Mean Wind 800 and Mean Wind 800 Loc) have been considered; this choice is motivated by the good performances highlighted by these two schemes in the simulations employed in the previous section. Moreover the forecast combinations proposed by Rapach and Strauss (2008) and Rapach *et al.* (2008), reviewed in section (3.2.1), have been examined; they are the Rapach and Strauss forecast combination (RS Mean), its trimmed version (RS Trim) and the Clark and McCraken combination (CM combined). Also in this analysis, the procedure based on an expanding window has been considered as benchmark. Again, the model confidence set has been used and the relative frequency a given forecast combination enters in the set of superior model has been determined.

Tables 20, 21 and 22 report the relative frequencies for the parameter specification M1, M2 and M3 for $T = 3000$. From their comparison, it is evident that all the considered forecast combinations have better performances with respect to the expanding window procedure. The RS Mean trimmed seems to have the best performance, for all the parameter specifications M1, M2, M3, when the break location is far from the end of the sample $\tau = (0.5T, 0.6T)$. The relative frequencies associated with these forecast combinations decrease when τ increases, especially for M2 and M3 specification. The RM mean seems to have a similar behaviour of the corresponding trimmed version but its performances are lower for $\tau = (0.5T, 0.6T)$ and greater for $\tau = (0.7T, 0.8T, 0.9T)$.

The CM combined seems to have good performances for all the values of τ in the case of M1 specification; however, for the other parameter specifications M2 and M3, it seems to have good performances for $\tau = (0.7T, 0.8T)$, it gets worse for $\tau = (0.5T, 0.6T)$ reaching low relative frequencies for $\tau = 0.9T$.

As shown in the previous subsection, the Mean Wind 800 has, in general, worse performances than the Mean Wind 800 Loc which outperforms all the other forecast combinations, especially when the break location is near to the end of the sample, for all the parameter specifications M1, M2 and M3. Moreover, it behaves well also in all the other cases.

The same parameter specifications for $T = 4000$ are reported in Tables 23, 24 and 25. For all the model specifications, the results confirm the behaviour of the competing forecast combinations observed when $T = 3000$. As expected, when T increases, the MCS procedure is able to better discriminate among forecast combinations; in this case there is even more evidence that the Mean Wind 800 Loc outperforms all the other forecast combinations, especially when τ increases.

3.5 An application to real data

In order to evaluate the effectiveness of the proposed forecast combinations on real data, we have considered the same data set used by Rapach and Strauss (2008). It consists of daily returns of the U.S. dollar against the currencies of Canada, Denmark, Germany, Japan, Norway, Switzerland and the daily returns corresponding to the U.S. trade-weighted exchange rate. The series (Figure 8) cover the period from 1/2/1908 to 8/13/2005 [11].

Following West and Cho (1995) the unconditional and conditional mean of these eight series can be considered zero while Rapach and Strauss (2008) have given evidence for modelling them as GARCH(1, 1) processes. Moreover, they have applied the iterated cumulative sum of squares algorithm (Inclan and Tiao, 1994) on the test proposed by Sansó *et al.* (2004) to detect possible breaks in the unconditional variance of the series. The exact locations of the structural breaks are reported in Table 26. In particular, the algorithm selects a single structural break in the unconditional variance of returns for Germany; two structural breaks for Japan, Norway and Switzerland; three structural breaks for Canada and the U.S.; four structural breaks for the U.K. As pointed out by Rapach and Strauss, they appear to correspond to significant economic events, giving evidence that the proposed test strategy is able to identify correctly the breaks. Moreover the GARCH (1,1) parameter estimates, for the full sample and for each sub-sample identified by the structural breaks, highlight significant shifts in the intercept of the GARCH (1,1) model. These shifts cause significant changes in the unconditional variance across regimes justifying the relevance of structural breaks in the analysis of the considered series and the use of forecasting techniques which account for them.

In the out-of-sample exercise, 1-step-ahead forecasts have been considered; they are based on the subsample formed of the last 300 observations. As in the simulation experiment, we have compared the forecasts of volatility generated by the forecast combinations Mean Wind 800, Mean Wind 800 Loc, and those proposed by Rapach and Strauss (2008) and Rapach *et al.* (2008) defined in the previous section and labelled RS Mean, RS Trim and CM combined. Again, the model confidence set has been used to construct the set of combinations which exhibits the same predictive ability, in terms of QLIKE loss function. Both the statistics pro-

[11] The data can be freely downloaded from `http://sites.slu.edu/rapachde/home/research`

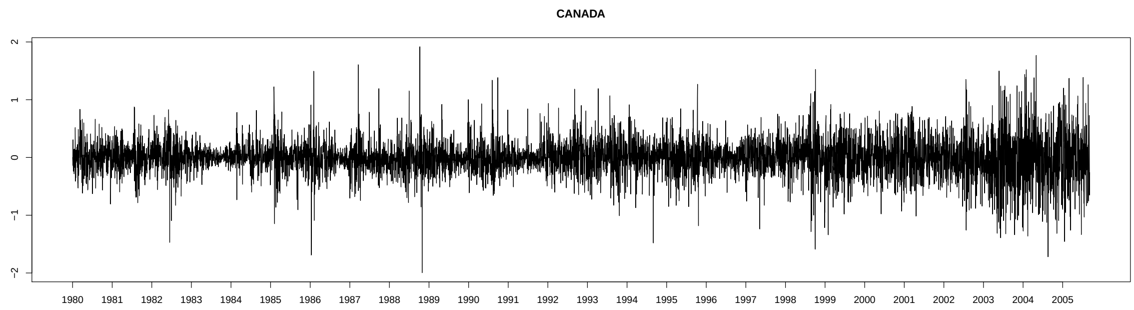
posed in the MCS procedure (see Appendix 3.8 for details) have been considered.
The results are reported in Table 27. With regard to Canada, the MCS reduces to a singleton containing only the RS Mean trimmed combination. Because the last surviving element is never eliminated, from Theorem 1 of Hansen *et al.* (2011), the associated probability approaches 1 as $T \to \infty$. In this case, the only information which could be deduced is that the loss associated to the RS Mean trimmed combination is the lowest. For all the other countries, there is strong evidence that accounting for structural breaks in the unconditional variance leads to out-of-sample forecasting gain: the expanding method is always out of the MCS. The composition of MCS changes across countries; this could be due to the different numbers, locations and magnitudes of the structural breaks in the series. However the Mean Wind 800 Loc forecast combination always enters in the set of superior model for both the statistics.

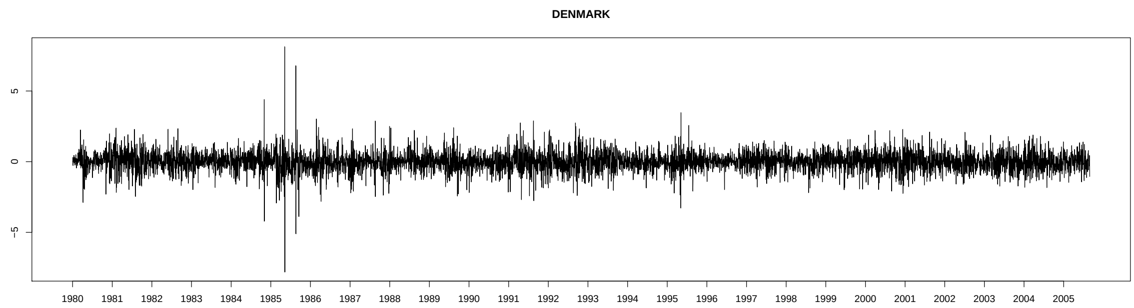
3.6 Concluding remarks

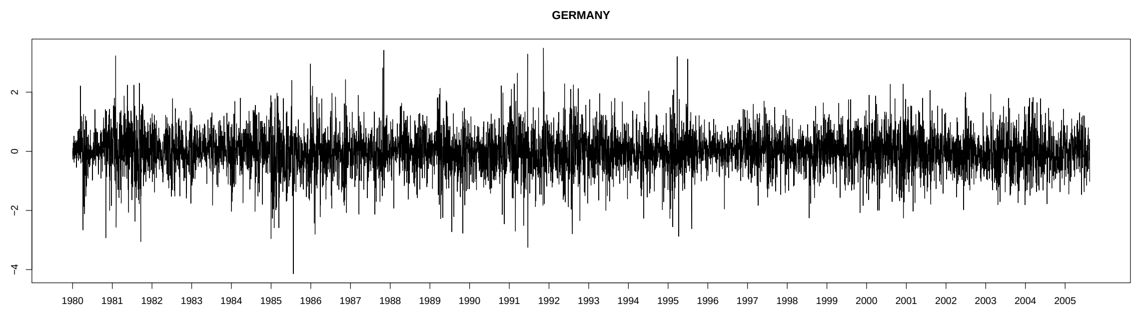
This paper has proposed three new combination schemes, which take into account for structural breaks in the unconditional variance of a GARCH(1, 1) model. They are obtained by averaging forecasts based on different estimation windows generated by recursively increasing an initial window of a fixed number of observations v. In the first scheme, the forecast combination is obtained by using equal weights to average the individual forecasts; the second weighting method simply assigns heavier weights to forecasts that use more recent information; the third is a trimmed version of the forecast combination with equal weights where a fixed fraction of forecasts with the worst performance are discarded.
The simulation experiment, carried out to evaluate the effect of the tuning parameter v on the performances of the proposed forecast combinations, has shown that the second weighting scheme based on an high value of v ($v = 800$) outperforms all the other forecast combinations with low or moderate values of v and with different weighting schemes. This result is particularly evident when the location of the structural break is near the end of the sample.
Through the Monte Carlo experiment, it has been verified that the forecast combination with location weights and with the value of v previously identified seems to better perform also with respect to some other forecast combinations proposed in the literature.
An empirical application, based on daily returns of the U.S. dollar against the currencies of

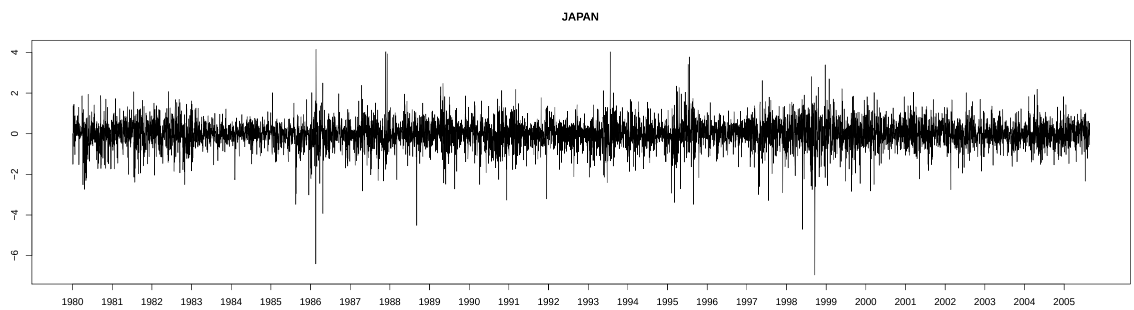
Canada, Denmark, Germany, Japan, Norway, Switzerland and the daily returns corresponding to the U.S. trade-weighted exchange rate, has confirmed these findings.
In any case, several different aspects should be further explored to get a better insight into the usage of the proposed forecasting combinations. From a computational point of view, a more extensive simulation experiment should be implemented in order to evaluate the proposed forecasting schemes when more than one break are present in the data generating process. From a statistical point of view, it should be assessed if different approaches, other then simulations, can be used to determine a plausible value for the tuning parameter.

