Essays in Financial Time Series

Submitted by

Seonhwi Lee

to

the University of Exeter

as a Thesis for the Degree of

Doctor of Philosophy

in

Economics

August 2015

Thesis Supervisors:

Andreea G. Halunga

Richard D.F. Harris

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ....

Acknowledgement

First of all, I would like to show my gratitude to my supervisors, Andreea Halunga and Richard Harris, who consistently provide the most helpful guidance and encouragement for the completion of this thesis. I would also like to thank to James Davidson for the useful comments and discussions.

I also owe a great deal of thanks to my family, especially to my father, mother, father-in-law and mother-in-law, whose dedication and everlasting love to me make it possible to accomplish this work successfully. Especially, I have not been able to make up this without greater sacrifice and love of my wife, Hyunjung, and children, Youngmin and Youngwon. And also, I offer my great regards to all my friends and colleagues who supported and encouraged me during the period of my doctorate course, special thank to Xiaoyu Li, Sarunas Girdenas, Lawrence Choo, Jungyeoun Lee, Dooruj Rambaccussing and Hyun Chang Yi.

In addition, I am greatly indebted to University of Exeter Business School for the financial supports. Also, I am greateful to participants of the following seminars, workshops and conferences: The Annual Conference of the Royal Economic Society at University of Manchester, 2015; The Symposium of Society for Nonlinear Dynamics and Econometrics at BI Norwegian Business School, 2015; University of Vienna, 2015; The Royal Economic Society Postgraduate Meetings & Job Market at University College London, 2015; Postgraduate Research Student Annual Conferences and Workshops at University of Exeter Business School, 2012-2015.

Abstract

This thesis consists of three essays on topics in financial time series with particular emphases on specification testing, structural breaks and long memory. The first essay develops an asymptotically valid specification testing framework for the Realised GARCH model of Hansen, Huang and Shek (2012). The misspecification tests account for the joint dependence between return and the realised measure of volatility and thus extend the existing literature for testing the adequacy of GARCH models. The testing procedure is constructed based on the conditional moment principle and the first-order asymptotic theory. Our Monte Carlo results reveal good finite sample size and power properties.

In the second essay, a Monte Carlo experiment is conducted to investigate the relative out-of-sample predictive ability of a class of conditional variance models when either a structural break or long memory is allowed. Our Monte Carlo results reveal that if the true volatility process is stationary short memory and its persistence level is not too high, but is contaminated by a structural break, the presence of the structural break is of importance in choosing a proper size of estimation window in the short-run forecast. If the persistence level is very high, spurious long memory may often dominate the true structural break in the longer-run forecast. For data generation processes without any structural break, the forecasting models, which can characterise the properties of the true conditional variance process, are favourable.

In the last essay, we analyse the properties of the S&P 500 stock index return volatility process using historical and realised measures of volatility. We investigate a true property of the stochastic volatility processes by means of econometric tests, which may disentangle true or spurious long memory. The realised variance and realised kernel of the US stock market return exhibit true long memory. However, the historical volatility process shows some evidence of spurious long memory. We examine relative out-of-sample performance of one-day-ahead forecasts, with emphasis on the predictive content of structural changes and long memory. A class of ARFIMA models consistently produces the best-performing forecasts compared to a class of GARCH models. Among the GARCH models, it is shown that a rolling window GARCH forecast and GARCH forecasts which account for breaks outperform the long memory-based GARCH models even with the long memory proxy process.

Contents

List of Tables

List of Figures

Chapter 1 Introduction

This thesis mainly concerns methodological issues on inference and prediction of conditional variance models and their applications in financial time series analysis. One of the topics is developing specification testing procedures for the parametric GARCHX type models when the joint dependence between the realised measure of volatility and the squared error process is allowed for the parametric conditional variance model specification. The other topics deal with evaluating relative predictive ability and financial economic benefits of the conditional volatility models with particular emphases on structural breaks and long memory properties.

In response to the increasing interest and need for analysis of the dynamics of financial asset return volatility, sophisticated econometric tools and techniques have been developed over several decades. As seminal studies, Engle (1982) introduced Autoregressive Conditional Heteroskedasticity (ARCH), and Bollerslev (1986) made an extension of ARCH, the so-called Generalised ARCH (GARCH). Both have made considerable contributions to the development of parametric conditional volatility models. Within the ARCH and GARCH frameworks, in general, a squared return series is used to measure a current level of volatility and predict future behaviour of volatility. In both academic and practical purposes, a variant of GARCH models has been widely utilised due to its simplicity in specification and its superiority for estimating and forecasting volatility, relative to traditional time series models.

Andersen and Bollerslev (1998) reveal that, on the other hand, the forecasting performance of the GARCH model can be improved using high frequency-based realised variance rather than the squared return, which is a noisy proxy of the true conditional variance. Andersen, Bollerslev, Diebold and Labys (2003) and Barndorff-Nielsen and Shephard (2002) introduced a new concept of volatility measurement, so-called realised volatility, as an alternative measure of squared return-based historical volatility. A development of a computerised database system makes it possible to collect and store intraday price data efficiently. Reducing sampling horizons from daily to intradaily may allow realised volatility to possibly be more informative than the squared return measure. In addition, Hansen and Lunde (2011) showed that the so-called GARCHX model, where the basic GARCH model is augmented by including an additional explanatory variable - in this case, the realised measure - outperforms the traditional GARCH model. In particular, a class of GARCHX models have been proposed by jointly specifying returns and realised measures of volatility, for example, the Multiplicative Error (MEM) model of Engle and Gallo (2006), the High-Frequency-Based Volatility (HEAVY) model of Shephard and Sheppard (2010) and the Realised GARCH model of Hansen, Huang and Shek (2012), among others. It is generally known that joint dependence might exist between financial asset return and realised measure of volatility. In this sense, approaches for the inference of such a modified GARCH model with a realised measure of volatility may vary depending on whether a model builder allows for such a dependence or not. If we allow for the joint dependence between the realised volatility process and the return error process, then the inference procedure would have to account for the joint dependence.

The second chapter of this thesis develops an asymptotically valid specification testing framework for the Realised GARCH model of Hansen et al. (2012), which resembles the GARCHX model for the conditional mean and variance equations but additionally captures the joint dependence of the realised measure of volatility and return. The misspecification tests proposed may thus extend the existing literature for testing the adequacy of GARCH models such as Bollerslev (1986), Li and Mak (1994), Engle and Ng (1993), Lundbergh and Teräsvirta (2002) and Halunga and Orme (2009).

The testing procedure extends the unifying framework of Halunga and Orme (2009) and the first-order asymptotic theory, based on the conditional moment tests of Newey (1985) and Tauchen (1985). Essentially, the test indicator for the correct specification of the conditional variance equation contains an additional component that accounts for the joint dependence between the realised measure and the error process. In addition, the recursive nature of the conditional variance is also taken into consideration. Although, the testing framework is designed to test the correct specification of the conditional variance equation, the unifying framework can be employed to test the correct specification of the realised variance equation as well. We propose test statistics for asymmetry in the Realised GARCH constructed based on an alternative the Realised Exponential GARCH models, where a leverage function is explicitly specified in the conditional variance equation.

Our Monte Carlo experiment reveals that the proposed test statistics have good finite sample size properties and high degrees of power against alternative data generation processes. In particular, the test statistic that accounts for the recursive nature of the conditional variance appears to be a powerful tool in the detection of potential misspecifications of the null model arising from asymmetry behaviour in financial asset returns. Along those lines, the empirical application also supports that the test statistic with the recursive nature of the processes works very well when the size of the asymmetry in the leverage effect is large enough. Specifically, the asymmetry test rejects the null at any significance level for the stock returns with a higher degree asymmetric volatility.

On the other hand, a large body of studies about long memory properties has been established to improve predictability when modelling and forecasting financial time series. The properties of long memory can be specified by dealing with slowly diminishing impacts of shocks. The fractional integrated framework has well been established to characterise such long memory properties.[1]

However, some studies have also pointed out that a non-linear specification which allows for structural changes or regime switching could be considered to avoid spurious estimation of a persistence parameter when a given time series process is not true long memory. As seminal works, Rappoport and Reichlin (1989) and Perron (1989) demonstrated that misspecification caused by overlooking shocks as structural breaks in the underlying deterministic trend might induce the erroneous non-rejection of the unit root hypothesis. Also, Leybourne, Mills and Newbold (1998) and Kim, Leybourne and Newbold (2000) have shown that if a time series process has true multiple breaks but only one break is considered, then neglected breaks that occurred earlier than the structural break taken into account would cause spurious rejection of the null for unit root in the Dickey-Fuller and Perron tests. Further, it has been argued that the interplay between long memory and structural breaks is very often observed in a variant of economic and financial time series, and allowing for structural breaks in a true long memory process can possibly reduce the persistence.[2]

[1] For further studies of long memory generation and related properties, see, for example, Granger (1980, 1981), Granger and Joyeux (1980), Geweke and Porter-Hudak (1983), Bollerslev and Engle (1993), Baillie (1996), Granger and Ding (1996), and references therein.

[2] For more specific reviews about the relationship between long memory and structural breaks, see, for

Some research have been developed to identify structural breaks and its relationship with long memory in financial time series. A class of approaches dealing with long memory and structural breaks has been specified by accounting for the changes in either mean or variance parameters in given specifications. For the changes in mean, Granger and Hyung (2004) have compared an occasional breaks model and a long memory $I(d)$ model using absolute S&P 500 stock index returns. Their empirical results point out that long memory can be partially caused by breaks even if the return forecasting performance of the break model is not better off than the long memory models. In addition, for stock market volatility, Perron and Qu (2010) have shown that the time series can generate highly similar properties to long memory if a short memory process cannot account for structural breaks in the mean. An employed semiparametric model using log-periodogram regression, and a random level shifts model, outperforms the GARCH and the ARFIMA models. Choi and Zivot (2007) and Choi, Yu and Zivot (2010) have also empirically found that structural breaks in the mean can be the part of long memory in cases with a forward discount rate and realised volatility, respectively.

Regarding the changes in conditional variance, some theoretical and empirical results have argued that the persistence in financial asset return volatility can be better characterised by a stationary short memory process with structural breaks rather than long memory, in terms of its in-sample and/or out-of-sample fit. Mikosch and Starica (2004) have documented that long memory might be due to non-stationarity, and that integrated models such as IGARCH could induce spurious estimation under the assumption of constant unconditional variance. Starica and Granger (2005) have found that non-stationary model with structural breaks in unconditional variance can perform relatively well in a longer horizon forecast. For more specific analysis in terms of a GARCH volatility, Mikosch and Starica (2004) and Hillebrand (2005) have demonstrated that neglecting structural breaks in GARCH parameters can possibly lead to misleading estimations of persistence with upward biases in a GARCH process. Moreover, Rapach and Strauss (2008) have revealed that allowing for the structural breaks in the unconditional variance of exchange rate returns may often improve the in-sample and out-of-sample performance of the GARCH volatility.

Along those lines, it might be desired to admit the fact that neglecting structural breaks can infer spurious long memory when modelling and forecasting volatility. However, a choice

instance, Banerjee and Urga (2005), Perron (2006) and Andreou and Ghysels (2009).

of long memory and structural break in a volatility process would still be quite arguable due to the difficulty in clearly distinguishing between long memory and structural breaks. Moreover, a proper choice of forecasting model is important to contribute to an accurate prediction because a misspecified econometric model is more likely to produce a poor in-sample or out-of-sample fit. Nevertheless, there is still a lack of information which is provided with simulation-based evidences on relative predictive ability of conditional variance forecasting models with particular emphases on empirical contents and economic benefits in and between structural break and long memory. In order to fill such a potential gap in the existing literature, we carry out a Monte Carlo simulation experiment under the assumption that we already know true properties under a wide variety of data generation processes. As a consequence, we may shed light on the views of the empirical relationship between structural break and long memory in terms of relative predictive ability of financial volatility models.

In the third chapter, a Monte Carlo experiment is conducted to investigate the relative out-of-sample predictive ability of a class of parsimonious conditional variance models when either a structural break or long-run dependence is allowed for a conditional variance process. We consider GARCH with various estimation windows, FIGARCH and short and long memory EWMA models to generate a model-based conditional variance forecast. Additionally, the forecast combinations are utilised against the single model-based forecasts. The aggregate average of Mean Squared Forecast Error and the Mean of Conditional Value at Risk loss functions are adopted to measure forecasting performance. In the statistical evaluation for the loss of multiple forecasts, we conduct the Reality Check test of White's (2000) reality check test and Hansen's (2005) superior predictive ability test of Hansen (2005). Further, the pairwise comparisons are accommodated across all of the generated forecasts, applying the modified Diebold-Mariano test of Harvey, Leybourne and Newbold (1997) with the pooled loss series.

In brief, our Monte Carlo results reveal that if the true process is short memory with a structural break and the persistence level is not very high, the presence of structural breaks is important to set a proper estimation window in order to achieve more accurate predictions of the conditional variance in the short-run. If the persistence level is very high, spurious long memory may often dominate the structural breaks in the longer-run forecast. For the data generation process without any artificial structural break, the forecasting models which can characterise the properties of the true conditional variance process are generally favourable. The econometric loss evaluation results support our findings fairly well.

In the fourth chapter, we analyse the properties of the S&P 500 stock index return volatility processes which are realised measures of volatility and historical volatility. We first investigate the true properties of given volatility measures by means of econometric tests which help us disentangle true long memory and structural breaks. We examine the relative out-of-sample performance of one-day-ahead forecasts, with emphasis on the predictive information content of structural changes and long memory. The log-realised measures of volatility are estimated and forecasted by means of a class of ARFIMA models. Also, a class of short and long memory GARCH models are utilised for the historical volatility estimation and prediction. The relative forecast performance is evaluated by accounting for some tests.

The main results of fourth chapter are as follows. The US stock market index, realised variance and realised kernel processes exhibit true long memory. However, the historical volatility process shows some evidence of spurious long memory subject to multiple structural breaks corresponding to stock market events. Once the structural breaks are adjusted to the squared daily return, the volatility process looks like a weak dependent stationary process, rather than a persistent process. In terms of relative predictive accuracy, a class of ARFIMA models consistently generates the best-performed forecasts relative to a class of GARCH models. Among GARCH models, it is shown that rolling window GARCH forecasts and GARCH forecasts which account for structural changes in their own specifications outperform long memory-based GARCH models, even with the long memory proxy process. Also, the sensitivity analysis of rolling window size for GARCH model supports that the appropriate choice of the rolling window size for the GARCH model is important to achieve relatively better predictive ability in the structural breaks even when the proxy of an actual volatility exhibits long memory.

Chapter 2 Misspecification Tests for Realised GARCH Models

2.1 Introduction

Over the past few decades, the GARCH model of Bollerslev (1986) that uses the squared daily (or lower frequency) returns has been popularly considered for modelling and forecasting volatility. However, it has been shown that the GARCH model exhibits poor out-of-sample fit (e.g. Jorion, 1995). Andersen and Bollerslev (1998) reveal that, on the other hand, the forecasting performance of the GARCH model can be improved using high frequency-based realised variance rather than the squared return, which is a noisy proxy of the true conditional variance. In addition, Hansen and Lunde (2011) showed that the so-called GARCHX model, where the basic GARCH model is augmented by including an additional explanatory variable - in this case, the realised measure - outperforms the traditional GARCH model. Their simulation analysis shows that it takes longer for the conditional variance within a standard GARCH framework to catch up to a new level of volatility during times with rapid volatility changes. Such models have been estimated by Engle (2002), Forsberg and Bollerslev (2002) and Fleming, Kirby and Ostdiek (2008), among others.

In the GARCH-X framework, the additional covariate is generally treated as an exogenous variable. However, a large body of empirical research, dealing with financial or macroeconomic variables, has shown that there possibly exists dependence between an augmented covariate and the original time series. In this sense, it would be less likely to obtain correct inference of a given specification under such a strong restriction of exogeneity. Also, that assumption might cause a lack of information for the interpretation of the conditional variance in practice.

Recent models have been proposed by jointly specifying returns and various realised measures of volatility such as the Multiplicative Error (MEM) model of Engle and Gallo (2006), the High-Frequency-Based Volatility (HEAVY) model of Shephard and Sheppard (2010) and the Realised GARCH model of Hansen et al. (2012), among others. The Realised GARCH model resembles the GARCHX model for the conditional mean and variance equations but additionally captures the joint dependence of the realised measure of volatility and return through a measurement of the volatility equation.

As a consequence, approaches for the inference of these modified GARCH models may vary depending on whether a model builder allows for such dependence or not. Under such a strong exogeneity assumption, the parameters in the stationary GARCHX model can be estimated without any consideration for the specification for the additional covariate because the added regressor is simply predetermined and treated as an observed value. Therefore, existing GARCH-based testing frameworks would be directly applicable, such as the score type test for testing a GARCH model against a higher order GARCH model of Bollerslev (1986), the portmanteau test statistic based on the null hypothesis of no autocorrelation for the squared standardised error process of Li and Mak (1994), asymmetry and/or non-linearity tests of Engle and Ng (1993) and Lundbergh and Teräsvirta (2002) or parametric constancy tests of Lundbergh and Teräsvirta (2002). In addition, Halunga and Orme (2009) showed that estimation effects from a conditional mean equation can be non-negligible when testing for asymmetry and/or non-linearity in GARCH models, such that the asymmetry tests of Engle and Ng (1993) and Lundbergh and Teräsvirta (2002) are asymptotically invalid. They proposed asymptotically valid tests for asymmetry and/or non-linearity that can have also improved power properties.

Nevertheless, if we allow for the joint dependence between the additional covariate and the error process, then the inference procedure would have to account for the joint dependence. However, there are no formal misspecification tests to check the adequacy of the GARCHX type model under the joint dependence, to our best knowledge. Therefore, developing a specification testing framework under the joint conditional density may bridge possible gaps between traditional GARCH models and the GARCH with a realised measure of volatility.

This chapter proposes a framework for the construction and analysis of misspecification tests in the Realised GARCH model of Hansen et al. (2012). The testing procedure extends the unifying framework developed by Halunga and Orme (2009) for pure GARCH models, based on the conditional moment tests of Newey (1985) and Tauchen (1985) as well as the first order asymptotic theory. Essentially, the test indicator for the correct specification of the conditional variance equation contains an additional component which accounts for the joint dependence of the realised measure and the error process. In addition, the recursive nature of the conditional variance is also taken into consideration. The unifying framework can also be employed to test the correct specification of the measurement equation. Moreover, we propose tests for asymmetry in the Realised GARCH based on the alternative Realised Exponential GARCH models of Hansen and Huang (2012) and Hansen, Lunde and Voev (2014), where a leverage function is explicitly specified in the conditional variance equation.

This chapter is organised as follows. Section 2.2 describes the null Realised GARCH model and the necessary assumptions for consistent estimation. In addition, the general framework of Quasi-Maximum Likelihood estimation is discussed. In Section 2.3, the asymptotically valid general misspecification testing framework of the Realised GARCH model is demonstrated. Giving higher attention to testing for the misspecification of the conditional variance equation, Section 2.4 details the new test statistics to test for the additional leverage effect with the alternative specification. Section 2.5 reports several Monte Carlo evidences for empirical size and size-adjusted power of the developed test statistics in support of principal theoretical findings. In Section 2.6, we also report the results of empirical applications using various financial asset returns and their realised volatility measures. Finally, the concluding remarks of this study are discussed in Section 2.7.

2.2 The Null Model and Estimation Framework

2.2.1 Realised GARCH Model

The Realised GARCH model for the return variable, y_t, is represented as

$$y_t = f(\mathbf{y}_{t-1}; \boldsymbol{\eta}) + \varepsilon_t,$$

where $f(\mathbf{y}_{t-1}; \boldsymbol{\eta})$ is any $\mathcal{F}_{t-1}$ measurable function which is also allowed to be linear or non-linear, $\mathbf{y}_{t-1} = (1, y_{t-1}, ..., y_{t-l})' \in \mathbb{R}^{l+1}$, and $\boldsymbol{\eta}$ is the parameter vector of the mean function above. The error process, $\{\varepsilon_t; \mathcal{F}_t\}$ is assumed to be a martingale difference sequence, where $\mathcal{F}_{t-1} = \sigma\{y_{t-1}, y_{t-2}, ...\}$ that is the σ-field of the past information up to and including time index $t-1$. The error process of y_t is given by

$$\varepsilon_t = \sqrt{h_t} z_t, \tag{1}$$

for $t = 1, \ldots, T$, where h_t is the conditional variance process of ε_t, $h_t = E\left(\varepsilon_t^2 | \mathcal{F}_{t-1}\right)$ and $z_t \sim i.i.d.\,(0,1)$. The general specification of the conditional variance and its realised measurement are given by

$$\log h_t = \omega + \sum_{i=1}^{p} \beta_i \log h_{t-i} + \sum_{j=1}^{q} \gamma_j \log x_{t-j}, \tag{2}$$

$$\log x_t = \xi + \varphi \log h_t + \boldsymbol{\delta}' a\left(z_t\right) + u_t, \tag{3}$$

where x_t is a realised measure of volatility, which is jointly dependent with the error process ε_t. The realised measure of volatility, x_t, may play a similar role as the squared error of return series, ε_t^2 in the standard GARCH framework. As shown in Engle (2002) and Barndorff-Nielsen and Shephard (2007), the squared lagged errors become insignificant once a realised measure of volatility is included in the standard GARCH specification. Namely, the lagged variables of x_t in (2) characterise the impact of new information arrivals in the same way as the lagged squared errors terms in the GARCH model. Therefore, the realised measurement equation may explain the time-dependent behaviour of x_t onto the conditional variance over the sample period. Since the past realisations of the realised volatility measurement are observable, all the past lags of x_t are treated as predetermined variables as well as exogenous in the context of GARCHX through estimation. For the stochastic error process in the measurement equation, (3), it is assumed that $u_t \sim i.i.d.\left(0, \sigma_u^2\right)$ and u_t is mutually independent of z_t. Over the sample period, the information set can be generalised as $\mathcal{F}_t = \sigma\left(\varepsilon_s, x_s\right)$ for $s \leq t$ with the additional information of the exogenous covariate. Thus, it is to say that h_t is adapted to $\mathcal{F}_{t-1}$ and x_t should be $\mathcal{F}_t$ measurable if γ_j is all non-zero. If $\xi = 0$ and $\varphi = 1$, x_t is a unbiased measure of h_t. Otherwise, the realised measure of volatility is likely to be biased of the true daily volatility. This is why a possible bias in high-frequency-based intradaily return data is subject to market microstructure noise and non-trading hours. Moreover, the null specification allows for the leverage function. The leverage effects in the measurement equation can be captured by $a(z_t) = (a_1(z_t), a_2(z_t), \ldots, a_k(z_t))^{'}$, assuming $E[a(z_t)] = 0$, without loss of generality. Following the setting of Hansen et al. (2012), we consider Hermite polynomials of z_t as a leverage function in this study. It may follow the fact that better predictive power of the GARCH model against competing GARCH-based models can be gained by allowing for the asymmetric feature of the conditional variance as pointed out by Hansen and Lunde (2005). In the null specification, the measurement equation captures the leverage effects jointly with the conditional variance even if the GARCH equation itself does not explicitly contain any term that can directly utilise the potential leverage effect. The positiveness condition for h_t is automatically satisfied with the log-linear specification of the null model. From Proposition 1 of Hansen et al. (2012), the level of conditional variance persistence can be explained by a following parameter:

$$\rho = \sum_{i}^{p} \beta_i + \varphi \sum_{j}^{q} \gamma_j.$$

The stationarity of $\log h_t$ of the Realised GARCH(1,1) is ensured if $\beta_1 + \gamma_1 \varphi_1 \in (-1, 1)$. For simplification of notation, denote the parameter vector $\boldsymbol{\theta}^{'} = \left(\boldsymbol{\eta}^{'}, \boldsymbol{\lambda}^{'}, \boldsymbol{\upsilon}^{'}, \sigma_u^2\right)$ where $\boldsymbol{\lambda}^{'} = \left(\omega, \boldsymbol{\beta}^{'}, \boldsymbol{\gamma}^{'}\right)$ and $\boldsymbol{\upsilon}^{'} = \left(\xi, \varphi, \boldsymbol{\delta}^{'}\right)$ where $\boldsymbol{\beta} = \left(\beta_1, \ldots, \beta_p\right)^{'}$, $\boldsymbol{\gamma} = \left(\gamma_1, \ldots, \gamma_q\right)^{'}$ and $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_k)^{'}$. Let $\mathbf{c}_{t-1} = (1, \log h_{t-1}, \ldots, \log h_{t-p}, \log x_{t-1}, \ldots, \log x_{t-q})^{'}$ and $\mathbf{m}_t^{'} = \left(1, \log h_t, a(z_t)^{'}\right)$, then the GARCH and the realised measurement equations are re-expressed as

$$\begin{aligned} \log h_t &= \boldsymbol{\lambda}^{'} \mathbf{c}_{t-1}, \\ \log x_t &= \boldsymbol{\upsilon}^{'} \mathbf{m}_t + u_t, \end{aligned}$$

respectively. The following assumptions ensure consistency and asymptotic normality of the quasi-maximum likelihood estimators.

Assumption A.

1. The $\boldsymbol{\Theta}$ is compact parameter space, and $\boldsymbol{\theta}_0$ lies in the interior of $\boldsymbol{\Theta}$.

2. $E\left|f(\mathbf{y}_{t-1}; \boldsymbol{\eta}) - f(\mathbf{y}_{t-1}; \boldsymbol{\eta}_0)\right|^2 > 0$, for all $\boldsymbol{\eta} \neq \boldsymbol{\eta}_0$.

3. The elements of y_t are strictly stationary and ergodic, and $f(\mathbf{y}_{t-1}; \boldsymbol{\eta})$ is at least twice continuously differentiable in $\boldsymbol{\eta}$ and $\mathcal{F}_{t-1}$ measurable for all $\boldsymbol{\eta} \in \Theta$.

4. $\{\varepsilon_t, h_t, x_t\}$ is stationary and ergodic with $E(z_t|\mathcal{F}_{t-1}) = 0$, $E\left(z_t^2|\mathcal{F}_{t-1}\right) = 1$, $E(u_t|\mathcal{F}_{t-1}) = 0$ and $E\left(u_t^2|\mathcal{F}_{t-1}\right) = \sigma_u^2$.

5. $E|z_t|^{4+s} < \infty$ and $E|u_t|^4 < \infty$ for some $s \geq 0$ and all t.

Assumption A.2-4 are general conditions to ensure stationarity and ergodicity of the conditional mean, variance and the realised variance process. In particular for the QMLE estimation with the Gaussian likelihood, Assumption A.4 is initially made by Hansen et al. (2012) for the robustness of the QMLE which can be reflected by the weak assumptions that make the score a martingale difference sequence. Without the rigorous derivation of stationarity and ergodicity, in this sense, we simply follow the conjecture of Hansen et al. (2012) through this study. Assumption A.5 implies that z_t and u_t have a finite fourth moment to ensure positive definiteness of the asymptotic covariance matrix of the score under the quadratic form of the leverage function. If the leverage function, $a(z_t)$, is more complicated, say it has a higher order Hermite polynomial form, a finiteness condition for additional moments of z_t is inevitably required since the first and second order partial derivatives of u_t with respect to the null parameters depend on the functional form of $a(z_t)$.

Hansen et al. (2012) conjectured that the QML estimators of the Realised GARCH model are asymptotically normal and consistent under the stationary assumption, adapting Theorem 7.1 of Straumann and Mikosch (2006). Since it is imposed that $f(\mathbf{y}_{t-1}; \boldsymbol{\eta}) = \mathbf{0}$ in Hansen et al. (2012), the conditional mean estimation effect can be simply assumed away in the estimation of the conditional variance and its realised volatility. However, the null model, embodying the general form of the conditional mean equation, exhibits a different specification from the original Realised GARCH model. Further, it has been shown that the mean estimation effect is important when constructing misspecification tests of the GARCH models as described in Halunga and Orme (2009). In this sense, this study extends the framework for estimation of the Realised GARCH model, also taking into account the analysis of the conditional mean estimation effect.

Since no particular distribution for ε_t and u_t is initially assumed, the quasi maximum likelihood estimation (QMLE) framework should be applicable to produce the consistent estimators of true parameters. For the purpose of the estimation, we allow for a Gaussian specification of the log-likelihood function for jointly distributed ε_t and u_t, assuming that

$z_t \sim i.i.d.\ \mathcal{N}(0,1)$ and $u_t \sim i.i.d.\ \mathcal{N}\left(0,\sigma_u^2\right)$. The quasi log-likelihood function is given by

$$L_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} l_t(\boldsymbol{\theta}),$$

where $l_t(\boldsymbol{\theta}) = -\frac{1}{2}\left(\log h_t + \frac{\varepsilon_t^2}{h_t} + \log\sigma_u^2 + \frac{u_t^2}{\sigma_u^2}\right)$.

Before deriving the score and Hessian of the log-likelihood function, the first-order partial derivatives of h_t with respect to $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ are provided by accounting for the recursive nature of the conditional variance, in advance. For the starting values of the conditional variance, Lemma 1 of Hansen et al. (2012) stated that $\frac{\partial \log h_t}{\partial \boldsymbol{\lambda}} = \mathbf{0}$ for $t \leq 0$, regardless of $(h_0, \ldots, h_{p-1})$ being treated as fixed or as a vector of unknown parameters. Indeed, all the past series of x_t is being treated as pre-determined, and the error process of the mean, ε_t, is omitted from the conditional variance equation because it is less likely to be empirically significant in the null model. Further, Halunga and Orme (2009) remarked that the unobserved sequence of the conditional variance is of order small enough (in probability) in the sums of the partial derivatives of h_t with respect to the conditional variance parameters, which follow from Berkes, Horváth and Kokoszka (2003) and Francq and Zakoian (2004). Based on the discussions above, we here suppose that pre-sample observations are available and that $h_t = 0$ for $t \leq 0$, for the sake of simplicity. Following this, define $\mathbf{h}_{\boldsymbol{\eta}t} = \frac{\partial \log h_t}{\partial \boldsymbol{\eta}}$ and $\mathbf{h}_{\boldsymbol{\lambda}t} = \frac{\partial \log h_t}{\partial \boldsymbol{\lambda}}$, then $\mathbf{h}_{\boldsymbol{\eta}s} = \mathbf{0}$ and $\mathbf{h}_{\boldsymbol{\lambda}s} = \mathbf{0}$, for $s \leq 0$. By exploiting stochastic recursion of the conditional variance itself, we are able to derive $\mathbf{h}_{\boldsymbol{\eta}t}$ and $\mathbf{h}_{\boldsymbol{\lambda}t}$:

$$\begin{aligned}\mathbf{h}_{\boldsymbol{\eta}t} &= \sum_{i=1}^{p}\beta_i \frac{\partial \log h_{t-i}}{\partial \boldsymbol{\eta}} \\ &= B(L)\,\mathbf{h}_{\boldsymbol{\eta}t},\end{aligned}$$

$$\begin{aligned}\mathbf{h}_{\boldsymbol{\lambda}t} &= \mathbf{c}_{t-1} + \sum_{i=1}^{p}\beta_i \frac{\partial \log h_{t-i}}{\partial \boldsymbol{\lambda}} \\ &= \mathbf{c}_{t-1} + B(L)\,\mathbf{h}_{\boldsymbol{\lambda}t},\end{aligned}$$

where $B(L) = \beta_1 L + \beta_2 L^2 + \ldots + \beta_p L^p$ with a lag operator L. Then, $\mathbf{h}_{\boldsymbol{\eta}t}$ and $\mathbf{h}_{\boldsymbol{\lambda}t}$ can be reformulated as

$$\begin{aligned}\{1-B(L)\}\,\mathbf{h}_{\boldsymbol{\eta}t} &= \mathbf{0},\\ \{1-B(L)\}\,\mathbf{h}_{\boldsymbol{\lambda}t} &= \mathbf{c}_{t-1},\end{aligned}$$

respectively. Now, let

$$\{1-B(L)\}_{+}^{-1} = \sum_{i=0}^{t-1}\psi_i L^i,$$

with $\psi_0 = 1, \psi_i > 0$ and satisfying $\psi_i = \sum_{j=1}^{p}\beta_j\psi_{i-j}$ with $\psi_s = 0, s < 0, 0 < \sum_{i=0}^{t-1}\psi_i = \left(1-\sum_{j=1}^{p}\beta_j\right)^{-1} < \infty$. This shows that $\{1-B(L)\}_{+}^{-1}$ must be non-zero and finite. Hence, the first derivatives of h_t with respect to the conditional mean and variance parameters are finally specified as

$$\mathbf{h}_{\boldsymbol{\eta}t} = \mathbf{0},$$

$$\begin{aligned}\mathbf{h}_{\boldsymbol{\lambda}t} &= \sum_{i=0}^{t-1}\psi_i\mathbf{c}_{t-1-i}\\ &= \sum_{i=0}^{t-1}\{B(L)^*\}^i\,\mathbf{c}_{t-1-i},\end{aligned}$$

where $B(L)^* = \beta_1 + \beta_2 L + \ldots + \beta_p L^{p-1}$, respectively. Particularly, for $\mathbf{h}_{\boldsymbol{\eta}t} = \mathbf{0}$, this resulted from the missing squared lagged error terms, ε_{t-j}^2, from the GARCH equation. In effect, the dynamics of the score and Hessian can be partly explained by the first derivatives of h_t. Thus, it is conjectured that $\mathbf{h}_{\boldsymbol{\eta}t}$ is possibly irrelevant to the dynamics of the moment of the score and Hessian.