3.7 Figures and Tables









continued on next page

continued from previous page

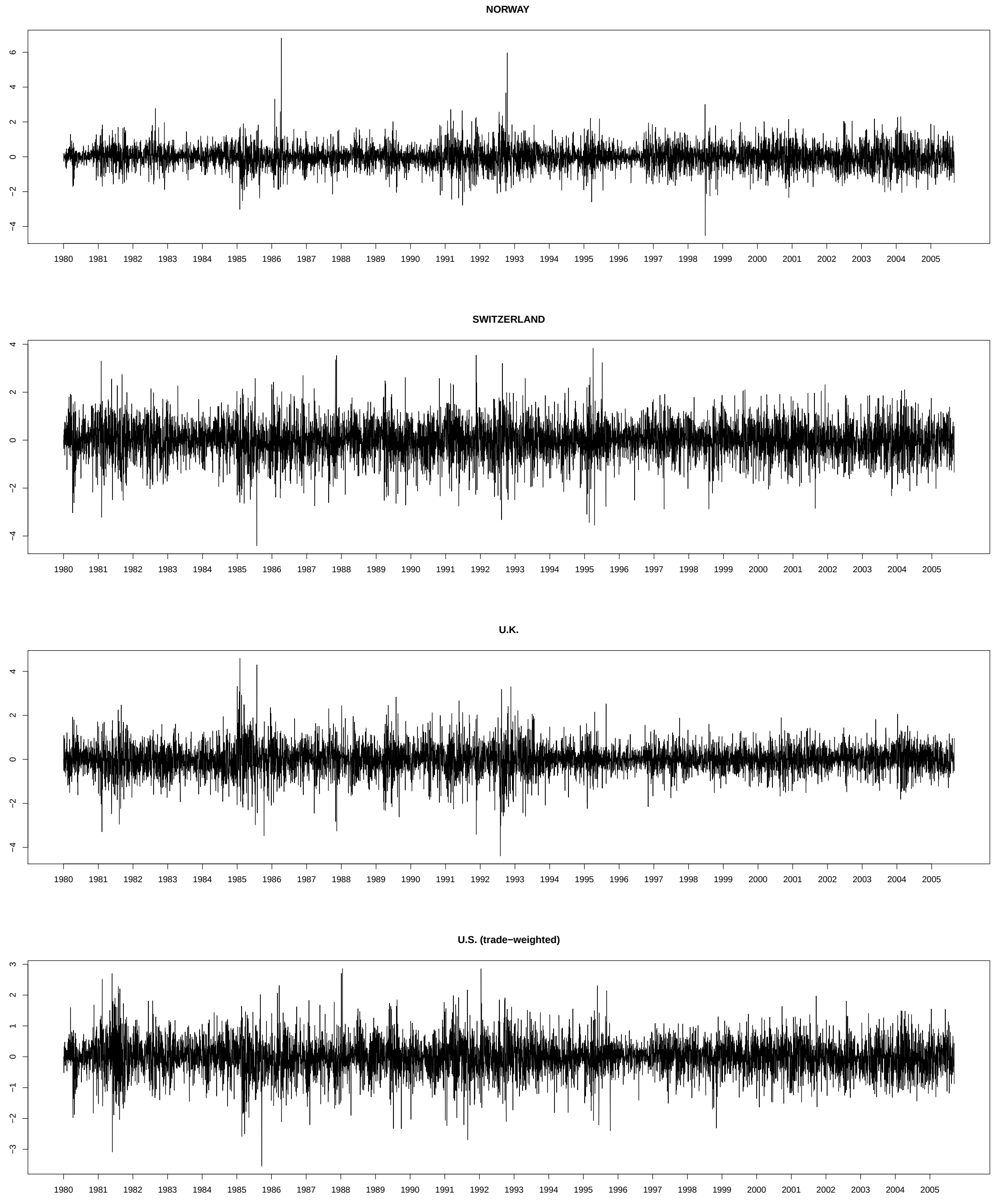


Figure 8: U.S. dollar exchange rate returns

Table 14: Evaluation of the proposed forecast combinations for parameter specification M1, T=3000.

MODEL M1	$\tau = 0.5T$				$\tau = 0.6T$				$\tau = 0.7T$				$\tau = 0.8T$				$\tau = 0.9T$			
	$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$	
T=3000	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ
Expanding Wind	0.96	0.88	0.77	0.75	0.92	0.87	0.73	0.71	0.93	0.84	0.80	0.67	0.83	0.72	0.60	0.49	0.93	0.83	0.66	0.67
Mean wind 50-Equ	0.95	0.93	0.82	0.78	0.92	0.92	0.79	0.79	0.93	0.86	0.85	0.74	0.88	0.78	0.68	0.58	0.89	0.78	0.71	0.66
Mean wind 50-Loc	0.97	0.99	0.92	0.94	0.94	0.96	0.89	0.92	0.94	0.92	0.87	0.82	0.91	0.88	0.78	0.74	0.91	0.84	0.76	0.75
Mean wind 50-Trim	0.95	0.92	0.84	0.79	0.92	0.89	0.74	0.71	0.92	0.85	0.84	0.68	0.86	0.74	0.66	0.51	0.89	0.76	0.72	0.63
Mean wind 100-Equ	0.96	0.93	0.83	0.76	0.93	0.92	0.81	0.78	0.94	0.88	0.85	0.75	0.86	0.79	0.66	0.57	0.90	0.77	0.72	0.67
Mean wind 100-Loc	0.98	0.99	0.92	0.95	0.95	0.96	0.87	0.91	0.95	0.94	0.88	0.83	0.91	0.89	0.79	0.75	0.92	0.85	0.77	0.76
Mean wind 100-Trim	0.97	0.92	0.83	0.79	0.93	0.87	0.75	0.75	0.93	0.86	0.84	0.72	0.84	0.74	0.62	0.52	0.88	0.76	0.72	0.63
Mean wind 200-Equ	0.96	0.92	0.80	0.78	0.93	0.91	0.82	0.80	0.94	0.87	0.85	0.75	0.88	0.83	0.69	0.56	0.91	0.78	0.73	0.65
Mean wind 200-Loc	0.99	0.99	0.92	0.94	0.95	0.97	0.88	0.94	0.95	0.95	0.89	0.82	0.94	0.93	0.79	0.78	0.92	0.86	0.82	0.77
Mean wind 200-Trim	0.97	0.92	0.81	0.79	0.93	0.90	0.78	0.76	0.92	0.87	0.84	0.70	0.86	0.77	0.66	0.51	0.88	0.77	0.73	0.67
Mean wind 300-Equ	0.94	0.91	0.81	0.81	0.94	0.93	0.79	0.79	0.94	0.88	0.87	0.76	0.89	0.83	0.70	0.61	0.90	0.80	0.73	0.67
Mean wind 300-Loc	0.99	0.99	0.94	0.96	0.96	0.99	0.90	0.92	0.97	0.95	0.91	0.87	0.95	0.92	0.83	0.81	0.92	0.88	0.79	0.77
Mean wind 300-Trim	0.92	0.90	0.81	0.79	0.93	0.91	0.79	0.77	0.94	0.87	0.84	0.70	0.86	0.74	0.66	0.53	0.88	0.79	0.71	0.67
Mean wind 400-Equ	0.97	0.93	0.82	0.75	0.94	0.91	0.81	0.84	0.92	0.89	0.82	0.77	0.90	0.85	0.71	0.62	0.88	0.78	0.72	0.67
Mean wind 400-Loc	0.98	0.99	0.93	0.94	0.98	0.99	0.90	0.94	0.96	0.95	0.90	0.89	0.98	0.96	0.84	0.84	0.90	0.88	0.81	0.77
Mean wind 400-Trim	0.97	0.95	0.83	0.77	0.92	0.90	0.76	0.80	0.93	0.88	0.81	0.71	0.87	0.78	0.71	0.54	0.86	0.77	0.66	0.68
Mean wind 500-Equ	0.97	0.95	0.84	0.79	0.91	0.92	0.83	0.81	0.94	0.89	0.85	0.81	0.96	0.87	0.71	0.63	0.87	0.81	0.69	0.65
Mean wind 500-Loc	0.99	0.99	0.95	0.95	0.97	0.99	0.92	0.92	0.98	0.98	0.92	0.91	1.00	0.97	0.87	0.88	0.93	0.92	0.78	0.83
Mean wind 500-Trim	0.97	0.93	0.84	0.79	0.93	0.92	0.79	0.76	0.94	0.88	0.85	0.77	0.94	0.82	0.71	0.59	0.90	0.79	0.67	0.66
Mean wind 600-Equ	0.97	0.95	0.80	0.80	0.93	0.91	0.80	0.77	0.94	0.91	0.86	0.79	0.90	0.82	0.69	0.60	0.90	0.84	0.73	0.70
Mean wind 600-Loc	0.97	0.98	0.91	0.94	0.97	0.99	0.89	0.93	0.98	0.98	0.91	0.92	0.97	0.98	0.89	0.88	0.96	0.91	0.86	0.84
Mean wind 700-Equ	0.94	0.95	0.81	0.82	0.92	0.95	0.81	0.84	0.95	0.95	0.90	0.83	0.96	0.90	0.78	0.75	0.92	0.88	0.75	0.71
Mean wind 700-Loc	0.98	0.98	0.94	0.95	0.97	0.99	0.92	0.97	1.00	1.00	0.95	0.96	1.00	0.99	0.96	0.98	0.97	0.97	0.91	0.91
Mean wind 800-Equ	0.98	0.96	0.81	0.85	0.93	0.92	0.81	0.79	0.94	0.95	0.87	0.82	0.95	0.90	0.77	0.75	0.90	0.88	0.73	0.69
Mean wind 800-Loc	0.97	0.99	0.92	0.95	0.98	1.00	0.94	0.96	0.99	1.00	0.97	0.99	1.00	0.99	0.98	0.99	0.97	0.97	0.90	0.93
Mean wind 900-Equ	0.96	0.94	0.81	0.82	0.94	0.95	0.84	0.78	0.94	0.93	0.89	0.84	0.95	0.88	0.75	0.67	0.90	0.86	0.72	0.66
Mean wind 900-Loc	0.98	0.96	0.90	0.91	0.96	1.00	0.92	0.95	0.98	1.00	0.94	0.97	0.99	0.99	0.95	0.96	0.96	0.97	0.88	0.91

Note: Entries denote the relative frequencies a given combination enters in the MCS at level α for the statistics R and SQ; τ indicates the location of the break.

Table 15: Evaluation of the proposed forecast combinations for parameter specification M2, T=3000.

MODEL M2	$\tau = 0.5T$				$\tau = 0.6T$				$\tau = 0.7T$				$\tau = 0.8T$				$\tau = 0.9T$			
	$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$	
T=3000	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ
Expanding Wind	0.87	0.85	0.74	0.64	0.83	0.71	0.60	0.41	0.71	0.48	0.36	0.23	0.62	0.33	0.35	0.17	0.53	0.43	0.26	0.25
Mean wind 50-Equ	0.92	0.87	0.83	0.75	0.89	0.79	0.72	0.56	0.76	0.58	0.48	0.33	0.75	0.40	0.39	0.23	0.60	0.49	0.32	0.29
Mean wind 50-Loc	0.99	0.97	0.93	0.92	0.98	0.93	0.88	0.85	0.88	0.68	0.58	0.49	0.82	0.64	0.54	0.41	0.70	0.63	0.50	0.47
Mean wind 50-Trim	0.93	0.88	0.82	0.73	0.86	0.71	0.66	0.49	0.74	0.49	0.43	0.26	0.67	0.34	0.36	0.19	0.59	0.45	0.29	0.28
Mean wind 100-Equ	0.93	0.87	0.83	0.78	0.89	0.78	0.73	0.58	0.76	0.58	0.47	0.35	0.76	0.39	0.42	0.24	0.62	0.49	0.30	0.27
Mean wind 100-Loc	0.99	0.97	0.94	0.94	0.98	0.94	0.89	0.86	0.90	0.69	0.60	0.49	0.85	0.63	0.56	0.41	0.70	0.65	0.50	0.48
Mean wind 100-Trim	0.93	0.88	0.83	0.72	0.88	0.73	0.66	0.48	0.73	0.53	0.43	0.27	0.69	0.33	0.37	0.19	0.58	0.45	0.30	0.25
Mean wind 200-Equ	0.94	0.87	0.83	0.79	0.90	0.80	0.76	0.58	0.84	0.59	0.48	0.34	0.76	0.38	0.44	0.22	0.61	0.48	0.33	0.25
Mean wind 200-Loc	0.98	0.97	0.93	0.94	0.98	0.96	0.89	0.86	0.90	0.73	0.66	0.51	0.86	0.72	0.58	0.44	0.73	0.65	0.48	0.50
Mean wind 200-Trim	0.94	0.88	0.83	0.73	0.87	0.72	0.70	0.53	0.73	0.50	0.44	0.29	0.68	0.33	0.36	0.22	0.62	0.45	0.34	0.24
Mean wind 300-Equ	0.95	0.90	0.83	0.79	0.90	0.79	0.73	0.60	0.81	0.61	0.53	0.38	0.75	0.41	0.45	0.24	0.64	0.50	0.37	0.34
Mean wind 300-Loc	0.98	0.97	0.93	0.95	0.98	0.96	0.90	0.86	0.92	0.75	0.69	0.58	0.87	0.68	0.66	0.50	0.79	0.73	0.55	0.54
Mean wind 300-Trim	0.94	0.89	0.80	0.79	0.89	0.76	0.74	0.57	0.74	0.58	0.45	0.35	0.71	0.36	0.38	0.22	0.60	0.50	0.36	0.29
Mean wind 400-Equ	0.95	0.91	0.84	0.82	0.91	0.82	0.79	0.61	0.84	0.60	0.53	0.36	0.76	0.49	0.47	0.22	0.62	0.50	0.33	0.27
Mean wind 400-Loc	0.99	0.99	0.93	0.94	0.96	0.96	0.92	0.91	0.90	0.83	0.74	0.61	0.87	0.75	0.66	0.58	0.76	0.73	0.54	0.56
Mean wind 400-Trim	0.97	0.91	0.86	0.81	0.91	0.75	0.74	0.55	0.75	0.54	0.45	0.31	0.72	0.40	0.39	0.20	0.63	0.48	0.32	0.27
Mean wind 500-Equ	0.95	0.89	0.86	0.82	0.90	0.82	0.79	0.62	0.84	0.65	0.60	0.41	0.76	0.50	0.49	0.29	0.58	0.49	0.31	0.28
Mean wind 500-Loc	0.98	0.99	0.95	0.96	0.95	0.95	0.89	0.91	0.93	0.88	0.78	0.71	0.90	0.81	0.73	0.61	0.77	0.74	0.58	0.62
Mean wind 500-Trim	0.96	0.91	0.86	0.80	0.92	0.81	0.76	0.59	0.86	0.59	0.59	0.38	0.77	0.44	0.44	0.27	0.62	0.50	0.32	0.28
Mean wind 600-Equ	0.90	0.91	0.82	0.80	0.89	0.81	0.74	0.61	0.83	0.66	0.58	0.40	0.77	0.52	0.47	0.24	0.64	0.55	0.39	0.33
Mean wind 600-Loc	0.99	0.99	0.92	0.97	0.97	0.97	0.89	0.90	0.93	0.87	0.78	0.70	0.93	0.86	0.68	0.71	0.84	0.82	0.66	0.68
Mean wind 700-Equ	0.97	0.93	0.85	0.84	0.93	0.84	0.81	0.72	0.88	0.76	0.67	0.53	0.83	0.62	0.55	0.37	0.70	0.65	0.42	0.39
Mean wind 700-Loc	0.99	1.00	0.93	0.96	0.99	1.00	0.97	0.98	1.00	1.00	0.96	0.95	0.97	0.98	0.91	0.89	0.96	0.97	0.81	0.82
Mean wind 800-Equ	0.91	0.91	0.81	0.79	0.94	0.90	0.76	0.69	0.86	0.72	0.66	0.56	0.81	0.61	0.53	0.34	0.68	0.59	0.43	0.35
Mean wind 800-Loc	0.99	1.00	0.92	0.97	0.99	1.00	0.98	0.98	1.00	1.00	0.98	0.99	1.00	1.00	0.97	0.97	0.98	0.98	0.90	0.89
Mean wind 900-Equ	0.90	0.91	0.77	0.76	0.93	0.86	0.76	0.63	0.89	0.74	0.64	0.53	0.85	0.60	0.55	0.34	0.69	0.60	0.46	0.35
Mean wind 900-Loc	0.95	0.97	0.82	0.91	0.97	0.99	0.93	0.93	1.00	1.00	0.94	0.94	0.95	0.96	0.87	0.86	0.92	0.95	0.80	0.86

Note: Entries denote the relative frequencies a given combination enters in the MCS at level α for the statistics R and SQ; τ indicates the location of the break.

Table 16: Evaluation of the proposed forecast combinations for parameter specification M3, T=3000.

MODEL M3	$\tau = 0.5T$				$\tau = 0.6T$				$\tau = 0.7T$				$\tau = 0.8T$				$\tau = 0.9T$			
	$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$	
T=3000	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ
Expanding Wind	0.82	0.77	0.62	0.58	0.80	0.66	0.54	0.35	0.63	0.40	0.26	0.20	0.41	0.23	0.15	0.06	0.10	0.10	0.04	0.03
Mean wind 50-Equ	0.89	0.82	0.73	0.64	0.84	0.72	0.66	0.50	0.69	0.44	0.34	0.26	0.68	0.34	0.31	0.10	0.23	0.18	0.08	0.07
Mean wind 50-Loc	0.97	0.97	0.92	0.94	0.90	0.90	0.83	0.75	0.80	0.57	0.46	0.36	0.82	0.60	0.48	0.38	0.43	0.40	0.31	0.32
Mean wind 50-Trim	0.84	0.78	0.68	0.59	0.80	0.66	0.51	0.39	0.66	0.41	0.29	0.20	0.56	0.29	0.26	0.10	0.15	0.15	0.06	0.07
Mean wind 100-Equ	0.89	0.82	0.74	0.63	0.83	0.74	0.68	0.53	0.68	0.44	0.36	0.27	0.68	0.34	0.32	0.11	0.22	0.18	0.07	0.07
Mean wind 100-Loc	0.98	0.98	0.91	0.96	0.90	0.94	0.84	0.76	0.81	0.63	0.48	0.38	0.82	0.61	0.51	0.39	0.43	0.41	0.34	0.32
Mean wind 100-Trim	0.84	0.79	0.69	0.60	0.82	0.66	0.53	0.43	0.65	0.41	0.33	0.21	0.58	0.29	0.26	0.10	0.18	0.16	0.05	0.07
Mean wind 200-Equ	0.90	0.84	0.75	0.66	0.86	0.74	0.67	0.54	0.72	0.45	0.37	0.29	0.70	0.36	0.34	0.11	0.21	0.18	0.08	0.07
Mean wind 200-Loc	0.98	0.98	0.94	0.96	0.91	0.93	0.87	0.79	0.84	0.66	0.52	0.43	0.85	0.66	0.57	0.45	0.49	0.45	0.37	0.32
Mean wind 200-Trim	0.85	0.81	0.70	0.60	0.82	0.68	0.56	0.45	0.65	0.41	0.30	0.25	0.61	0.32	0.27	0.11	0.21	0.17	0.08	0.06
Mean wind 300-Equ	0.89	0.86	0.73	0.66	0.84	0.75	0.67	0.56	0.73	0.50	0.42	0.29	0.73	0.38	0.35	0.15	0.23	0.18	0.10	0.10
Mean wind 300-Loc	0.98	0.99	0.95	0.96	0.94	0.95	0.90	0.80	0.84	0.73	0.59	0.48	0.87	0.73	0.63	0.54	0.48	0.51	0.37	0.34
Mean wind 300-Trim	0.86	0.85	0.72	0.64	0.83	0.70	0.61	0.51	0.68	0.43	0.36	0.26	0.69	0.35	0.33	0.14	0.20	0.18	0.10	0.13
Mean wind 400-Equ	0.93	0.89	0.80	0.68	0.85	0.78	0.71	0.56	0.75	0.55	0.41	0.35	0.76	0.41	0.36	0.18	0.20	0.19	0.05	0.05
Mean wind 400-Loc	0.98	0.99	0.97	0.97	0.93	0.94	0.86	0.82	0.87	0.76	0.65	0.52	0.86	0.80	0.65	0.55	0.48	0.50	0.37	0.38
Mean wind 400-Trim	0.88	0.83	0.72	0.64	0.85	0.72	0.67	0.50	0.72	0.47	0.35	0.25	0.71	0.35	0.31	0.14	0.20	0.19	0.10	0.07
Mean wind 500-Equ	0.95	0.89	0.81	0.76	0.86	0.81	0.75	0.56	0.76	0.55	0.48	0.35	0.76	0.43	0.37	0.21	0.19	0.17	0.06	0.07
Mean wind 500-Loc	0.99	0.99	0.97	0.98	0.94	0.95	0.87	0.81	0.91	0.79	0.70	0.60	0.90	0.89	0.69	0.66	0.46	0.52	0.30	0.39
Mean wind 500-Trim	0.94	0.89	0.81	0.69	0.82	0.77	0.68	0.49	0.73	0.50	0.44	0.32	0.75	0.44	0.38	0.17	0.19	0.17	0.07	0.04
Mean wind 600-Equ	0.89	0.87	0.79	0.78	0.87	0.73	0.65	0.58	0.79	0.58	0.45	0.37	0.77	0.40	0.38	0.18	0.21	0.18	0.08	0.08
Mean wind 600-Loc	0.98	0.99	0.94	0.98	0.96	0.96	0.88	0.81	0.91	0.88	0.73	0.70	0.90	0.88	0.69	0.62	0.56	0.58	0.46	0.47
Mean wind 700-Equ	0.87	0.94	0.77	0.76	0.91	0.88	0.75	0.70	0.84	0.66	0.55	0.45	0.83	0.55	0.47	0.33	0.27	0.26	0.14	0.13
Mean wind 700-Loc	0.99	1.00	0.92	0.99	0.99	0.99	0.95	0.96	0.98	0.98	0.97	0.95	1.00	1.00	0.97	0.98	0.90	0.90	0.75	0.77
Mean wind 800-Equ	0.88	0.87	0.74	0.69	0.89	0.88	0.76	0.71	0.82	0.64	0.58	0.43	0.79	0.48	0.42	0.24	0.22	0.22	0.09	0.09
Mean wind 800-Loc	0.97	0.99	0.87	0.93	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.97	0.98	0.90	0.92
Mean wind 900-Equ	0.87	0.86	0.74	0.67	0.89	0.86	0.80	0.63	0.86	0.65	0.55	0.43	0.77	0.43	0.42	0.22	0.20	0.21	0.08	0.09
Mean wind 900-Loc	0.93	0.95	0.85	0.90	1.00	0.99	0.95	0.95	0.97	0.99	0.95	0.95	0.99	1.00	0.78	0.87	0.68	0.80	0.47	0.52