Assuming that $L_T(\boldsymbol{\theta})$ is at least twice continuously differentiable in $\boldsymbol{\theta}$ for any t and also holding Assumptions A, the score vector of the log-likelihood function is defined as

$$\mathbf{S}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}) = \frac{\partial L_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}),$$

where $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}) = \frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$, partitioned as $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}) = \left(\mathbf{s}_{\boldsymbol{\eta}t}', \mathbf{s}_{\boldsymbol{\lambda}t}{}', \mathbf{s}_{\boldsymbol{\upsilon}t}{}', \mathbf{s}_{\sigma t}\right)'$, corresponding to the entire set of the parameters to be estimated. The individual score vector is given by

$$\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}) = -\frac{1}{2}\left\{-\frac{2\varepsilon_t}{h_t}\mathbf{f}_t, \left(1 - z_t^2 + \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\mathbf{h}_{\boldsymbol{\lambda}t}, -\frac{2u_t}{\sigma_u^2}\mathbf{m}_t, \frac{1}{\sigma_u^2}\left(1 - \frac{u_t^2}{\sigma_u^2}\right)\right\}',$$

where $\mathbf{f}_t = \frac{\partial f(\mathbf{y}_{t-1};\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}$ and $\dot{u}_t = \frac{\partial u_t}{\partial \log h_t} = -\varphi + \frac{1}{2}\boldsymbol{\delta}' z_t \dot{a}_t$ with $\dot{a}_t = \frac{\partial a(z_t)}{\partial z_t}$. In addition, the second-order derivative of the log-likelihood, $\frac{\partial \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$, is given by

$$\begin{pmatrix} -\frac{1}{h_t}\mathbf{f}_t\mathbf{f}_t' & \cdot & \mathbf{0} & \mathbf{0} \\ -\frac{\varepsilon_t}{h_t}\mathbf{f}_t\mathbf{h}_{\boldsymbol{\lambda}t}' & -\frac{1}{2}\left(1 - z_t^2 + \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\frac{\partial \mathbf{h}_{\boldsymbol{\lambda}t}}{\partial \boldsymbol{\lambda}'} - \frac{1}{2}\left\{z_t^2 + \frac{2\left(\dot{u}_t^2 + u_t\ddot{u}_t\right)}{\sigma_u^2}\right\}\mathbf{h}_{\boldsymbol{\lambda}t}\mathbf{h}_{\boldsymbol{\lambda}t}' & \cdot & \cdot \\ \mathbf{0} & \frac{1}{\sigma_u^2}\left(\dot{u}_t\mathbf{m}_t + u_t\dot{\mathbf{m}}_t\right)\mathbf{h}_{\boldsymbol{\lambda}t}' & -\frac{1}{\sigma_u^2}\mathbf{m}_t\mathbf{m}_t' & \cdot \\ \mathbf{0} & \frac{u_t\dot{u}_t}{\sigma_u^4}\mathbf{h}_{\boldsymbol{\lambda}t}' & \frac{u_t}{\sigma_u^4}\mathbf{m}_t' & \frac{\left(\sigma_u^2 - 2u_t^2\right)}{2\sigma_u^6} \end{pmatrix}, \quad (4)$$

where $\dot{\mathbf{m}}_t = \frac{\partial \mathbf{m}_t}{\partial \log h_t} = \left(0, 1, -\frac{1}{2}z_t\dot{a}_t'\right)'$ and $\ddot{u}_t = \frac{\partial^2 u_t}{\partial \log h_t^2} = -\frac{1}{4}\boldsymbol{\delta}'\left(z_t\dot{a}_t + z_t^2\ddot{a}_t\right)$ with $\ddot{a}_t = \frac{\partial^2 a(z_t)}{\partial z_t^2}$. For example, if $\boldsymbol{\delta}' a(z_t) = \delta_1 z_t + \delta_2\left(z_t^2 - 1\right)$ then, $\dot{u}_t = -\varphi + \frac{1}{2}\delta_1 z_t + \delta_2 z_t^2$, $\ddot{u}_t = -\frac{1}{4}\delta_1 z_t - \delta_2 z_t^2$ and $\dot{\mathbf{m}}_t = \left(0, 1, -\frac{1}{2}z_t, -z_t^2\right)'$.

The consistent estimator maximises the quasi-likelihood function such that

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_T(\boldsymbol{\theta}),$$

so that it has to be satisfied with $\mathbf{S}_{\boldsymbol{\theta}T}\left(\widehat{\boldsymbol{\theta}}\right) = \mathbf{0}$. The first-order Taylor expansion of the score, about $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, yields

$$\mathbf{0} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{s}_{\boldsymbol{\theta}t}\left(\widehat{\boldsymbol{\theta}}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0) + \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \mathbf{s}_{\boldsymbol{\theta}t}\left(\overline{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'}\sqrt{T}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right),$$

where $\boldsymbol{\theta}_0$ denotes a true parameter vector which lies in the interior of the parameter space $\boldsymbol{\Theta}$, and $\overline{\boldsymbol{\theta}}$ is a mean value lying between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Within stationary and ergodic conditional variance process, Central Limit Theorem (CLT) for the martingale difference sequence with a suitable Uniform Law of Large Numbers (ULLN) may yield that the score of the Gaussian log-likelihood function has the limiting distribution:

$$\sqrt{T}\mathbf{S}_{\boldsymbol{\theta}T}(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathcal{J}_{\boldsymbol{\theta\theta}}\right),$$

where $\mathcal{J}_{\boldsymbol{\theta\theta}} = \frac{1}{T}\sum_{t=1}^{T} E\left[\mathbf{s}_{\boldsymbol{\theta}t}\left(\boldsymbol{\theta}\right)\mathbf{s}_{\boldsymbol{\theta}t}\left(\boldsymbol{\theta}\right)^{'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$

$$= \begin{pmatrix} E\left(\frac{1}{h_t}\mathbf{f}_t\mathbf{f}_t^{'}\right) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{4}E\left(1 - z_t^2 + \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)^2 E\left(\mathbf{h}_{\boldsymbol{\lambda}t}\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right) & \cdot & \cdot \\ \mathbf{0} & -\frac{1}{\sigma_u^2}E\left(\dot{u}_t\mathbf{m}_t\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right) & \frac{1}{\sigma_u^2}E\left(\mathbf{m}_t\mathbf{m}_t^{'}\right) & \cdot \\ \mathbf{0} & -\frac{E\left(u_t^3\right)E\left(\dot{u}_t\right)}{2\sigma_u^6}E\left(\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right) & \frac{E\left(u_t^3\right)}{2\sigma_u^6}E\left(\mathbf{m}_t^{'}\right) & \frac{1}{4\sigma_u^4}E\left(1 - \frac{u_t^2}{\sigma_u^2}\right)^2 \end{pmatrix}.$$

In the same manner, the moment of Hessian is derived as

$$-\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \mathbf{s}_{\boldsymbol{\theta}t}\left(\overline{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}^{'}} \xrightarrow{p} \mathcal{I}_{\boldsymbol{\theta\theta}},$$

where $\mathcal{I}_{\boldsymbol{\theta\theta}} = -E\left[\frac{\partial \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{'}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$

$$= \begin{pmatrix} E\left(\frac{1}{h_t}\mathbf{f}_t\mathbf{f}_t^{'}\right) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \left\{\frac{1}{2} + \frac{E\left(\dot{u}_t^2\right)}{\sigma_u^2}\right\}E\left(\mathbf{h}_{\boldsymbol{\lambda}t}\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right) & \cdot & \mathbf{0} \\ \mathbf{0} & -\frac{1}{\sigma_u^2}E\left\{\left(\dot{u}_t\mathbf{m}_t + u_t\dot{\mathbf{m}}_t\right)\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right\} & \frac{1}{\sigma_u^2}E\left(\mathbf{m}_t\mathbf{m}_t^{'}\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{2\sigma_u^4} \end{pmatrix}.$$

Given Assumption A. and with required conditions for Theorem 7.1 of Straumann and Mikosch (2006), $\widehat{\boldsymbol{\theta}}$ is able to converge to $\boldsymbol{\theta}_0$ in probability. Finally, this implies that

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{a} \mathcal{N}\left(\mathbf{0}, \mathcal{I}_{\boldsymbol{\theta\theta}}^{-1}\mathcal{J}_{\boldsymbol{\theta\theta}}\mathcal{I}_{\boldsymbol{\theta\theta}}^{-1}\right),$$

where both $\mathcal{I}_{\boldsymbol{\theta\theta}}$ and $\mathcal{J}_{\boldsymbol{\theta\theta}}$ are finite and positive definite. This asymptotic normality for the QML estimators should be persistently employed for the construction of misspecification test statistics.

2.2.2 Analysis of Estimation Effect

Within the framework of QML, the estimation effect arises from the property of asymptotic correlations indicated in an information matrix. As studied in Engle (1982) and Bollerslev (1986), the conditional mean and variance parameters in the standard GARCH model can be separately estimated without loss of asymptotic efficiency. This naturally works as a conditional expectation of the off-diagonal block of Hessian of the GARCH log-likelihood is zero. Along these lines, the analysis of the estimation effects from the conditional mean specification have often been ruled out in the inference of GARCH under the conditional symmetry assumption of z_t, i.e. $E\left(z_t^3\right) = 0$, see Halunga and Orme (2009). However, the Realised GARCH specification does not include the terms of squared lagged error process in its GARCH equation. As well as this, the model allows for joint dependence between h_t and x_t. Therefore, further statistical analysis of the estimation effect between the estimated parameters will accommodate a distinctive contribution for estimating and testing the Realised GARCH specification compared to the standard GARCH inference. In this sense, we are willing to pay greater attention to identifying the properties of asymptotic independence (or dependence) of the off-diagonal elements in the information matrix.

Given that the estimated conditional mean and variance parameters are asymptotically orthogonal, since $\mathbf{h}_{\boldsymbol{\eta}t} = \mathbf{0}$, as proved earlier, $E\left[\mathbf{h}_{\boldsymbol{\lambda}t}\mathbf{h}_{\boldsymbol{\eta}t}^{'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is indeed zero, then the off-diagonals that involve $E\left[\mathbf{h}_{\boldsymbol{\lambda}t}\mathbf{h}_{\boldsymbol{\eta}t}^{'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ in $\mathcal{J}_{\boldsymbol{\theta\theta}}$ and $\mathcal{I}_{\boldsymbol{\theta\theta}}$ should be zero as well. Namely, if the conditional variance equation does not include squared lagged errors of the conditional mean, ε_{t-j}^2, in the specification, it automatically satisfies $E\left[\mathbf{h}_{\boldsymbol{\lambda}t}\mathbf{h}_{\boldsymbol{\eta}t}^{'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}$ without any assumption of the conditional symmetry. Also, it can be shown that excluding the squared lagged error term ensures the asymptotic orthogonality between the estimate parameters of λ and η. With $E\left(\varepsilon_t|\mathcal{F}_{t-1}\right) = 0$ and $E\left(\varepsilon_t^2|\mathcal{F}_{t-1}\right) = h_t$, then

$$\begin{aligned}\mathcal{I}_{\boldsymbol{\lambda\eta}} &= \frac{1}{T}\sum_{t=1}^{T} E\left[E\left\{\left.\frac{\partial \mathbf{s}_{\boldsymbol{\lambda}t}}{\partial \boldsymbol{\eta}^{'}}\right| \mathcal{F}_{t-1}\right\}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= -\frac{1}{T}\sum_{t=1}^{T} E\left[\frac{1}{h_t}\mathbf{f}_t\mathbf{h}_{\boldsymbol{\lambda}t}^{'} E\left(\varepsilon_t|\,\mathcal{F}_{t-1}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \mathbf{0}.\end{aligned}$$

This implies that asymptotically, the conditional variance parameters can be estimated separately from the mean parameters within the null specification. Further, for the relation between $\boldsymbol{\eta}$ and $\boldsymbol{\upsilon}$, the derivatives of $\mathbf{s}_{\boldsymbol{\upsilon}t}$ and $\mathbf{s}_{\sigma t}$ with respect to $\boldsymbol{\eta}$ are given by

$$\begin{aligned}\frac{\partial \mathbf{s}_{\boldsymbol{\upsilon}t}}{\partial \boldsymbol{\eta}^{'}} &= \frac{1}{\sigma_u^2}\left(\dot{u}_t\mathbf{h}_{\boldsymbol{\eta}t}\mathbf{m}_t^{'} + u_t\mathbf{h}_{\boldsymbol{\eta}t}\dot{\mathbf{m}}_t^{'}\right), \\ \frac{\partial \mathbf{s}_{\sigma t}}{\partial \boldsymbol{\eta}^{'}} &= \frac{u_t\dot{u}_t}{\sigma_u^4}\mathbf{h}_{\boldsymbol{\eta}t},\end{aligned}$$

respectively. It is also automatically satisfied that the conditional mean parameters are asymptotically orthogonal to the parameters in the measurement equation x_t as $\mathbf{h}_{\boldsymbol{\eta}t} = \mathbf{0}$, since

$$\mathcal{I}_{\boldsymbol{\upsilon\eta}} = E\left[\frac{\partial \mathbf{S}_{\boldsymbol{\upsilon}T}(\boldsymbol{\theta})}{\partial \boldsymbol{\eta}'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0},$$

$$\mathcal{I}_{\sigma\boldsymbol{\eta}} = E\left[\frac{\partial \mathbf{S}_{\sigma T}(\boldsymbol{\theta})}{\partial \boldsymbol{\eta}'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}.$$

On the other hand, as specified in $\mathcal{J}_{\boldsymbol{\theta\theta}}$, allowing for the leverage function, $a(z_t)$, can induce the asymptotic correlation in estimated parameters of $\boldsymbol{\lambda}$ and $\boldsymbol{\upsilon}$, along with the nuisance parameter of the stochastic innovation of x_t such that σ_u^2. Therefore, the estimation effects of the specification x_t on the GARCH equation, h_t, is subject to the dependence between ε_t and x_t, regardless of the fact that the conditional mean specification has no impact on the estimation of the conditional variance.

Overall, we have uncovered the properties of the asymptotic independence that are involved in allowing for the conditional mean to the Realised GARCH model. Specifically, the conditional mean specification has no asymptotic influences in the estimation of the conditional variance and the realised measurement, regardless of particular conditions about the moment property of standardised error processes of h_t and x_t. Moreover, we have demonstrated the possible existence of an estimation effect between the realised measure of volatility and the conditional variance. However, it would be still arguable whether or not the conditional mean effects can be assumed away when testing for misspecification of the null model. We deal with this matter in a later discussion.

2.3 A Generic Specification Test Statistic

In this section, the misspecification testing procedure for the Realised GARCH(p, q) model is developed. We design a class of parametric test statistics to detect any potential misspecification of the conditional variance specification. The generic form of the test statistic is obtained on the basis of the conditional moment (score) principle and limit distribution theory that should be able to advocate the construction of asymptotically valid conditional moment test statistics. The conditional moment testing framework can be established having $E\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta})\middle|\mathcal{F}_{t-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}$ in mind. In other words, the conditional expectation of the score vector ought to be zero when any measurable function of the past information set (σ-field) is uncorrelated with any elements of the score vector under the assumption that $\{\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}), \mathcal{F}_t\}$ is a martingale difference vector sequence. In this context, the correct specification of the Realised GARCH model implies that the following conditions are satisfied simultaneously:

$$E\left(\frac{\varepsilon_t^2}{h_t} - 1 \middle| \mathcal{F}_{t-1}\right) = 0, \quad E\left(u_t | \mathcal{F}_{t-1}\right) = 0, \quad \text{and} \quad E\left(\frac{u_t^2}{\sigma_u^2} - 1 \middle| \mathcal{F}_{t-1}\right) = 0. \tag{5}$$

The first condition infers that the squared standardised error process is serially uncorrelated with its past history. In accordance with the conditional moment principle, this condition is indeed the baseline property when constructing misspecification tests of typical GARCH-nested conditional variance models. In addition, since the null model of this study is designed to capture the dependence between h_t and x_t, the latter conditions which are presumed as the distributional properties of u_t should also be taken into account to detect potential misspecification of the conditional variance equation that particularly arises from incorrect specification of the realised measurement equation.

Let $d_{\boldsymbol{\lambda}t} = z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}$, where $d_{\boldsymbol{\lambda}t}$ is the typical element of the score function which corresponds to the score contribution from the parameters of the GARCH equation, $\boldsymbol{\lambda}$. Indeed, $d_{\boldsymbol{\lambda}t}$ consists of two terms $z_t^2 - 1$ and $u_t\dot{u}_t$, where $\dot{u}_t$ is defined in (4). Thus, infringement of the conditional moment condition for $d_{\boldsymbol{\lambda}t}$ may be driven by joint causes of potential misspecification in $\log h_t$ and $\log x_t$. More specifically, $E\left[z_t^2 - 1 \middle| \mathcal{F}_{t-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0$ and $E\left[u_t\dot{u}_t | \mathcal{F}_{t-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0$ have to be simultaneously satisfied if the null model is correctly specified. Otherwise, either one or both of them could be non-zero. It implies that incorrect specification of $\log x_t$ may also cause misspecification of the conditional variance even if $\log h_t$ has a correct specification. Namely, it can be seen that a test indicator to be constructed for testing adequacy of the Realised GARCH conditional variance is being tied onto the conditional moment conditions, presented in (5). In accordance with the conditional moment principle for the score contribution under the martingale difference sequence, the general form of the null hypothesis to detect misspecification of the Realised GARCH model can be stated as

$$\mathbf{H}_0 : E\left[d_{\boldsymbol{\lambda}t} | \mathcal{F}_{t-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0, \tag{6}$$

where $\boldsymbol{\theta}_0$ is a true parameter vector under the null. Simply, a failure in rejection of the null implies that there is no evidence of misspecification of the Realised GARCH model. A conditional moment test can be constructed based on

$$E\left[d_{\boldsymbol{\lambda}t}\mathbf{w}_t\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0},$$

where $\mathbf{w}_t$ is an $\mathcal{F}_{t-1}$ measurable function and can be identified with test variables implicated in any particular alternative conditional variance specification against the null. The test variable in $\mathbf{w}_t$ should be able to accommodate the relevant past history of the process h_t. For example, consider one alternative specification whose conditional variance equation, denoted as h_t^a, is specified with unknown parameter vector, $\boldsymbol{\pi}$, of omitted variables. Then, the rejection of $\mathbf{H}_0 : \boldsymbol{\pi} = \mathbf{0}$ implies misspecification of the null model in terms of h_t. In this light, testing for (6) can be utilised by the following generic conditional moment test indicator with the unknown parameter vector, $\boldsymbol{\pi}$. It is defined as

$$\mathbf{S}_{\boldsymbol{\pi}T}\left(\widehat{\boldsymbol{\theta}}\right) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{s}_{\boldsymbol{\pi}t}\left(\widehat{\boldsymbol{\theta}}\right), \tag{7}$$

where $\mathbf{s}_{\boldsymbol{\pi}t}\left(\widehat{\boldsymbol{\theta}}\right) = \widehat{d}_{\boldsymbol{\lambda}t}\widehat{\mathbf{w}}_t$ and $\widehat{\boldsymbol{\theta}}$ is the consistent QML estimator of the null parameters. The test indicator will be adapted to the construction of the generic form of the misspecification test statistics for the Realised GARCH conditional variance. The first-order mean value expansion of the test indicator about $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ yields

$$\sqrt{T}\mathbf{S}_{\boldsymbol{\pi}T}\left(\widehat{\boldsymbol{\theta}}\right) = \sqrt{T}\mathbf{S}_{\boldsymbol{\pi}T}\left(\boldsymbol{\theta}_0\right) + \frac{\partial \mathbf{S}_{\boldsymbol{\pi}T}\left(\overline{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'}\sqrt{T}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right),$$

where the mean value $\overline{\boldsymbol{\theta}}$ lies between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, so that $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + o_p(1)$ under the null hypothesis. We devise Assumption B to ensure necessary moment conditions for $\mathbf{w}_t$, so as to satisfy the finite sample variance-covariance matrix property of $\mathbf{S}_{\boldsymbol{\pi}T}\left(\widehat{\boldsymbol{\theta}}\right)$, which is to be defined in the following section.

Assumption B.

1. $E\sup_{\boldsymbol{\theta}}\left\|\mathbf{w}_t\right\| < \infty$ for all t.

2. $E\sup_{\boldsymbol{\theta}}\left\|\mathbf{w}_t\right\|^2 < \infty$ for all t.

3. $E\sup_{\boldsymbol{\theta}}\left\|\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\theta}}\right\| < \infty$ for all t.

Under Assumptions A and B, a suitable ULLN ensures that $\frac{\partial \mathbf{S}_{\boldsymbol{\pi}T}(\overline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^{'}} - E\left[\frac{\partial \mathbf{S}_{\boldsymbol{\pi}T}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{'}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{p} \mathbf{0}$. Then, exploiting a CLT, the limiting distribution of $\sqrt{T}\mathbf{S}_{\boldsymbol{\pi}T}\left(\widehat{\boldsymbol{\theta}}\right)$ can be derived as

$$\sqrt{T}\mathbf{S}_{\boldsymbol{\pi}T}\left(\widehat{\boldsymbol{\theta}}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathcal{V}_T\right),$$

where $\mathcal{V}_T$ is a positive definite and finite matrix, given by

$$\mathcal{V}_T = \mathcal{A}_T \mathcal{J}_T \mathcal{A}_T^{'},$$

with

$$\mathcal{A}_T = \left[-\mathcal{I}_{\boldsymbol{\pi\theta}}\mathcal{I}_{\boldsymbol{\theta\theta}}^{-1} : \mathbf{I}_r\right] \quad \text{and} \quad \mathcal{J}_T = \begin{pmatrix} \mathcal{J}_{\boldsymbol{\theta\theta}} & \mathcal{J}_{\boldsymbol{\pi\theta}}^{'} \\ \mathcal{J}_{\pi\boldsymbol{\theta}} & \mathcal{J}_{\boldsymbol{\pi\pi}} \end{pmatrix},$$

provided that, $\mathcal{J}_{\boldsymbol{\pi\theta}} = \frac{1}{T}\sum_{t=1}^{T} E\left[\mathbf{s}_{\boldsymbol{\pi}t}\left(\boldsymbol{\theta}\right)\mathbf{s}_{\boldsymbol{\theta}t}\left(\boldsymbol{\theta}\right)^{'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \left[\mathcal{J}_{\boldsymbol{\pi\eta}} : \mathcal{J}_{\boldsymbol{\pi\lambda}} : \mathcal{J}_{\boldsymbol{\pi\upsilon}} : \mathcal{J}_{\boldsymbol{\pi}\sigma}\right]$ with

$$\begin{aligned}
\mathcal{J}_{\boldsymbol{\pi\eta}} &= \mathbf{0}, \\
\mathcal{J}_{\boldsymbol{\pi\lambda}} &= \frac{1}{2}E\left\{\left(1 - z_t^2 + \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)^2\right\} E\left(\mathbf{w}_t\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right), \\
\mathcal{J}_{\boldsymbol{\pi\upsilon}} &= -\frac{2}{\sigma_u^2}E\left(\dot{u}_t\mathbf{w}_t\mathbf{m}_t^{'}\right), \\
\mathcal{J}_{\boldsymbol{\pi}\sigma} &= -\frac{1}{\sigma_u^6}E\left(u_t^3\right)E\left(\dot{u}_t\mathbf{w}_t\right),
\end{aligned}$$

and $\mathcal{J}_{\boldsymbol{\pi\pi}} = \frac{1}{T}\sum_{t=1}^{T} E\left[\mathbf{s}_{\boldsymbol{\pi}t}\left(\boldsymbol{\theta}\right)\mathbf{s}_{\boldsymbol{\pi}t}\left(\boldsymbol{\theta}\right)^{'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$,

$$= E\left\{\left(1 - z_t^2 + \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)^2\right\} E\left(\mathbf{w}_t\mathbf{w}_t^{'}\right),$$

and $\mathcal{I}_{\boldsymbol{\pi}\theta} = -E\left[\frac{\partial \mathbf{S}_{\boldsymbol{\pi}t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{'}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \left[\mathcal{I}_{\boldsymbol{\pi\eta}} : \mathcal{I}_{\boldsymbol{\pi\lambda}} : \mathcal{I}_{\boldsymbol{\pi\upsilon}} : \mathcal{I}_{\boldsymbol{\pi}\sigma}\right]$ with

$$\begin{aligned}
\mathcal{I}_{\boldsymbol{\pi\eta}} &= \mathbf{0}, \\
\mathcal{I}_{\boldsymbol{\pi\lambda}} &= E\left\{\left(1 + \frac{2\dot{u}_t^2}{\sigma_u^2}\right)\mathbf{w}_t\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right\}, \\
\mathcal{I}_{\boldsymbol{\pi\upsilon}} &= -\frac{2}{\sigma_u^2}E\left(\dot{u}_t\mathbf{w}_t\mathbf{m}_t^{'} + u_t\mathbf{w}_t\dot{\mathbf{m}}_t^{'}\right), \\
\mathcal{I}_{\boldsymbol{\pi}\sigma} &= \mathbf{0},
\end{aligned}$$

and $\mathbf{I}_r$ is the identity matrix with $r = rank\left(\mathcal{J}_{\boldsymbol{\pi}\boldsymbol{\pi}}\right).$ From the preceding results, the asymptotically valid generic conditional moment test statistic is constructed as

$$T \times \mathbf{S}_{\boldsymbol{\pi}T}^{'}\left(\widehat{\boldsymbol{\theta}}\right)\widehat{\mathcal{V}}_T^{-1}\mathbf{S}_{\boldsymbol{\pi}T}\left(\widehat{\boldsymbol{\theta}}\right), \tag{8}$$

which has a χ_r^2 limiting distribution under the null. $\widehat{\mathcal{V}}_T$ is any consistent estimator of $\mathcal{V}_T$, i.e. $\widehat{\mathcal{V}}_T = \mathcal{V}_T + o_p\left(1\right).$

Against the frameworks of the asymmetry test of Engle and Ng (1993) and the nonlinearity test of Lundbergh and Teräsvirta (2002) for GARCH models, Halunga and Orme (2009) showed that estimation effects from the conditional mean specification may not be asymptotically negligible even under conditional symmetry of the standardised error process. On the other hand, the Realised GARCH equation employs the realised measure of volatility instead of lagged squared returns, to estimate and predict conditional variance. This automatically guarantees asymptotic orthogonality between $\boldsymbol{\eta}$ and either $\boldsymbol{\lambda}$ or $\boldsymbol{\upsilon}$ in the QML framework without any conditional symmetry assumptions, holding $\mathbf{h}_{\boldsymbol{\eta}t} = \mathbf{0}$. Moreover, in the construction of the generic misspecification test statistic, all the effects from the conditional mean parameters are ruled out from the dynamics of variance matrix estimators in (8) under the null, since $\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\eta}} = \mathbf{0}$ and $\mathcal{J}_{\boldsymbol{\pi}\boldsymbol{\eta}} = \mathbf{0}$ along with $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\eta}} = \mathbf{0}$ and $\mathcal{J}_{\boldsymbol{\theta}\boldsymbol{\eta}} = \mathbf{0}$. Therefore, it can be suggested that the test of the adequacy of the Realised GARCH will not be influenced by the estimation of conditional mean parameter, $\boldsymbol{\eta}$. Namely, potential misspecification of the conditional mean can be automatically negligible in both the estimation and testing of the conditional variance specification. However, the orthogonality in the conditional moments between the parameters of omitted variables, $\boldsymbol{\pi}$, and the conditional variance, $\boldsymbol{\lambda}$, still needs to be identified. In the previous section, we found that an estimation effect exists between the parameters in h_t and x_t. Hence, this might suggest that the estimation effect of the variance estimator of the test statistic with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\upsilon}$ would be asymptotically non-negligible under the joint dependence even if lagged variables of x_t are still exogenous and observable. Along these lines, the following section will propose the test statistic to test conditional variance specifications with particular alternatives. Also, the dynamics of variance matrix estimator in (8) is verified, incorporating the analysis of the estimation effect with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\upsilon}$ over $\boldsymbol{\pi}$.

2.4 Misspecification Tests for Conditional Variance

2.4.1 Test Indicator with Alternative Conditional Variance

In order to construct misspecification test statistics for potential misspecification in conditional variance particularly, we here consider a possible alternative GARCH equation in which additional stylised factors or some candidates of latent factors of the true conditional variance are also specified. Those additional variables that are not explicitly involved in the null specification can be accommodated by the test variables in $\mathbf{w}_t$. In this sense, an alternative specification should be able to achieve more accurate estimation and prediction by accounting for some additional latent factors of volatility that are less likely to be captured by the null specification.

Consider the error process of return series within the alternative:

$$\varepsilon_t = z_t \sqrt{h_t^a},$$

where h_t^a is a conditional variance under the alternative. The log-linear specification of the null model might suggest us to have the following specifications for h_t^a in mind.

$$h_t^a = h_t g_t \quad \text{or} \quad h_t^a = h_t \exp\left(g_t\right),$$

where $g_t = g\left(\mathbf{v}_{t-1}; \boldsymbol{\pi}\right)$ is any $\mathcal{F}_{t-1}$ measurable linear or non-linear function that characterises possible misspecification of the null model. $\mathbf{v}_{t-1}$ is the vector of omitted variables from h_t. $\boldsymbol{\pi}$ denotes the corresponding unknown parameter vector to $\mathbf{v}_{t-1}$. $\mathbf{v}_{t-1}$ could be derived from a (quasi) score principle, and could be a latent volatility variable which particularly specified in any alternative model. For example, asymmetry, non-linearity etc. The specific test variables of $\mathbf{v}_{t-1}$ for the additional leverage effect in volatility will be detailed in the following section. If $\mathbf{v}_{t-1}$ is specified with log-linear form in the alternative model, then it would prefer to take the first form of h_t^a. Otherwise, the latter could be more suitable than the former. With this in mind, an alternative test for misspecification of the null specification is now constructed with the following alternative specification in mind:

$$\log h_t^a = \boldsymbol{\lambda}' \mathbf{c}_{t-1}^a + g_t, \tag{9}$$

$$\log x_t^a = \boldsymbol{\upsilon}' \mathbf{m}_t^a + u_t, \tag{10}$$

where $\mathbf{c}_{t-1}^a = \left(1, \log h_{t-1}^a, \ldots, \log h_{t-p}^a, \log x_{t-1}^a, \ldots, \log x_{t-q}^a\right)'$ and $\mathbf{m}_t^a = \left(1, \log h_t^a, a(z_t)\right)'$. Hence, under the null such that

$$\mathbf{H}_0 : \boldsymbol{\pi} = \mathbf{0},$$

the test indicator is of the form (7) with the test variables that is constructed as

$$\widehat{\mathbf{w}}_t = \left[\frac{\partial \log h_t^a}{\partial \boldsymbol{\pi}}\right]_{\boldsymbol{\pi}=\mathbf{0}, \boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}.$$

For the sake of simplicity, suppose that g_t has a form of a linear polynomial, i.e. $g_t = \boldsymbol{\pi}' \mathbf{v}_{t-1}$. Denote $\mathbf{h}_{\boldsymbol{\pi}t}^a = \frac{\partial \log h_t^a}{\partial \boldsymbol{\pi}}$, then

$$\begin{aligned} \mathbf{h}_{\boldsymbol{\pi}t}^a &= \frac{\partial \boldsymbol{\lambda}' \mathbf{c}_{t-1}^a}{\partial \boldsymbol{\pi}} + \frac{\partial \boldsymbol{\pi}' \mathbf{v}_{t-1}}{\partial \boldsymbol{\pi}} \\ &= \boldsymbol{\lambda}' \mathbf{c}_{\pi t-1}^a + \mathbf{v}_{t-1} + \boldsymbol{\pi} \dot{\mathbf{v}}_{t-1}' \mathbf{h}_{\boldsymbol{\pi}t-1}^a, \end{aligned}$$

where $\mathbf{c}_{\pi t-1}^a = \frac{\partial \mathbf{c}_{t-1}^a}{\partial \boldsymbol{\pi}} = \left(\mathbf{0}_{n\times 1}, \mathbf{h}_{\boldsymbol{\pi}t-1}^a, ..., \mathbf{h}_{\boldsymbol{\pi}t-p}^a, \mathbf{0}_{n\times q}\right)'$ is a $(p+q+1) \times n$ matrix, n is row length of the unknown parameter vector $\boldsymbol{\pi}$ and $\dot{\mathbf{v}}_{t-1}' = \frac{\partial \mathbf{v}_{t-1}'}{\partial \log h_{t-1}^a}$. Under the null, $\mathbf{h}_{\boldsymbol{\pi}t}^a$ can be formulised with the recursion of the conditional variance as follows.

$$\mathbf{h}_{\boldsymbol{\pi}t}^a = \boldsymbol{\lambda}' \mathbf{c}_{\pi t-1}^a + \mathbf{v}_{t-1}$$

$$\begin{aligned} &= \sum_{i=1}^{p} \beta_i \mathbf{h}_{\boldsymbol{\pi}t-i}^a + \mathbf{v}_{t-1} \\ &= \sum_{i=0}^{t-1} \psi_i \mathbf{v}_{t-1-i} \\ &= \sum_{i=0}^{t-1} \left\{B(L)^*\right\}^i \mathbf{v}_{t-1-i}. \end{aligned}$$

So as to detect misspecification of the conditional variance equation, the generic test variables therefore take the form with the recursive nature of the conditional variance as follows.

$$\widehat{\mathbf{w}}_t = \sum_{i=0}^{t-1} \left\{ \widehat{B}(L)^* \right\}^i \mathbf{v}_{t-1-i}, \tag{11}$$

where $\widehat{B}(L)^* = \widehat{\beta}_1 + \widehat{\beta}_2 L + \ldots + \widehat{\beta}_p L^{(p-1)}$, and "hats" denotes the everything is evaluated at the consistent null parameter estimator, $\widehat{\boldsymbol{\theta}}$. For example, the test variable for the null of Realised GARCH$(1,1)$ is formed as

$$\widehat{\mathbf{w}}_t = \sum_{i=0}^{t-1} \widehat{\beta}^i \mathbf{v}_{t-1-i}.$$

Because of the fact that the errors in time series are serially correlated in general, it is expected that the test statistic with recursion can possibly improve testing performance in terms of the size and power of the test against the one without recursion, $\mathbf{v}_{t-1}$. In accordance with the null model, however, the lagged variables of the ex-post realised measure of volatility are used to model conditional volatility instead of the squared past error process. In addition, as noted earlier, the realised measurement is not necessarily an unbiased estimator of the latent volatility. For these reasons, we hypothesise carefully that none of the two different test variables shows strong dominance in testing performance.