Note: Entries denote the relative frequencies a given combination enters in the MCS at level α for the statistics R and SQ; τ indicates the location of the break.

Table 17: Evaluation of the proposed forecast combinations for parameter specification M1, T=4000.

MODEL M1	$\tau = 0.5T$				$\tau = 0.6T$				$\tau = 0.7T$				$\tau = 0.8T$				$\tau = 0.9T$			
T=4000	$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$	
	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ
Expanding Wind	0.89	0.89	0.74	0.77	0.93	0.93	0.77	0.78	0.93	0.84	0.75	0.57	0.89	0.79	0.70	0.49	0.79	0.66	0.59	0.44
Mean wind 50-Equ	0.94	0.91	0.80	0.82	0.95	0.96	0.85	0.89	0.93	0.89	0.80	0.73	0.88	0.83	0.70	0.54	0.81	0.68	0.61	0.52
Mean wind 50-Loc	0.99	0.99	0.91	0.96	0.99	0.99	0.91	0.96	0.96	0.93	0.87	0.87	0.93	0.92	0.79	0.70	0.88	0.81	0.70	0.57
Mean wind 50-Trim	0.94	0.92	0.82	0.81	0.95	0.95	0.82	0.83	0.92	0.89	0.81	0.63	0.89	0.79	0.70	0.52	0.82	0.64	0.58	0.45
Mean wind 100-Equ	0.93	0.90	0.82	0.83	0.95	0.96	0.88	0.88	0.93	0.89	0.83	0.72	0.88	0.83	0.70	0.55	0.83	0.69	0.59	0.48
Mean wind 100-Loc	0.99	0.99	0.91	0.94	0.99	0.99	0.93	0.96	0.96	0.94	0.87	0.87	0.94	0.92	0.80	0.75	0.90	0.85	0.71	0.60
Mean wind 100-Trim	0.94	0.93	0.81	0.82	0.95	0.96	0.83	0.83	0.91	0.88	0.80	0.63	0.89	0.79	0.70	0.50	0.83	0.63	0.57	0.44
Mean wind 200-Equ	0.95	0.92	0.82	0.81	0.97	0.96	0.87	0.88	0.94	0.91	0.83	0.70	0.90	0.83	0.71	0.56	0.85	0.71	0.58	0.46
Mean wind 200-Loc	0.98	0.99	0.91	0.94	0.99	0.99	0.92	0.97	0.96	0.96	0.91	0.89	0.96	0.93	0.84	0.75	0.90	0.87	0.74	0.64
Mean wind 200-Trim	0.93	0.94	0.80	0.81	0.94	0.96	0.84	0.83	0.92	0.86	0.78	0.63	0.91	0.80	0.71	0.51	0.82	0.63	0.56	0.44
Mean wind 300-Equ	0.95	0.90	0.81	0.81	0.98	0.97	0.85	0.87	0.94	0.91	0.84	0.73	0.90	0.86	0.74	0.60	0.81	0.69	0.61	0.47
Mean wind 300-Loc	0.97	0.99	0.91	0.95	1.00	0.99	0.93	0.98	0.96	0.94	0.90	0.90	0.94	0.94	0.86	0.82	0.88	0.89	0.72	0.64
Mean wind 300-Trim	0.94	0.94	0.80	0.80	0.94	0.96	0.84	0.84	0.92	0.87	0.78	0.67	0.89	0.79	0.72	0.51	0.84	0.65	0.61	0.45
Mean wind 400-Equ	0.93	0.91	0.79	0.81	0.97	0.96	0.88	0.90	0.93	0.89	0.82	0.73	0.89	0.85	0.70	0.60	0.82	0.70	0.59	0.44
Mean wind 400-Loc	0.97	0.99	0.91	0.95	1.00	0.99	0.93	0.96	0.96	0.96	0.90	0.90	0.95	0.95	0.85	0.81	0.91	0.88	0.78	0.73
Mean wind 400-Trim	0.94	0.94	0.83	0.81	0.96	0.97	0.85	0.85	0.91	0.88	0.80	0.69	0.89	0.81	0.73	0.55	0.83	0.68	0.60	0.45
Mean wind 500-Equ	0.95	0.93	0.84	0.84	0.96	0.98	0.83	0.89	0.95	0.92	0.86	0.74	0.91	0.87	0.73	0.66	0.85	0.73	0.64	0.50
Mean wind 500-Loc	0.97	0.99	0.93	0.96	0.99	0.99	0.90	0.98	0.96	0.97	0.93	0.92	0.96	0.97	0.91	0.86	0.95	0.92	0.80	0.75
Mean wind 500-Trim	0.95	0.93	0.86	0.87	0.92	0.96	0.84	0.84	0.92	0.89	0.81	0.64	0.90	0.84	0.69	0.61	0.85	0.70	0.63	0.52
Mean wind 600-Equ	0.95	0.93	0.82	0.80	0.96	0.98	0.86	0.89	0.95	0.93	0.88	0.76	0.91	0.89	0.77	0.65	0.84	0.75	0.66	0.52
Mean wind 600-Loc	0.96	0.99	0.92	0.95	0.99	0.98	0.91	0.96	0.98	0.97	0.95	0.94	0.96	0.98	0.89	0.89	0.92	0.91	0.79	0.80
Mean wind 700-Equ	0.94	0.92	0.81	0.83	0.98	0.97	0.83	0.90	0.93	0.92	0.84	0.72	0.92	0.88	0.79	0.66	0.83	0.74	0.64	0.51
Mean wind 700-Loc	0.97	0.99	0.90	0.95	1.00	0.99	0.91	0.97	0.97	0.97	0.95	0.93	0.96	0.98	0.90	0.91	0.92	0.96	0.78	0.81
Mean wind 800-Equ	0.95	0.97	0.84	0.85	0.98	0.98	0.85	0.94	0.95	0.93	0.87	0.79	0.92	0.90	0.80	0.69	0.83	0.82	0.67	0.55
Mean wind 800-Loc	0.98	0.98	0.93	0.96	1.00	1.00	0.94	0.99	1.00	1.00	0.98	1.00	1.00	1.00	0.98	0.99	0.96	0.99	0.94	0.97
Mean wind 900-Equ	0.93	0.91	0.79	0.82	0.97	0.96	0.84	0.89	0.94	0.91	0.83	0.72	0.93	0.90	0.83	0.73	0.83	0.77	0.67	0.55
Mean wind 900-Loc	0.97	0.99	0.90	0.94	0.98	0.99	0.85	0.95	0.98	0.99	0.96	0.96	0.99	0.99	0.93	0.96	0.93	0.94	0.81	0.82

Note: Entries denote the relative frequencies a given combination enters in the MCS at level α for the statistics R and SQ; τ indicates the location of the break.

Table 18: Evaluation of the proposed forecast combinations for parameter specification M2, T=4000.

MODEL M2	$\tau = 0.5T$				$\tau = 0.6T$				$\tau = 0.7T$				$\tau = 0.8T$				$\tau = 0.9T$			
	$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$	
T=4000	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ
Expanding Wind	0.91	0.91	0.72	0.70	0.83	0.76	0.62	0.50	0.83	0.53	0.40	0.24	0.41	0.22	0.16	0.07	0.37	0.19	0.10	0.08
Mean wind 50-Equ	0.95	0.95	0.83	0.77	0.90	0.84	0.76	0.65	0.91	0.67	0.59	0.38	0.58	0.30	0.27	0.13	0.46	0.24	0.17	0.13
Mean wind 50-Loc	0.96	0.99	0.92	0.95	0.95	0.97	0.86	0.91	0.94	0.89	0.80	0.65	0.62	0.46	0.42	0.25	0.55	0.42	0.31	0.26
Mean wind 50-Trim	0.94	0.94	0.80	0.73	0.91	0.78	0.69	0.57	0.87	0.59	0.43	0.29	0.50	0.25	0.20	0.12	0.45	0.23	0.15	0.09
Mean wind 100-Equ	0.95	0.96	0.83	0.76	0.92	0.84	0.76	0.64	0.92	0.68	0.59	0.38	0.58	0.30	0.27	0.14	0.46	0.24	0.17	0.11
Mean wind 100-Loc	0.95	0.99	0.92	0.96	0.94	0.98	0.88	0.92	0.94	0.92	0.82	0.68	0.62	0.44	0.46	0.26	0.59	0.43	0.31	0.26
Mean wind 100-Trim	0.94	0.94	0.81	0.75	0.90	0.79	0.70	0.57	0.86	0.59	0.43	0.29	0.52	0.26	0.21	0.13	0.44	0.22	0.16	0.09
Mean wind 200-Equ	0.95	0.97	0.82	0.76	0.90	0.85	0.79	0.66	0.92	0.73	0.62	0.41	0.57	0.33	0.27	0.15	0.48	0.24	0.19	0.14
Mean wind 200-Loc	0.96	0.99	0.91	0.94	0.96	0.99	0.89	0.93	0.94	0.95	0.81	0.70	0.63	0.50	0.47	0.26	0.61	0.45	0.36	0.28
Mean wind 200-Trim	0.95	0.95	0.79	0.75	0.90	0.81	0.74	0.59	0.88	0.60	0.48	0.31	0.52	0.26	0.21	0.13	0.46	0.23	0.17	0.10
Mean wind 300-Equ	0.95	0.96	0.82	0.77	0.91	0.83	0.76	0.69	0.91	0.72	0.64	0.40	0.58	0.32	0.28	0.15	0.48	0.23	0.19	0.13
Mean wind 300-Loc	0.96	0.99	0.91	0.95	0.96	1.00	0.87	0.94	0.97	0.95	0.86	0.76	0.64	0.52	0.48	0.28	0.63	0.49	0.37	0.25
Mean wind 300-Trim	0.93	0.95	0.80	0.74	0.90	0.82	0.74	0.62	0.89	0.59	0.48	0.35	0.53	0.27	0.23	0.13	0.46	0.24	0.17	0.10
Mean wind 400-Equ	0.95	0.96	0.83	0.81	0.93	0.86	0.78	0.67	0.92	0.76	0.68	0.46	0.57	0.35	0.33	0.17	0.48	0.25	0.19	0.14
Mean wind 400-Loc	0.97	0.99	0.91	0.96	0.97	1.00	0.94	0.97	0.98	0.95	0.88	0.76	0.67	0.58	0.51	0.34	0.64	0.50	0.38	0.28
Mean wind 400-Trim	0.94	0.96	0.84	0.79	0.92	0.86	0.77	0.62	0.90	0.67	0.57	0.38	0.57	0.29	0.23	0.14	0.46	0.26	0.20	0.13
Mean wind 500-Equ	0.95	0.97	0.83	0.78	0.90	0.86	0.79	0.68	0.94	0.74	0.64	0.46	0.59	0.34	0.34	0.17	0.51	0.24	0.23	0.12
Mean wind 500-Loc	0.97	0.99	0.91	0.94	0.97	1.00	0.89	0.95	0.96	0.95	0.87	0.78	0.68	0.61	0.54	0.38	0.66	0.59	0.42	0.31
Mean wind 500-Trim	0.95	0.97	0.83	0.79	0.92	0.85	0.77	0.67	0.90	0.64	0.54	0.36	0.54	0.28	0.23	0.13	0.46	0.24	0.18	0.11
Mean wind 600-Equ	0.96	0.95	0.82	0.78	0.92	0.90	0.81	0.73	0.94	0.73	0.68	0.49	0.63	0.38	0.37	0.19	0.51	0.28	0.22	0.13
Mean wind 600-Loc	0.97	0.99	0.91	0.95	0.97	1.00	0.95	0.97	0.97	0.96	0.88	0.78	0.75	0.64	0.54	0.42	0.73	0.63	0.47	0.41
Mean wind 700-Equ	0.94	0.95	0.83	0.78	0.93	0.87	0.79	0.72	0.93	0.74	0.58	0.42	0.62	0.38	0.35	0.19	0.51	0.24	0.26	0.16
Mean wind 700-Loc	0.96	0.99	0.88	0.94	0.99	1.00	0.97	0.99	0.95	0.95	0.87	0.79	0.74	0.67	0.60	0.48	0.73	0.67	0.54	0.42
Mean wind 800-Equ	0.95	0.99	0.86	0.89	0.94	0.92	0.82	0.75	0.93	0.80	0.72	0.53	0.63	0.43	0.41	0.27	0.54	0.32	0.28	0.18
Mean wind 800-Loc	0.98	1.00	0.95	0.98	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98
Mean wind 900-Equ	0.95	0.99	0.82	0.79	0.95	0.85	0.76	0.65	0.93	0.79	0.73	0.48	0.65	0.41	0.42	0.26	0.52	0.29	0.22	0.16
Mean wind 900-Loc	0.97	1.00	0.92	0.98	0.98	1.00	0.90	0.94	0.99	0.98	0.93	0.88	0.89	0.89	0.75	0.68	0.80	0.81	0.63	0.64

Note: Entries denote the relative frequencies a given combination enters in the MCS at level α for the statistics R and SQ; τ indicates the location of the break.

Table 19: Evaluation of the proposed forecast combinations for parameter specification M3, T=4000.

MODEL M3	$\tau = 0.5T$				$\tau = 0.6T$				$\tau = 0.7T$				$\tau = 0.8T$				$\tau = 0.9T$			
	$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$	
T=4000	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ
Expanding Wind	0.86	0.83	0.67	0.63	0.84	0.72	0.60	0.41	0.60	0.45	0.34	0.22	0.39	0.25	0.16	0.10	0.02	0.01	0.00	0.00
Mean wind 50-Equ	0.91	0.87	0.80	0.73	0.88	0.78	0.70	0.59	0.74	0.60	0.50	0.34	0.53	0.33	0.29	0.18	0.06	0.03	0.01	0.00
Mean wind 50-Loc	0.98	0.98	0.95	0.92	0.98	0.98	0.86	0.87	0.85	0.81	0.70	0.62	0.64	0.58	0.44	0.38	0.16	0.09	0.06	0.03
Mean wind 50-Trim	0.89	0.86	0.75	0.67	0.83	0.70	0.64	0.45	0.69	0.53	0.35	0.25	0.50	0.29	0.21	0.14	0.03	0.02	0.00	0.00
Mean wind 100-Equ	0.92	0.87	0.82	0.72	0.90	0.78	0.72	0.58	0.77	0.61	0.53	0.36	0.52	0.33	0.29	0.18	0.05	0.03	0.01	0.00
Mean wind 100-Loc	0.98	0.98	0.94	0.94	0.96	0.98	0.85	0.88	0.85	0.83	0.73	0.62	0.65	0.58	0.43	0.38	0.15	0.09	0.06	0.03
Mean wind 100-Trim	0.91	0.87	0.78	0.68	0.84	0.74	0.64	0.47	0.69	0.54	0.38	0.26	0.51	0.29	0.26	0.14	0.04	0.02	0.00	0.00
Mean wind 200-Equ	0.93	0.85	0.82	0.74	0.89	0.80	0.73	0.61	0.80	0.63	0.58	0.39	0.54	0.34	0.32	0.20	0.05	0.03	0.02	0.00
Mean wind 200-Loc	0.98	0.98	0.96	0.94	0.97	1.00	0.92	0.93	0.88	0.85	0.79	0.65	0.65	0.59	0.47	0.40	0.18	0.11	0.06	0.04
Mean wind 200-Trim	0.90	0.84	0.73	0.71	0.84	0.75	0.66	0.43	0.72	0.54	0.43	0.26	0.52	0.30	0.24	0.14	0.04	0.02	0.00	0.00
Mean wind 300-Equ	0.91	0.88	0.79	0.72	0.89	0.80	0.73	0.63	0.81	0.66	0.59	0.39	0.55	0.36	0.33	0.20	0.07	0.03	0.03	0.01
Mean wind 300-Loc	0.97	0.98	0.93	0.94	0.98	1.00	0.90	0.95	0.89	0.84	0.81	0.71	0.70	0.65	0.49	0.45	0.19	0.18	0.07	0.05
Mean wind 300-Trim	0.92	0.86	0.78	0.70	0.86	0.75	0.63	0.51	0.71	0.55	0.45	0.27	0.52	0.30	0.27	0.17	0.06	0.03	0.01	0.00
Mean wind 400-Equ	0.92	0.88	0.80	0.72	0.91	0.81	0.75	0.60	0.77	0.65	0.57	0.42	0.56	0.37	0.33	0.22	0.05	0.02	0.02	0.00
Mean wind 400-Loc	0.98	0.98	0.94	0.95	0.99	1.00	0.95	0.97	0.91	0.87	0.85	0.73	0.70	0.67	0.54	0.49	0.18	0.16	0.07	0.07
Mean wind 400-Trim	0.91	0.90	0.79	0.72	0.86	0.78	0.68	0.59	0.72	0.59	0.49	0.30	0.52	0.34	0.32	0.16	0.06	0.03	0.02	0.00
Mean wind 500-Equ	0.94	0.88	0.86	0.77	0.91	0.84	0.73	0.65	0.80	0.62	0.57	0.37	0.57	0.38	0.35	0.22	0.07	0.04	0.03	0.01
Mean wind 500-Loc	0.98	0.99	0.97	0.96	1.00	1.00	0.95	0.99	0.90	0.88	0.82	0.71	0.75	0.68	0.53	0.51	0.27	0.25	0.15	0.14
Mean wind 500-Trim	0.93	0.90	0.84	0.77	0.87	0.77	0.70	0.52	0.72	0.59	0.46	0.28	0.52	0.34	0.30	0.17	0.07	0.04	0.02	0.01
Mean wind 600-Equ	0.92	0.90	0.79	0.73	0.90	0.85	0.78	0.69	0.83	0.63	0.62	0.45	0.60	0.41	0.35	0.29	0.10	0.06	0.03	0.00
Mean wind 600-Loc	0.97	0.98	0.91	0.92	1.00	1.00	0.98	1.00	0.93	0.88	0.82	0.75	0.80	0.74	0.57	0.60	0.36	0.31	0.22	0.19
Mean wind 700-Equ	0.89	0.87	0.75	0.70	0.89	0.83	0.76	0.67	0.80	0.65	0.59	0.41	0.57	0.39	0.34	0.24	0.07	0.02	0.04	0.01
Mean wind 700-Loc	0.92	0.96	0.87	0.91	1.00	1.00	0.98	1.00	0.89	0.87	0.82	0.75	0.77	0.72	0.53	0.57	0.28	0.24	0.09	0.11
Mean wind 800-Equ	0.93	0.93	0.80	0.74	0.92	0.88	0.76	0.71	0.82	0.77	0.66	0.47	0.61	0.49	0.37	0.27	0.08	0.04	0.04	0.02
Mean wind 800-Loc	0.98	0.99	0.96	0.99	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.99	0.99
Mean wind 900-Equ	0.93	0.93	0.83	0.80	0.89	0.81	0.69	0.58	0.85	0.73	0.67	0.44	0.62	0.43	0.37	0.28	0.08	0.03	0.03	0.02
Mean wind 900-Loc	0.99	1.00	0.97	0.97	0.94	0.98	0.83	0.93	0.98	0.97	0.96	0.91	0.87	0.87	0.73	0.77	0.45	0.44	0.25	0.26

Note: Entries denote the relative frequencies a given combination enters in the MCS at level α for the statistics R and SQ; τ indicates the location of the break.