2.4.2 Testing for Additional Asymmetry

In the case of the Realised GARCH$(1,1)$ model, once we replace $\log x_{t-1}$ in (2) with (3), the leverage function that is able of capturing asymmetric behaviour of h_t explicitly appears in the conditional variance equation as follows:

$$\log h_t = \omega_1 + \gamma_1 \xi_1 + (\beta_1 + \gamma_1 \varphi_1) \log h_{t-1} + \gamma_1 \boldsymbol{\delta}' a(z_{t-1}) + \gamma_1 u_{t-1}.$$

The asymmetric dynamics of the conditional variance are explained by a leverage function $a(z_{t-1})$ that accommodates the dependence between the return at time $t-1$ and the future volatility at time t. Note that we consider a quadratic specification for the leverage function which corresponds to the second order Hermite polynomial, $\boldsymbol{\delta}' a(z_{t-1}) = \delta_1 z_{t-1} + \delta_2 \left(z_{t-1}^2 - 1\right)$, in convenience for the construction of the test statistic. Then, it can be seen that the transformed GARCH equation above resembles the specification of the Exponential GARCH

(EGARCH) model of Nelson (1991). As noted earlier, $\boldsymbol{\delta}^{'} a(z_{t-1})$ may play a role to capture leverage effects in the volatility process. Moreover, u_{t-1} utilises a stochastic volatility innovation, since we do not know u_t at time $t-1$ whereas we know u_{t-1} at time t. The drawback of the given transformed specification is that the same parameter γ_1 is implicitly imposed as a coefficient of both $a(z_{t-1})$ and u_t. In this case, the Realised GARCH specification has a limitation to identify exact partial effects of each $a(z_{t-1})$ and u_{t-1} onto the dynamics of conditional volatility. Thus, if we assume that the null model is insufficient to capture the leverage effect, then one particular alternative specification would be able to provide a better guidance by detecting additional asymmetry which is potentially implemented into the true volatility process.

In this chapter, the Realised Exponential GARCH (Realised EGARCH) of Hansen and Huang (2012) is utilised as one of a special case of possible alternatives against the null specification. The Realised EGARCH model is given by

$$\begin{aligned} \log h_t &= \breve{\omega} + \breve{\beta} \log h_{t-1} + \breve{\boldsymbol{\tau}}^{'} a(z_{t-1}) + \breve{\gamma} u_{t-1}, & (12) \\ \log x_t &= \breve{\xi} + \breve{\varphi} \log h_t + \breve{\boldsymbol{\delta}}^{'} a(z_t) + u_t, & (13) \end{aligned}$$

where $\breve{\boldsymbol{\tau}}^{'} a(z_t) = \breve{\tau}_1 z_t + \breve{\tau}_2 \left(z_t^2 - 1\right)$ is a new leverage function that directly includes within the GARCH equation. It is worth noting that $\breve{\gamma} u_{t-1}$ in (12) is the main channel by which the realised measures drive expectations of future volatility movement. It would be also seen that Realised GARCH$(1,1)$ is implicitly nested to Realised EGARCH with single realised measure of volatility when $\breve{\boldsymbol{\tau}}^{'} a(z_t) = \gamma_1 \boldsymbol{\delta}^{'} a(z_t)$. The empirical analysis conducted in Hansen and Huang (2012) shows that holding equality of parameters between $\breve{\boldsymbol{\tau}}$ and $\gamma_1 \boldsymbol{\delta}$ is less likely to be significant in estimation. Further, the Realised EGARCH has superiority to the Realised GARCH in both the log-likelihood and the partial log-likelihood obtained throughout their empirical results. In this sense, we are willing to take the Realised EGARCH model as an alternative model against the null model. Moreover, before carrying out specification testing, the parameter constraint for $\breve{\boldsymbol{\tau}} \neq \gamma_1 \boldsymbol{\delta}$ is presumed based on the discussions above. Then, we may believe that the Realised GARCH and the Realised EGARCH models have different specification each other.

As the alternative specification of conditional variance to be tested, we employ the Realised

EGARCH model for market returns of Hansen et al. (2014), rather than the more general Realised EGARCH specification given in (12) and (13). In the GARCH equation of (12), the dependence between h_t and x_t can be captured by allowing for u_{t-1}, the one-step past stochastic innovation of x_t. However, it is natural to restrict that $\breve{\boldsymbol{\tau}} \neq \gamma_1 \boldsymbol{\delta}$,as noted earlier. Then, the null specification is not nested in the given specification of (12), in general. In this sense, keeping existing variables and parameters in the null model, the alternative specification for testing additional asymmetry in the conditional variance is given by

$$\begin{aligned} \log h_t^a &= \boldsymbol{\lambda}' \mathbf{c}_{t-1}^a + \pi_1 z_{t-1} + \pi_2 \left(z_{t-1}^2 - 1\right), \\ \log x_t^a &= \boldsymbol{\upsilon}' \mathbf{m}_t^a + u_t. \end{aligned} \tag{14}$$

Indeed, $g_t = \boldsymbol{\pi}' \mathbf{v}_{t-1}$ should be a quadratic function of z_t that implicitly capture additional asymmetric feature of volatility. Taking recursion to the conditional variance, the test variable can be derived as

$$\widehat{\mathbf{w}}_t = \sum_{i=0}^{t-1} \left\{ \widehat{B}(L)^* \right\}^i \begin{pmatrix} z_{t-1-i} \\ z_{t-1-i}^2 - 1 \end{pmatrix}. \tag{15}$$

2.4.2.1 Orthogonality in Variance Estimator

Following the generic testing procedures, demonstrating orthogonality in $\mathcal{I}_{\boldsymbol{\pi}\theta}$ across the given parameters plays a role to explain potential asymptotically non-negligible estimation effect in the variance estimator of the test statistic, (8). Since it has been turned out that the estimation effect from the conditional mean parameters onto the other parameters can be automatically ignorable within the null specification, we pay greater attention to verify whether the estimation effect over either the conditional variance parameter, $\boldsymbol{\lambda}$ or the realised measurement parameter, $\boldsymbol{\upsilon}$ is asymptotically negligible or not, under the alternative for the construction of a variance estimator, by investigating $\mathcal{I}_{\boldsymbol{\pi\lambda}}$ and $\mathcal{I}_{\boldsymbol{\pi\upsilon}}$. As noted earlier, we take the quardratic form of the leverage function in the alternative specification, $\boldsymbol{\delta}' a(z_t) = \delta_1 z_t + \delta_2 \left(z_t^2 - 1\right)$. Then, denote that $\dot{u}_t = -\varphi + \frac{1}{2}\delta_1 z_t + \delta_2 z_t^2$.

With respect to conditional variance parameters, $\mathcal{I}_{\boldsymbol{\pi\lambda}}$ is derived as

$$E\left[\left(1 + \frac{2\dot{u}_t^2}{\sigma_u^2}\right) \sum_{i=0}^{t-1} \sum_{j=0}^{t-1} \psi_i \psi_j \begin{pmatrix} z_{t-1-i} \\ z_{t-1-i}^2 - 1 \end{pmatrix} \mathbf{c}_{t-1-j}' \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \tag{16}$$

As $\left|\sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\psi_i\psi_j\right| = \left|\sum_{i=0}^{t-1}\psi_i^2\right| \leq \left(\sum_{i=0}^{t-1}|\psi_i|\right)^2 < \infty$, the typical expectation of (16) exists obviously, and is given by

$$E\left[\left(1+\frac{2\dot{u}_t^2}{\sigma_u^2}\right)\begin{pmatrix} z_{t-l} \\ z_{t-l}^2 - 1 \end{pmatrix}\mathbf{c}_{t-m}^{'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \tag{17}$$

for all $l, m < t$, where $\mathbf{c}_{t-m} = (1, \log h_{t-m}, \ldots, \log h_{t-p-m+1}, \log x_{t-m}, \ldots, \log x_{t-q-m+1})^{'}$. It is worth noting that x_{t-k}, $k = 1, \ldots, t-1$, is initially set to be predetermined through the estimation. Depending on the size of l and m, some of elements in $\mathbf{c}_{t-m}$ are counted as either random or non-random, when taking the conditional expectation given $\mathcal{F}_{t-l-1}$. Hence, the typical expectation, (17), can be examined for the cases of $l = m$, $l < m$ and $l > m$, respectively. Indeed, those of the random component in $\mathbf{c}_{t-m}$ are only subject to the dynamics of the conditional variance over given lags l and m. However, by initial setting of the error process, z_t is conditionally and unconditionally uncorrelated with h_t, for any time t, if the null model is correctly specified. Therefore, it can be said that none of the random elements in $\mathbf{c}_{t-m}$ is asymptotically correlated with z_{t-l}, regardless of the size of m against l. Furthermore, $\dot{u}_t$ is a function of z_t whose lagged process is assumed to be independent itself, thus, $\dot{u}_t$ and z_{t-l} must be asymptotically uncorrelated.

Conditioning on $\mathcal{F}_{t-l-1}$, the typical expectation the above can be expressed as

$$E\left[E\left\{\left(1+\frac{2\dot{u}_t^2}{\sigma_u^2}\right)\begin{pmatrix} z_{t-l} \\ z_{t-l}^2 - 1 \end{pmatrix}\mathbf{c}_{t-m}^{'}\middle|\mathcal{F}_{t-l-1}\right\}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \tag{18}$$

which is zero if the conditional expectation given $\mathcal{F}_{t-l-1}$ is zero. The latter can be naturally established by the discussions noted above. Specifically, it is always valid that

$$E\left[\begin{pmatrix} z_{t-l} \\ z_{t-l}^2 - 1 \end{pmatrix}(\log h_{t-m}, \log h_{t-m-1}, \ldots, \log h_{t-p-m+1})\middle|\mathcal{F}_{t-l-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0},$$

under Assumption A.4, since $E\left[z_{t-l}|\,\mathcal{F}_{t-l-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ and $E\left[z_{t-l}^2 - 1|\,\mathcal{F}_{t-l-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ are zero. Also, there is no asymptotic correlation between z_t and z_{t-l}. Finally, it brings us to show that the terms in the typical expectation are asymptotically zero, in general. As a consequence, the estimation effect of the conditional variance equation may be asymptotically negligible when testing for additional asymmetry in the conditional variance of the null, say $\mathcal{I}_{\boldsymbol{\pi\lambda}} = \mathbf{0}$, a.s., in general.

On the other hand, we investigate whether estimation effects from the realised measurement equation parameters, $\boldsymbol{\upsilon}$, are negligible or not. With the alternative model for additional leverage effect, $\mathcal{I}_{\boldsymbol{\pi\upsilon}}$ is given by

$$-\frac{2}{\sigma_u^2} E\left[\sum_{i=0}^{t-1} \psi_i \begin{pmatrix} z_{t-1-i} \\ z_{t-1-i}^2 - 1 \end{pmatrix} \left(\dot{u}_t \mathbf{m}_t' + u_t \dot{\mathbf{m}}_t'\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

Since $0 < \sum_{i=0}^{t-1} \psi_i < \infty$, the typical expectation is given by

$$\begin{aligned} & E\left[\begin{pmatrix} z_{t-l} \\ z_{t-l}^2 - 1 \end{pmatrix} \left(\dot{u}_t \mathbf{m}_t' + u_t \dot{\mathbf{m}}_t'\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ = \; & E\left[E\left\{\begin{pmatrix} z_{t-l} \\ z_{t-l}^2 - 1 \end{pmatrix} \left(\dot{u}_t \mathbf{m}_t' + u_t \dot{\mathbf{m}}_t'\right) \middle| \mathcal{F}_{t-l-1}\right\}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \end{aligned}$$

which is zero if the conditional expectation given $\mathcal{F}_{t-l-1}$ is zero. Recall $\mathbf{m}_t = \left(1, \log h_t, z_t, z_t^2 - 1\right)'$ and $\dot{\mathbf{m}}_t = \left(0, 1, -\frac{1}{2} z_t, -z_t^2\right)'$. In addition, we initially assumed that u_t and z_t are mutually independent over time t. Therefore, since $E\left[z_{t-l} \middle| \mathcal{F}_{t-l-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ and $E\left[z_{t-l}^2 - 1 \middle| \mathcal{F}_{t-l-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ are zero, it can be shown that the conditional expectation given $\mathcal{F}_{t-l-1}$ is asymptotically zero, in general. Consequently, the estimation effect of the parameters in the realised measurement equation onto the variance estimator can be asymptotically negligible in this context.

Based on all these discussions that $\mathcal{I}_{\boldsymbol{\pi\lambda}} = \mathbf{0}$ and $\mathcal{I}_{\boldsymbol{\pi\upsilon}} = \mathbf{0}$, the asymptotically valid test statistic for additional asymmetry can be constructed as

$$\mathcal{T}_A = T \times \mathbf{S}_{\boldsymbol{\pi}T}'\left(\hat{\theta}\right) \widehat{\mathcal{V}}_{AT}^{-1} \mathbf{S}_{\boldsymbol{\pi}T}\left(\hat{\theta}\right), \tag{19}$$

which is asymptotically distributed as χ^2 with r degrees of freedom that is $\dim\left(\mathbf{v}_{t-1}\right) = 2$ in the case. The variance estimator, $\widehat{\mathcal{V}}_{AT}$, can be specified by taking the diagonal elements only from $\widehat{\mathcal{V}}_T$, given in the generic form of the misspecification test statistic.

2.5 Monte Carlo Evidences

In this section, the Monte Carlo simulation evidence is presented with the evaluation results of the finite sample size and size-adjusted power performance of the proposed misspecification tests for the Realised GARCH models. Over all of the experiments considered, the data generation process (DGP) has the following specification:

$$\begin{aligned} y_t &= \varepsilon_t, \\ \varepsilon_t &= \sqrt{h_t} z_t. \end{aligned}$$

Specifically, we do not take into account lagged dependent regressors in the return equation because the conditional mean effects on estimation and specification testing are asymptotically negligible. Further, for the simplification of the analysis, the (conditional and unconditional) mean of return series is set to be zero. In addition, through the experiment, we assume that there is no misspecification in the realised measurement equation for the generation of the synthetic error process of the return as well as realised measure of volatility process itself. We generate artificial series of standardised error process, z_t and stochastic innovation term of realised measure of volatility, u_t, independently. We assume that the standardised error process of the return series follows $z_t \sim \mathcal{N}(0,1)$, or $z_t \sim t(v)$, standardised student t-distribution with v degrees of freedom with $v \in \{12, 10, 7\}$. The stochastic innovation of the realised measurement equation is assumed to be $u_t \sim \mathcal{N}\left(0, \sigma_u^2\right)$, where the unconditional variance of u_t is set to 0.15 over all the DGPs generated, without loss of generality. For each DGP, a series of initial 5000 realisations is randomly produced via GAUSS12. The first 3000 realisations are discarded in order to remove initialisation effects, which leaves a total of 2000 observations to be used. The Monte Carlo simulations are performed using 1000 replications. The null generation model is given by the Realised GARCH$(1,1)$ specification, assuming that a measurement equation is correctly specified. The Realised GARCH$(1,1)$ model is estimated by QML. The test statistic for the detection of potential misspecification of the additional asymmetry is specified with two different types of test variables. One is that $\widehat{\mathbf{w}}_t = \left(z_{t-1}, z_{t-1}^2 - 1\right)'$, without taking recursive nature of the elements in $\widehat{\mathbf{w}}_t$ into account. Then, this statistic is denoted as $\mathcal{T}_A^n$. The other is $\mathcal{T}_A$ of (19) with $\widehat{\mathbf{w}}_t = \sum_{i=0}^{t-1} \left\{\widehat{B}(L)^*\right\}^i \left(z_{t-i-1}, z_{t-i-1}^2 - 1\right)'$. It is worthy to note that some test statistics for detecting leverage effect or non-linearity within the context of the conventional GARCH models are not adopted here for the comparison purpose, because the null model does not specify squared lagged error process in the conditional variance equation. Instead, we adopt specification-free testing frameworks, such as the Ljung-Box (LB) test statistic of Ljung and Box (1978) and Lagrange Multiplier (LM) test statistic of Engle (1982), to test for autocorrelations and remaining ARCH effects in the squared standardised residuals, respectively, with 20 lags. With these settings for the experiment, we investigate the empirical size and size-adjusted power of the given misspecification test statistic at 5% significance level.

2.5.1 Finite Sample Size of the Test

In order to check the empirical size of the test statistic, firstly, we arbitrarily choose the parameter values of generating processes. The DGPs could be generally classified by accounting for different levels of persistence and leverage effect. As described earlier, the level of persistence is determined by $\rho = \beta_1 + \gamma_1 \varphi_1$, when $p = 1$ and $q = 1$. We impose a restriction that $\varphi_1 \simeq 1$, reasoning the fact that a realised measure of volatility is proportional to logarithmically transformed quantity of conditional volatility. In this setting, the persistence level now depends on the estimated values of β_1 and γ_1. Considering four levels of ρ, each **L** (0.85), **M1** (0.90) , **M2** (0.95) and **H** (0.99) yields low-, medium-low-, medium-high- and high-persistent dynamics of conditional variance, respectively. The value in the bracket corresponds to the size of ρ. The data generation processes for the conditional variance and realised measurement equations are:

$$\mathbf{L} \quad : \quad \begin{array}{l} \log h_t = 0.10 + 0.50 \log h_{t-1} + 0.35 \log x_{t-1} \\ \log x_t = -0.20 + 1.01 \log h_t + \boldsymbol{\delta}' a\left(z_{t-1}\right) + u_t \end{array} ,$$

$$\mathbf{M1} \quad : \quad \begin{array}{l} \log h_t = 0.05 + 0.50 \log h_{t-1} + 0.40 \log x_{t-1} \\ \log x_t = -0.20 + 1.01 \log h_t + \boldsymbol{\delta}' a\left(z_{t-1}\right) + u_t \end{array} ,$$

$$\mathbf{M2} \quad : \quad \begin{array}{l} \log h_t = 0.04 + 0.60 \log h_{t-1} + 0.35 \log x_{t-1} \\ \log x_t = -0.10 + 1.01 \log h_t + \boldsymbol{\delta}' a\left(z_{t-1}\right) + u_t \end{array} ,$$

$$\mathbf{H} \quad : \quad \begin{array}{l} \log h_t = 0.02 + 0.60 \log h_{t-1} + 0.40 \log x_{t-1} \\ \log x_t = -0.10 + 0.99 \log h_t + \boldsymbol{\delta}' a\left(z_{t-1}\right) + u_t \end{array} .$$

To specify level of asymmetry across the DGPs considered, we now introduce the news impact curve (NIC) of the null specification from Hansen et al. (2012).

$$\text{NIC} = \gamma_1 \boldsymbol{\delta}' a\left(z\right),$$

where the leverage function is set to be $\boldsymbol{\delta}' a\left(z\right) = \delta_1 z + \delta_2 \left(z^2 - 1\right)$. Controlling the parameters of the leverage function under the baseline DGPs above, if γ_1 is fixed, δ_1 and δ_2 would allow the DGPs to exhibit different levels of asymmetry. Plugging the following leverage functions into the baseline DGPs finally yields twelve data generation processes, in total.

$$\mathbf{A1} \quad : \quad -0.02 z_t + 0.07\left(z_t^2 - 1\right),$$

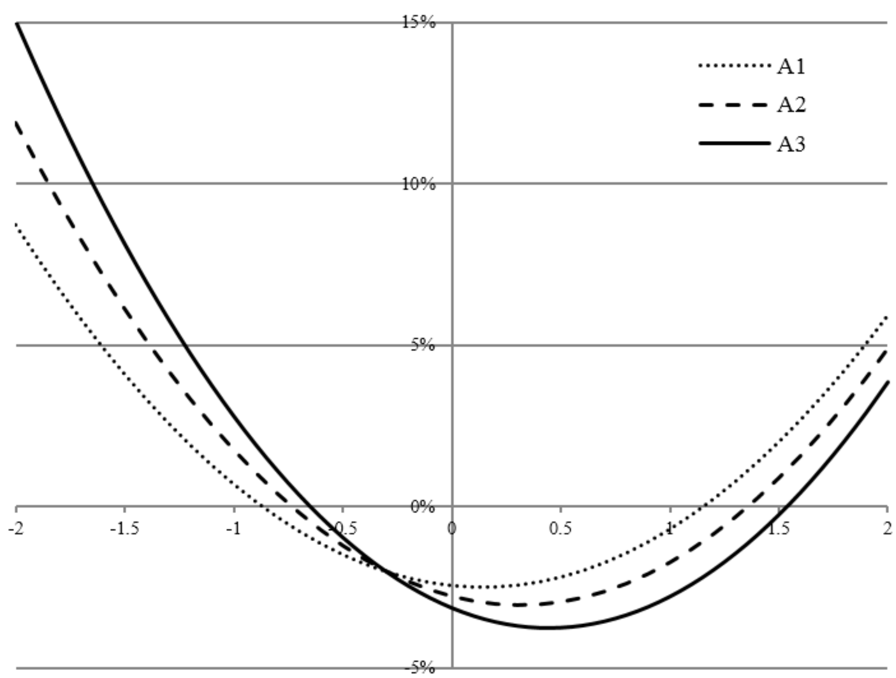
$$\mathbf{A2} \quad : \quad -0.05 z_t + 0.08\left(z_t^2 - 1\right),$$

$$\mathbf{A3} \quad : \quad -0.08 z_t + 0.09\left(z_t^2 - 1\right).$$

For example, implied news impact curves for the **A1**, **A2** and **A3** are drawn in Figure 1, exhibiting relative leverage effects, when restricting $\gamma_1 = 0.35$. It can be seen that **A3** is less symmetric about zero than that of **A1** or **A2**. Therefore, **A1** is subject to weak asymmetry relative to the others. **A2** and **A3** are treated as the cases that stand for the mid- and high-level relative leverage effects, respectively.

Table 1 reports the actual rejection frequency under the true null. When $z_t \sim \mathcal{N}\left(0,1\right)$, the empirical sizes for $\mathcal{T}_A^n$ and $\mathcal{T}_A$ are close to the nominal size of 5%. More specifically, $\mathcal{T}_A^n$ reveals better empirical size than $\mathcal{T}_A$ for the strong asymmetric DGPs (**A3**), regardless of the level of persistence. $\mathcal{T}_A^n$ also results in relatively good empirical size, compared to $\mathcal{T}_A$ for higher persistence DGPs (**H**). When both persistence and asymmetry levels are medium, the empirical sizes of $\mathcal{T}_A$ get closer to the nominal size (**M1/M2-A2**). Further, $\mathcal{T}_A$ is slightly oversized for the low leverage models (**A1**), but is undersized for the high leverage models (**A3**). On the other hand, for $z_t \sim t\left(v\right)$, both $\mathcal{T}_A^n$ and $\mathcal{T}_A$ are undersized, in general, as the degrees of freedom decrease. When $z_t \sim t(12)$, it is not easy to find any evidence that there is a significant difference of the empirical size of $\mathcal{T}_A^n$ and $\mathcal{T}_A$ from the case of normal distribution. For $t(10)$ or $t(7)$, the lower persistent DGPs (**L**, **M1**) report quite poor-sized fits against a 5% nominal size. The case of the more persistent DGPs (**M2**, **H**) show better empirical sizes compared to the former case. In addition, for every distribution considered in the experiment, the empirical size appears to decrease in most cases of $\mathcal{T}_A^n$ and $\mathcal{T}_A$, as the magnitude of asymmetry increases.

Figure 1. News Impact Curve



Additionally, the experiment evaluates the empirical size of the synthetic data processes, generated by the parameters that are estimated from several real data used in Hansen et al. (2012). We estimate the Realised GARCH$(1,1)$ model with open-to-close returns and realised kernel series of twelve NYSE stocks. The estimates for Realised GARCH$(1,1)$ are presented in Table 7, which can be found in the Appendix. Table 2 illustrates the simulation results for the empirical size of the test statistic for this case. When $z_t \sim \mathcal{N}(0,1)$ and $t(12)$, the empirical sizes for $\mathcal{T}_A^n$ and $\mathcal{T}_A$ are closed to the nominal size. For $t(10)$ and $t(7)$, it can be seen that $\mathcal{T}_A^n$ and $\mathcal{T}_A$ are significantly undersized against the nominal 5% size. For $t(10)$, the empirical size of $\mathcal{T}_A$ is worse than $\mathcal{T}_A^n$. In reverse, the empirical size of $\mathcal{T}_A$ is better than $\mathcal{T}_A^n$, for $t(7)$. In general, it reveals consistent outcomes with the former arbitrary DGP case.

Table 1: Empirical size of the test I

	$N(0,1)$				$t(12)$				$t(10)$				$t(7)$			
	T_A^n	T_A	LB	LM	T_A^n	T_A	LB	LM	T_A^n	T_A	LB	LM	T_A^n	T_A	LB	LM
L-A1	4.6	5.4	5.3	4.9	5.0	5.4	4.3	4.7	3.9	3.9	5.6	5.7	3.2	3.7	6.8	6.8
L-A2	4.7	4.7	5.2	4.9	4.8	4.9	4.3	4.8	3.9	3.2	5.6	5.6	3.0	3.2	6.7	6.8
L-A3	4.5	4.4	5.2	4.9	4.7	4.4	4.4	4.8	3.8	2.9	5.5	5.6	2.8	3.0	6.6	6.8
M1-A1	4.6	5.4	5.1	4.7	4.9	5.2	4.3	4.6	3.9	3.9	5.7	5.8	3.2	3.6	6.7	6.9
M1-A2	4.8	4.9	5.1	4.8	4.7	4.9	4.3	4.6	3.9	3.5	5.5	5.6	3.1	3.2	6.7	6.9
M1-A3	4.6	4.4	5.1	4.8	4.7	4.3	4.5	4.7	3.8	2.8	5.3	5.6	2.8	3.0	6.6	6.9
M2-A1	4.8	5.1	4.8	4.6	4.8	5.8	4.5	4.5	4.2	3.7	5.8	6.1	3.2	3.7	6.5	6.7
M2-A2	4.7	4.9	4.8	4.7	4.6	4.8	4.6	4.6	4.0	3.4	5.7	6.1	2.9	3.7	6.4	6.7
M2-A3	4.6	4.6	4.9	4.8	4.7	4.3	4.5	4.7	3.9	3.4	5.6	5.9	2.6	3.7	6.6	6.7
H-A1	5.2	4.8	5.1	4.6	4.8	5.6	4.7	4.8	4.4	4.0	5.4	5.9	3.4	4.3	6.7	6.3
H-A2	5.0	4.4	5.1	4.5	4.9	4.7	4.7	4.6	4.2	3.9	5.6	5.9	3.3	3.7	6.8	6.4
H-A3	5.0	4.6	5.1	4.6	4.8	4.5	4.6	4.5	4.5	3.8	5.8	5.9	3.2	3.7	6.7	6.3

Note: DGPs with abitrarily selected parameters.

Table 2: Empirical size of the test II

	$N(0,1)$				$t(12)$				$t(10)$				$t(7)$			
	T_A^n	T_A	LB	LM	T_A^n	T_A	LB	LM	T_A^n	T_A	LB	LM	T_A^n	T_A	LB	LM
BAC	4.9	5.0	4.9	4.7	4.7	5.2	4.7	4.6	3.9	3.7	5.7	5.8	3.0	3.5	6.4	6.6
CVX	4.6	5.2	4.8	4.6	4.8	4.6	4.7	4.8	4.0	3.6	5.7	5.9	3.0	3.2	6.4	6.6
DIS	4.7	4.5	5.1	4.9	4.7	4.3	4.8	4.6	3.9	3.3	5.3	5.5	2.8	3.5	6.7	6.7
GE	5.1	4.9	5.0	4.7	4.8	4.7	4.6	4.6	4.5	3.6	5.0	5.3	2.9	3.9	6.8	6.3
GM	4.9	4.4	4.8	4.8	4.5	4.1	4.5	4.6	4.2	2.9	5.7	5.5	2.7	3.2	6.7	6.6
IBM	4.8	4.5	5.0	4.8	4.4	4.8	4.7	4.6	4.0	3.3	5.7	5.6	3.0	3.6	6.7	6.7
KO	5.0	5.1	4.9	4.7	4.5	5.3	4.7	4.6	4.3	3.6	5.8	5.6	3.1	3.6	6.6	6.6
MCD	5.1	4.6	4.8	4.7	4.5	4.5	4.7	4.7	4.5	2.8	5.7	5.6	2.9	3.6	6.8	6.7
MMM	5.1	4.7	5.0	4.7	4.8	5.5	4.5	4.5	4.4	3.6	5.9	5.9	3.3	3.9	6.6	6.7
MRK	4.9	5.4	5.0	4.7	4.7	5.3	4.5	4.5	4.6	4.2	5.9	5.9	3.2	3.5	6.5	6.5
PG	4.8	4.5	5.0	4.8	4.7	5.0	4.4	4.7	3.8	3.4	5.9	5.9	3.1	3.7	6.5	6.7
SPY	4.6	4.9	4.9	4.9	4.6	4.5	4.9	4.6	3.9	3.5	5.6	5.5	2.9	3.3	6.5	6.7

Note: The parameters for DGPs are based on the estimates of the Realised GARCH(1,1) model, given in Table 2.

2.5.2 Finite Sample Power of the Test

In checking the empirical size-adjusted power of the test statistic, we utilise the alternative specification, provided in (14), against the null of the Realised GARCH$(1,1)$ model. As shown in Hansen and Huang (2012), imposing zero mean restriction reveals better performance in both the in-sample and out-of-sample fit of the Realised EGARCH model. In this sense, we shall impose that the expectation of return is zero for the entire set of alternative DGPs. The alternative data generating process is given by

$$\begin{aligned} \log h_t &= \omega + \beta \log h_{t-1} + \gamma \log x_{t-1} + \pi_1 z_{t-1} + \pi_2 \left(z_{t-1}^2 - 1 \right), \\ \log x_t &= \xi + \varphi \log h_t + \delta_1 z_t + \delta_2 \left(z_t^2 - 1 \right) + u_t. \end{aligned} \tag{20}$$

In order to demonstrate the effects of asymmetry in terms of the empirical power, we firstly consider the following five baseline DGPs: **A**(DIS), **B**(KO), **C**(CAT), **D**(MRK), **F**(SPY), where the parameter values are from the estimates of the Realised GARCH$(1,1)$ model with open-to-close returns and realised kernel series, presented in Table 7. The choice of stocks relies on the persistence level of the estimated conditional volatility. For the parameters of the omitted variables, we arbitrarily take two pairs of $\boldsymbol{\pi}$ such that $(\pi_1, \pi_2) = (-0.03, 0.01)$ and $(-0.01, 0.01)$, which are denoted as **G1** and **G2**, respectively. We expect that the alternative DGP generated by **G1** involves in a conditional volatility process with a greater magnitude of leverage effect against the null process. The alternative DGP with **G2** exhibits the volatility process with low asymmetry relative to **G1**. Combining additional leverage functions with the estimated null model, additional asymmetric volatility components are explicitly specified into every single alternative DGP. The results for assessing the size-adjusted power of the test are displayed in Table 3, with a nominal size of 5%. At any distribution of z_t, the test statistic performs very well in the case of **G1**, regardless of the level of persistence. In the case of a smaller size of leverage effect, we confirm that the power of the test is much weaker than before. Moreover, it is also seen that $\mathcal{T}_A$ generally dominates $\mathcal{T}_A^n$ over the alternative DGPs. Particularly, when the level of asymmetry is lower, $\mathcal{T}_A$ would be remarkably effective to detect misspecification of the null model.

We also evaluate the size-adjusted power of the test statistic based on the parameters estimated using real data. To decide the parameter values of the DGP, we take the estimates

Table 3: Empirical size-adjusted power of the test I

	$N(0,1)$				$t(12)$				$t(10)$				$t(7)$			
	T_A^n	T_A	LB	LM	T_A^n	T_A	LB	LM	T_A^n	T_A	LB	LM	T_A^n	T_A	LB	LM
A-G1	98.0	100.0	6.5	5.9	98.8	100.0	6.4	6.1	97.9	100.0	7.0	6.7	97.9	100.0	8.5	8.0
B-G1	96.2	99.7	6.6	6.1	97.2	99.8	6.4	6.1	96.3	99.8	7.2	6.7	95.9	99.7	8.3	8.1
C-G1	98.9	100.0	6.7	5.7	99.6	100.0	6.0	5.6	98.8	100.0	7.2	6.1	99.5	100.0	8.3	8.0
D-G1	98.7	100.0	6.7	6.2	99.7	100.0	6.6	6.2	99.2	100.0	8.4	7.6	99.6	100.0	8.9	8.7
E-G1	98.1	99.8	6.5	5.9	99.0	100.0	6.3	5.8	98.2	99.9	6.7	6.2	98.4	99.9	7.8	7.8
A-G2	47.4	70.6	6.1	5.5	54.3	79.8	6.2	5.9	54.5	79.1	6.9	6.2	54.6	81.4	8.4	8.3
B-G2	43.7	60.5	5.9	5.6	49.1	69.8	6.3	5.9	49.7	70.9	7.1	6.4	50.8	74.2	8.2	8.1
C-G2	52.6	76.5	6.1	5.3	60.4	85.6	6.2	5.8	61.1	84.3	6.7	6.4	59.8	85.7	8.1	8.2
D-G2	56.1	79.6	6.4	5.5	66.1	87.8	6.4	5.9	68.3	88.4	7.8	7.4	67.5	90.6	8.9	8.7
E-G2	49.3	64.1	6.0	5.6	56.8	74.5	6.2	5.8	58.0	74.4	6.8	6.1	57.9	77.4	8.1	7.6

Note: The DGPs are generated with abitrarily selected parameters of π_1 and π_2. **G1** reflects a higher degree of asymmetry in leverage effect within the alternative conditional variance specification. Against **G1**, **G2** exhibits the lower level of asymmetry. The parameter values for π_1 and π_2 are abitrarily chosen.

Table 4: Empirical size-adjusted power of the test II

	$N(0,1)$				$t(12)$				$t(10)$				$t(7)$			
	T_A^n	T_A	LB	LM	T_A^n	T_A	LB	LM	T_A^n	T_A	LB	LM	T_A^n	T_A	LB	LM
BAC	100.0	100.0	9.5	8.9	100.0	100.0	9.4	9.7	100.0	100.0	11.4	11.1	100.0	100.0	12.8	12.1
CVX	100.0	100.0	12.8	11.6	100.0	100.0	11.7	11.2	100.0	100.0	13.5	12.6	99.8	99.9	16.0	16.0
DIS	97.8	100.0	6.0	5.3	98.8	100.0	5.7	5.7	98.1	100.0	6.3	6.1	98.4	100.0	7.6	7.4
GE	100.0	100.0	23.7	21.2	100.0	100.0	22.1	20.2	100.0	100.0	23.8	22.8	99.9	100.0	24.6	23.6
GM	100.0	100.0	32.8	30.4	100.0	100.0	31.4	28.7	99.8	99.8	34.4	31.4	99.6	99.9	32.4	31.0
IBM	100.0	100.0	10.0	9.3	100.0	100.0	9.3	9.6	100.0	100.0	11.2	11.0	100.0	100.0	12.4	11.8
KO	78.4	94.4	5.8	5.3	78.5	95.0	6.0	6.0	80.0	94.7	6.6	6.1	76.6	94.2	7.9	7.7
MCD	67.8	93.9	5.5	4.8	65.2	93.1	5.5	5.4	67.4	92.6	5.5	5.5	61.3	91.1	7.2	7.2
MMM	100.0	100.0	5.7	5.1	100.0	100.0	5.4	4.9	100.0	100.0	6.4	6.5	100.0	100.0	7.5	7.4
MRK	91.7	99.0	5.8	5.3	90.5	99.2	4.1	4.8	89.5	99.2	5.8	5.5	90.2	98.4	6.6	6.4
PG	99.8	100.0	8.0	7.1	100.0	100.0	7.9	7.6	100.0	100.0	9.6	8.9	100.0	100.0	10.6	10.2
SPY	100.0	100.0	16.9	15.6	100.0	100.0	15.9	14.6	100.0	100.0	17.8	16.5	100.0	100.0	18.7	17.8

Note: The parameters for DGPs are based on the estimates for the alternative Realised EGARCH model with single realised measure of volatility, given in Table 5.

for the Realised EGARCH with single realised measure of volatility (realised kernel), reported in Hansen and Huang (2012, Table 3). However, the alternative specification is not exactly the same as the specification of Hansen and Huang (2012). Instead, we employ the Realised EGARCH specification of Hansen et al. (2014) as described in (20). Hence, it is inevitably needed to recalculate the parameter values of (20) using the estimates from Hansen and Huang (2012), assuming the estimates of the realised measurement equations remains the same in both cases. Additionally, in order to verify the consistency of the estimates between the two different EGARCH specifications, we also estimate (20) using the same real stock return and realised kernel data and compare the estimates with the recalculated values. Then, it is confirmed that there is no significant difference between them.[3]

Table 8 presents the recalculated parameters for our alternative specification with the persistence level, which can be found in the Appendix. As discussed in Hansen and Huang (2012), a degree of asymmetry in the leverage effect of the conditional variance can be roughly captured by the coefficient of z_t. A higher (less negative) valued π_1 may stand for a lower degree of asymmetry of volatility for returns, and vice versa. In our case, three data processes generated using the estimates from KO, MCD and MRK should represent lower-asymmetry synthetic return series.

In Table 4, we can observe that the DGPs that exhibit a relatively low degree of asymmetry with a smaller value of coefficient π_1 appear to have weaker power of performance in the tests by $\mathcal{T}_A^n$. However, the empirical power of $\mathcal{T}_A$ for those DGPs reveals better properties than that of $\mathcal{T}_A^n$. Overall, the results of the Monte Carlo experiments confirm that the new test statistics, $\mathcal{T}_A^n$ and $\mathcal{T}_A$, have fairly good size properties and very good power to detect strong asymmetry. We may also suggest that $\mathcal{T}_A^n$ is often less effective in terms of its testing power, if the return series follows the student t-distribution with a relatively small number of degrees of freedom. Thus, it can be said that $\mathcal{T}_A$ - the test statistic with stochastic recursion in the test variable - should be more applicable to test for potential misspecification of the null model, which could arise from strong asymmetric latent volatility.

[3] The results of the estimates for the Realised EGARCH used in Hansen, Lunde and Voev (2014) are available from authors upon request.

2.6 Empirical Applications

We carry out some empirical analysis for the log-linear Realised GARCH$(1,1)$ model as well. In the simulation experiment, we assumed that the specification of the realised measurement equation is correct to generate the series of ε_t and x_t. In this sense, the empirical analysis helps us check the validity of the constructed test statistic when detecting potential misspecification of the conditional variance process, relaxing the assumption of the correct specification of x_t. Several real financial data taken from Hansen et al. (2012) and Shephard and Sheppard (2010) are used to examine the test statistic for detecting any misspecification in leverage effects of conditional volatility.

Table 5: p-values of Hansen et al. (2012)

	$\mathcal{T}_A^n$	$\mathcal{T}_A$	LB	LM
BAC	0.000	0.000	0.972	1.000
CVX	0.000	0.000	0.242	0.999
DIS	0.000	0.001	0.027	0.000
GE	0.000	0.000	0.974	1.000
GM	0.000	0.000	0.217	0.817
IBM	0.000	0.000	0.450	1.000
KO	0.167	0.036	0.617	0.946
MCD	0.000	0.034	0.444	0.874
MMM	0.000	0.000	0.491	1.000
MRK	0.067	0.038	0.000	0.000
PG	0.000	0.000	0.214	0.347
SPY	0.000	0.000	0.095	1.000

First, Table 5 reports the p-values of the test statistic for the twelve stock and portfolio returns with realised kernel data which are the same as what we analysed in the simulation experiment. For every return process, $\mathcal{T}_A$ rejects the null hypothesis that there is no additional leverage effect, at 5% significance level. While, $\mathcal{T}_A^n$ is not rejected for KO and MRK at the same significance level. As discussed earlier, we realised the fact that the estimated coefficients for z_t in the alternative GARCH equation for KO and MRK are not bigger than the others. In this respect, it would be natural to conjecture that the KO and MRK stock return volatility processes exhibit a low level of asymmetry relative to the other stock return volatility. In addition, $\mathcal{T}_A^n$ presents relatively weak performance power when the true volatility is subject to weak asymmetry in the leverage effect of the returns, according to the experiment results. In this light, we can see that the test statistic can perform well to detect additional asymmetry in leverage effects that is not captured by the orginal Realised GARCH specification.

Since the realised measure of volatility is indeed treated as a noise proxy of the true conditional volatility, we would like to empirically examine whether the test statistic is still working properly across the different types of the realised measurement. In this part, the log-linear Realised GARCH$(1,1)$ model is estimated for some stock market index and foreign exchange rate returns with two popularly used realised measures such as realised variance (RV) and realised kernel (RK). We briefly look at the differences in the estimates of Realised GARCH volatility and report the estimated parameter values in Table 9, which can be found in the Appendix.

Table 6: p-values of Shephard and Sheppard (2010)

	$x_t = RV$				$x_t = RK$			
	T_A^n	T_A	LB	LM	T_A^n	T_A	LB	LM
DJI	0.000	0.000	0.495	1.000	0.000	0.000	0.487	1.000
IXIC	0.000	0.000	0.000	0.752	0.000	0.000	0.000	0.512
SPX	0.000	0.000	0.097	0.999	0.000	0.000	0.138	0.999
FTSE	0.000	0.000	0.247	0.986	0.000	0.000	0.259	0.981
N225	0.000	0.000	0.016	1.000	0.000	0.000	0.002	1.000
GBP	0.000	0.000	0.475	0.693	0.000	0.000	0.312	0.798
EUR	0.000	0.000	0.518	0.997	0.000	0.000	0.369	0.998
CHF	0.000	0.000	0.476	0.040	0.000	0.000	0.397	0.148
JPY	0.000	0.000	0.166	0.839	0.000	0.000	0.212	0.987

For the same return series, but with different realised measures of volatility, a relatively large difference in estimates is often found in the intercept terms, ω and ξ. It is conjectured that such a difference might come from the different level of unconditional variance of u_t between RV and RK. However, it would be difficult to say that such a difference is significantly great on average. In the result of the misspecification test, both $\mathcal{T}_A^n$ and $\mathcal{T}_A$ infer strong evidences for the rejection of the null hypothesis at any significance level, as shown in Table 6. Namely, the Realised GARCH$(1,1)$ model is not well specified to capture the leverage effect properly. However, it is hard to see any discrepancy between RV and RK in terms of the statistical decision made by the test. Thus, we may be able to suggest that the test statistic produces empirically consistent performance for testing additional asymmetry of Realised GARCH volatility, regardless of the type of realised volatility measurement.

2.7 Concluding Remarks

In this chapter, we provide a unifying and generic class of misspecification testing frameworks for the Realised GARCH(p, q) model. The proposed test statistics are constructed based on the conditional moment principle, having an asymptotic chi-square distribution under the null. The misspecification test procedures can be simply applicable in practice without further bootstrapping procedures etc., and help reduce the cost of time and computation load. In addition, our analysis of the conditional mean effect in estimation and testing has provided theoretical soundness for the test statistic to make it robust to the conditional heteroskedasticity of the return process. As shown in a series of seminal studies, the conditional mean specification might be asymptotically non-negligible when testing the specification of conventional GARCH models. However, we have shown that the conditional mean parameter effects in the estimation and testing for the Realised GARCH are effectively negligible in the absence of the squared error process in the GARCHX specification. Therefore, the Realised GARCH model with various types of mean specifications can be tested using the proposed generic framework without any loss of generality.

Our Monte Carlo experiment reveals that the proposed test statistics have good finite sample size properties and high degrees of power against alternative DGPs. In particular, the test statistic that accounts for the recursive nature of the conditional variance appears to be a powerful tool in the detection of the potential misspecification of the null model arising from asymmetry behaviour in financial asset returns. The empirical application also supports that the test statistic with the recursive nature of the processes works very well when the size of the asymmetry in the leverage effect is large enough. Specifically, the asymmetry test rejects the null at any significance level for the stock returns with a higher degree of asymmetry.

In the Realised GARCH process, including the squared error term would not be significant empirically, as shown in Engle (2002) and Barndorff-Nielsen and Shephard (2007). However, the lagged squared error-term also has often played an important role in estimating the current level of volatility, alongside additional variables that are significant for predicting latent volatility. Han and Park (2013) showed that using the information provided by squared returns, the realised measures and implied volatility performs the best in terms of the in-sample fit in their analysis. Therefore, the unifying and generic framework proposed in this study can extensively be employed for the misspecification tests for more general GARCH-type models, which include other components of latent volatility, such as the squared returns, jump component, implied volatility and multiple realised measure of volatility etc. The further discussions would be of interest, for example, to investigate the asymptotic properties of the proposed misspecification tests of the Realised GARCH model. Moreover, the misspecification testing framework can be extended to dealing with a non-stationary covariate case such as long memory, as analysed in Han and Kristensen (2014).

Appendix 2.1 Additional Tables

Table 7: Estimates for log-linear Realised GARCH model

	ω	β	γ	ξ	φ	δ_1	δ_2	σ_u^2	$\beta+\gamma\varphi$
BAC	-0.006	0.527	0.438	0.003	1.000	-0.041	0.077	0.177	0.965
CAT	0.071	0.646	0.289	-0.151	1.086	-0.027	0.092	0.145	0.960
CVX	0.041	0.570	0.290	-0.098	1.326	-0.077	0.079	0.152	0.954
DIS	0.026	0.644	0.307	-0.055	1.100	-0.043	0.090	0.170	0.982
GE	-0.001	0.703	0.291	0.004	0.979	-0.003	0.080	0.171	0.988
GM	0.138	0.657	0.301	-0.318	1.021	-0.007	0.123	0.223	0.964
IBM	0.000	0.638	0.359	0.010	0.938	-0.039	0.082	0.152	0.975
KO	-0.084	0.601	0.399	0.188	0.929	-0.019	0.075	0.149	0.972
MCD	0.015	0.718	0.262	-0.017	0.989	-0.046	0.111	0.204	0.977
MMM	-0.008	0.563	0.380	0.025	0.972	-0.018	0.073	0.175	0.933
MRK	0.084	0.633	0.241	-0.230	1.283	0.007	0.068	0.229	0.943
PG	-0.087	0.553	0.367	0.182	1.048	-0.051	0.081	0.173	0.938
SPY	0.061	0.550	0.410	-0.183	1.036	-0.067	0.073	0.146	0.975

Note: The entries are from the estimates for the Realised GARCH(1,1) model using real data set provided by Hansen, Huang, and Shek (2012), while they report the estimates of Realised GARCH(1,2) only.

Table 8: Estimates for alternative Realised EGARCH model

	ω	β	γ	π_1	π_2	ξ	φ	δ_1	δ_2	σ_u^2	$\beta+\gamma\varphi$
BAC	0.007	0.591	0.397	-0.073	0.019	-0.016	0.972	-0.032	0.077	0.170	0.977
CVX	0.045	0.592	0.272	-0.034	0.026	-0.101	1.293	-0.078	0.073	0.145	0.944
DIS	0.023	0.643	0.306	-0.031	0.007	-0.040	1.097	-0.039	0.088	0.166	0.979
GE	0.002	0.719	0.264	-0.045	0.030	0.000	0.996	0.004	0.077	0.166	0.982
GM	0.110	0.732	0.230	-0.042	0.041	-0.321	1.050	-0.024	0.118	0.217	0.973
IBM	-0.004	0.676	0.307	-0.058	0.019	0.022	0.960	-0.022	0.080	0.145	0.971
KO	-0.063	0.633	0.356	-0.021	0.008	0.158	0.953	-0.012	0.076	0.151	0.972
MCD	0.015	0.712	0.264	-0.023	0.004	-0.010	0.993	-0.025	0.109	0.200	0.974
MMM	-0.008	0.590	0.326	-0.052	0.004	0.036	1.062	-0.019	0.071	0.166	0.936
MRK	0.089	0.629	0.239	-0.023	-0.005	-0.256	1.340	-0.009	0.070	0.230	0.949
PG	-0.078	0.600	0.319	-0.039	0.016	0.185	1.070	-0.032	0.078	0.168	0.941
SPY	0.029	0.671	0.272	-0.084	0.031	-0.161	1.096	-0.076	0.073	0.132	0.969

Note: The entries are recalculated values from the estimates for the Realised EGARCH with realised kernel of Hansen and Huang (2012, Table 3).

Table 9: Estimates for log-linear Realised GARCH model with RV and RK

	ω	β	γ	ξ	φ	δ_1	δ_2	σ_u^2	$\beta+\gamma\varphi$
RV									
DJI	-0.033	0.587	0.383	-0.653	0.998	-0.135	0.065	0.225	0.969
IXIC	-0.264	0.584	0.339	0.211	1.161	-0.166	0.051	0.238	0.978
SPX	-0.221	0.599	0.348	-0.119	1.069	-0.152	0.052	0.244	0.971
FTSE	0.272	0.668	0.337	-1.237	0.940	-0.136	0.079	0.189	0.985
N225	0.082	0.648	0.325	-1.163	0.978	-0.072	0.071	0.235	0.966
GBP	-0.199	0.715	0.267	0.228	1.018	-0.010	0.080	0.097	0.987
EUR	-0.421	0.725	0.234	1.107	1.108	0.001	0.096	0.116	0.984
CHF	-0.396	0.747	0.214	1.055	1.105	-0.020	0.085	0.110	0.983
JPY	-0.915	0.596	0.315	1.262	1.122	-0.041	0.099	0.139	0.949
Average	-0.233	0.652	0.307	0.077	1.055	-0.081	0.075	0.177	0.975
RK									
DJI	0.045	0.574	0.403	-0.808	0.982	-0.135	0.057	0.207	0.970
IXIC	-0.165	0.597	0.342	-0.094	1.112	-0.170	0.060	0.257	0.977
SPX	-0.241	0.605	0.342	-0.050	1.072	-0.151	0.052	0.251	0.972
FTSE	0.209	0.672	0.327	-1.080	0.954	-0.140	0.085	0.203	0.984
N225	0.020	0.658	0.311	-1.032	0.988	-0.073	0.071	0.259	0.965
GBP	-0.221	0.713	0.264	0.272	1.033	-0.007	0.092	0.117	0.986
EUR	-0.414	0.726	0.230	0.999	1.111	0.005	0.109	0.146	0.982
CHF	-0.415	0.764	0.192	1.133	1.130	-0.023	0.105	0.151	0.981
JPY	-1.015	0.597	0.299	1.482	1.155	-0.039	0.110	0.173	0.942
Average	-0.244	0.656	0.301	0.091	1.060	-0.081	0.082	0.196	0.973

Note: The entries are from the estimates for the Realised GARCH(1,1) model using selected real data set from Shephard and Sheppard (2010).

Appendix 2.2 Derivation of $\mathcal{J}_{\boldsymbol{\pi\theta}}$ and $\mathcal{I}_{\boldsymbol{\pi\theta}}$

The score vector of the individual log-likelihood function, given by

$$\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}) = -\frac{1}{2}\left\{-\frac{2\varepsilon_t}{h_t}\mathbf{f}_t, \left(1 - z_t^2 + \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\mathbf{h}_{\boldsymbol{\lambda}t}, -\frac{2u_t}{\sigma_u^2}\mathbf{m}_t, \frac{1}{\sigma_u^2}\left(1 - \frac{u_t^2}{\sigma_u^2}\right)\right\}',$$

and the typical element of the score for the GARCH parameter is $d_{\boldsymbol{\lambda}t} = z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}$. We hold $\mathbf{h}_{\boldsymbol{\eta}t} = \mathbf{0}$, under the stochastic recursion of the conditional variance. In addition, we note that $E\left(z_t|\,\mathcal{F}_{t-1}\right) = 0, E\left(z_t^2\middle|\,\mathcal{F}_{t-1}\right) = 1, E\left(z_t^2 - 1\middle|\,\mathcal{F}_{t-1}\right) = 0$ and $E\left(u_t\dot{u}_t|\,\mathcal{F}_{t-1}\right) = 0$.Here, we show the details how to derive $\mathcal{J}_{\boldsymbol{\pi\theta}}$ and $\mathcal{I}_{\boldsymbol{\pi\theta}}$. First, we have defined $\mathcal{J}_{\boldsymbol{\pi\theta}} = \frac{1}{T}\sum_{t=1}^{T} E\left[\mathbf{s}_{\boldsymbol{\pi}t}(\boldsymbol{\theta})\,\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta})'\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \left[\mathcal{J}_{\boldsymbol{\pi\eta}} : \mathcal{J}_{\boldsymbol{\pi\lambda}} : \mathcal{J}_{\boldsymbol{\pi\upsilon}} : \mathcal{J}_{\boldsymbol{\pi}\sigma}\right]$. The each of the elements of $\mathcal{J}_{\boldsymbol{\pi\theta}}$ can be derived with respect to the given parameters as follows.

$$\begin{aligned}
\mathcal{J}_{\boldsymbol{\pi\eta}} &= E\left[\left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\frac{\varepsilon_t}{h_t}\mathbf{w}_t\mathbf{f}_t'\right] \\
&= E\left[E\left(z_t^2 - 1\middle|\,\mathcal{F}_{t-1}\right)\frac{\varepsilon_t}{h_t}\mathbf{w}_t\mathbf{f}_t'\right] - \frac{2}{\sigma_u^2}E\left[E\left(u_t\dot{u}_t|\,\mathcal{F}_{t-1}\right)\frac{\varepsilon_t}{h_t}\mathbf{w}_t\mathbf{f}_t'\right] \\
&= \mathbf{0}.
\end{aligned}$$

$$\begin{aligned}
\mathcal{J}_{\boldsymbol{\pi\lambda}} &= \frac{1}{2}E\left[\left(1 - z_t^2 + \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)^2\mathbf{w}_t\mathbf{h}_{\boldsymbol{\lambda}t}'\right] \\
&= \frac{1}{2}E\left\{\left(1 - z_t^2 + \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)^2\right\}E\left(\mathbf{w}_t\mathbf{h}_{\boldsymbol{\lambda}t}'\right).
\end{aligned}$$

$$\begin{aligned}
\mathcal{J}_{\boldsymbol{\pi\upsilon}} &= E\left[\frac{1}{\sigma_u^2}\left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)u_t\mathbf{w}_t\mathbf{m}_t'\right] \\
&= E\left[\frac{1}{\sigma_u^2}\left(z_t^2 - 1\right)u_t\mathbf{w}_t\mathbf{m}_t'\right] - \frac{2}{\sigma_u^4}E\left(u_t^2\dot{u}_t\mathbf{w}_t\mathbf{m}_t'\right) \\
&= -\frac{2}{\sigma_u^2}E\left(\dot{u}_t\mathbf{w}_t\mathbf{m}_t'\right).
\end{aligned}$$

$$\begin{aligned}
\mathcal{J}_{\boldsymbol{\pi}\sigma} &= E\left[-\frac{1}{2}\left(\frac{\sigma_u^2 - u_t^2}{\sigma_u^4}\right)\left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\mathbf{w}_t\right] \\
&= -\frac{1}{2\sigma_u^4}E\left[\left(\sigma_u^2 - u_t^2\right)\left(z_t^2 - 1\right)\mathbf{w}_t\right] + \frac{1}{\sigma_u^6}E\left[\left(\sigma_u^2 - u_t^2\right)u_t\dot{u}_t\mathbf{w}_t\right] \\
&= \frac{1}{\sigma_u^4}E\left(u_t\dot{u}_t\mathbf{w}_t\right) - \frac{1}{\sigma_u^6}E\left(u_t^3\dot{u}_t\mathbf{w}_t\right) \\
&= -\frac{1}{\sigma_u^6}E\left(u_t^3\right)E\left(\dot{u}_t\mathbf{w}_t\right).
\end{aligned}$$

Next we have defined $\mathcal{I}_{\boldsymbol{\pi}\theta} = -E\left[\frac{\partial \mathbf{S}_{\boldsymbol{\pi}t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{'}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \left[\mathcal{I}_{\boldsymbol{\pi\eta}} : \mathcal{I}_{\boldsymbol{\pi\lambda}} : \mathcal{I}_{\boldsymbol{\pi\upsilon}} : \mathcal{I}_{\boldsymbol{\pi}\sigma}\right]$. The each of the elements of $\mathcal{I}_{\boldsymbol{\pi\theta}}$ can be derived with respect to given parameters as follows.

For $\mathcal{I}_{\boldsymbol{\pi\eta}}$, we firstly calculate

$$\begin{aligned}
\frac{\partial d_{\boldsymbol{\lambda}t}\mathbf{w}_t}{\partial \boldsymbol{\eta}^{'}} &= \left\{\frac{\partial z_t^2}{\partial \boldsymbol{\eta}^{'}} - \frac{2}{\sigma_u^2}\left(\dot{u}_t\frac{\partial u_t}{\partial \boldsymbol{\eta}^{'}} + u_t\frac{\partial \dot{u}_t}{\partial \boldsymbol{\eta}^{'}}\right)\right\}\mathbf{w}_t + \left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\eta}^{'}} \\
&= \left\{-z_t^2\mathbf{h}_{\boldsymbol{\eta}t}^{'} - \frac{2}{\sigma_u^2}\left(\dot{u}_t^2 + u_t\ddot{u}_t\right)\mathbf{h}_{\boldsymbol{\eta}t}^{'}\right\}\mathbf{w}_t + \left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\eta}^{'}} + \left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\eta}^{'}}.
\end{aligned}$$

We take unconditional expectation for both sides of the equation above,

$$\begin{aligned}
E\left[\frac{\partial d_{\boldsymbol{\lambda}t}\mathbf{w}_t}{\partial \boldsymbol{\eta}^{'}}\right] &= E\left[\left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\eta}^{'}}\right] \\
&= E\left[\left(z_t^2 - 1\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\eta}^{'}}\right] - \frac{2}{\sigma_u^2}E\left[u_t\dot{u}_t\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\eta}^{'}}\right] \\
&= E\left[E\left(z_t^2 - 1\middle|\mathcal{F}_{t-1}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\eta}^{'}}\right] - \frac{2}{\sigma_u^2}E\left[E\left(u_t\dot{u}_t\middle|\mathcal{F}_{t-1}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\eta}^{'}}\right] \\
&= \mathbf{0}.
\end{aligned}$$

For $\mathcal{I}_{\boldsymbol{\pi\lambda}}$, we compute

$$\begin{aligned}
\frac{\partial d_{\boldsymbol{\lambda}t}\mathbf{w}_t}{\partial \boldsymbol{\lambda}^{'}} &= \frac{\partial}{\partial \boldsymbol{\lambda}^{'}}\left\{\left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\mathbf{w}_t\right\} \\
&= \left\{\frac{\partial}{\partial \boldsymbol{\lambda}^{'}}\left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\right\}\mathbf{w}_t + \left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\lambda}^{'}} \\
&= -\left\{z_t^2 + \frac{2}{\sigma_u^2}\left(\dot{u}_t^2 + u_t\ddot{u}_t\right)\right\}\mathbf{w}_t\mathbf{h}_{\boldsymbol{\lambda}t}^{'} + \left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\lambda}^{'}}.
\end{aligned}$$

Then,

$$\begin{aligned}
-E\left[\frac{\partial d_{\boldsymbol{\lambda}t}\mathbf{w}_t}{\partial \boldsymbol{\lambda}^{'}}\right] &= E\left[\left\{z_t^2 + \frac{2}{\sigma_u^2}\left(\dot{u}_t^2 + u_t\ddot{u}_t\right)\right\}\mathbf{w}_t\mathbf{h}_{\boldsymbol{\lambda}t}^{'} - \left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\lambda}^{'}}\right] \\
&= E\left(z_t^2\mathbf{w}_t\mathbf{h}_{\lambda t}^{'}\right) + \frac{2}{\sigma_u^2}E\left[\left(\dot{u}_t^2 + u_t\ddot{u}_t\right)\mathbf{w}_t\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right] - E\left[\left(z_t^2 - 1\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\lambda}^{'}}\right] + \frac{2}{\sigma_u^2}E\left(u_t\dot{u}_t\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\lambda}^{'}}\right) \\
&= E\left[E\left(z_t^2\middle|\mathcal{F}_{t-1}\right)\mathbf{w}_t\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right] + \frac{2}{\sigma_u^2}E\left(\dot{u}_t^2\mathbf{w}_t\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right) + \frac{2}{\sigma_u^2}E\left[E\left(u_t\ddot{u}_t\middle|\mathcal{F}_{t-1}\right)\mathbf{w}_t\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right] \\
&\quad -E\left[E\left(z_t^2 - 1\middle|\mathcal{F}_{t-1}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\lambda}^{'}}\right] + \frac{2}{\sigma_u^2}E\left[E\left(u_t\dot{u}_t\middle|\mathcal{F}_{t-1}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\lambda}^{'}}\right] \\
&= E\left(\mathbf{w}_t\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right) + \frac{2}{\sigma_u^2}E\left(\dot{u}_t^2\mathbf{w}_t\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right) \\
&= E\left[\left(1 + \frac{2\dot{u}_t^2}{\sigma_u^2}\right)\mathbf{w}_t\mathbf{h}_{\boldsymbol{\lambda}t}^{'}\right].
\end{aligned}$$

For $\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\upsilon}}$, we calculate

$$\begin{aligned}
\frac{\partial d_{\boldsymbol{\lambda}t}\mathbf{w}_t}{\partial \boldsymbol{\upsilon}'} &= \frac{\partial}{\partial \boldsymbol{\upsilon}'}\left\{\left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\mathbf{w}_t\right\} \\
&= \left\{\frac{\partial}{\partial \boldsymbol{\upsilon}'}\left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\right\}\mathbf{w}_t + \left(z_t^2 - 1 - \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\upsilon}'} \\
&= \frac{2}{\sigma_u^2}\left(\dot{u}_t\mathbf{m}_t' + u_t\dot{\mathbf{m}}_t\right)\mathbf{w}_t + \left(1 - z_t^2 + \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\upsilon}'}.
\end{aligned}$$

Then, taking expectation for both sides of the equation above,

$$\begin{aligned}
-E\left[\frac{\partial d_{\boldsymbol{\lambda}t}\mathbf{w}_t}{\partial \boldsymbol{\upsilon}'}\right] &= -\frac{2}{\sigma_u^2}E\left[\left(\dot{u}_t\mathbf{m}_t + u_t\dot{\mathbf{m}}_t\right)\mathbf{w}_t\right] - E\left[\left(1 - z_t^2 + \frac{2u_t\dot{u}_t}{\sigma_u^2}\right)\frac{\partial \mathbf{w}_t}{\partial \boldsymbol{\upsilon}'}\right] \\
&= -\frac{2}{\sigma_u^2}E\left[\left(\dot{u}_t\mathbf{m}_t + u_t\dot{\mathbf{m}}_t\right)\mathbf{w}_t\right] \\
&= -\frac{2}{\sigma_u^2}E\left(\dot{u}_t\mathbf{w}_t\mathbf{m}_t' + u_t\mathbf{w}_t\dot{\mathbf{m}}_t'\right).
\end{aligned}$$

Finally, $\mathcal{I}_{\boldsymbol{\pi}\boldsymbol{\sigma}}$ can be obtained by calculating

$$\begin{aligned}
E\left[\frac{\partial d_{\boldsymbol{\lambda}t}\mathbf{w}_t}{\partial \sigma_u^2}\right] &= \frac{2}{\sigma_u^4}E\left[E\left(u_t\dot{u}_t \middle| \mathcal{F}_{t-1}\right)\mathbf{w}_t\right] \\
&= 0.
\end{aligned}$$

Chapter 3 Structural Breaks versus Long Memory on Relative Predictive Ability of GARCH Models: A Simulation Study

3.1 Introduction

Over the past few decades, the properties of long-range dependence have been established in a large body of econometric literature. Some approaches have also been suggested and developed to model long memory conditional variance. In general, a long memory behaviour of a time series can be explained by dealing with slowly diminishing impact of shocks which may accommodate non-summable autocovariances of a time series process. Typically, the hyperbolically decaying autocorrelation structure may reflect such a distinctive characteristic of long memory against a short memory process. In this sense, allowing for a fractional differencing parameter contributes to making it more convenient to capture long-range dependence for traditional time series models. Naturally, it has also come along with the introduction of long memory-based conditional heteroskedastic models and recently, a class of the fractionally integrated conditional heteroskedasticity models has been widely used to model the persistence of conditional variance (e.g. Baillie, Bollerslev and Mikkelsen, 1996; Bollerslev and Mikkelsen, 1996; Davidson, 2004).