Table 20: Comparison among forecast combinations for parameter specification M1, T=3000.

MODEL M1	$\tau = 0.5T$				$\tau = 0.6T$				$\tau = 0.7T$				$\tau = 0.8T$				$\tau = 0.9T$			
	$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$	
T=3000	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ
Expanding Wind	0.89	0.87	0.71	0.75	0.88	0.86	0.77	0.71	0.86	0.85	0.69	0.63	0.67	0.64	0.45	0.40	0.79	0.78	0.59	0.60
RS Mean	0.99	1.00	0.97	0.97	0.96	0.97	0.90	0.93	0.95	0.92	0.85	0.85	0.88	0.92	0.71	0.70	0.85	0.87	0.69	0.73
RS Mean Trim	0.98	0.99	0.92	0.99	0.97	0.99	0.92	0.96	0.94	0.93	0.84	0.82	0.77	0.77	0.62	0.57	0.84	0.81	0.72	0.69
CM combined	0.93	0.95	0.90	0.89	0.96	0.94	0.89	0.90	0.99	0.97	0.87	0.86	0.99	0.99	0.94	0.91	0.94	0.93	0.82	0.84
Mean wind 800-Equ	0.97	0.97	0.88	0.88	0.93	0.96	0.85	0.86	0.97	0.94	0.81	0.83	0.87	0.90	0.69	0.67	0.85	0.89	0.65	0.71
Mean wind 800-Loc	0.98	0.98	0.90	0.96	0.96	0.96	0.91	0.92	0.99	0.98	0.96	0.96	0.96	0.97	0.89	0.89	0.96	0.97	0.90	0.91

Note: Entries denote the relative frequencies a given combination enters in the MCS at level α for the statistics R and SQ; τ indicates the location of the break.

Table 21: Comparison among forecast combinations for parameter specification M2, T=3000.

MODEL M2	$\tau = 0.5T$				$\tau = 0.6T$				$\tau = 0.7T$				$\tau = 0.8T$				$\tau = 0.9T$			
	$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$	
T=3000	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ
Expanding Wind	0.82	0.80	0.57	0.54	0.73	0.71	0.42	0.36	0.54	0.47	0.31	0.23	0.33	0.28	0.15	0.14	0.40	0.34	0.21	0.22
RS Mean	0.97	0.97	0.86	0.85	0.90	0.91	0.82	0.83	0.98	0.95	0.81	0.77	0.63	0.69	0.40	0.40	0.62	0.60	0.41	0.46
RS Mean Trim	0.96	0.97	0.92	0.94	0.99	0.99	0.97	0.96	0.88	0.83	0.66	0.60	0.51	0.47	0.30	0.30	0.52	0.50	0.38	0.34
CM combined	0.94	0.91	0.74	0.71	0.89	0.86	0.76	0.72	0.99	0.99	0.97	0.94	0.94	0.95	0.85	0.86	0.89	0.88	0.73	0.72
Mean wind 800-Equ	0.92	0.91	0.76	0.74	0.85	0.83	0.73	0.68	0.88	0.90	0.63	0.62	0.58	0.65	0.33	0.34	0.56	0.58	0.35	0.40
Mean wind 800-Loc	0.98	0.96	0.82	0.83	0.94	0.94	0.85	0.85	1.00	1.00	0.97	0.93	0.96	0.96	0.91	0.91	0.91	0.93	0.78	0.82

Note: Entries denote the relative frequencies a given combination enters in the MCS at level α for the statistics R and SQ; τ indicates the location of the break.

Table 22: Comparison among forecast combinations for parameter specification M3, T=3000.

MODEL M3	$\tau = 0.5T$				$\tau = 0.6T$				$\tau = 0.7T$				$\tau = 0.8T$				$\tau = 0.9T$			
	$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$	
T=3000	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ
Expanding Wind	0.73	0.70	0.48	0.49	0.72	0.68	0.39	0.33	0.48	0.37	0.22	0.18	0.07	0.06	0.01	0.01	0.02	0.05	0.01	0.01
RS Mean	0.93	0.92	0.82	0.79	0.98	0.98	0.85	0.84	0.94	0.93	0.68	0.69	0.39	0.42	0.13	0.11	0.30	0.29	0.14	0.14
RS Mean Trim	0.99	0.99	0.97	0.98	1.00	1.00	0.99	0.99	0.79	0.73	0.45	0.44	0.25	0.22	0.07	0.05	0.18	0.18	0.12	0.12
CM combined	0.88	0.84	0.66	0.68	0.92	0.90	0.74	0.72	0.99	0.99	0.93	0.93	0.84	0.83	0.63	0.62	0.66	0.66	0.52	0.55
Mean wind 800-Equ	0.89	0.87	0.69	0.65	0.88	0.89	0.65	0.67	0.80	0.82	0.49	0.51	0.46	0.51	0.24	0.24	0.25	0.28	0.05	0.09
Mean wind 800-Loc	0.94	0.93	0.82	0.77	0.96	0.96	0.83	0.84	0.98	0.95	0.89	0.88	0.97	0.97	0.95	0.94	0.90	0.90	0.82	0.82

Note: Entries denote the relative frequencies a given combination enters in the MCS at level α for the statistics R and SQ; τ indicates the location of the break.

Table 23: Comparison among forecast combinations for parameter specification M1, T=4000.

MODEL M1	$\tau = 0.5T$				$\tau = 0.6T$				$\tau = 0.7T$				$\tau = 0.8T$				$\tau = 0.9T$			
	$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$	
T=4000	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ
Expanding Wind	0.85	0.87	0.66	0.69	0.89	0.89	0.72	0.68	0.84	0.81	0.63	0.58	0.71	0.70	0.49	0.47	0.66	0.63	0.47	0.41
RS Mean	0.94	0.96	0.82	0.84	0.99	0.99	0.91	0.92	0.95	0.97	0.87	0.90	0.84	0.85	0.64	0.66	0.70	0.70	0.53	0.52
RS Mean Trim	0.97	0.98	0.92	0.97	1.00	1.00	0.96	0.97	0.96	0.98	0.94	0.93	0.79	0.82	0.60	0.55	0.71	0.67	0.58	0.54
CM combined	0.92	0.95	0.73	0.74	0.96	0.99	0.80	0.81	0.94	0.95	0.86	0.84	0.92	0.91	0.76	0.71	0.82	0.81	0.65	0.60
Mean wind 800-Equ	0.92	0.95	0.81	0.83	0.92	0.97	0.83	0.86	0.94	0.93	0.85	0.82	0.93	0.94	0.74	0.72	0.78	0.78	0.59	0.58
Mean wind 800-Loc	0.94	0.95	0.87	0.88	0.98	1.00	0.88	0.94	0.96	0.98	0.90	0.93	0.99	1.00	0.99	1.00	0.98	0.99	0.96	0.96

Note: Entries denote the relative frequencies a given combination enters in the MCS at level α for the statistics R and SQ; τ indicates the location of the break.

Table 24: Comparison among forecast combinations for parameter specification M2, T=4000.

MODEL M2	$\tau = 0.5T$				$\tau = 0.6T$				$\tau = 0.7T$				$\tau = 0.8T$				$\tau = 0.9T$			
	$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$	
T=4000	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ
Expanding Wind	0.85	0.82	0.61	0.56	0.77	0.74	0.40	0.42	0.59	0.57	0.30	0.25	0.22	0.23	0.07	0.06	0.20	0.15	0.08	0.09
RS Mean	0.98	0.97	0.92	0.88	0.95	0.96	0.86	0.88	0.95	0.95	0.88	0.84	0.58	0.48	0.33	0.30	0.38	0.34	0.23	0.22
RS Mean Trim	0.99	1.00	0.97	0.98	0.99	1.00	0.95	0.98	1.00	1.00	0.97	0.99	0.44	0.41	0.25	0.27	0.37	0.28	0.19	0.21
CM combined	0.95	0.91	0.80	0.76	0.95	0.90	0.76	0.75	0.95	0.93	0.82	0.79	0.62	0.62	0.40	0.38	0.57	0.57	0.35	0.37
Mean wind 800-Equ	0.96	0.95	0.85	0.84	0.90	0.91	0.71	0.74	0.84	0.84	0.60	0.61	0.61	0.60	0.33	0.37	0.43	0.44	0.20	0.19
Mean wind 800-Loc	0.98	0.97	0.91	0.93	0.96	0.98	0.89	0.91	0.97	0.98	0.87	0.85	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.99

Note: Entries denote the relative frequencies a given combination enters in the MCS at level α for the statistics R and SQ; τ indicates the location of the break.

Table 25: Comparison among forecast combinations for parameter specification M3, T=4000.

MODEL M3	$\tau = 0.5T$				$\tau = 0.6T$				$\tau = 0.7T$				$\tau = 0.8T$				$\tau = 0.9T$			
	$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.10$		$\alpha = 0.25$	
T=4000	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ	R	SQ
Expanding Wind	0.81	0.81	0.57	0.56	0.73	0.69	0.41	0.40	0.51	0.53	0.22	0.24	0.14	0.17	0.03	0.06	0.01	0.01	0.00	0.00
RS Mean	0.95	0.97	0.85	0.86	0.98	0.97	0.86	0.84	0.93	0.91	0.81	0.76	0.35	0.38	0.22	0.19	0.08	0.07	0.04	0.04
RS Mean Trim	1.00	1.00	0.96	0.97	0.99	1.00	0.95	0.95	0.99	1.00	0.98	0.98	0.35	0.34	0.14	0.13	0.06	0.07	0.04	0.03
CM combined	0.88	0.88	0.73	0.73	0.94	0.91	0.78	0.78	0.90	0.86	0.72	0.69	0.48	0.47	0.32	0.32	0.19	0.19	0.08	0.09
Mean wind 800-Equ	0.90	0.91	0.76	0.78	0.89	0.90	0.69	0.69	0.74	0.75	0.55	0.51	0.48	0.47	0.28	0.29	0.07	0.08	0.03	0.03
Mean wind 800-Loc	0.96	0.98	0.85	0.88	0.97	1.00	0.92	0.92	0.91	0.94	0.79	0.76	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: Entries denote the relative frequencies a given combination enters in the MCS at level α for the statistics R and SQ; τ indicates the location of the break.