A growing literature suggests that long memory could be spurious due to neglected structural breaks present in the time series. Diebold (1986) and Lamoureux and Lastrapes (1990) showed that neglected structural breaks may cause misleading inference of the persistence in conditional variance. More recently, Mikosch and Starica (2004) showed that spurious long memory dynamics might be presented due to the non-linearity of the conditional variance process, and an integrated conditional heteroskedastic model could induce a spurious estimation under the constant unconditional variance assumption. In terms of a predictive ability comparison, Diebold and Inoue (2001) argued that long memory may be a useful description, even if the data generating process exhibits structural breaks and weak dependence. In addition, Morana and Beltratti (2004) described that neglecting breaks could be trivial in very short term forecasting, once it allowed for a long memory component in the volatility model. A superior forecast can be obtained at longer horizons by modelling both long memory and structural changes. On the other hand, Starica and Granger (2005) found that non-linear models particularly with structural breaks in unconditional variance can produce better performance in longer horizon forecasts. For generalised autoregressive conditional heteroskedasticity (GARCH), Mikosch and Starica (2004) and Hillebrand (2005) demonstrated that neglecting structural breaks in GARCH parameters possibly brings in a misleading estimation of persistence with upward biases in the GARCH process. Moreover, Rapach and Strauss (2008) have revealed that allowing for the structural breaks in the unconditional variance of exchange rate returns may often improve the in-sample and out-of-sample performance of the GARCH volatility.

As described in the discussions above, in general, some of theoretical and empirical results support that the persistence in conditional variance can be better characterised by a short memory process with structural breaks than stationary long memory process. Thus, it could be admitted that neglecting structural breaks can infer spurious long memory when modelling and forecasting volatility. However, a choice of long memory and structural break in conditional variance modelling would still be quite arguable due to the difficulty distinguishing between long memory and structural breaks. Along these lines, it has been less likely to obtain consistent and completed empirical results which deal with an issue of relative predictive ability between long memory and structural break forecasting models. This empirical feature is indeed involved in the fact that an interplay between long memory and breaks is obviously present in time series as pointed out by Perron (2006). In this sense, we consider the following competing arguments through this study: long memory in variance cannot be fully explained by structural breaks, whilst the presence of structural breaks in the true conditional variance process may also infer spurious long memory.

A proper choice of forecasting models is important to obtain accurate predictions since a misspecified econometric model is more likely to produce a poor in-sample or out-of-sample fit. For this reason, empirical researchers or practitioners want to make a proper selection of forecasting models when true properties of given time series data are uncovered by rigorous statistical tests or reliable empirical evidences. Nevertheless, there is a lack of information which is provided with comprehensive simulation-based evidence on the relative predictive ability of GARCH-type forecasting models with particular emphases on confused arguments in and between structural break and long memory. In this sense, we shed light on such a predictive content via a Monte Carlo simulation experiment with potentially misspecified short or long memory conditional variance models. Specifically, this chapter aims to investigate relative forecasting performance across a class of parsimonious GARCH-based models under the assumption that the true memory properties of the underlying volatility process are already known.

The data generation process (DGP) is broadly categorised with respect to the memory properties through the experiment. We consider error processes exhibiting stationary short memory conditional heteroskedasticity with or without structural change. The other accounts for stationary long memory without structural break. The standard GARCH$(1,1)$ model of Bollerslev (1986) is adopted to generate a set of short memory DGPs with or without structural breaks. We classify the GARCH-based DGPs as follows: lower/medium/higher persistent stationary GARCH; lower/medium/higher persistent GARCH with the single break in the intercept term; GARCH with the single break in the persistent parameter. Also, the experiment accounts for three different locations of the single structural break, which correspond to $30\%, 50\%$ and 70% of the in-sample period, respectively. The fractionally integrated GARCH$(1,d,1)$ (FIGARCH) model of Baillie et al. (1996) is used to produce a synthetic error series for the stationary long memory conditional heteroskedastic process. We consider two stationary FIGARCH DGPs with two different persistence levels, which explicitly stand for the different degrees of long-range dependence in conditional variance.

We utilise a class of parsimonious conditional variance models in the context of GARCH and exponentially weighted moving average (EWMA). Specifically, the estimation models are as follows: GARCH$(1,1)$, FIGARCH$(1,d,1)$, RiskMetrics EWMA of J.P. Morgan (1994) and Long Memory EWMA of Zumbach (2006). Basically, the out-of-sample forecasts of these models are produced by means of a recursive window scheme. Particularly for the GARCH forecast, we also apply the following estimation windows: a rolling window with one-half and one-quarter lengths of the in-sample period and a post-break sample window. In addition, mean and trimmed-mean forecast combinations are utilised to compare with the single model-based forecasts. We consider $1, 5$ and 22-step-ahead forecast horizons. Each step size may imply daily, weekly and monthly forecast of conditional volatility, respectively. The evaluations for relative performance of forecasts are carried out by means of Mean Squared Forecasting Error (MSFE) and the mean of conditional Value at Risk (MVaR) loss functions across the employed forecasting models. The MSFE loss function is mainly used for further econometric loss evaluations due to its convenience in computation. We investigate the rejection frequency using White's (2000) reality check and Hansen's (2005) superior predictive ability. In addition, the pairwise comparisons are considered across the forecasting models applying the modified Diebold and Mariano test statistic of Harvey, Leybourne and Newbold (1997) with pooled loss series.

Our Monte Carlo experiment reveals some interesting findings with respect to the memory property of DGP, the forecast horizon and the level of persistence. It has been shown that the forecasting models which can capture the given properties of the true conditional variance process are generally favourable in the absence of structural breaks. When an artificial break is present in the DGP, in general, the presence of the break is important to set a proper estimation window size for the short memory models in the shorter-run forecasting cases. In the case of a high persistent short memory process with structural break, spurious long memory may often dominate the structural breaks in the longer-run forecasts. When the artificial break is located at a relatively close point of the end of the in-sample, it is hard to find any consistent feature or pattern in terms of forecast superiority between spurious long memory and true long memory. Nevertheless, it can be seen that long memory-based forecasts are generally better off than the competing forecasts. On the other hand, it has also been found that two combined forecasts work well in the presence of a structural break, regardless of the forecast horizon and the level of persistence. The results of the econometric evaluation for the MSFE loss may also support the findings fairly well.

The chapter is organised as follows. Section 3.2 describes the specification of the forecasting models and forecasting methods. Section 3.3 covers a brief introduction to econometric evaluation tools and methods. In Section 3.4, we first detail a design of the Monte Carlo simulation and specific settings for the experiment. Further, the full-sample estimation and out-of sample forecasting results with their statistical evaluations are analysed. Finally, the conclusion of this study is given in Section 3.5.

3.2 Forecasting Conditional Variance

Denote that y_t is a time series process. Under the assumption that the conditional and unconditional means of y_t are zero,

$$\begin{aligned} y_t &= \varepsilon_t \\ \varepsilon_t &= \sqrt{h_t} z_t, \end{aligned}$$

for $t = 1, \ldots, T$, where $z_t \sim \mathcal{N}(0,1)$ is the standardised error process of ε_t, and $h_t = E\left(\varepsilon_t^2 \middle| \mathcal{F}_{t-1}\right)$ is the conditional variance of ε_t, where the conditioning set $\mathcal{F}_{t-1}$ is the σ-field of all available past information set, up to and including time $t-1$.

A class of GARCH models are estimated using the Quasi-Maximum Likelihood (QML) framework. The specifications of the parsimonious conditional variance models are briefly introduced with their s-step ahead of the conditional variance forecast equations. Next, the estimation windows for the forecasting models are specified, followed by the details of the forecast combinations.

3.2.1 Conditional Volatility Models

3.2.1.1 GARCH(1, 1)

The GARCH(1, 1) model is given by

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}.$$

The non-negativeness conditions are $\omega > 0, \alpha \geq 0$ and $\beta \geq 0$. $\alpha + \beta < 1$ which also ensure covariance stationarity of the GARCH conditional variance process. The persistence of conditional variance is measured by $\alpha+\beta$ in GARCH. As $\alpha+\beta$ gets closer to the unity, the GARCH process is more likely to exhibit high persistence. However, it is generally acknowledged that the sample autocorrelations of the GARCH process is decaying exponentially, so that the model might characterise stationary short memory conditional volatility even better. Indeed, $h_t = E\left(\varepsilon_t^2 \middle| \mathcal{F}_{t-1}\right)$, so that h_t is $\mathcal{F}_{t-1}$ measurable. When $\omega = (1-\alpha-\beta)\,\sigma^2$, where σ^2 is an unconditional variance of GARCH process, the s-step ahead forecast of the GARCH$(1,1)$ can be therefore expressed as

$$E\left(h_{t+s}\middle|\mathcal{F}_{t-1}\right) = \sigma^2 + (\alpha+\beta)^s\left(h_t - \sigma^2\right).$$

3.2.1.2 FIGARCH$(1, d, 1)$

The main difference of the FIGARCH model from GARCH is that the model incorporates a slow hyperbolic decay for the autocorrelation of lagged squared errors in the conditional variance with the long memory parameter d. It may be able to explain the long-run dynamics of past shocks to the current level of conditional variance. The ARMA representation of FIGARCH$(1, d, 1)$ is given by

$$h_t = \omega + \left[1 - (1-\beta L)^{-1}(1-\phi L)(1-L)^d\right]\varepsilon_t^2,$$

where L is lag operator[4]. As noted by Baillie et al. (1996), for $0 < d \leqslant 1$, the FIGARCH process does not have a finite unconditional variance, and is not covariance stationary which are a shared feature with the IGARCH process. However, by a direct extension of the corresponding proof for the IGARCH model, they showed that h_t may be also strictly stationary and ergodic, but it all depends on the distribution of ε_t. If $d = 0$, the FIGARCH model reduces to the GARCH process. In addition, following Baillie et al. (1996), a sufficient condition for the non-negativity of the conditional variance for FIGARCH$(1, d, 1)$ model is $\omega > 0$ and $0 \leqslant \beta < d < 1$. The s-step ahead forecast of the FIGARCH$(1, d, 1)$ conditional variance is modelled as

$$E\left(h_{t+s}\middle|\mathcal{F}_{t-1}\right) = \omega + \left[1 - (1-\beta L)^{-1}(1-\phi L)(1-L)^d\right]\varepsilon_{t+s-1}^2.$$

[4] The truncation lag is set at 2000 for all of the estimation and generation of forecasts followed in the experiment.

3.2.1.3 RiskMetrics EWMA

Given data available at time t, for $t = 1, \ldots, T$, the one-step-ahead conditional volatility forecast of the RiskMetrics EWMA model is given by

$$\begin{aligned} \sigma^2_{t+1|t} &= (1-\lambda)\sum_{i=0}^{\infty}\lambda^i \varepsilon^2_{t-i} \\ &= \lambda\sigma^2_{t|t-1} + (1-\lambda)\,\varepsilon^2_t, \end{aligned}$$

where λ is the decay factor, $0 < \lambda < 1$ and set $\lambda = 0.94$ for daily financial asset return, as is usually recommended. The weight λ decays geometrically and plays a role in generating a short memory process. The larger value of λ implies the higher persistence in h_t and the lower response to a shock. It is worth to note that σ^2_t denotes the exponentially weighted moving average variance measurement at a given time t, but it does not hold the same conditional variance assumption as the case of the GARCH model. "$t+1|t$" is read "the time $t+1$ forecast given information up to and including time t." Following J.P. Morgan (1994), based on the idea that the variance forecasts for two consecutive periods are the same, the s-period forecast is defined as

$$\begin{aligned} \sigma^2_{t+s|t} &= \sum_{k=1}^{s} E\left(\sigma^2_{t+k}\middle|\,\mathcal{F}_t\right) \\ &= s \times E\left(\sigma^2_{t+1}\middle|\,\mathcal{F}_t\right) \\ &= s \times \sigma^2_{t+1|t}. \end{aligned}$$

3.2.1.4 Long Memory EWMA

Zumbach (2006, 2011) considers a class of processes where the conditional variance is a linear function of the past squared error series, $\sigma^2_t = \sum_{i=0}^{\infty}\lambda(i)\,\varepsilon^2_{t-i}$, with $\sum_{i=0}^{\infty}\lambda(i) = 1$ and $\lambda(i) > 0$.[5] The long memory conditional variance is defined as the weighted average of R short memory EWMA processes, given by

$$\sigma^2_t = \sum_{r=1}^{R} w_r h_{r,t},$$

[5] $\lambda(i)$ is the geometrical decay factor in the framework of the RiskMetrics EWMA of J.P. Morgan (1994), yielding a short memory process.

where $\sigma_{r,t}^2 = \mu_r \sigma_{r,t-1}^2 + (1-\mu_r)\,\varepsilon_t^2$. The decay factor μ_r of the rth EWMA process is defined as $\mu_r = \exp(-1/\tau_r)$, where a characteristic time structure, $\tau_r = \tau_1 \rho^{r-1}$ for $r = 1, \cdots, R$. The value of ρ is set to $\sqrt{2}$ as recommended by Zumbach (2006). The logarithmically decayed weights is computed by

$$w_r = \frac{1}{C}\left(1 - \frac{\ln(\tau_r)}{\ln(\tau_0)}\right),$$

where the normalisation constant, C, is defined as $C = R - \sum_r^R \frac{\ln(\tau_r)}{\ln(\tau_0)}$ such that $\sum_r^R w_r = 1$. Zumbach (2006) sets the optimal parameter values of $\tau_0 = 1560$, $\tau_1 = 4$ and $\tau_R = 512$, respectively. This is equivalent to $R = 15$. Within this framework, 1-day ahead forecast of conditional variance is defined as

$$\sigma_{t+1|t}^2 = \sum_{i=0}^{\infty}\sum_{r=1}^{R} w_r\,(1-\mu_r)\,\mu_r^i \varepsilon_{t-i}^2.$$

Furthermore, the s-step-ahead cumulative (s-period) forecast can be derived as

$$\sigma_{t+s|t}^2 = s\sum_{i=0}^{T} \lambda(s,i)\,\varepsilon_{t-i}^2,$$

where $\lambda(s,i) = \frac{1}{s}\sum_{r=1}^{R}\sum_{j=1}^{s-1} w_{j,r}\frac{(1-\mu_r)}{1-\mu_r^T}\mu_r^i$ in which $w_{j,r}$ is the rth element of vector $w_j = w'\left(M + (\iota - \mu)\,w'\right)^j$, $w = (w_1, \cdots, w_R)'$, μ is the vector of μ_r, M is the diagonal matrix consisting of μ_r, and ι is a unit vector. As initially conditioned that $\sum_r^R w_r = 1$, it also naturally satisfies that $\sum_{i=0}^{T} \lambda(s,i) = 1$.

3.2.2 Estimation Window

For out-of-sample forecasting, we firstly divide the generated synthetic error series into the in-sample and the out-of-sample period. Denote that s is a forecast horizon and p is a size of out-of-sample. $T-p$ is the length of the in-sample period. For expanding window forecasting, the in-sample observations are used to generate the first out-of-sample forecast. Namely, the initial set of observations spans from the first realisation up to the $(T-p)$th observation. Once we obtain a new forecast, we expand the estimation window by one observation to forecast conditional variance for the next period, say the first observation through observation $T-p+1$. By repeating this procedure up to the end of the available out-of-sample period, we can finally obtain p numbers of out-of-sample forecasts for every single expanding window-based models. GARCH$(1,1)$, FIGARCH$(1,d,1)$ and two EWMA-based models are utilised within the framework of expanding window forecasting. We note that the GARCH$(1,1)$ expanding window forecast is treated as the benchmark for latter use in the evaluation of the MSFE and MVaR loss functions.

We also consider two rolling window forecasts for the standard GARCH model. The model is estimated with two different rolling window sizes that are one-half and one-quarter the lengths of the in-sample period. Let v denote the rolling window size. Thus, $v =$ 0.50 and 0.25 in our case. Then the initial sample size used for the estimation is from $round\left[(1-v)\times(T-p)\right]+1$ to $T-p$. Once we obtain a new forecast, we roll over the estimation window by one observation to forecast conditional variance for the next period. Specifically, the new estimation window covers the observations from $round\left[(1-v)\times(T-p)\right]+2$ to $T-p+1$. We repeat this procedure up to the end of the available out-of-sample period. Corresponding to the given window sizes, those of forecasting models are denoted as GARCH(0.50) and GARCH(0.25), respectively. In using a shorter estimation window, the forecast model has a relatively smaller number of observations available to estimate GARCH parameters, but it is more likely to reduce an overlapping part in data between different regimes. Since the experiment imposes some of GARCH DGPs to have three different locations of the single artificial break point, we would anticipate that the rolling window models can show relatively accurate predictive ability of conditional variance to the others when the regime change point is quite close to the rolling window size. For example, GARCH with 0.50 rolling window could be selected as the best-performing model when the artificial break is placed in the middle of the in-sample period.

Additionally, this study takes a post-break estimation window into account for the GARCH forecast. We adopt the framework used in Rapach and Strauss (2008) (henceforth, RS) which is based on the CUSUM test statistic of Inclan and Tiao (1994) for the detection of multiple structural breaks in unconditional variance against the stationary short memory process. In fact, the asymptotic distribution of the test statistic is obtained under the assumption that a time series follows a Gaussian *i.i.d.* process. However, the test statistic may suffer from upward size distortions as the sample size increases, when a given sequence of observations is a dependent process such as a GARCH process, as shown in Andreou and Ghysels (2002) and Sansó, Aragó and Carrion (2004). To complement this drawback, we utilise the non-parametric-adjusted CUSUM statistic using the Bartlett kernel estimator to test the null of a constant unconditional variance against the alternative that the structural breaks are present in the unconditional variance, based on initial setting of the asymptotically valid test statistic of Sansó et al. (2004). Denote $C_k = \sum_{t=1}^{k} y_t^2$ which is the cumulative sum of the squared return from 1 up to time k. By the assumption that $E\left(\varepsilon_t | \mathcal{F}_{t-1}\right) = 0$, we test the null hypothesis, using the following CUSUM test statistic:

$$Q_T(k) = \sup_k \left| \frac{1}{\sqrt{\hat{\lambda}}} \left(\frac{1}{\sqrt{T}} C_k - \frac{k}{T\sqrt{T}} C_T \right) \right|,$$

where $\hat{\lambda} = \hat{\gamma}_0 + 2\sum_{l=1}^{m} \left[1 - l\,(m+1)^{-1}\right] \hat{\gamma}_l$, in which $\hat{\gamma}_l = T^{-1} \sum_{t=l+1}^{T} \left(y_{i,t}^2 - \hat{\sigma}^2\right)\left(y_{i,t-l}^2 - \hat{\sigma}^2\right)$. $\hat{\sigma}^2 = T^{-1} C_T$, is unconditional variance of the return. m is a lag truncation parameter that can be determined using the procedure in Newey and West (1994). k is the estimate of variance change point which can maximise the function in given statistic. The asymptotic distribution of $Q_T(k)$ is given by

$$Q_T(k) \rightarrow \sup_k \left| W^0(k) \right|,$$

where $W^0(k) = W(k) - kW(1)$ is a Brownian Bridge and $W(k)$ is standard Brownian Motion. We use the same finite sample critical value as the one generated via simulation, provided by Sansó et al. (2004). To test for multiple structural breaks in unconditional variance and estimate the break points, the modified iterated Cumulative Sum of Squares (ICSS) algorithm of Inclan and Tiao (1994) is applied to the $Q_T(k)$ statistic at 5% significance level. Note that the modified ICSS algorithm is applied to the in-sample period, not including out-of-sample observations. If we can detect significant evidence of single or multiple structural breaks, the last break point among all of the estimated break points is used to determine the estimation window size for the GARCH with break forecasts. Specifically, the GARCH model could be estimated using the part of the in-sample observations from $k_f + 1$ up to $T - p$, where k_f is the final structural break point detected. Then, we can obtain the first out-of-sample forecast. If no break is detected, the generated forecasts must be equivalent to the forecast from the GARCH with expanding window. After that, the second out-of-sample forecast can be generated using the observations from the new break point by the modified ICSS to $T - p + 1$. By repeating the described procedure up to the end of the full-sample, we finally obtain p number of out-of-sample forecasts which may account for the potential (final) structural breaks through the entire sample. We denote the forecasts produced by means of this framework as GARCH with breaks or GARCH(break). However, GARCH(break) is likely to suffer from an issue, related to the number of observations to be used for reasonably reliable estimates of the GARCH parameters. Namely, if the detected break point is located too close around the forecast date, then a short sample would be available for estimation.

3.2.3 Forecast Combination

It is generally known that determining the optimal estimation window size for forecasting is not easy in the presence of unknown structural breaks. As emphasised by Pesaran and Timmermann (2007) with a linear regression-based analysis, the trade-off between bias and forecast error variance should be considered to properly select an estimation window size for out-of-sample forecasting. If structural break points are unknown, it has to be concerned whether size, timing and number of breaks are precisely estimated in order to optimally exploit the bias-variance trade-off. Therefore, it would be ideal if we can well characterise those uncertainty when selecting the optimal estimation window in forecasting. However, specifying the structural breaks might inevitably require substantial cost in the simulation experiment, and also it might be beyond the topic of this study, we would rather take an alternative practical way, instead. Pesaran and Timmermann (2007) and Clark and McCracken (2009) suggest one practical way that is combining individual forecasts, generated in various sizes of estimation window. Delivering the Monte Carlo simulation for the linear regression model, they showed that a forecast combination can outperform in the presence of structural breaks relative to the single model-based forecast with the expanding window that may neglect the potential effect of structural breaks. RS also conducted empirical analysis using exchange rate return volatility that forecast combinations of GARCH models with various window sizes can produce a better out-of-sample fit in the presence of structural breaks, particularly in a longer-run forecast.

In this study, mean and trimmed mean forecast combination methods are utilised for the entire set of individual forecasts. The mean combination can be obtained by taking an average of every single model-based forecast in each step of the simulation replication. To obtain trimmed mean combination, we discard each of the best and the worst forecasts from a set of individual forecasts in each step of the replication, then we calculate the average.

3.3 Forecasting Performance Measurement

3.3.1 Loss Functions

We investigate the forecast performance across employed conditional variance forecasting models based on two forecasting loss functions. We use the squared error process as a proxy of the true underlying variance process, following Awartani and Corradi (2005). They demonstrated that squared errors are allowed to be a valid proxy in assessing the relative predictive performance of various asymmetric GARCH models, while the true volatility process is unobservable. For measuring forecasting accuracy, we utilise two loss functions, MSFE and MVaR, to evaluate forecasting errors across the individual forecasts as well as their combinations.

As supported by the empirical investigation of Awartani and Corradi (2005) and Hansen and Lunde (2006) amongst others, the MSFE loss function may produce a consistent empirical ranking of forecasting models when squared errors are used as a proxy for the latent financial asset volatility. Also, Patton (2011) showed that the MSFE is a robust loss function on evaluating relative predictive accuracy of volatility models, so that it may provide a consistent ranking among the competing forecasts. However, Andersen and Bollerslev (1998) argued that the realised squared daily returns could be poor conditional volatility estimators in the presence of large idiosyncratic noise for financial asset return. Therefore, the simpler form of MSFE would be less preferred to the aggregated MSFE which allows users to partly offset some of the idiosyncratic error in ε_t^2 by taking average. In this study, the aggregated version of MSFE loss function used in Starica and Granger (2005) is applied here to measure forecasting performance. It is given by

$$\text{MSFE}_s = \frac{1}{n} \sum_{t=T-p+s}^{T} \left(\tilde{\varepsilon}_t^2 - \tilde{h}_{t|t-s} \right)^2 ,$$

where $\tilde{\varepsilon}_t^2 = \sum_{j=1}^{s} \varepsilon_{t-(j-1)}^2$ and $\tilde{h}_{t|t-s,i} = \sum_{j=1}^{s} \hat{h}_{t-(j-1)|t-s}$, $n = p - s + 1$, is the number of forecasts, p is a number of out-of-sample observations, so that $T - p$ corresponds to the in-sample size, and s denotes the forecast horizon. A smaller MSFE_s indicates a better out-of-sample fit relative to other competing forecasts.

In addition to MSFE, we account for MVaR to evaluate the goodness-of-fit of the out-of-sample forecasts. Specifically, the conditional VaR loss function of González-Rivera, Lee and

Mishra (2004) is adopted for this study. The conditional quantile in the error distribution at time $t+1$ is defined as

$$\Pr\left(\varepsilon_{t+1} \leq \mathrm{VaR}_{t+1}^{\alpha} \middle| \mathcal{F}_t\right) = \delta.$$

The one-step-ahead VaR at given conditional quantile δ can be estimated as

$$\mathrm{VaR}_{t+1}^{\delta} = \rho_{t+1} + \Phi_{t+1}^{-1}(\delta)\, h_{t+1},$$

where Φ_{t+1} is the forecast cumulative distribution function of the standardised error process, $\rho_{t+1} = E\left(\varepsilon_{t+1} \middle| \mathcal{F}_t\right)$ is the conditional mean forecast, and $h_{t+1} = E\left(\varepsilon_{t+1}^2 \middle| \mathcal{F}_t\right)$ is the one-step ahead conditional variance forecast. Within these settings, the mean loss function of the conditional VaR with δ quantile of the cumulative distributions for the cumulative error process is given by

$$\mathrm{MVaR}_s^{\delta} = \frac{1}{n} \sum_{t=T-p+s}^{T} \left(\delta - I_t^{\delta}\right)\left(\tilde{\varepsilon}_t - \mathrm{VaR}_{t|t-s}^{\delta}\right),$$

where $\mathrm{VaR}_{t|t-s}^{\delta}$ is the forecast of the cumulative distribution function of $\tilde{\varepsilon}_t$ at δ quantile, in which $\tilde{\varepsilon}_t = \sum_{j=1}^{s} \varepsilon_{t-(j-1)}$ is the cumulative error process, formed at time $t-s$. $I_t^{\delta} = I\left(\tilde{\varepsilon}_t < \mathrm{VaR}_{t|t-s}^{\delta}\right)$, where I is an indicator function that takes a value of unity when $\tilde{\varepsilon}_t < \mathrm{VaR}_{t|t-s}^{\delta}$, otherwise zero. If $\tilde{\varepsilon}_t$ is less than $\mathrm{VaR}_{t|t-s}^{\delta}$, the absolute value of weight to the difference between $\tilde{\varepsilon}_t$ and $\mathrm{VaR}_{t|t-s}^{\delta}$ is $1-\delta$. In contrast, a bigger value of $\tilde{\varepsilon}_t$ than $\mathrm{VaR}_{t|t-s}^{\delta}$ is associated with much smaller weight, δ, to that difference. Thus, this kind of asymmetric feature of MVaR may be driven by allowing for the indicator function for lager cumulative losses. Besides, we expect that the mean loss of the conditional VaR may partly rule out a potential effect from latent volatility which is often observed in macroeconomic and financial time series variables. We consider 5% conditional quantile, $\delta = 0.05$. With $z_t \sim \mathcal{N}(0,1)$, we generate the single series of ε_t using the estimates of the conditional variance at each time of forecast, then calculate $\tilde{\varepsilon}_t$ at a given time period. Replicating this procedure 2000 times, we obtain the empirical distribution of $\tilde{\varepsilon}_t$, then $\mathrm{VaR}_{t|t-s}^{0.05}$ is collected at the 100th element in the ordered sequence. In comparison analysis, a smaller MVaR stands for a better goodness-of-fit for the out-of-sample forecasts to the actual series of y_t.

3.3.2 Forecasting Loss Evaluation

Although a forecasting model ranking based on a loss comparison could be informative in a selection of the best-performing (the smallest loss) forecasting model, it does not mean that the loss difference between forecasting models is statistically significant. Moreover, it does not indicate whether such a rank order is consistently robust in a different sample. In this sense, we briefly introduce several econometric tests to examine the MSFE loss of the conditional variance forecast. Firstly, we apply the equality predictive ability test of Diebold and Mariano (1995) and West (1996) (henceforth DMW) to analyse pairwise comparison of the competing forecasting models. Within a generic framework of DMW, the null hypothesis is $\mathbf{H}_0 : E[d_t] = 0$,where $d_t = L\left(h_{t,i}, \varepsilon_t^2\right) - L\left(h_{t,j}, \varepsilon_t^2\right)$, that is the MSFE loss differential between forecasting model i and j for $i \neq j$. The original DMW test statistic is given by

$$T^{DMW} = \overline{d}\widehat{V}^{-1/2}, \tag{21}$$

where $\overline{d}_n$ is the sample mean of d_t which is $\overline{d}_n = \sum_{t=(T-p)+s}^{T} d_t/n$. $\widehat{V}$ is the asymptotic long-run variance of $\sqrt{n}\overline{d}_n$. The DM statistic is asymptotically distributed standard normal for non-nested model comparison as shown in West (1996). Further, we also utilise the modification version of (21) for improving finite sample performance, proposed by Harvey et al. (1997). The modified DMW test statistic (henceforth MDM) is given by

$$T^{MDM} = \left[\frac{n+1-2s+n^{-1}s\left(s-1\right)}{n}\right]^{1/2} T^{DMW}.$$

T^{MDM} is evaluated against the critical values from the Student's t distribution with $(n-1)$ degrees of freedom, rather than from the standard normal distribution in the case of T^{DMW}. To save space, we only report the results of the MDM tests in this chapter. The rejection of the null hypothesis indicates that the i model outperforms the j model when $T^{MDM} < 0$, and vice versa. However, the MDM has some drawbacks when it applies to a large set of competing forecasting models. For example, White (2000) points out a data snooping problem, and Hansen, Lunde and Nason (2011) concerns a high-dimensionality issue in the estimation of a covariance matrix with a large number of competing forecasting models, amongst others.

Further, White's (2000) reality check (henceforth RC) and Hansen's (2005) superior predictive ability (henceforth SPA) tests are conducted for the multiple comparison of competing forecasts. As the general aspects of RC and SPA are identical, we are interested in the null hypothesis that one particular model fixed as the benchmark is not worse than any of the competing forecasts in terms of expected loss. The null hypothesis is given by $\mathbf{H}_0 : E\left[\mathbf{d}_t\right] \leq \mathbf{0}$,where $\mathbf{d}_t = \left(d_{t,1}, \ldots, d_{t,j}\right)$, in which $d_{t,i} = L\left(h_{t,0}, \varepsilon_t^2\right) - L\left(h_{t,i}, \varepsilon_t^2\right)$, for $i = 1, \ldots, j$, j is the number of competing models ($j = 8$ in this study). Let $d_{t,i}$ denote the loss differential of forecast i relative to the benchmark forecast. The RC and SPA test statistics are given by

$$\begin{aligned} T^{RC} &= \max_i \left(\sqrt{n}\overline{d}_1, \ldots, \sqrt{n}\overline{d}_i\right) \\ T^{SPA} &= \max\left[\max_i \frac{\sqrt{n}\overline{d}_i}{\widehat{\sigma}_i}, 0\right] \end{aligned}$$

respectively, where $\overline{d}_i = \sum_{t=(T-p)+s}^{T} d_{t,i}/n$ and $\widehat{\sigma}_i^2$ is a consistent estimator of the asymptotic variance of $\sqrt{n}\overline{d}_i$. In our applications, we set each of individual forecasts as the benchmark and the rest of the others as an alternative for the comparison. The loss function for both RC and SPA test statistics are derived using a simple mean squared error metric. The p-values of the given test statistics are computed using the stationary bootstrap of Politis and Romano (1994) with 1000 bootstrap replications.[6] A high p-value implies that we cannot reject that a benchmark model does not outperform competing models.

3.4 Monte Carlo Experiment

3.4.1 Simulation Design

We design the Monte Carlo simulation experiment with respect to the properties of the true conditional variance data generation process: the level of persistence, the location of structural breaks and the memory property. Each DGP is generated with 5600 observations from the standard normal distribution with zero mean and unit variance.[7] We discard the first 3000 observations to remove the initialisation effect, so that the total number of observations is 2600 for each replication. The forecasting models are estimated and replicated 1000 times. The size of the out-of-sample forecasts is set to 100 observations. We generate 1-step, 5-step and 22-step-ahead forecasts to compare the relative predictive accuracy with regard to the

[6] For SPA, we report the consistent p-value of T^{SPA}.

[7] Random numbers are generated via GAUSS12.

length of the forecast horizon.

The synthetic error process with stationary short memory conditional heteroskedasticity is generated by the standard GARCH$(1, 1)$ model. The level of persistence is determined by a GARCH persistence parameter, $\alpha + \beta$. Without any loss of generality, the unconditional variance of ε_t is imposed to be the unity over the entire cases. Next, the GARCH with structural break DGPs are produced by dealing with the level of unconditional variance or persistence level. We set the single structural break at $30\%, 50\%$ and 70% of the in-sample period. Each of the structural break points corresponds to 750th, 1250th and 1750th observation, respectively. Holding other parameter values fixed, the structural break is imposed to the intercept or the persistence parameter of the GARCH$(1, 1)$ model by changing the value of ω or β, respectively. Particularly, for the change in ω, the magnitude of a change in the unconditional variance is set to 5, equivalent to 2.24-standard deviation. Namely, the unconditional variance of the post-break sample period is 5 times greater than the pre-break sample period. For the GARCH DGP with structural break in persistence, we consider a change in the value of the persistence parameter from 0.95 to 0.99 without any changes in the intercept parameter. Overall, the level of the unconditional variance is bounded from 0.2 to 1.0 for all the GARCH-Break DGPs. On the other hand, for the generation of stationary long memory process, we use the FIGARCH$(1, d, 1)$ model, and account for two different degrees of long memory parameter.[8] The initial parameter values of the FIGARCH DGPs are taken from Rapach and Strauss (2008) and Baillie et al. (1996). The parameter values used for the data generation are presented in Table 10.

3.4.2 QML Estimates for GARCH and FIGARCH

The average of full-sample QML estimates for GARCH$(1, 1)$ and FIGARCH$(1, d, 1)$ models are displayed in Table 11. We also report the average values of the stationarity and the positiveness conditions for the GARCH and the FIGARCH models.[9] For the short memory stationary GARCH DGPs (**LP**, **MP**, **HP**), it can be seen that the average estimate of ω is

[8] Indeed, the FIGARCH process may present a hyperbolic memory decaying feature, but its sum of the autocovariance is finite. Strictly speaking, thus, the FIGARCH process is not a true long memory by definition. In this paper, however, we admit the fact that the FIGARCH volatility can exhibit even "longer" memory process relative to a stationary short memory volatility process, and we account for the FIGARCH process as long memory in a wide sense, as generally represented in a variant of empirical studies.

[9] Although we report here the average values only, the result of the estimates for each of the individual Monte Carlo replications is available from authors upon request.

Table 10: Parameter values for data generation

	GARCH			FIGARCH		GARCH-Break			
	LP	**MP**	**HP**	**LM1**	**LM2**	**LP-B**	**MP-B**	**HP-B**	**BinP**
ω_0	0.20	0.05	0.01	0.10	0.02	0.04	0.01	0.002	0.01
ω_1						0.20	0.05	0.01	
$\alpha(\phi)$	0.05	0.05	0.05	0.25	0.10	0.05	0.05	0.05	0.05
β_0	0.75	0.90	0.94	0.60	0.68	0.75	0.90	0.94	0.90
β_1									0.94
d				0.45	0.65				
$\alpha+\beta$	0.80	0.95	0.99			0.80	0.95	0.99	0.95/0.99

Note: **LP**, **MP** and **HP** indicate GARCH(1,1) with low, medium and high persistence, respectively. **LP-B**, **MP-B** and **HP-B** allow the single artificial break in the intercept parameter ω of GARCH(1,1) process. **BinP** has the break in the parameter β, exhibiting the structural break in persistence level of GARCH(1,1). **LM1** and **LM2** are stationary FIGARCH(1,d,1) processes with a different value of d.

slightly overestimated, whereas the average estimate of β is underestimated. Specifically, $\overline{\omega}$ is underestimated by 0.04 when the DGP exhibits lower persistent process. It may contribute to the overestimation of $\overline{\beta}$ by 0.05 along with a decrease in persistent level by 0.04, compared to the true parameter values. For the **MP** and **HP**, there is no big difference between the estimated and the true parameter values. The average long memory parameter estimates, $\overline{d}$, for the FIGARCH model increases holding the stationarity condition, as the level of persistence of the true short memory GARCH process goes up.

For the stationary long memory DGPs (**LM1**, **LM2**), it seems that the GARCH estimates mirror a change in d of the true processes. Say, the persistent level of GARCH in **LM1** is slightly smaller than that in **LM2**, by 0.01. It can also be observed that the long memory dynamics may lead to a rise in α, on average, rather than ω or β in the GARCH estimates, holding the level of the unconditional variance equal. In this sense, we infer that the squared lagged errors would show larger contributions to explain the dynamics of current GARCH conditional variance in the stationary long memory environment compared to **MP** and **HP.** For the FIGARCH estimates, on the other hand, we can find that the average of the estimated intercept parameters for both of the stationary FIGARCH DGPs is slightly biased upwards.

In the case of the full-sample estimation for the GARCH-Break DGPs (**LP-B**, **MP-B**, **HP-B, BinP**), the GARCH full-sample estimates generally assign extremely high persistence $\left(\overline{\alpha}+\overline{\beta}\approx 1\right)$ with the large values of $\overline{\beta}$, regardless of the locations of the artificial structural break. We conjecture that such a large integration is spuriously driven by the presence of the structural break. Given a change in the unconditional variance, $\overline{\omega}$ is quite close to zero, and $\overline{\alpha}$

Table 11: Average of the estimates for GARCH and FIGARCH models

	GARCH - *Lower regime*				GARCH - *Upper regime*				GARCH - *Full sample*				FIGARCH - *Full sample*					
	$\bar{\omega}$	$\bar{\alpha}$	$\bar{\beta}$	$\bar{\alpha}+\bar{\beta}$	$\bar{\omega}$	$\bar{\alpha}$	$\bar{\beta}$	$\bar{\alpha}+\bar{\beta}$	$\bar{\omega}$	$\bar{\alpha}$	$\bar{\beta}$	$\bar{\alpha}+\bar{\beta}$	$\bar{\omega}$	$\bar{\phi}$	$\bar{\beta}$	$\bar{d}$	$\bar{\phi}+\bar{d}$	$1-2\bar{\phi}$
LP									0.241	0.052	0.707	0.759	0.606	0.181	0.197	0.066	0.247	0.637
MP									0.056	0.051	0.893	0.944	0.248	0.252	0.409	0.205	0.458	0.495
HP									0.012	0.050	0.937	0.988	0.073	0.245	0.673	0.475	0.720	0.510
LM1									0.046	0.089	0.890	0.979	0.126	0.227	0.581	0.450	0.677	0.546
LM2									0.021	0.109	0.883	0.992	0.082	0.085	0.689	0.674	0.758	0.831
Break point: 0.3																		
LP-B	0.043	0.058	0.730	0.788	0.236	0.051	0.714	0.764	0.001	0.039	0.961	0.999	0.052	0.301	0.619	0.380	0.682	0.398
MP-B	0.012	0.056	0.884	0.940	0.067	0.051	0.883	0.933	0.002	0.054	0.945	0.999	0.041	0.273	0.650	0.432	0.705	0.454
HP-B	0.003	0.055	0.934	0.989	0.029	0.041	0.931	0.972	0.001	0.057	0.942	0.999	0.027	0.189	0.757	0.612	0.801	0.621
BinP	0.007	0.054	0.918	0.972	0.030	0.041	0.930	0.971	0.002	0.055	0.944	0.999	0.035	0.240	0.698	0.506	0.746	0.520
Break point: 0.5																		
LP-B	0.135	0.042	0.825	0.867	0.756	0.038	0.812	0.849	0.004	0.036	0.963	0.999	0.201	0.299	0.614	0.379	0.678	0.401
MP-B	0.010	0.053	0.900	0.953	0.079	0.051	0.871	0.922	0.001	0.053	0.946	0.999	0.031	0.270	0.650	0.435	0.705	0.460
HP-B	0.002	0.054	0.940	0.994	0.034	0.040	0.926	0.966	0.001	0.056	0.943	0.999	0.020	0.177	0.767	0.636	0.813	0.646
BinP	0.005	0.051	0.928	0.979	0.035	0.039	0.926	0.965	0.002	0.053	0.945	0.998	0.029	0.255	0.676	0.470	0.725	0.490
Break point: 0.7																		
LP-B	0.141	0.040	0.820	0.860	0.636	0.034	0.840	0.873	0.004	0.035	0.964	0.999	0.180	0.302	0.599	0.362	0.664	0.395
MP-B	0.009	0.051	0.904	0.955	0.091	0.051	0.858	0.909	0.002	0.052	0.945	0.997	0.027	0.272	0.639	0.424	0.696	0.456
HP-B	0.002	0.053	0.940	0.993	0.046	0.037	0.917	0.954	0.001	0.055	0.943	0.998	0.017	0.172	0.769	0.643	0.815	0.655
BinP	0.005	0.050	0.927	0.977	0.039	0.036	0.923	0.960	0.002	0.051	0.945	0.996	0.027	0.271	0.649	0.430	0.701	0.459

is biased downwards by 20% in **LP-B**, whereas $\overline{\alpha}$ is estimated around its true values in **MP-B** and **HP-B.** The parameter of lagged conditional variance is considerably over-estimated in **LP-B** and **MP-B,** similar to the empirical results of Mikosch and Starica (2004) and Hillebrand (2005). Therefore, a substantial overestimation of β should stem from the presence of a structural break in the unconditional variance, and may also cause a misleading result of persistence level in GARCH estimation. On the other hand, it is less likely to have a clear look for the effect of the structural change in the persistence level in the DGP of **BinP**. The estimated coefficient for the lagged dependent variable is around its true parameter value of the upper-variance (second) regime at 0.95. It implies that the GARCH model could be less capable of distinguishing a type of structural break in the unconditional variance, in terms of the break in either intercept or persistence parameter. In the results of the FIGARCH estimates, we can observe that the degree of $\overline{d}$ is generally large relative to that of $\overline{d}$ in the stationary short memory GARCH DGPs, regardless of the persistence level. It may arise from allowing for the structural break in the unconditional variance. Since a short memory GARCH process is contaminated by the structural break, it may suffer from spurious long memory with a higher degree of the fractional integration.

To view clearly the effect of structural break on the GARCH estimates, we separately estimate the standard GARCH model using a different part of the observations which belong to the lower-variance (first) regime, and which belong to the upper-variance (second) regime, respectively. As the most distinctive feature, in general, it can be seen that the persistence levels of the estimated GARCH processes from both of the regimes are considerably lower than the estimate from the full-sample. And also, the estimated persistence parameters are similar with the true persistence level. In this sense, we can confirm spurious increases in the magnitude of persistence in the full-sample GARCH estimation. In the case of the lower regime, the estimated parameters are very similar with their true values used for the data generation for **MP-B** and **HP-B**, regardless of the break location. For the same DGP, however, it can be seen that $\overline{\omega}$ tends to be overestimated, but $\overline{\beta}$ tends to be underestimated in the second regime estimation, as the structural break point is getting closer to the end of the in-sample period. In particular, $\overline{\alpha}$ is biased downwards too, in the high persistent DGP (**HP-B**). This would be due to the number of observations available in each regime to be used for the estimation which varies depending on the location of the structural break. Namely, the undersized α or β represent relatively small contributions to the current conditional variance, explained by the past information content in the dynamics of the return or the conditional volatility. Thus, ω can possibly be overestimated to maintain the given level of unconditional variance and persistence. As a consequence, the largest magnitude of misleading estimation is observed in the most recent break DGP cases. For **LP-B**, the estimated GARCH process seems not to mirror its true process well, viewing overestimation features in $\overline{\omega}, \overline{\beta}$and $\overline{\alpha} + \overline{\beta}$, and even worse in the upper regime cases. Moreover, the effect of the structural break on the persistence level has moderated in the estimated GARCH process of **BinP**, so that it cannot clearly make up the regime differences in terms of persistence level. As a result, the estimated levels of persistence in both regimes are around the average of the pre- and post-break levels.

3.4.3 Out-of-Sample Forecast Results

3.4.3.1 Non-Break DGPs

We evaluate the forecasting performance using the MSFE and the MVaR metric for the stationary short memory GARCH DGPs and the stationary long memory FIGARCH DGPs. We report the average MSFE and MVaR ratios to the benchmark GARCH$(1, 1)$ expanding window model in Table 12. First, for the stationary short memory cases, the benchmark GARCH expanding window model is preferred to the competing models regardless of the length of forecast horizon as well as the level of persistence. In the DGP with lower persistence (**LP**), the benchmark is followed by either FIGARCH or GARCH with break forecast, even though long memory or changes in regime are not presumed for the true data generation. When the level of persistence is 0.99 (**HP**), it is seen that the forecast combinations outperform the single model-based forecasts. Evaluating the MVaR loss function, a set of individual forecasts which reveals poor MSFE performance remains inferior in each of the corresponding DGPs. However, the best-performing individual forecasts in terms of MVaR are not exactly matched with the MSFE results. For **LP**, the benchmark is still the most favourable in the short forecast horizon. The FIGARCH forecast is chosen as the most superior in the longer-run horizon. Further, we can see that the forecast combinations for **MP** and **HP** generate more accurate forecasts than any other single model-based forecasts in terms of the MVaR performance.

Over the long memory data generation processes, both the long memory-based forecast and

Table 12: Average MSFE and MVaR ratios of GARCH and FIGARCH DGPs

	LP			**MP**			**HP**			**LM1**			**LM2**		
	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$
MSFE ratio															
GARCH	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
GARCH(0.50)	1.0015	1.0055	1.0233	1.0013	1.0047	1.0189	1.0015	1.0069	1.0327	1.0008	1.0040	1.0258	1.0011	1.0064	1.0367
GARCH(0.25)	1.0037	1.0122	1.0563	1.0046	1.0173	1.0640	1.0051	1.0225	1.0969	1.0028	1.0106	1.0559	1.0030	1.0119	1.0634
GARCH(break)	1.0004	1.0017	1.0141	1.0012	1.0039	1.0167	1.0098	1.0415	1.1367	1.0076	1.0280	1.0988	1.0084	1.0314	1.0842
FIGARCH	1.0005	1.0011	1.0100	1.0019	1.0080	1.0297	1.0056	1.0246	1.0892	0.9979	0.9891	0.9788	0.9988	0.9965	1.0008
EWMA	1.0241	1.1285	1.7719	1.0130	1.0688	1.4189	1.0029	1.0159	1.1146	1.0009	0.9980	1.0296	0.9996	0.9897	0.9150
LM-EWMA	1.0136	1.0887	1.6309	1.0054	1.0351	1.2836	1.0037	1.0163	1.0808	0.9951	0.9729	0.9501	1.0096	1.0164	0.9185
Comb.Mean	1.0014	1.0110	1.0973	1.0007	1.0041	1.0447	1.0008	1.0035	1.0226	0.9956	0.9798	0.9527	0.9951	0.9786	0.9370
Comb.T-Mean	1.0008	1.0065	1.0630	1.0004	1.0026	1.0323	1.0006	1.0032	1.0205	0.9957	0.9801	0.9517	0.9954	0.9802	0.9391
MVaR ratio															
GARCH	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
GARCH(0.50)	1.0010	1.0009	1.0001	1.0001	0.9998	1.0002	1.0002	0.9999	1.0006	0.9997	0.9996	1.0026	1.0001	1.0001	1.0048
GARCH(0.25)	1.0026	1.0010	1.0007	1.0026	1.0022	1.0005	1.0032	1.0031	1.0030	1.0017	1.0009	1.0007	1.0022	1.0013	1.0035
GARCH(break)	1.0001	1.0006	1.0009	1.0011	1.0005	0.9993	1.0064	1.0047	1.0061	1.0058	1.0046	1.0068	1.0076	1.0059	1.0075
FIGARCH	1.0005	0.9994	0.9958	1.0006	1.0015	0.9980	1.0014	1.0017	1.0007	0.9984	0.9979	0.9903	0.9978	0.9975	0.9913
EWMA	1.0173	1.0182	1.0264	1.0079	1.0092	1.0174	1.0029	1.0019	1.0035	1.0058	1.0018	1.0016	1.0100	1.0057	0.9988
LM-EWMA	1.0096	1.0116	1.0202	1.0029	1.0032	1.0083	1.0022	1.0010	1.0018	1.0010	0.9973	0.9951	1.0122	1.0076	1.0006
Comb.Mean	1.0004	1.0006	1.0012	0.9991	0.9990	0.9980	0.9994	0.9988	0.9978	0.9977	0.9962	0.9926	0.9991	0.9974	0.9939
Comb.T-Mean	1.0000	1.0001	1.0003	0.9991	0.9988	0.9976	0.9993	0.9988	0.9978	0.9976	0.9962	0.9926	0.9987	0.9974	0.9942

combined forecasts generally show superiority relative to the other short memory GARCH-based forecasts in terms of both the MSFE and MVaR loss measures. For the MSFE loss ratio of **LM1**, the best forecast is generated by LM-EWMA over the entire forecast horizons, followed by combined forecasts. In the results of MVaR, the forecast combinations are still favourable relative to both the GARCH-based forecasts and the FIGARCH forecast. For higher persistent DGP of **LM2**, on the other hand, it can be found that the combined forecasts and FIGARCH forecasts are better off than the others in terms of MSFE and MVaR for $s = 1$, whereas the performance of LM-EWMA forecast is even poorer than before. The RiskMetrics EWMA model mostly dominates the other model-based forecasts in relatively long-run forecasting in terms of MSFE, as partly evidenced by the empirical study of Harris and Nguyen (2013). They showed that a class of EWMA models could be counted as the proficient conditional variance model in the covariance forecasts of a financial asset portfolio. On the average ratio of MVaR loss, it is observed that the most accurate forecast may be generated by either FIGARCH or forecasting combination. Based on the results obtained above, we can extrapolate that the model which mirrors properties of the true DGP seems to offer superior predictive accuracy in terms of the average MSFE and MVaR ratios. Forecast combinations are generally favourable, regardless of the property of the true data generation.

3.4.3.2 GARCH with Single Artificial Structural Break DGPs

We now turn to the analysis of forecasting performance when the artificial structural break is allowed for the short memory GARCH process. The results of the average MSFE and MVaR loss ratios are reported in Table 13 and 14, respectively. The analysis is followed up with respect to the persistence level.

Lower-Level of Persistence with a Structural Break Considering the MSFE loss function, it can be seen that single model-based forecasts produced by GARCH rolling window and GARCH with a break consistently preferred. The combined forecasts also take relatively higher places in terms of the performance ranking, whereas the benchmark and two EWMA-type models produce even poorer forecasts than other competing forecasts. The relative performance of the GARCH rolling window forecast depends on the size of the estimation window under the structural break. Specifically, if the synthetic break is placed at earlier times

Table 13: Average MSFE ratio of GARCH-Break DGPs

	LP-B			MP-B			HP-B			BinP		
	$s = 1$	$s = 5$	$s = 22$	$s = 1$	$s = 5$	$s = 22$	$s = 1$	$s = 5$	$s = 22$	$s = 1$	$s = 5$	$s = 22$
Break point: 0.3												
GARCH	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
GARCH(0.50)	0.9869	0.9354	0.7678	0.9887	0.9486	0.8261	0.9881	0.9932	0.9989	0.9985	0.9952	1.0066
GARCH(0.25)	0.9895	0.9426	0.7843	0.9924	0.9606	0.8525	1.0019	1.0082	1.0432	1.0021	1.0010	1.0501
GARCH(break)	0.9862	0.9351	0.7724	0.9902	0.9538	0.8434	1.0092	1.0367	1.0952	1.0100	1.0389	1.1039
FIGARCH	0.9960	0.9822	0.8855	0.9926	0.9639	0.8463	0.9997	0.9961	0.9734	0.9991	0.9946	0.9737
EWMA	1.0058	1.0335	1.1402	0.9997	1.0004	1.0102	0.9993	0.9970	0.9975	0.9997	0.9991	1.0068
LM-EWMA	0.9964	1.0003	1.0686	0.9924	0.9716	0.9395	1.0001	1.0004	0.9982	1.0005	1.0023	1.0062
Comb.Mean	0.9887	0.9516	0.8335	0.9899	0.9543	0.8397	0.9975	0.9885	0.9591	0.9978	0.9899	0.9643
Comb.T-Mean	0.9889	0.9522	0.8322	0.9898	0.9544	0.8400	0.9973	0.9878	0.9588	0.9975	0.9891	0.9643
Break point: 0.5												
GARCH	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
GARCH(0.50)	0.9869	0.9356	0.7753	0.9897	0.9526	0.8429	0.9981	0.9931	1.0012	0.9994	0.9993	1.0302
GARCH(0.25)	0.9890	0.9430	0.7944	0.9934	0.9649	0.8709	1.0024	1.0106	1.0550	1.0035	1.0166	1.0854
GARCH(break)	0.9873	0.9383	0.7831	0.9935	0.9666	0.8766	1.0117	1.0477	1.1322	1.0140	1.0576	1.1720
FIGARCH	0.9967	0.9862	0.9113	0.9936	0.9682	0.8607	0.9997	0.9965	0.9760	0.9992	0.9951	0.9770
EWMA	1.0066	1.0389	1.1709	1.0007	1.0049	1.0301	0.9997	0.9993	1.0053	1.0010	1.0056	1.0337
LM-EWMA	0.9965	1.0039	1.0943	0.9933	0.9758	0.9578	1.0004	1.0019	1.0057	1.0017	1.0082	1.0337
Comb.Mean	0.9888	0.9530	0.8416	0.9908	0.9580	0.8526	0.9977	0.9898	0.9651	0.9987	0.9942	0.9845
Comb.T-Mean	0.9888	0.9526	0.8377	0.9907	0.9578	0.8521	0.9975	0.9889	0.9638	0.9984	0.9932	0.9827
Break point: 0.7												
GARCH	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
GARCH(0.50)	1.0076	1.0415	1.1824	1.0055	1.0284	1.1309	1.0032	1.0151	1.0621	1.0043	1.0219	1.1044
GARCH(0.25)	0.9903	0.9491	0.8151	0.9951	0.9736	0.9105	1.0041	1.0212	1.1113	1.0053	1.0254	1.1391
GARCH(break)	0.9903	0.9500	0.8171	0.9993	0.9901	0.9534				1.0184	1.0771	1.2488
FIGARCH	0.9981	0.9942	0.9593	0.9953	0.9765	0.8897	1.0003	0.9985	0.9833	0.9997	0.9974	0.9861
EWMA	1.0080	1.0457	1.1987	1.0025	1.0139	1.0711	1.0011	1.0060	1.0278	1.0028	1.0147	1.0769
LM-EWMA	0.9976	1.0099	1.1178	0.9949	0.9837	0.9933	0.9992	1.0016	1.0233	1.0027	1.0142	1.0699
Comb.Mean	0.9932	0.9745	0.9180	0.9946	0.9756	0.9190	0.9985	0.9950	0.9892	1.0002	1.0012	1.0137
Comb.T-Mean	0.9941	0.9788	0.9315	0.9948	0.9768	0.9238	0.9989	0.9965	0.9936	0.9998	1.0000	1.0117

Table 14: Average MVaR ratio of GARCH-Break DGPs

	LP-B			MP-B			HP-B			BinP		
	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$
Break point: 0.3												
GARCH	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
GARCH(0.50)	0.9928	0.9926	0.9960	0.9937	0.9937	0.9978	0.9983	0.9991	1.0034	0.9984	0.9993	1.0035
GARCH(0.25)	0.9947	0.9937	0.9944	0.9965	0.9961	0.9968	1.0012	1.0021	1.0051	1.0065	1.0023	1.0052
GARCH(break)	0.9934	0.9936	0.9956	0.9957	0.9953	0.9981	1.0064	1.0058	1.0095	1.0065	1.0063	1.0097
FIGARCH	0.9977	1.0007	0.9939	0.9960	0.9971	0.9908	0.9991	1.0001	0.9968	0.9988	1.0002	0.9974
EWMA	1.0037	1.0051	1.0070	1.0009	1.0006	1.0016	1.0007	1.0001	1.0010	1.0009	1.0003	1.0014
LM-EWMA	0.9975	0.9993	1.0021	0.9961	0.9957	0.9968	1.0003	0.9998	1.0024	1.0004	0.9999	1.0026
Comb.Mean	0.9925	0.9928	0.9914	0.9936	0.9931	0.9915	0.9977	0.9977	0.9980	0.9977	0.9979	0.9981
Comb.T-Mean	0.9927	0.9929	0.9912	0.9937	0.9931	0.9913	0.9975	0.9977	0.9979	0.9976	0.9979	0.9981
Break point: 0.5												
GARCH	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
GARCH(0.50)	0.9902	0.9917	0.9908	0.9932	0.9936	0.9993	0.9976	0.9986	1.0037	0.9976	0.9987	1.0054
GARCH(0.25)	0.9917	0.9924	0.9918	0.9959	0.9959	0.9984	1.0008	1.0019	1.0021	1.0008	1.0059	1.0076
GARCH(break)	0.9903	0.9914	0.9913	0.9967	0.9964	1.0014	1.0074	1.0071	1.0135	1.0076	1.0076	1.0164
FIGARCH	0.9988	1.0008	0.9895	0.9963	0.9981	0.9923	0.9993	1.0005	0.9975	0.9988	1.0009	0.9990
EWMA	1.0043	1.0057	1.0066	1.0003	1.0004	1.0027	1.0003	0.9999	1.0015	1.0003	1.0001	1.0034
LM-EWMA	0.9970	0.9999	1.0030	0.9956	0.9955	0.9981	0.9998	0.9996	1.0029	0.9998	0.9998	1.0047
Comb.Mean	0.9913	0.9926	0.9900	0.9933	0.9931	0.9925	0.9974	0.9975	0.9985	0.9973	0.9977	1.0000
Comb.T-Mean	0.9913	0.9926	0.9899	0.9933	0.9930	0.9923	0.9972	0.9976	0.9984	0.9971	0.9977	0.9999
Break point: 0.7												
GARCH	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
GARCH(0.50)	1.0028	1.0049	1.0089	0.9992	1.0005	1.0087	1.0001	1.0002	1.0033	0.9985	0.9994	1.0080
GARCH(0.25)	0.9918	0.9931	0.9957	0.9952	0.9955	1.0027	0.9999	1.0002	1.0067	0.9996	1.0010	1.0108
GARCH(break)	0.9919	0.9936	0.9955	0.9983	0.9976	1.0064				1.0072	1.0067	1.0173
FIGARCH	1.0003	1.0038	0.9939	0.9966	0.9991	0.9960	0.9998	0.9986	0.9941	0.9985	1.0016	1.0016
EWMA	1.0044	1.0063	1.0098	0.9996	0.9998	1.0063	1.0010	1.0004	1.0016	0.9991	0.9993	1.0069
LM-EWMA	0.9972	1.0007	1.0059	0.9950	0.9951	1.0015	0.9985	0.9997	1.0020	0.9987	0.9988	1.0074
Comb.Mean	0.9937	0.9955	0.9948	0.9940	0.9939	0.9965	0.9972	0.9972	0.9975	0.9965	0.9969	1.0015
Comb.T-Mean	0.9943	0.9960	0.9955	0.9942	0.9941	0.9967	0.9975	0.9974	0.9977	0.9964	0.9970	1.0015

such that 30% or 50% of the in-sample period, the GARCH with 0.50 rolling window model reveals more accurate predictive ability. Otherwise, the GARCH with 0.25 rolling window model produces substantially better forecasts in the latest break DGP (70%) . As analysed in Sansó et al. (2004), the modified ICSS framework is more powerful and statistically reliable for the detection of structural breaks in the unconditional variance when a GARCH process is weakly persistent. In this sense, it can be explained why the GARCH with break model can also beat the other individual forecasting models. It implies that effective testing for unknown structural breaks should be able to lead to less biased forecasting errors in such a case.

In the evaluation of the MVaR loss function, on the other hand, a general feature of the forecasting model ranking is quite similar with that of the MSFE evaluation, for $s = 1, 5$. Further, it can be seen that the FIGARCH forecast is consistently counted as the best single model-based forecast in the long-run, for $s = 22$, regardless of the location of the break. It is however worth noting that the FIGARCH forecast has been treated as one of the less accurate forecasts in terms of the MSFE ratio, even if it beats the benchmark. It is conjectured that such a difference in relative performance could be driven by the different structure of given loss functions. Specifically, the MSFE measure is directly computed by dealing with the loss differential between the conditional variance forecasts and the squared errors. Unlike the MSFE measure, the MVaR measure takes the loss of the error series which is indirectly associated with the conditional variance forecast. Nevertheless, it could be still concluded that the FIGARCH model may show relatively accurate predictability to the benchmark or the EWMA-type models, although the structural break is allowed in the weakly persistent GARCH process. This result is also consistent with the fact that the degrees of the estimated GARCH persistence show a tendency toward spurious long memory in every GARCH-Break DGPs, as noted by our QML estimates analysis.

Medium-Level of Persistence with a Structural Break It is observed that most forecasts except EWMA tend to outperform the benchmark, in general. By the MSFE evaluation, the GARCH 0.50 rolling window is overall favourable to the competing single model-based forecasts as well as the forecast combinations over the forecast horizons, only if the structural break is at either 30% or 50% point of the in-sample period. Although the rank of the GARCH model with a break decreases slightly because of a weakened power in detection of the breaks as the persistence level increases, the GARCH with break forecast is still more accurate than the benchmark in the shorter-run forecast. The forecast combinations are also working effectively, dominating most individual model-based forecasts in terms of the MSFE ratio. In the MVaR evaluation, the shorter-run forecasts of the GARCH 0.50 rolling window reveal better accuracy beating the individual forecasting models in the earlier break DGPs. When $s = 22$, it can be suggested that the long memory-based forecasts such as FIGARCH and LM-EWMA may be favourable. We can also see that the forecast combinations perform comparably with the best single model-based forecast.

Considering the latest structural break DGP, the most conflicting feature from the earlier break DGP cases is that the best single model-based forecast, GARCH 0.50 rolling window, is no longer effective enough to be chosen as the most accurate forecasts. This would be why the rolling window size is not suitable to take any potential effects of the recent structural break into account. Rather, the GARCH 0.25 rolling window model outperforms over the forecast horizons in terms of MSFE. Overall, the forecast combinations are still working very well here, more than offsetting losses from poorer forecasts. Further, it is viewed that the long memory-based models are superior relative to the short memory models in the presence of the recent structural break. On the other hand, the GARCH with break model gets worse in its predictive ability for the latest break DGP. If the last structural break point detected is quite near to the end of the in-sample period, a relatively small number of post-break observations could be used to estimate the GARCH model for out-of-sample forecasting. In this respect, the post-break sample is more likely to be less informative. As a result, it could happen to result in misleading forecast outcomes in such cases.

Higher-Level of Persistence with a Structural Break The high persistence feature of **HP-B** is mainly driven by a larger contribution of the lagged conditional variance to the current volatility dynamics, as compared to **LP-B** and **MP-B**, regardless of the presence of the structural break. In this respect, the relative predictive accuracy of FIGARCH and a class of EWMA models seem to be slightly improved under the higher persistence environment, whereas the GARCH with break model is ranked in the lowest overall, even when the structural break is allowed. It is also worth noting that the forecast combinations show their superiority consistently over the DGPs, regardless of the location of the break and the forecast horizon.

In the evaluation of the single model-based forecasts, the GARCH 0.50 rolling window still dominates the competing forecasts for $s = 1, 5$ in both the MSFE and the MVaR average ratios when the artificial structural break is placed at either 30% or 50% point of the in-sample period. The short memory EWMA model beats the benchmark as well, even though it used to show the poorest performance for **LP-B** and **MP-B**. In the long-run forecasts, we may advocate that the FIGARCH model may outperform the short memory GARCH-based forecasts in the spurious long memory environment. In former DGP cases, it has been addressed that modelling long memory would be beneficial in the longer-term forecast, although the true structural break is possessed in the volatility process, particularly at an earlier time of the in-sample period. A pattern in favour of a long memory-based model under spurious long memory continuously appears in **HP-B**, as well. Specifically, it can be viewed that either LM-EWMA or FIGARCH forecasts are favourable in terms of MSFE and MVaR, following two combined forecasts, even when it comes to the shorter forecast horizons. For $s = 22$, the FIGARCH model remains the most accurate, followed by the forecast combinations. Unfortunately, the results of GARCH with break cannot be provided, as the estimation was not feasible in some of the replications through the simulation.

Break in the Persistence Level In the DGPs analysed earlier, the structural break in the unconditional variance is imposed by changing the intercept parameter of the GARCH specification. For the purposes of comparison, we deal with the case that the break is formed by the parameter change of the lagged dependent variable of the GARCH equation. We assume the same magnitude and location of the break point as for the previous of a break in the unconditional variance. Similar to the former cases, the forecast combinations consistently perform very well in every break location and every forecast horizon. We can additionally see that the GARCH with break forecast performs quite poorly. Such an outcome would be natural because the change in the persistence is formed at relatively higher levels from 0.95 to 0.99. In turn, we now look up the performance of the single model-based forecasts when the break is set at 30% of the in-sample period. The GARCH 0.50 rolling window forecast presents superiority as compared to the competing individual forecasts for $s = 1$. However, the model performance seems to get worse as the forecast horizon increases. In contrast, the FIGARCH model consistently produces favourable forecasts relative to the single model-based forecasts in any lengths of the forecast horizon. This pattern maintains the case of the DGP with a structural break at 50%. The FIGARCH predictive ability seems to be strengthened overall. The relative performance of the GARCH rolling window forecast varies as if it depends on the forecast horizon. On the other hand, we find that the GARCH 0.25 rolling window model is no longer preferred when the true volatility process is subject to a recent change in persistence. Rather, the benchmark GARCH expanding window model produces a comparable forecast with the best-performing FIGARCH forecast in terms of the MSFE metric. The relative accuracy of the two EWMA class models seem to be improved as well. However, it is less likely to draw a strong tendency across the forecasts in terms of two different loss measures. For example, completely controversial results are made in the comparison of GARCH and FIGARCH for $s = 1, 5$.