Table 26: Break dates identified by Rapach and Strauss (2008) for the U.S. dollar exchange rate returns series.

Canada	**Denmark**	**Germany**	**Japan**
02/01/1980	02/01/1980	02/01/1980	02/01/1980
03/08/1998	20/09/1993	26/12/1996	05/05/1997
10/03/2003	29/08/2000		30/03/2000
26/04/2004	06/04/2001		
31/08/2005	31/08/2005	31/08/2005	31/08/2005

Norway	**Switzerland**	**U.K.**	**U.S.(t.w.)**
02/01/1980	02/01/1980	02/01/1980	02/01/1980
17/12/1990	03/10/1995	05/02/1984	19/09/1995
31/08/1992	24/08/1998	25/09/1985	26/12/1996
		05/03/1991	13/04/2000
		14/10/1995	
31/08/2005	31/08/2005	31/08/2005	31/08/2005

Table 27: MCS p-values for the U.S. dollar exchange rate returns series

	Canada		**Denmark**		**Germany**		**Japan**	
	R	SQ	R	SQ	R	SQ	R	SQ
Expanding wind	0.00	0.00	0.10	0.10	0.09	0.09	0.11	0.11
RS Mean	0.00	0.00	0.22	0.22	0.22	0.22	0.02	0.02
RS Mean Trim	1.00	1.00	0.44	0.44	1.00	1.00	0.02	0.02
CM combined	0.00	0.00	0.14	0.14	0.09	0.09	0.64	0.64
Mean wind 800-Equ	0.00	0.00	0.33	0.33	0.26	0.26	0.17	0.17
Mean wind 800-Loc	0.00	0.00	1.00	1.00	0.91	0.91	1.00	1.00

	Norway		**Switzerland**		**U.K.**		**U.S. (tw)**	
	R	SQ	R	SQ	R	SQ	R	SQ
Expanding wind	0.01	0.01	0.00	0.00	0.21	0.21	0.06	0.06
RS Mean	0.33	0.33	0.01	0.01	0.28	0.28	0.71	0.71
RS Mean Trim	0.63	0.63	1.00	1.00	1.00	1.00	1.00	1.00
CM combined	0.06	0.06	0.00	0.00	0.28	0.28	0.90	0.90
Mean wind 800-Equ	0.25	0.25	0.00	0.00	0.25	0.25	0.14	0.14
Mean wind 800-Loc	1.00	1.00	0.42	0.42	0.25	0.25	1.00	1.00

3.8 Appendix. The model confidence set

The objective of this procedure is to determine which methods, from an initial set M^0 of methods indexed by $i = 1, \dots, M^0$, exhibit the same predictive ability in term of a loss function, given a level of confidence. In this approach, the comparison is made by using a loss function which is used to rank competing methods in term of forecasting accuracy.
Let us consider $\hat{M}*$ as the collection of the best methods, M^0 the initial collection of all the methods and $L_{i,t}$ the loss function associated with the method i in period t.
Define the relative performance variables as $d_{ij,t} = L_{i,t} - L_{j,t} \quad \forall i, j \in M^0$ and assumes that $E(d_{ij,t})$ is finite and does not depend on t. The set of the best models is defined by:

$$\hat{M}* = \left\{ i \in M^0 : E(d_{ij,t}) \leq 0 \quad \forall j \in M^0 \right\} \tag{103}$$

In order to determine $\hat{M}*$, a sequence of significance tests is made and the models that result to be significantly inferior to other element of M^0 are eliminated. The hypothesis that is being tested is:

$$H_{0,M} : E(d_{ij,t}) = 0 \quad \text{for all } i, j \in M \subset M^0 \tag{104}$$

$$H_{A,M} = E(d_{ij,t}) \neq 0 \quad \text{for some } i, j \in M \subset M^0$$

Alternatively, the previous hypothesis can be also formulated as:

$$H_{0,M} : E(d_{i.,t}) = 0 \quad \text{for all } i \in M \subset M^0 \tag{105}$$

$$H_{A,M} = E(d_{i.,t}) \neq 0 \quad \text{for some } i \in M \subset M^0$$

where $d_{i.,t} = (m-1)^{-1} \sum_{j \in M} d_{ij,t}$.
The MSC is a stepwise procedure which starts by setting $M = M_0$. The test $H_{0,M}$ is then implemented at level α. If $H_{0,M}$ is not rejected $\hat{M}^*_{1-\alpha} = M$; if $H_{0,M}$ is rejected an object from M is eliminatee and the procedure is repeated until $H_{0,M}$ is not rejected. The set $\hat{M}^*_{1-\alpha}$ is defined as the ”superior set of models” and it contains the surviving method.
Despite its sequential nature, the MCS procedure does not accumulate type I error. This is due to the fact that the test stops when the first hypothesis is not rejected.
Let define the following two t-statistics that form the basis of tests of hypothesis (104) and

(105):

$$t_{ij} = \frac{\overline{d}_{ij}}{\sqrt{v\hat{a}r(\overline{d}_{ij})}} \qquad\qquad t_{i.} = \frac{\overline{d}_{i.}}{\sqrt{v\hat{a}r(\overline{d}_{i.})}} \tag{106}$$

where:

$$\overline{d}_{ij} = \frac{1}{n}\sum_{t=1}^{n} d_{ij,t} \qquad\qquad \overline{d}_{i.} = \frac{1}{m}\sum_{j\in M} \overline{d}_{ij} \tag{107}$$

are, respectively, the relative sample loss between ith and jth models and the sample loss of the i-th model relative to the average across models; $v\hat{a}r(\overline{d}_{ij})$ and $v\hat{a}r(\overline{d}_{i.})$ denote estimates of the variance of $\overline{d}_{ij}$ and $\overline{d}_{i.}$ respectively.

The two hypothesis (104) and (105) respectively, map naturally into the two test statistics:

$$T_{R,M} = \max_{i,j\in M} |t_{ij}| \qquad\qquad T_{SQ,M} = \max_{i\in M} t_{i.} \tag{108}$$

References

[1] Armstrong J.S. 1989. Combining forecasts: The end of the beginning or the beginning of the end?. *International Journal of Forecasting* **5(4)**: 585-588.

[2] Aggarwal R, Inclan C, Leal R. 1999. Volatility in emerging stock markets. *Journal of Financial and Quantitative Analysis* **34(1)**: 33-55.

[3] Andersen TG, Bollerslev T, Diebold FX. 2007. Roughing It Up: Including Jump Components in the Measurement, Modeling anf Forecasting of Returns Volatility. *The Review of Economics and Statistics* **89(4)**: 701-720.

[4] Andreou E, Ghysels E. 2009. Structural breaks in financial time serie. In Anderson TG, Davis RA, Kreiss JP, Mikosch T (Eds.),*Handbook of financial time series*: 839-870. Springer Berlin Heidelberg.

[5] Andreou E, Ghysels E, Kourouyiannis C. 2012. *Robust volatility forecasts in the presence of structural breaks* (No. 08-2012). University of Cyprus Department of Economics.

[6] Andrews DWK. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society* **59(3)**: 817-858.

[7] Andrews DWK, Fair RC. 1988. Inference in Nonlinear Econometric Models with Structural Change. *The Review of Economic Studies* **55(4)**: 615-639

[8] Asai M, McAleer M, Medeiros MC. 2012. Asymmetry and long memory in volatility modeling. *Journal of Financial Econometrics* **10(3)**: 495-512.

[9] Assenmacher-Wesche K, Pesaran MH. 2008. Forecasting the Swiss economy using VECX models: An exercise in forecast combination across models and observation windows. *National Institute Economic Review* **203(1)**: 91-108.

[10] Bai J, Perron P. 1998 . Estimating and testing linear models with multiple structural changes. *Econometrica* **66(1)**: 47-78.

[11] Bai J, Perron P. 2003a. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* **18(1)**: 1-22.

[12] Bai J, Perron P. 2003b. Critical values for multiple structural change tests. *Econometrics Journal* **6(1)**: 72-78.

[13] Bates JM, Granger CW. 1969. The combination of forecasts. *Journal of the Operational Research Society*, **20(4)**: 451-468.

[14] Billingsley P. 1999. *Convergence of Probability Measures second edition* John Wiley & Sons, NewYork.

[15] Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31(3)**: 307-327.

[16] Bougerol P, Picard N. 1992: Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* **52(1)**: 115-127.

[17] Braun JV, Braun RK, Muller HG. 2000. Multiple change-point fitting via quasi-likelihood, with application to DNA sequence segmentation. *Biometrika*, **87(2)**: 301-314.

[18] Brown RL, Durbin J, Evans JM. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)* **37(2)**: 149-192.

[19] Cai Y, Shintani M. 2006. On the alternative long-run variance ratio test for a unit root. *Econometric Theory* **22(3)**: 347-372.

[20] Carlstein E. 1986. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics* **14(3)**: 1171-1179.

[21] Carrasco M, Chen X. 2002. Mixing and moment properties of various GARCH and Stochastic Volatility models. *Econometric Theory* **18(1)**: 17-39.

[22] Catania L, Bernardi M. 2015. MCS: Model Confidence Set Procedure. R package. http://CRAN.R-project.org/package=MCS

[23] Chow GC. 1960. Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica* **28(3)**: 591-605.

[24] Chu CSJ, Hornik K, Kuan CM. 1995a. MOSUM Tests for Parameter Constancy. *Biometrika* **82(3)**: 603-617.

[25] Chu CSJ, Hornik K, Kuan CM. 1995b. The Moving-Estimates Test for Parameter Stability. *Econometric Theory* **11(4)**: 669-720.

[26] Clark TE, McCracken MW. 2009. Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review*, **50(2)**: 363-395.

[27] Clemen RT. 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, **5(4)**: 559-583.

[28] Clements MP, Hendry DF. 2006. Forecasting with breaks. *In Handbook of Economic Forecasting*, ed. G. Elliot, C.W.J. Granger, and A. Timmermann (Elsevier) 605-657.

[29] Corsi F. 2004. A Simple Long Memory Model of Realized Volatility. *Available at SSRN 626064.*

[30] Corsi F. 2009. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* **7(2)**: 174-196.

[31] Corsi F, Renó R. 2010. HAR Volatility Modelling with Heterogeneous Leverage and Jumps. *Available at SSRN 1316953*.

[32] Davis RA, Lee TCM, Rodriguez-Yam GA. 2006. Structural break estimation for non stationary time series models. *Journal of the American Statistical Association*, **101(473)**: 223-239.

[33] Diebold FX. 1986. Modelling the Persistence of Conditional Variance: A Comment. *Econometric Reviews*, **5(1)**: 51-56.