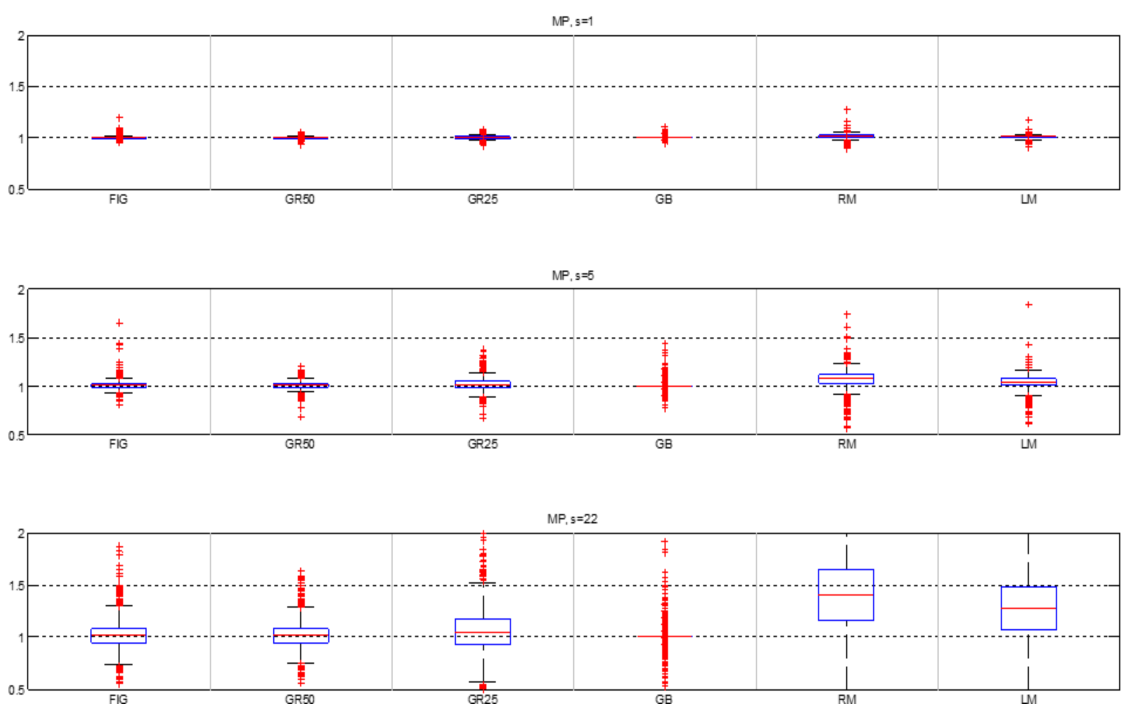
3.4.4 MSFE Loss Evaluation

We examine the MSFE loss series particularly by means of some econometric tests. First of all, we search how many times each of the competing forecasts beats the benchmark GARCH expanding window model. The results are displayed in Table 15 and 16. Next, the RC and SPA tests have been applied to every single replication through the simulation. We investigate the rejection frequency of the null hypothesis that the benchmark forecast is not inferior to any of the alternative forecasts. Each of the individual forecasts has been tested by taking itself as a benchmark against other competing forecasts. By looking at the rejection rate of the null, we expect to see that an individual forecast is relatively inferior to the others overall. It is presumed that a certain individual forecast which reports greater rejection frequency than any of the others is more likely to be selected as the most inferior one. In this sense, we may utilise the results of RC and SPA to figure out whether the poorly performed forecast in terms of the average MSFE metric can still be statistically inferior. The results of RC and SPA are presented in Table 17 and 18. We also carry out a pairwise comparison across the forecasts by means of the MDM test. The input data for MDM are obtained by pooling the MSFE-based loss differentials produced in each of the replications. Due to space limitations, we report the most superior forecast only among the entire pairwise comparison results in Table 19, 20 and 21. A positive (negative) sign of the MDM test statistic implies that the model in the column is better (worse) than the model in the row. Moreover, we only report the results for the model (in the row index) which is selected as the most superior forecast by the pairwise comparison, not by the average MSFE performance. We apply 5% significance level for all the tests employed.

3.4.4.1 Short Memory GARCH DGPs

When the DGP is stationary short memory, we find that some of the forecasts with poor MSFE performance generally show relatively small amounts of the beating rates to more superior forecasts. Also, the fact that most of the forecasts cannot gain a beating rate larger than 60% may be evident to ensure the superiority of the benchmark model, which has consistently presented the lowest average MSFE ratio at every forecast horizon and persistence level. However, the GARCH with break forecast seems not to follow such a pattern. Although we do not allow for structural breaks, the GARCH with break model has not been treated as one of the candidates for the most inferior forecast. Nevertheless, it marks the smallest beating frequency over **LP** and **MP**, and the second smallest in **HP**. This might be due to its large dispersion in the average loss ratio with much lower density to a central tendency, as shown in Figure 2.

Figure 2: Box-plot of the average MSFE ratios in **MP** for $s = 1, 5, 22$.



Note: FIG (FIGARCH), GR50 (GARCH 0.50 rolling), GR25 (GARCH 0.25 rolling), GB (GARCH w/break), RM (EWMA), LM (Long Memory EWMA)

In terms of the RC and the SPA tests, the highest rejection frequency is reported by either EWMA or LM-EWMA for **LP** and **MP**. We recall that those models have had a larger average MSFE ratio relative to the other competing models. However, the difference in the rejection frequency does not seem to be big enough to clearly justify forecasting inferiority. For **HP**, it can be generally seen that the rejection rate of the null is not such a big number, and also there is just a small difference in the rejection rate among them. Despite, the pairwise comparison of the MDM test shows that the GARCH expanding window forecast which gained the lowest average MSFE ratio can be still considered as the most accurate forecast.

3.4.4.2 Long Memory FIGARCH DGPs

For **LM1**, the LM-EWMA forecast has had the smallest average MSFE loss ratio over the forecast horizons. It also achieves a 66% beating rate over the replications, for $s = 1, 5$. For $s = 22$, the beating rate of LM-EWMA is slightly lower than FIGARCH and the forecast combinations, but the differences in the rate between them do not appear to be significantly large. Both combined forecasts are comparable to the best model in the beating frequency in any horizons, as are in the MSFE ratio evaluation. Overall, we can see that the long memory-based models and the forecast combinations are generally better off than the short memory-based models, similarly with the preceding discussions. Further, it seems that the RC and SPA tests consistently support the MSFE-based evaluation results when $s = 1, 5$. LM-EWMA, FIGARCH and forecast combinations are chosen as relatively accurate models overall, whereas the short memory-based forecasts are inferior. In the MDM pairwise comparison, the FIGARCH forecast is mostly preferred. However, it would be difficult to say that the FIGARCH model can gain strongly significant dominance against the comparable models such as LM-EWMA and combined forecasts. For **LM2,** on the other hand, the statistical evaluation results are generally similar with the case for **LM1**. The forecasts with relatively smaller average MSFE ratios for $s = 1, 5$ show higher beating rates as well as lower rejection frequencies in RC and SPA. Particularly for $s = 22$, the inferior model ranking of RC and SPA seems not to follow a similar pattern which has been observed in **LM1**. Further, the MDM reveals consistent results with the MSFE ratio for the selection of superior forecasts such that the combination mean forecast for $s = 1, 5$ and LM-EWMA for $s = 22$.

Table 15: Percentage on MSFE when it beats GARCH(1,1) expanding window: GARCH and FIGARCH DGPs

	LP			**MP**			**HP**			**LM1**			**LM2**		
	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$
GARCH(0.50)	40.1	42.3	42.2	41.4	42.2	42.8	43.7	43.8	45.2	48.0	50.2	51.1	48.4	50.0	50.1
GARCH(0.25)	35.3	38.7	38.2	34.8	37.0	38.8	38.1	38.9	41.0	45.6	48.2	52.0	46.8	49.4	52.1
GARCH(break)	4.4	3.9	3.9	9.8	10.1	10.1	29.1	29.2	31.3	37.0	37.1	38.9	39.6	40.2	42.9
FIGARCH	51.6	46.7	40.2	46.6	46.7	44.2	34.6	33.9	40.7	56.5	59.6	60.6	54.3	53.7	58.6
EWMA	11.2	9.6	10.0	17.0	16.3	17.6	27.4	28.2	28.8	48.5	48.6	44.7	58.1	57.6	64.8
LM-EWMA	18.1	14.8	11.5	26.0	23.4	20.3	42.5	39.0	35.3	66.2	66.7	58.8	61.6	64.8	72.0
Comb.Mean	34.1	29.1	26.7	35.5	34.4	31.8	44.7	44.5	41.9	63.2	62.9	60.9	72.6	69.7	66.6
Comb.T-Mean	35.7	31.3	28.1	37.7	35.5	32.3	45.6	44.2	42.4	65.3	63.8	60.9	71.5	69.1	67.5

Table 16: Percentage on MSFE when it beats GARCH(1,1) expanding window: GARCH-Break DGPs

	LP-B			**MP-B**			**HP-B**			**BinP**		
	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$
Break point: 0.3												
GARCH(0.50)	79.1	83.7	85.7	79.7	80.5	80.2	67.8	69.1	69.2	66.2	68.0	68.2
GARCH(0.25)	75.1	82.2	84.7	71.4	75.2	77.9	54.2	56.7	58.6	53.6	56.3	58.4
GARCH(break)	80.2	83.5	85.1	76.5	78.3	77.9	53.0	54.2	54.8	52.6	53.6	55.8
FIGARCH	55.8	58.4	77.9	70.3	75.9	85.3	50.6	54.4	65.9	53.2	57.0	68.1
EWMA	22.3	15.4	14.7	50.6	47.8	46.0	53.0	53.1	49.8	48.2	47.7	44.7
LM-EWMA	53.4	47.3	37.1	70.8	71.8	71.6	58.0	58.9	59.4	56.2	56.5	57.7
Comb.Mean	88.6	90.7	90.8	85.8	88.1	87.9	68.9	69.6	71.0	67.7	68.3	70.2
Comb.T-Mean	87.8	91.2	90.7	85.8	87.2	87.4	70.5	72.0	72.0	69.8	69.6	71.9
Break point: 0.5												
GARCH(0.50)	76.8	83.8	84.9	78.5	80.5	80.0	65.6	68.5	68.4	62.7	63.7	64.4
GARCH(0.25)	75.1	82.8	83.5	70.5	74.1	77.3	53.5	56.8	57.9	50.4	52.9	55.0
GARCH(break)	77.6	83.2	84.2	72.9	74.2	76.4	48.8	49.7	53.0	45.7	46.1	49.1
FIGARCH	52.5	55.7	71.1	66.2	72.1	83.0	48.9	53.1	65.7	50.7	54.9	67.5
EWMA	22.1	14.5	11.7	42.9	40.0	36.5	50.8	50.4	45.8	38.1	37.8	33.8
LM-EWMA	53.0	46.6	33.9	68.4	69.8	67.8	57.6	56.8	59.8	53.1	52.4	53.3
Comb.Mean	86.8	89.8	90.3	84.2	86.6	86.6	67.0	68.3	71.9	63.9	65.8	68.0
Comb.T-Mean	85.7	89.8	90.1	84.7	86.5	87.0	69.0	70.7	73.2	64.6	65.8	68.5
Break point: 0.7												
GARCH(0.50)	15.2	10.5	11.4	22.3	21.4	25.0	30.1	31.6	35.3	30.7	31.1	36.0
GARCH(0.25)	73.8	80.2	82.7	69.4	73.2	75.0	52.1	54.4	58.7	47.7	50.8	53.1
GARCH(break)	73.9	79.8	82.2	64.2	67.5	71.3				41.7	43.8	50.7
FIGARCH	49.4	51.1	62.4	62.2	67.3	80.4	48.6	50.3	61.0	50.2	53.2	66.6
EWMA	17.9	11.4	9.0	32.2	29.2	27.1	39.8	38.4	37.9	34.0	31.8	32.6
LM-EWMA	49.7	42.4	29.6	65.2	67.0	62.5	56.8	57.1	60.0	50.1	50.8	52.4
Comb.Mean	74.4	77.8	81.3	77.5	81.6	82.8	63.6	62.3	67.7	57.9	61.1	63.1
Comb.T-Mean	70.9	72.1	78.6	77.6	80.8	82.1	58.9	59.7	64.2	58.0	61.1	63.0

Table 17: Rejection frequency of RC and SPA tests on MSFE loss series: GARCH and FIGARCH DGPs

	LP			**MP**			**HP**			**LM1**			**LM2**		
	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$
RC															
GARCH	0.2	0.6	2.2	0.8	2.7	7.2	2.5	7.5	17.8	5.0	11.4	23.6	5.6	11.7	27.6
GARCH(0.50)	0.1	0.2	2.6	1.1	2.4	7.1	3.2	7.9	19.2	6.1	12.3	23.9	6.7	12.5	27.9
GARCH(0.25)	0.1	0.3	2.9	2.1	4.2	8.5	5.3	9.0	18.0	5.1	9.7	21.3	5.8	10.6	25.6
GARCH(break)	0.4	0.9	3.0	2.0	3.0	7.2	7.5	11.3	21.9	8.9	13.4	22.1	9.3	12.6	25.9
FIGARCH	0.1	0.2	1.9	1.7	3.4	7.9	5.5	12.7	23.9	4.0	8.0	20.4	5.1	9.8	25.5
EWMA	15.6	32.5	59.5	8.4	18.5	42.5	3.6	7.9	23.8	3.2	5.5	18.7	2.4	3.8	10.1
LM-EWMA	10.8	30.6	59.6	5.4	16.7	41.1	4.7	10.9	27.2	2.9	4.2	13.3	8.8	9.9	13.6
Comb.Mean	0.5	0.9	3.8	0.7	2.5	8.1	2.7	6.4	18.2	2.6	5.3	16.8	2.9	6.9	18.4
Comb.T-Mean	0.3	0.5	1.9	0.8	2.7	7.4	2.9	7.1	18.0	2.8	6.1	16.2	3.4	7.1	19.0
SPA															
GARCH	3.8	11.2	26.7	5.0	12.3	32.0	6.4	12.9	35.4	13.7	22.4	46.4	14.5	22.6	48.9
GARCH(0.50)	7.5	16.7	36.0	6.8	16.1	38.5	11.2	19.6	42.5	15.0	24.6	47.0	18.4	25.2	50.8
GARCH(0.25)	10.0	17.7	38.5	10.7	19.4	43.0	14.5	22.8	46.1	14.7	23.1	46.6	16.3	23.9	48.1
GARCH(break)	4.5	12.7	27.9	6.1	12.7	34.1	12.9	17.6	38.9	16.6	21.5	43.1	17.5	24.1	45.1
FIGARCH	3.4	8.9	36.6	6.3	13.3	35.2	12.4	22.5	43.8	9.7	20.3	40.7	8.7	19.9	51.1
EWMA	22.8	39.2	63.5	13.9	27.4	55.4	8.9	20.3	50.8	10.0	13.5	41.4	8.8	11.9	30.8
LM-EWMA	12.5	34.8	65.9	11.4	28.0	55.8	8.7	22.4	55.3	3.5	9.5	46.1	11.2	13.1	27.4
Comb.Mean	10.0	25.3	56.4	8.1	22.2	55.5	11.0	23.5	56.6	9.1	19.3	53.8	7.8	16.7	50.5
Comb.T-Mean	8.2	22.9	56.7	6.7	21.8	56.4	10.1	24.2	59.1	9.1	19.9	57.5	9.7	18.8	55.4

Table 18: Rejection frequency of RC and SPA tests on MSFE loss series: GARCH-Break DGPs

	LP-B			MP-B			HP-B			BinP		
	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$	$s=1$	$s=5$	$s=22$
Break point: 0.3												
GARCH	15.4 (8.5)	28.4(21.9)	62.2(46.4)	15.9 (8.9)	26.6(19.3)	59.8(42.8)	12.2 (4.8)	22.1 (9.2)	53.9(26.0)	11.4 (4.0)	20.8 (8.5)	54.2(25.4)
GARCH(0.50)	4.4 (0.5)	10.0 (0.7)	29.3 (4.0)	6.1 (0.9)	12.5 (2.1)	31.2 (7.1)	8.0 (3.2)	13.9 (7.2)	35.1(18.4)	7.8 (3.4)	15.1 (6.9)	35.4(19.0)
GARCH(0.25)	7.2 (0.5)	13.2 (0.8)	35.1 (4.4)	8.2 (2.3)	15.2 (3.1)	36.1 (8.0)	13.5 (4.9)	20.4 (7.6)	43.6(17.1)	13.1 (4.7)	20.7 (7.7)	45.7(17.9)
GARCH(break)	2.4 (0.6)	8.8 (1.3)	24.8 (4.4)	6.2 (1.7)	11.1 (4.3)	32.0 (9.1)	14.2 (9.1)	18.8(12.3)	39.6(24.2)	14.2 (9.2)	17.9(12.1)	39.2(23.8)
FIGARCH	7.3 (6.7)	21.9(19.4)	55.1(32.7)	5.0 (3.7)	15.9 (9.6)	48.4(20.4)	7.4 (3.2)	16.6 (9.3)	46.0(21.0)	6.3 (2.1)	14.5 (8.6)	40.1(19.7)
EWMA	20.9(10.8)	42.9(30.4)	76.6(57.5)	12.7 (8.5)	25.6(19.2)	62.5(41.4)	8.3 (4.5)	18.1 (8.5)	49.5(24.3)	8.2 (4.1)	19.6 (9.1)	51.2(24.8)
LM-EWMA	11.9 (7.7)	42.3(27.1)	78.5(54.7)	9.3 (4.0)	29.7(13.7)	65.9(37.9)	7.0 (4.3)	19.4 (8.6)	52.2(24.9)	8.2 (4.1)	21.0 (9.2)	55.1(25.3)
Comb.Mean	6.0 (2.4)	19.0(10.0)	53.6(21.6)	6.4 (2.0)	18.3 (6.1)	52.2(20.2)	8.0 (1.7)	17.6 (6.3)	53.5(19.4)	7.5 (1.8)	18.1 (5.7)	53.9(20.1)
Comb.T-Mean	5.8 (2.0)	19.3(10.2)	51.4(22.0)	6.4 (1.9)	19.3 (6.6)	54.1(20.1)	8.0 (1.9)	19.2 (6.4)	55.2(19.0)	7.6 (1.9)	19.8 (5.3)	57.0(19.3)
Break point: 0.5												
GARCH	14.3 (7.9)	27.4(20.4)	63.1(46.6)	12.9 (8.2)	23.5(17.2)	58.0(41.5)	11.7 (4.0)	18.9 (9.0)	51.6(24.4)	8.1 (3.9)	15.6 (7.0)	48.3(22.5)
GARCH(0.50)	3.7 (0.4)	8.8 (0.4)	26.1 (3.6)	5.3 (1.3)	11.2 (2.5)	31.4 (7.9)	8.3 (3.4)	13.9 (6.3)	33.9(18.2)	8.3 (2.9)	14.3 (6.7)	36.2(19.1)
GARCH(0.25)	4.9 (0.2)	11.6 (0.5)	29.8 (4.6)	8.4 (2.1)	14.4 (2.9)	33.8 (8.0)	12.8 (4.9)	21.2 (7.4)	43.4(17.1)	13.6 (5.0)	21.9 (7.5)	45.7(18.7)
GARCH(break)	5.0 (0.7)	8.3 (0.4)	25.4 (4.2)	6.9 (2.7)	13.1 (3.5)	32.9(11.2)	16.0(10.6)	20.7(11.9)	39.9(24.2)	16.2(11.4)	20.6(13.2)	41.4(26.9)
FIGARCH	6.5 (5.5)	23.2(19.2)	53.9(30.9)	5.1 (3.5)	15.8 (9.5)	43.9(20.0)	7.2 (2.9)	17.6 (9.6)	45.8(22.4)	4.9 (2.2)	14.1 (6.9)	38.4(18.4)
EWMA	22.5(12.7)	45.7(31.4)	80.3(57.1)	13.4 (8.8)	26.8(17.7)	65.6(41.5)	9.1 (4.2)	19.3 (9.0)	52.3(24.6)	9.9 (3.7)	20.8 (7.6)	53.1(24.8)
LM-EWMA	10.2 (7.3)	36.9(27.4)	79.3(58.1)	10.3 (3.9)	27.2(13.3)	69.1(39.4)	7.4 (3.8)	20.9 (9.7)	54.1(24.8)	9.3 (4.6)	24.7 (9.3)	59.5(24.8)
Comb.Mean	4.9 (1.8)	19.2 (9.4)	55.1(20.8)	6.2 (1.9)	18.0 (6.1)	51.5(20.5)	8.6 (1.5)	18.7 (5.0)	52.1(19.4)	8.0 (2.1)	18.8 (4.7)	55.1(17.9)
Comb.T-Mean	4.2 (1.3)	20.4 (9.4)	54.5(19.1)	5.6 (1.8)	18.2 (6.4)	54.1(20.3)	8.1 (1.8)	19.1 (5.4)	56.8(18.6)	6.8 (2.6)	18.2 (4.6)	60.0(18.2)
Break point: 0.7												
GARCH	8.8 (3.9)	19.8(13.9)	58.0(37.2)	7.1 (4.9)	17.1(11.8)	59.9(34.3)	6.4 (3.9)	14.1 (8.5)	45.4(27.4)	3.9 (2.0)	10.2 (6.0)	44.0(18.8)
GARCH(0.50)	26.8(11.1)	45.4(28.9)	73.8(49.3)	20.3 (9.5)	34.5(20.2)	66.6(42.6)	17.4 (6.4)	27.5(13.1)	56.4(33.4)	13.7 (4.4)	25.0 (8.0)	57.9(24.6)
GARCH(0.25)	3.0 (0.3)	9.6 (0.6)	26.1 (4.3)	6.8 (1.9)	11.0 (3.1)	28.2 (7.8)	12.1 (6.0)	15.8(10.2)	26.6(19.6)	12.3 (4.6)	19.7 (7.3)	42.1(17.3)
GARCH(break)	3.5 (0.9)	7.0 (1.0)	19.6 (4.5)	8.3 (3.7)	11.8 (4.9)	26.5(10.9)				17.3(12.4)	21.2(14.8)	38.1(25.9)
FIGARCH	3.9 (3.7)	18.2(14.6)	45.0(22.8)	3.3 (2.5)	11.4 (6.4)	34.6(15.3)	6.8 (4.1)	13.9(10.0)	36.4(21.8)	2.6 (1.7)	11.4 (5.8)	33.1(18.4)
EWMA	22.4(11.2)	46.0(29.4)	80.7(55.0)	13.4 (7.3)	23.5(15.7)	67.9(39.3)	9.4 (5.1)	20.7(10.7)	55.5(29.8)	10.1 (3.5)	19.5 (7.7)	59.9(23.3)
LM-EWMA	7.9 (6.4)	31.4(25.2)	75.7(54.6)	7.1 (3.6)	21.0(11.0)	64.3(33.1)	5.2 (3.1)	17.7(10.1)	48.7(27.6)	8.3 (4.3)	23.8 (9.3)	57.4(23.4)
Comb.Mean	3.5 (3.0)	15.9(14.7)	47.7(28.4)	6.8 (2.4)	14.5 (7.9)	46.6(22.2)	5.4 (3.4)	12.9 (9.1)	47.3(27.6)	6.8 (2.0)	17.1 (4.7)	48.7(17.1)
Comb.T-Mean	5.3 (3.0)	22.3(15.9)	59.0(31.0)	7.3 (3.0)	16.9 (8.2)	46.6(22.6)	6.6 (3.6)	17.2 (9.6)	55.3(27.4)	6.1 (2.3)	16.8 (4.6)	58.1(15.4)

Note: A rejection frequency of RC is reported in the parentheses.

Table 19: MDM test results for pooled MSFE loss series: GARCH and FIGARCH DGPs

	LP(1)	**LP**(5)	**LP**(22)	**MP**(1)	**MP**(5)	**MP**(22)	**HP**(1)	**HP**(5)	**HP**(22)
	GARCH	**GARCH**	**GARCH**	**GARCH**	**GARCH**	**GARCH**	**GARCH**	**GARCH**	**GARCH**
GARCH	-	-	-	-	-	-	-	-	-
GARCH(0.50)	4.644 [0.000]	5.093 [0.000]	4.790 [0.000]	4.145 [0.000]	4.100 [0.000]	3.326 [0.001]	2.247 [0.025]	2.303 [0.021]	2.547 [0.011]
GARCH(0.25)	8.072 [0.000]	8.119 [0.000]	7.049 [0.000]	8.441 [0.000]	7.889 [0.000]	6.737 [0.000]	3.876 [0.000]	3.532 [0.000]	3.574 [0.000]
GARCH(break)	2.472 [0.013]	1.992 [0.046]	2.730 [0.006]	5.399 [0.000]	3.532 [0.000]	2.384 [0.017]	6.099 [0.000]	5.282 [0.000]	4.328 [0.000]
FIGARCH	3.373 [0.001]	4.473 [0.000]	4.001 [0.000]	4.401 [0.000]	4.928 [0.000]	4.325 [0.000]	3.497 [0.000]	3.246 [0.001]	2.240 [0.025]
EWMA	24.17 [0.000]	24.39 [0.000]	26.48 [0.000]	12.77 [0.000]	13.25 [0.000]	14.95 [0.000]	3.461 [0.001]	3.109 [0.002]	3.244 [0.001]
LM-EWMA	17.02 [0.000]	21.03 [0.000]	27.45 [0.000]	7.229 [0.000]	9.154 [0.000]	12.47 [0.000]	2.715 [0.007]	2.544 [0.011]	3.039 [0.002]
Comb.Mean	6.728 [0.000]	8.870 [0.000]	12.31 [0.000]	2.426 [0.015]	2.534 [0.011]	3.182 [0.001]	1.590 [0.112]	1.476 [0.140]	1.583 [0.114]
Comb.T-Mean	5.116 [0.000]	7.151 [0.000]	9.927 [0.000]	1.624 [0.104]	1.878 [0.060]	2.606 [0.009]	0.964 [0.335]	0.913 [0.361]	1.489 [0.137]

	LM1(1)	**LM1**(5)	**LM1**(22)	**LM2**(1)	**LM2**(5)	**LM2**(22)
	FIGARCH	FIGARCH	FIGARCH	**Mean**	**Mean**	LM-EWMA
GARCH	0.608 [0.543]	0.769 [0.442]	1.410 [0.159]	0.122 [0.903]	0.293 [0.769]	0.106 [0.915]
GARCH(0.50)	1.300 [0.194]	1.803 [0.071]	2.717 [0.007]	0.451 [0.652]	0.623 [0.533]	0.332 [0.740]
GARCH(0.25)	2.259 [0.024]	2.166 [0.030]	2.367 [0.018]	1.011 [0.312]	1.192 [0.233]	0.605 [0.545]
GARCH(break)	4.680 [0.000]	3.915 [0.000]	2.505 [0.012]	1.438 [0.151]	0.722 [0.470]	0.988 [0.323]
FIGARCH	-	-	-	0.482 [0.630]	0.724 [0.479]	0.208 [0.835]
EWMA	3.114 [0.002]	2.513 [0.012]	2.519 [0.012]	1.712 [0.087]	1.766 [0.077]	0.724 [0.469]
LM-EWMA	1.049 [0.294]	0.756 [0.450]	0.899 [0.369]	0.529 [0.597]	0.370 [0.711]	-
Comb.Mean	0.536 [0.592]	0.896 [0.370]	1.333 [0.182]	-	-	0.192 [0.848]
Comb.T-Mean	0.065 [0.948]	0.497 [0.619]	1.191 [0.234]	0.111 [0.911]	0.438 [0.661]	0.112 [0.911]

Note: A value in a braket with DGP corresponds to the forecast horizon. We only report the results for the model (in the row index) which is selected as the most superior forecast in the pairwise comparison. 'Mean' indicates mean forecasting combination. A model in bold implies its forecast achieves the lowest average MSFE ratio as well. A positive value of the test statistic implies that the model in the column is better than the model in the row, and vice versa. A p-value of given statistic is reported in the squared braket. The full results of the Modified DM tests for the pairwise comparison are available upon the request from authors.

Table 20: MDM test results for pooled MSFE loss series: GARCH-Break DGPs I

	Break point: 0.3			*Break point*: 0.5			*Break point*: 0.7		
	LP-B(1)	**LP-B**(5)	**LP-B**(22)	**LP-B**(1)	**LP-B**(5)	**LP-B**(22)	**LP-B**(1)	**LP-B**(5)	**LP-B**(22)
	GAR(break)	**GAR(break)**	**GAR(0.50)**	**GAR(0.50)**	**GAR(0.50)**	**GAR(0.50)**	GAR(break)	GAR(break)	**GAR(0.25)**
GARCH	15.79 [0.000]	17.88 [0.000]	21.08 [0.000]	16.74 [0.000]	19.09 [0.000]	20.70 [0.000]	13.00 [0.000]	15.80 [0.000]	19.48 [0.000]
GARCH(0.50)	1.913 [0.056]	1.010 [0.313]	-	-	-	-	16.48 [0.000]	18.75 [0.000]	21.39 [0.000]
GARCH(0.25)	5.908 [0.000]	6.362 [0.000]	6.825 [0.000]	4.938 [0.000]	6.675 [0.000]	6.183 [0.000]	1.183 [0.237]	0.173 [0.863]	-
GARCH(break)	-	-	0.775 [0.438]	1.420 [0.156]	2.589 [0.010]	2.392 [0.017]	-	-	0.497 [0.619]
FIGARCH	12.16 [0.000]	15.70 [0.000]	18.15 [0.000]	13.18 [0.000]	17.25 [0.000]	19.87 [0.000]	10.60 [0.000]	14.26 [0.000]	18.53 [0.000]
EWMA	18.44 [0.000]	22.17 [0.000]	27.19 [0.000]	19.55 [0.000]	22.87 [0.000]	25.86 [0.000]	16.93 [0.000]	20.55 [0.000]	25.11 [0.000]
LM-EWMA	10.92 [0.000]	17.89 [0.000]	27.02 [0.000]	11.32 [0.000]	18.79 [0.000]	26.56 [0.000]	8.514 [0.000]	15.40 [0.000]	25.34 [0.000]
Comb.Mean	5.084 [0.000]	9.059 [0.000]	13.65 [0.000]	5.778 [0.000]	11.04 [0.000]	16.28 [0.000]	6.183 [0.000]	10.70 [0.000]	17.44 [0.000]
Comb.T-Mean	5.382 [0.000]	9.279 [0.000]	13.73 [0.000]	5.752 [0.000]	10.94 [0.000]	16.15 [0.000]	7.064 [0.000]	11.48 [0.000]	18.28 [0.000]

	Break point: 0.3			*Break point*: 0.5			*Break point*: 0.7		
	MP-B(1)	**MP-B**(5)	**MP-B**(22)	**MP-B**(1)	**MP-B**(5)	**MP-B**(22)	**MP-B**(1)	**MP-B**(5)	**MP-B**(22)
	GAR(0.50)	**GAR(0.50)**	**GAR(0.50)**	**GAR(0.50)**	**GAR(0.50)**	**GAR(0.50)**	LM-EWMA	**GAR(0.25)**	GAR(0.25)
GARCH	11.22 [0.000]	11.87 [0.000]	13.76 [0.000]	11.10 [0.000]	11.83 [0.000]	13.70 [0.000]	12.14 [0.000]	8.802 [0.000]	12.20 [0.000]
GARCH(0.50)	-	-	-	-	-	-	15.16 [0.000]	11.21 [0.000]	14.40 [0.000]
GARCH(0.25)	7.777 [0.000]	7.116 [0.000]	6.740 [0.000]	7.689 [0.000]	7.131 [0.000]	6.760 [0.000]	0.899 [0.369]	-	-
GARCH(break)	5.050 [0.000]	4.048 [0.000]	3.802 [0.000]	9.261 [0.000]	7.455 [0.000]	4.952 [0.000]	5.875 [0.000]	7.044 [0.000]	5.697 [0.000]
FIGARCH	6.172 [0.000]	6.716 [0.000]	7.188 [0.000]	6.923 [0.000]	7.639 [0.000]	8.196 [0.000]	4.976 [0.000]	4.035 [0.000]	6.359 [0.000]
EWMA	11.38 [0.000]	12.38 [0.000]	14.93 [0.000]	11.36 [0.000]	12.41 [0.000]	14.93 [0.000]	14.85 [0.000]	9.571 [0.000]	13.40 [0.000]
LM-EWMA	4.912 [0.000]	7.413 [0.000]	12.21 [0.000]	4.900 [0.000]	7.487 [0.000]	12.25 [0.000]	-	2.854 [0.004]	9.522 [0.000]
Comb.Mean	2.900 [0.004]	3.854 [0.000]	5.962 [0.000]	3.257 [0.001]	4.281 [0.000]	6.194 [0.000]	1.696 [0.090]	2.505 [0.012]	6.388 [0.000]
Comb.T-Mean	2.732 [0.006]	3.884 [0.000]	6.110 [0.000]	3.027 [0.002]	4.245 [0.000]	6.338 [0.000]	2.382 [0.017]	2.836 [0.005]	6.886 [0.000]

Note: A value in a braket with DGP corresponds to the forecast horizon. We only report the results for the model (in the row index) which is selected as the most superior forecast in the pairwise comparison. 'GAR(.)' indicates GARCH with a particular type of forecasting window. A model in bold implies its forecast achieves the lowest average MSFE ratio as well. A positive value of the test statistic implies that the model in the column is better than the model in the row, and vice versa. A p-value of given statistic is reported in the squared braket. The full results of the Modified DM tests for the pairwise comparison are available upon the request from authors.