[34] Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13(3)**: 253-263.

[35] Doukhan P. 1994. Mixing. In *Mixing*:15-23. Springer New York.

[36] Edwards AWF, Caavalli-Sforza LL. 1965. A method for cluster analysis. *Biometrics*, **21**: 362-375.

[37] Engle RF, Bollerslev T. 1986. Modelling the persistence of conditional variances. *Econometric Reviews* **5(1)**: 1-50.

[38] Efron B. 1979. Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7(1)**: 1-26.

[39] Gadea MD, Gómez-Loscos A, Montañés, A. 2016. Oil price and economic growth: a long story?. *Econometrics* **4(4)**:4-41.

[40] Giraitis L, Kapetanios G, Price S. 2013. Adaptive forecasting in the presence of recent and ongoing structural change. *Journal of Econometrics* **177(2)**: 153–170.

[41] Hall P. 1985. Resampling a coverage pattern. *Stochastic Processes and their Applications* **20(2)**: 231-246.

[42] Hansen BE. 2001. The new econometrics of structural change: Dating breaks in US labor productivity. *The Journal of Economic Perspectives* **15(4)**: 117-128.

[43] Hansen PR. 2005. A test for superior predictive ability. *Journal of Business & Economic Statistics* **23(4)**: 365-380.

[44] Hansen PR, Lunde A. 2005. A forecast comparison of volatility models: does anything beat a GARCH (1,1)? *Journal of Applied Econometrics* **20(7)**: 873-889.

[45] Hansen PR, Lunde A, Nason JM. 2003. Choosing the best volatility models: The model confidence set approach. *Oxford Bulletin of Economics and Statistics* **65(1)**: 839-861

[46] Hansen PR, Lunde A, Nason JM. 2005. Model confidence sets for forecasting models. *Working Paper, Federal Reserve Bank of Atlanta*.

[47] Hansen PR, Lunde A, Nason JM. 2011. The model confidence set. *Econometrica* **79(2)**: 453-497.

[48] Harvey A, Ruiz E, Shephard N. 1994. Multivariate stochastic variance models. *The Review of Economic Studies* **61(2)**: 247-264.

[49] He C, Teräsvirta T. 1999a. Fourth moment structure of the GARCH(p,q) process. *Econometric Theory* **15(06)**: 824-846.

[50] He C, Teräsvirta T. 1999b. Properties of moments of a family of GARCH processes. *Journal of Econometrics* **92(1)**: 173-192.

[51] Herrndorf N. 1984. A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability* **12(1)**: 141-153.

[52] Hillebrand E. 2005. Neglecting parameter changes in GARCH models. *Journals of Econometrics* **129(1)**: 121-138.

[53] Huang BN, Yang CW. 2001. The impact of settlement time on the volatility of stock market revisited: an application of the iterated cumulative sums of squares detection method for changes of variance. *Applied Economics Letters* **8(10)**: 665-668.

[54] Hwang S, Valls Pereira PL. 2006. Small sample properties of GARCH estimates and persistence. *The European Journal of Finance* **12(6-7)**: 473-494.

[55] Kiefer NM, Vogelsang TJ. 2002. Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory* **18(6)**: 1350-1366.

[56] Kiefer NM, Vogelsang TJ. 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* **21(6)**: 1130-1164.

[57] Kokoszka P., Leipus R. 2000. Change-point estimation in ARCH models. *Bernoulli* **6(3)**: 513-539.

[58] Kuan CM, Hornik K. 1995. The generalized fluctuation test: A unifying view. *Econometric Reviews* **14(2)**: 135-161.

[59] Künsch H.R. 1989. The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics* **17**: 1217-1241.

[60] Inclan C, Tiao GC. 1994. Use of Cumulative Sums of Squares for Retrospective Detection of Changes in Variance. *Journal of American Statistic Association* **89(427)**: 913-923.

[61] Inoue A, Jin L, Rossi B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics* **196(1)**: 55-67.

[62] Lamoureux CG, Lastrapes WD. 1990. Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics* **8(2)**: 225-234.

[63] Lindner A.M. 2009. *Stationarity, mixing, distributional properties and moments of GARCH (p, q)–processes*. In Anderson TG, Davis RA, Kreiss JP, Mikosch T (Eds.),*Handbook of financial time series*: 43-69. Springer Berlin Heidelberg.

[64] Liu C, Maheu JM. 2008. Are there structural breaks in Realized Volatility?. *Journal of Financial Econometrics* **6(3)**: 326-360.

[65] Liu LY, Patton AJ, Sheppard K. 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics* **187(1)**: 293-311.

[66] MacKinnon J. 1994. Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business & Economic Statistics* **12(2)**: 167-176.

[67] Maheu JM, Gordon S. 2008. Learning, forecasting and structural breaks. *Journal of Applied Econometrics* **23(5)**: 553-583.

[68] Marcellino M. (2004). Forecast pooling for short time series of macroeconomic variables. *Oxford Bulletin of Economic and Statistics* **66**: 91-112.

[69] Makridakis S, Hibon M. 2000. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting* **16(4)**: 451-476.

[70] Mensi W, Hammoudeh S, Nguyen DK, Kang SH. 2016. Global financial crisis and spillover effects among the US and BRICS stock markets. *International Review of Economics & Finance* **42**: 257-276.

[71] Mikosh T, Stărică C. 2004. Nonstationarities in Financial Time Series, the Long -range Dependence and the IGARCH Effects. *Review of Economics and Statistics* **86(1)**: 378-390.

[72] Nelson DB. 1990. Stationarity and persistence in the GARCH (l, l) model. *Econometric theory* **6(3)**: 318-334.

[73] Newey WK, West KD. 1994. Automatic Lag Selection in Covariance Matrix estimation. *Review of Economic Studies* **61(4)**: 631-654

[74] Ni J, Wohar ME, Wang B. 2016. Structural Breaks in Volatility: The Case of Chinese Stock Returns. *The Chinese Economy* **49(2)**: 81-93.

[75] Patton AJ. 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* **160(1)**: 246-256.

[76] Perron P. 2006. Dealing with structural breaks. *Palgrave handbook of econometrics vol I*: 278-352.

[77] Pesaran MH, Pettenuzzo D, Timmermann A. 2006. Forecasting time series subject to multiple structural breaks. *Review of Economic Studies* **73(4)**: 1057-1084.

[78] Pesaran MH, Pick A. 2011. Forecast combination across estimation windows. *Journal of Business and Economic Statistics* **29(2)**: 307-318.

[79] Pesaran MH, Pick A, Pranovich M. 2013. Optimal forecasts in the presence of structural breaks. *Journal of Econometrics* **177(2)**: 134-152.

[80] Pesaran MH, Schuermann T, Smith LV. 2009. Forecasting economic and financial variables with global VARs. *International Journal of Forecasting* **25(4)**: 642-675.

[81] Pesaran MH, Timmermann A. 2002. Market Timing and Return Prediction under Model Instability. *Journal of Empirical Finance* **9(5)**: 495-510.

[82] Pesaran MH, Timmermann A. 2007. Selection of estimation windows in the presence of breaks. *Journal of Econometrics* **137(1)**: 134-161.

[83] Ploberger W, Krämer W. 1992. The CUSUM Test With OLS Residuals. *Econometrica* **60(2)**: 271-285.

[84] Ploberger W, Krämer W, Kontrus K. 1989. A new test for structural stability in the linear regression model. *Journal of Econometrics* **40(2)**: 307-318.

[85] Politis DN, Romano JP. 1994. The stationary bootstrap. *Journal of the American Statistical association* **89(428)**: 1303-1313.

[86] Politis DN, White H. 2004. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* **23(1)**: 53-70.

[87] Rapach DE, Strauss JK. 2008. Structural breaks and GARCH models of exchange rate volatility. *Journal of Applied Econometrics* **23(1)**: 65-90.

[88] Rapach DE, Strauss JK, Wohar, ME. 2008. Forecasting stock return volatility in the presence of structural breaks. In: Rapach DE, Wohar, ME (Eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Elsevier-Emerald Series Frontiers of Economics and Globalization, Bingley, UK: 381-416.

[89] Rodrigues PM, Rubia A. 2007. On the Finite-Sample Biases in Nonparametric Testing for Variance Constancy. *Documentos de Trabajo FUNCAS* 304.

[90] Ross GJ. 2013. Modelling financial volatility in the presence of abrupt changes. *Physica A: Statistical Mechanics and its Applications* **392(2)**: 350-360.

[91] Sansó A, Arragó V, Carrion-i-Silvestre J.L. 2004. Testing for Change in the Unconditional Variance of Financial Time Series. *Revista de Economia Financiera* **4(1)**: 32-53.

[92] Schrimpf A, Wang Q. 2010. A reappraisal of the leading indicator properties of the yield curve under structural instability. *International Journal of Forecasting* **26(4)**: 836-857.

[93] Stărică C, Herzel S, Nord T. (2005). Why does the GARCH(1,1) model fail to provide sensible longer-horizon volatility forecasts? *Manuscript, Chalmers University of Technology*.

[94] Stock JH, Watson, MW. 2001. Vector autoregressions. *The Journal of Economic Perspectives* **14(4)**: 101-115.

[95] Stock JH, Watson, MW. 2004. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* **23(6)**: 405-430.

[96] Taylor S.J. 2007. *Modelling Financial Time Series (second edition)*. World Scientific Publishing.

[97] Tian J, Anderson HM. 2014. Forecast Combinations under Structural Break Uncertainty. *International Journal of Forecasting* **30(1)**: 161-175.

[98] Timmermann A. 2006. Forecast combinations. *In Handbook of Economic Forecasting*, ed. G. Elliot, C.W.J. Granger, and A. Timmermann (Elsevier) 135-196.

[99] Tsay RS. 2010. *Analysis of financial time series* John Wiley & Sons, New York.

[100] Vortelinos DI. 2015. Forecasting realized volatility: HAR against Principal Components Combining, neural networks and GARCH. *Research in international business and finance*, **39(2)**: 824–839.

[101] West KD. 1996. Asymptotic inference about predictive ability. *Econometrica* **64(5)**: 1067-1084.

[102] West KD, Cho D. 1995. The predictive ability of several models of exchange rate volatility. *Journal of Econometrics*, **69(2)**: 367-391.

[103] White H. 2000. A reality check for data snooping. *Econometrica* **68(5)**: 1097-1126.

[104] Yang K, Chen L, Tian F. 2015. Realized volatility forecast of stock index under structural breaks. *Journal of Forecasting*, **34(1)**: 57-82.

[105] Zeileis A, Kleiber C, Kramer W, Hornik K. 2003. Testing and dating of structural changes in practice. *Computational Statistics and Data Analysis* **44(1)**: 109-123.