Table 21: MDM test results for pooled MSFE loss series: GARCH-Break DGPs II

	Break point: 0.3			*Break point*: 0.5			*Break point*: 0.7		
	HP-B(1)	**HP-B**(5)	**HP-B**(22)	**HP-B**(1)	**HP-B**(5)	**HP-B**(22)	**HP-B**(1)	**HP-B**(5)	**HP-B**(22)
	T-Mean	**T-Mean**	FIGARCH	**T-Mean**	**T-Mean**	LM-EWMA	LM-EWMA	LM-EWMA	LM-EWMA
GARCH	4.525 [0.000]	4.264 [0.000]	1.967 [0.049]	3.807 [0.000]	3.659 [0.000]	2.028 [0.043]	2.919 [0.004]	2.738 [0.006]	4.491 [0.000]
GARCH(0.50)	1.173 [0.241]	0.864 [0.388]	0.201 [0.840]	0.807 [0.419]	0.397 [0.692]	0.366 [0.715]	3.816 [0.000]	3.745 [0.000]	5.592 [0.000]
GARCH(0.25)	4.747 [0.000]	4.223 [0.000]	2.186 [0.029]	4.856 [0.000]	4.369 [0.000]	2.473 [0.013]	3.799 [0.000]	3.713 [0.000]	4.268 [0.000]
GARCH(break)	6.638 [0.000]	6.302 [0.000]	3.340 [0.001]	8.053 [0.000]	4.649 [0.000]	3.830 [0.000]			
FIGARCH	1.872 [0.061]	1.335 [0.182]	-	2.529 [0.011]	2.428 [0.015]	0.821 [0.412]	2.758 [0.006]	2.090 [0.037]	2.124 [0.034]
EWMA	4.442 [0.000]	3.892 [0.000]	1.902 [0.057]	4.026 [0.000]	3.719 [0.000]	2.192 [0.028]	3.035 [0.002]	2.885 [0.004]	4.417 [0.000]
LM-EWMA	1.240 [0.215]	0.893 [0.372]	0.199 [0.842]	0.903 [0.367]	0.557 [0.578]	-	-	-	-
Comb.Mean	1.566 [0.117]	1.993 [0.046]	0.325 [0.746]	1.726 [0.084]	1.536 [0.124]	0.660 [0.509]	1.465 [0.143]	1.383 [0.167]	2.617 [0.009]
Comb.T-Mean	-	-	0.232 [0.817]	-	-	0.524 [0.600]	1.857 [0.063]	1.765 [0.078]	3.186 [0.001]

	Break point: 0.3			*Break point*: 0.5			*Break point*: 0.7		
	BinP(1)	**BinP**(5)	**BinP**(22)	**BinP**(1)	**BinP**(5)	**BinP**(22)	**BinP**(1)	**BinP**(5)	**BinP**(22)
	T-Mean	**T-Mean**	FIGARCH	**T-Mean**	**T-Mean**	FIGARCH	T-Mean	T-Mean	LM-EWMA
GARCH	4.453 [0.000]	3.964 [0.000]	1.882 [0.060]	3.280 [0.001]	3.148 [0.002]	1.628 [0.097]	3.286 [0.001]	4.062 [0.000]	1.791 [0.073]
GARCH(0.50)	1.454 [0.146]	1.048 [0.295]	0.531 [0.595]	1.478 [0.140]	1.121 [0.262]	0.759 [0.448]	4.175 [0.000]	4.212 [0.000]	2.712 [0.007]
GARCH(0.25)	4.822 [0.000]	4.154 [0.000]	2.225 [0.026]	4.761 [0.000]	4.157 [0.000]	2.264 [0.024]	4.844 [0.000]	4.297 [0.000]	2.561 [0.010]
GARCH(break)	6.761 [0.000]	6.408 [0.000]	3.450 [0.001]	8.689 [0.000]	4.833 [0.000]	3.863 [0.000]	9.685 [0.000]	7.740 [0.000]	5.026 [0.000]
FIGARCH	1.167 [0.243]	0.848 [0.396]	-	1.497 [0.134]	0.969 [0.333]	-	1.645 [0.100]	1.419 [0.156]	1.014 [0.311]
EWMA	4.412 [0.000]	3.807 [0.000]	1.908 [0.056]	3.865 [0.000]	3.546 [0.000]	1.924 [0.054]	3.689 [0.000]	3.811 [0.000]	2.171 [0.030]
LM-EWMA	1.381 [0.167]	0.993 [0.321]	0.840 [0.401]	1.171 [0.242]	0.843 [0.399]	0.964 [0.335]	0.805 [0.421]	0.629 [0.530]	-
Comb.Mean	1.557 [0.119]	2.022 [0.043]	0.524 [0.600]	2.504 [0.012]	1.844 [0.065]	0.707 [0.479]	2.554 [0.011]	2.122 [0.034]	0.731 [0.465]
Comb.T-Mean	-	-	0.439 [0.661]	-	-	0.567 [0.571]	-	-	0.668 [0.504]

Note: A value in a braket with DGP corresponds to the forecast horizon. We only report the results for the model (in the row index) which is selected as the most superior forecast in the pairwise comparison. 'T-Mean' indicates trimmed mean forecasting combination. A model in bold implies its forecast achieves the lowest average MSFE ratio as well. A positive value of the test statistic implies that the model in the column is better than the model in the row, and vice versa. A p-value of given statistic is reported in the squared braket. The full results of the Modified DM tests for the pairwise comparison are available upon the request from authors.

3.4.4.3 GARCH-Break DGPs

At a glance, in the presence of the true structural break, we can find one apparent rule that a set of superior (inferior) forecasts in terms of the average MSFE ratio is consistently being considered as the superior (inferior) one in the statistical evaluation. Two of the forecast combinations and the best-performed individual forecasts may still admit their relative superiority with the highest beating rate, regardless of the level of persistence and the length of forecast horizon. Further, a set of forecasts with poor MSFE performance generally has had to bear high-level rejection frequency in the RC and the SPA evaluations. Specifically, a class of the EWMA models and the benchmark GARCH model are commonly inferior when the true volatility process exhibits lower- and medium-level persistence short memory with a structural break. For **HP-B** or **BinP**, such a general pattern remains when $s = 1, 5$. If $s = 22$, however, the RC and SPA tests are less likely to admit the same pattern because the highest rate of rejection has been reported several times by the best MSFE forecast. In the MDM pairwise comparison, one of the superior forecasts in terms of the average MSFE ratio generally shows its dominance to the competing forecasts. When the break is located at 30% or 50% of the in-sample period, it turns out that the best-performing forecasts are still significantly superior at a lower- and a medium-level persistence, over the forecast horizon. For **HP-B**, the pairwise comparison suggests that the long memory-based forecasts are better off when $s = 22$. In the shorter-run for $s = 1, 5$, the combined mean forecast is chosen as the most accurate one. When the break point is located at the 70% of the in-sample period, it seems that we are less likely to confirm whether the best-performed forecasts by the MSFE evaluation are still significantly superior in the pairwise comparison. However, it can be seen that the LM-EWMA forecast and the trimmed mean forecast are generally superior relative to the others over the forecast horizons, except in the case of **LP-B**.

3.5 Concluding Remarks

In this chapter, we have investigated the relative predictive ability of a class of GARCH-based parsimonious conditional variance models via the Monte Carlo simulation experiment. This study has mainly aimed to evaluate and compare these models' forecasting performance in several different data generations of the synthetic error process, which are associated with the memory properties of the conditional variance. Specifically, the synthetic error processes are generated by the stationary short memory GARCH, the short memory GARCH with the single artificial structural break, and the stationary long memory FIGARCH. The simulations also consider various different locations of the structural break and the persistence levels. Through the simulation experiment, we expect that we can provide comprehensive and unified insights into the economic benefits of the employed models in terms of volatility forecasting, even if the forecasting models are potentially misspecified against the true volatility properties.

Our experiment results reveal some supporting evidence of the discussions of the existing relevant literature. If the conditional variance process is stationary short or long memory in the absence of a structural break, the forecasting models, which are allowed to capture the properties of the true process are more favourable than any other misspecified models. When the true short memory process is contaminated by the structural break, the detection of the break may play an important role in choosing a proper window size for the short-run forecasting. Further we have found that spurious long memory may strongly dominate the true structural break in the long-run forecasting when the true short memory process is highly persistent. However, it has not been easy to justify any consistent features or patterns in forecast superiority among the individual forecasting models when the structural break is located around the end of the in-sample period. Nevertheless, it can be seen that the long memory-based forecasts are generally better off than the short memory-based competing forecasts in the presence of the most recent break. On the other hand, two forecast combinations are very favourable in the presence of a structural break, regardless of the forecast horizon and the level of persistence. At a general point of view, the statistical tests have also provided consistent and robust results, supporting the findings above.

The outcomes may also address some practical implications when modelling and forecasting macroeconomic or financial time series. Firstly, the standard GARCH model with expanding window can be a candidate to produce a superior forecast when a volatility process is stationary without any evidence of structural breaks or long memory. When true structural breaks are specified, and their locations are not too close to the end of the in-sample period, the GARCH rolling window can be a good forecast candidate if the persistence level is not extremely high. We suggest that the size of rolling window is effectively determined using the information of the last break point detected. Otherwise, the long memory-based forecasting models are more likely to generate superior forecasts in the following situations: (1) lack of statistically significant evidence of the structural break; (2) longer-term forecasting under spurious long memory; and (3) true stationary long memory process. Finally, it is worth noting that forecast combinations can be better or very comparable alternatives for a volatility process with a certain memory property which we have considered through this chapter.

A number of extensions would be possible, based on the limitations of this study. For example, the baseline data generation processes considered in this study would be somewhat restricted to dealing with structural break and long memory separately. Our simulation design can be naturally extended to accounting for more general non-stationary volatility processes which are subject to structural break and long memory simultaneously or other non-linearities. Moreover, since we adopt for a class of parsimonious GARCH-type conditional volatility models in the analysis, it would be more informative in a general sense if we additionally take more various conditional volatility models such as stochastic volatility models and markov-switching models into account, as analysed in Lux and Morales-Arias (2013).

Chapter 4 Long Memory and Structural Changes in Forecasting Daily Return Variability of S&P 500 Stock Index with Historical and Realised Measures of Volatility

4.1 Introduction

It is generally acknowledged that financial asset return volatility is time-varying, and tends to exhibit high persistent dynamics. Basically, such volatility dynamics of persistency can be modelled well by taking long memory into account. Some financial econometric literature has also assessed spurious long memory property, particularly associated with the analysis of the effect of structural breaks on volatility persistence, and have pointed out that neglected structural breaks can lead to spurious high persistence in conditional volatility, such as Lamoureux and Lastrapes (1990), Hamilton and Susmel (1994), Mikosch and Starica (2004) and Hillebrand (2005), amongst others. Hence, a choice between long memory and spurious long memory is important for modelling volatility in a variant of financial time series applications such as risk management, asset allocation, option pricing etc.

There have been several studies which accommodate that either long memory or structural break does matter when evaluating relative volatility forecasting accuracy. As a recent contribution with emphasis on structural break as a factor of spurious persistence, Starica and Granger (2005) found that a non-stationary model with breaks in unconditional variance can produce better performance in longer horizon forecasts. The empirical study of Rapach and Strauss (2008) pointed out that the presence of structural breaks in the unconditional variance can reliably generate more accurate forecasts in real-time exchange rate return volatility in a class of GARCH models. In contrast, there has been a competing argument in favour of long memory in terms of forecasting performance as well. Diebold and Inoue (2001) argued that long memory may be a useful description for a forecasting purpose even if the data generating process exhibits structural breaks and weak dependence. In addition, Morana and Beltratti (2004) showed that neglecting breaks is not important for very short term forecasting once it allows for a long memory component in the model, whilst superior forecasts can be obtained at longer horizons by modelling both long memory and structural changes. Choi et al. (2010) found that the persistence of volatility can be partly explained by structural breaks in mean of log-realised volatility, but the proposed long memory model can produce robust forecasting even if the true volatility process is subject to structural changes. However, it is generally known that structural changes and long memory are not easily disentangled, and even worse it can often be seen that both are coexisting and interplaying.

The main problem for measuring volatility is that a true volatility process is not observed and latent. Hence, it is inevitably required to model volatility using an underlying return process. In this study, we consider typical measurements of volatility such as daily return-based historical volatility and high-frequency intradaily return-based realised measures of volatility. Although both the historical and the realised measurement are an unbiased estimator of true volatility, it is acknowledged that the latter is more robust to noise potentially existing in asset return. In this sense, the realised measures of volatility have been widely accepted as a proxy of an actual volatility process, when evaluating volatility model performance. In terms of identifying structural changes and long memory, on the other hand, a variant of theoretical and empirical investigations has often resulted in controversial outcomes between those measurements for the same underlying return series. Particularly for US stock market index return volatility, it is evident that historical volatility processes contain structural breaks, and induce spurious increases in volatility persistence as discussed by Lu and Perron (2010), Perron and Qu (2010) and McCloskey and Perron (2013) among others. For realised measures of volatility, a large body of literature suggests that the realised variance of the US stock market return can be effectively modelled with the apparent long memory characteristic, as supported by the hyperbolic decay rate of its sample autocorrelations. For instance, Martens and Zein (2004) and Koopman, Jungbacker and Hol (2005), amongst others.

Along the same lines of the empirical findings above, we here suppose that different unbiased volatility estimates from the same underlying return series can possibly hold inconsistent memory properties in terms of long memory or spurious long memory. It also motivates us to study whether explicitly distinguishing structural breaks and long memory can provide beneficial information in terms of the predictive ability of different volatility measurements. In this study, we analyse the properties of the S&P 500 stock index return volatility processes which are realised measures of volatility (daily realised variance and realised kernel) and historical volatility (squared daily return). Before modelling volatility, we first investigate the true memory properties of given volatility measures using the econometric tests of Qu (2011) and Baek and Pipiras (2014), which help us to disentangle true long memory and structural breaks. Further, if it is evident that a volatility process is subject to structural breaks, then some tests could be applied to identify the number and locations of structural changes. Although the memory properties are identified, it would be hard to say that the relative performance of forecasting models is directly and entirely dependent upon the memory properties only. As a further step, we examine the relative out-of-sample performance of one-day-ahead forecasts, with emphasis on the predictive content of structural changes and long memory. We use realised kernel series as the proxy of an unknown true volatility process, and realised variance, realised kernel and squared returns are utilised for parametric or semi-parametric volatility modelling. The log-realised measures of volatility are estimated and forecasted by means of a class of ARFIMA models: ARFIMA$(1, d, 1)$ and Markov-Switching ARFIMA$(0, d, 0)$ models. Also, a class of short and long memory GARCH models, GARCH$(1, 1)$, EGARCH$(1, 1)$, FIGARCH$(1, d, 1)$, FIEGARCH$(1, d, 1)$, Adaptive-FIGARCH$(1, d, 1, 1)$, GARCHX$(1, 1)$and Realised GARCH$(1, 1)$, are utilised for the historical volatility estimation and prediction. The relative forecast performance is evaluated by accounting for some tests in terms of equal predictive ability, superior predictive ability and model confidence set with respect to MSE, MAE and QLIKE forecast loss functions.

The main results of this chapter are as follows. In a given sample period of the US stock market index, realised variance and realised kernel processes exhibit true long memory. However, the historical volatility process shows some evidence of spurious long memory subject to multiple structural breaks corresponding to stock market events. Once the structural breaks are adjusted to the squared daily return, the volatility process looks like a weak dependent stationary process rather than a persistent process. In terms of relative predictive accuracy, a class of ARFIMA models consistently generates the best-performed forecasts relative to a class of GARCH models over the loss functions under the realised kernel volatility proxy. Among GARCH models, it is shown that a rolling window GARCH forecast and GARCH forecasts, which account for structural changes in its own specification, outperform long memory-based GARCH models even with the long memory proxy process. Also, the sensitivity analysis of rolling window size for the GARCH model reveals that the appropriate choice of rolling window size for the GARCH model is important to achieve relatively better predictive ability in the structural breaks even when the proxy of an actual volatility exhibits long memory. In addition, using some pre-break data can be an effective way when forecasting GARCH volatility.

The rest of this chapter is organised as follows. Section 4.2 describes return, volatility measures and their properties. We carry out tests to distinguish long memory and structural breaks in Section 4.3. The structural breaks in the historical volatility process are detected and our analysis of the effect of breaks is followed up in Section 4.4. Section 4.5 introduces the forecasting methodology including the estimation model specification, the estimation window, the loss functions and the loss evaluation criteria. The empirical results of the out-of sample forecasting are presented with their statistical evaluations in Section 4.6. Section 4.7 concludes this chapter.

4.2 Return, Volatility and Data Description

We consider the daily return series which is obtained by the logarithm difference of the S&P 500 index prices. The demeaned return model is given by

$$r_t = \sigma_t z_t,$$

for $t = 1, \ldots, T$, where $z_t \sim i.i.d.\,(0,1)$, the standardised process of the daily return innovation and σ_t is the latent volatility. Conditioning on $\mathcal{F}_{t-1}$, the past information set up to $t-1$, then the conditional volatility is given by $\sigma_t^2 = E\left[r_t^2 \middle| \mathcal{F}_{t-1} \right]$. In a discrete time series model, the conditional variance is usually estimated by means of modelling the squared innovation of the return process.

Indeed, true volatility is unobservable. Some volatility measures may provide an unbiased estimate of latent volatility and are widely used to estimate and predict volatility. It is generally known that realised volatility is used by means of a relatively accurate proxy of true volatility to squared returns, as pointed out by Andersen and Bollerslev (1998). Since a stochastic innovation of the return is unobservable, such a component would necessarily be accounted for unbiased and consistent estimation when modelling and forecasting volatility. Intradaily return can be counted as virtually observable. In the rationale of realised volatility, Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002) noted that the daily realised variance of intradaily return converges in probability to the true integrated (daily) volatility under continuous semimartingale processes, such as diffusion processes. As a time interval approaches zero, realised variance can be a model free measure of integrated volatility. In this sense, the realised measure of volatility can be considered for being able to provide relatively close approximation, controlling for such noise driven by an idiosyncractic error of the return. Through this study, we particularly focus on realised variance and realised kernel to utilise them as the proxies of actual daily return volatility.

Let $p_{n,t}$ denote the logarithmic intradaily price process, the intradaily return is defined as

$$r_{n,t} = p_{n,t} - p_{n-1,t},$$

for $n = 1, \ldots, N$, where N denotes the number of intradaily returns in a trading day. Then, realised variance can be obtained by summing up intra-daily squared returns.

$$RV_t = \sum_{n=1}^{N} r_{n,t}^2.$$

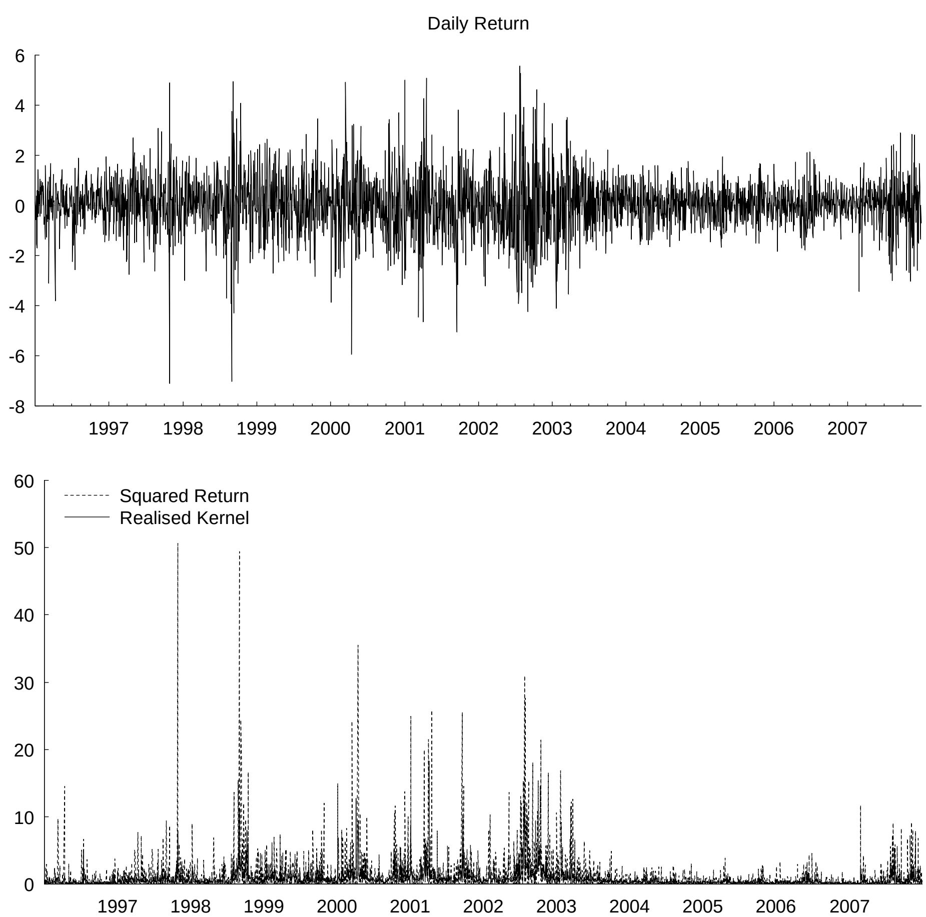
In addition, this study employs the realised kernel estimator introduced by Barndorff-Nielsen, Hansen, Lunde and Shephard (2008). Realised kernel is considered as a more robust estimator to the noise of market microstructure effects than realised variance. We utilise realised kernel in a form of a heteroskedasticity and autocorrelation consistent (HAC) type estimator, computing

$$
\begin{aligned}
RK_t &= \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h, \\
\gamma_h &= \sum_{n=|h|+1}^{N} r_{n,t} r_{n-|h|,t},
\end{aligned}
$$

where H is the bandwidth of the kernel estimator. The precise choice of the bandwidth has been made by referring to Barndorff-Nielsen, Hansen, Lunde and Shephard (2009). $k(x)$ is the Parzen kernel function:

$$
k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & , \quad 0 \leq x < \frac{1}{2} \\ 2(1-x)^3 & , \quad \frac{1}{2} \leq x < 1 \\ 0 & , \quad x \geq 1 \end{cases} .
$$

Figure 3: Daily return and volatility



The sample data for the empirical research consists of S&P 500 stock index price quotes during the period from 3 January 1996 to 27 February 2009 (2971 trading days). The original realised volatility data are available at the database, "Oxford-Man Institute's realized library", produced by Heber, Lunde, Shephard and Sheppard (2009)[10]. The data employed in this study is part of the cleaned data used in Shephard and Sheppard (2010)[11]. The daily realised variance series is obtained using 5-minute intradaily returns with subsampling. The opening and closing 15 minutes of the trading records are excluded from the original data set to control for overnight effects. Consistently, we use open-to-close daily returns, ignoring overnight effects. See Shephard and Sheppard (2010) and references therein for further details of the data characteristics, if necessary.

The daily return and volatility series are displayed in Figure 3. We can observe several negative shocks to S&P 500 daily return over the sample period. Particularly, two very large peaks in volatility can be found at Fall 1997 and Fall 1998 which may be associated with the Asian and Russian financial crisis, respectively. The effects from such large negative shocks may lead to negative skewness and positive and strong skewness in the return and squared return distribution, respectively. In addition, it seems that the US stock market return series has also shown quite a volatile period, reflecting the Dot-com bubble and its collapse in the early 2000s, towards the end of 2003. After that period, the US stock market has continued moving in a relatively stable direction for a while. Around the end of the sample, the volatility level slightly increases more than before.

Table 22: Summary statistics for returns and volatilities

	Mean	Median	St.Dev.	Skewness	Kurtosis
r_t	0.005	0.053	1.313	-0.258	11.025
$r_{t,rv}$	0.130	0.079	1.396	0.268	3.235
$r_{t,rk}$	0.127	0.078	1.372	0.248	3.308
r_t^2	1.723	0.385	5.457	11.152	178.778
RV_t	0.964	0.491	2.089	10.976	202.195
RK_t	1.003	0.515	2.141	10.406	180.991
$\log r_t^2$	-1.338	-0.954	2.473	-1.182	6.048
$\log RV_t$	-0.657	-0.711	0.993	0.556	3.825
$\log RK_t$	-0.621	-0.663	1.001	0.530	3.772

[10] See http://realized.oxford-man.ox.ac.uk/.

[11] The cleaned data are obtained from Journal of Applied Econometrics Data Archive, adopting daily return and the realised covariance series which are already computed and provided by Shephard and Sheppard (2010).

Table 22 contains the summary statistics of daily returns, daily standardised return, realised variance, realised kernel and logarithmic transformed volatilities for the full-sample period. The distribution of the daily return process is negatively skewed and strongly positively leptokurtic. The demeaned return series is standardised by the standard deviations of the daily realised variance and the daily realised kernel, denoted as $r_{t,rv}$ and $r_{t,rk}$, respectively. Both of the standardised daily returns are approximately standard normal distribution[12]. The raw series of the volatility measures commonly exhibit positive skewness and severely large excess kurtosis which may be driven by the large shocks to return over the sample period. The logarithmic transformed realised measures of volatility appear approximately Gaussian, but with a slightly fatter tail to the left and positive leptokurticity. The log-squared return series is negatively skewed and more leptokurtic than log-realised measures of volatility.

Table 23: Long memory tests and parameter estimation

	$\mathbf{H}_0$: $I(0)$		$\hat{d}$			
	V/S	HML	GPH	LW	ELW	2ELW
r_t^2	0.000	0.000	0.336	0.311	0.311	0.314
RV_t	0.000	0.000	0.484	0.473	0.472	0.477
RK_t	0.000	0.000	0.487	0.476	0.475	0.479
$\log r_t^2$	0.000	0.000	0.268	0.257	0.271	0.274
$\log RV_t$	0.000	0.000	0.538	0.519	0.518	0.530
$\log RK_t$	0.000	0.000	0.536	0.518	0.516	0.529

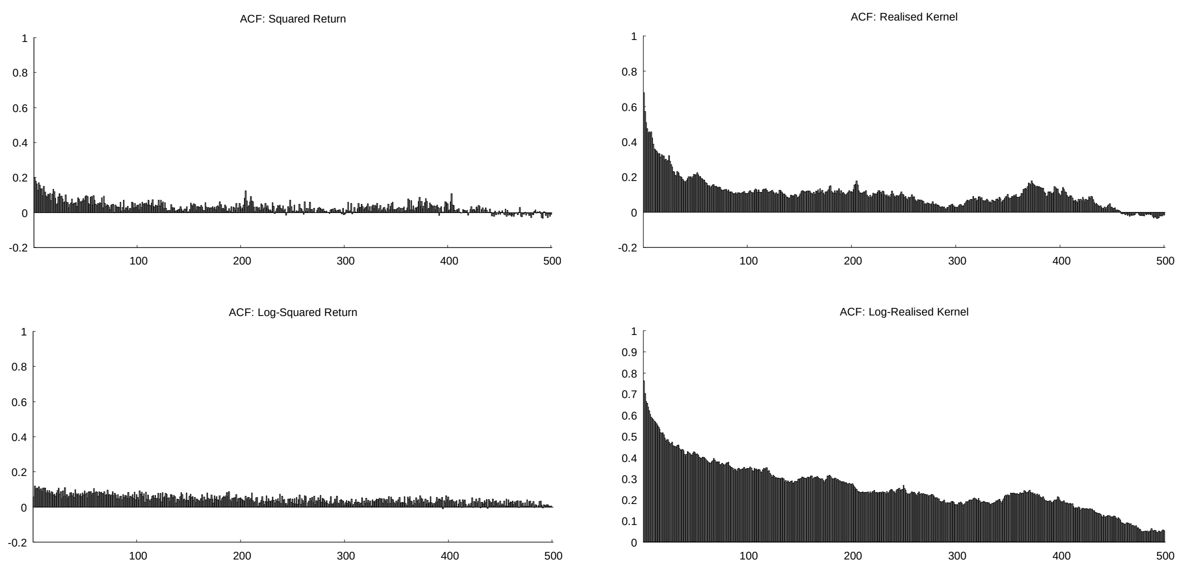
Note: The p-values of $I(0)$ tests against $I(d)$ are reported.

Further, we implement some tests to examine the persistence property of daily return volatility processes in terms of short memory stationarity and long memory. In order of testing $I(0)$ against $I(d)$, we employ the rescaled variance test statistic (V/S) of Giraitis, Kokoszka, Leipus and Teyssière (2003) and the long-range autocovariance-based test statistic (HML) of Harris, McCabe and Leybourne (2008). The p-values of the V/S statistic are analytically calculated. The truncation parameter and the bandwidth truncation of the variance parameter of HML are set to 1 and 0.66, respectively. The long-range dependence of the volatility series is also investigated by means of the semiparametric long memory estimators. We employ the narrow band log-periodogram (GPH) estimator of Geweke and Porter-Hudak (1983), the

[12]It is worth noting that a financial asset return series standardised by realised volatility seems to exhibit a weak volatility clustering feature as compared to its raw return series. And also, the distribution of the standardised returns which are obtained by an one-day-ahead conditional variance estimate from a parametric ARCH or stochastic volatility model is typically quite leptokurtic as discussed in Andersen, Bollerslev, Diebold and Ebens (2001a); Andersen, Bollerslev, Diebold and Labys (2001b); Andersen et al. (2003).

local Whittle (LW) Gaussian maximum likelihood estimator of Robinson (1995) and Kunsch (1987), the exact local Whittle (ELW) estimators of Shimotsu and Phillips (2005). The ELW offers a general-purpose estimation of long memory parameter over both stationary and non-stationary regions of d, whereas the LW estimator is discontinuous at $d = \frac{3}{4}$ and $d = 1$. We also apply the two-step ELW estimator (2ELW) of Shimotsu (2010) which is allowed to accommodate an unknown mean and a polynomial time trend for the ELW. For the estimation of semiparametric long memory estimators, the size of the bandwidth is chosen to $T^{0.70}$ for all of the estimators.

Figure 4: Autocorrelations for daily return volatility



As displayed in Table 23, both V/S and HML strongly reject the null of the short memory stationarity for all of the daily variance processes at any significance levels. It might conclude that none of the volatiliy series are a pure stationary short memory. The semi-parametric analysis suggests that the persistence level of the raw and log-volatility series are well-bounded within the stationarity condition, $-\frac{1}{2} < d < \frac{1}{2}$, except $\log RV_t$ and $\log RK_t$ which exhibit non-stationary long memory property, presenting higher persistent of $\widehat{d}$, slightly over $\frac{1}{2}$. Interestingly, the (log-) squared return series exhibits shorter memory relative to the (log-) realised measures of volatility. Moreover, it can be seen that a sample autocorrelation function of the log-realised kernel series is decaying very slowly relative to others, as displayed in Figure 4. The level of the sample autocorrelations for the log-squared return are much smaller than that for the log-realised kernel, even though it does not look like it is decaying exponentially. The raw series of squared return and realised kernel show faster decaying autocorrelations than their logarithmic transformed series. These features in the graphic can act as supporting evidence for longer memory of the log-variance series, rather than stationary short memory. Nevertheless, the sample autocorrelation function possibly exhibits a slow rate of decay, and $\widehat{d}$ is likely to be biased away from 0 when the process is truly stationary short memory, but is apparently subject to any significant non-linear component (spurious long memory) such as regime changes, as pointed out in Diebold and Inoue (2001) and Granger and Hyung (2004), amongst others.

4.3 True or Spurious Long Memory?

It is generally acknowledged that the structural changes and long memory are easily confused and even difficult to clearly distinguish from each other when they are co-existing. Spurious long memory is likely to be induced in the presence of true structural breaks. Otherwise, spurious structural changes may also cause misleading inference of time series models to a true long memory process. Also, it is therefore inevitable to verify a true property of a volatility process in terms of memory property and the effect of structural breaks in order of a correct choice of an estimation model or an estimation window for forecasting. In this sense, we attempt to disentangle such a confusion in a given volatility series by applying a variant of econometric tests which helps to uncover properties of the return variance. We apply some tests to the log-volatility series since the raw volatility series itself such as $r_t^2, RV_t$and RK_t suffers from a non-negativity constraint. Also, there would be no significant loss relative to using the raw variance in the identification of true or spurious long memory because log-variance is a monotonic transformation, as stated in Lu and Perron (2010). Moreover, the log-volatility process is even closer to being Gaussian distribution, so that the logarithmic series are more appropriate than the raw series in terms of meeting the required conditions of the applied test procedures.

Although the semiparametric long memory estimates present such a high persistent feature in the log-volatility series, we cannot simply neglect potential structural changes. If structural breaks are evidently significant, neglected breaks could spuriously lead to an increase in an estimate of the long memory parameter. In order to statistically identify a true memory property of given volatility series, we implement the test for distinguishing long memory against spurious long memory which is the local Whittle estimation-based test of Qu (2011) (henceforth, Qu). The Qu test is based on the frequency domain and the derivatives of the profiled local Whittle likelihood function in a degenerating neighbourhood of the origin. Its test statistic can be implemented relaxing the Gaussianity assumption usually assumed in other long memory tests. Say, it allows for non-Gaussianity or conditional heteroskedasticity as well. Following the recommended trimming portion for financial data applications, a trimming size is set to 2% in a large sample and the bandwidths for local Whittle estimator are $T^{0.60}$ and $T^{0.70}$. The alternative hypothesis of Qu considers an $I\left(0\right)$ short memory process but is contaminated by level shifts or a smoothly varying trend. Therefore, a rejection of the null hypothesis may address the fact that non-linearity possibly due to the presence of structural breaks may strongly be present, rather than long memory. However, the rejection of the null of the long memory test cannot directly accommodate whether a volatility process is only subject to the structural changes. In order to obtain a clear clue in terms of structural breaks among other possible alternatives, we additionally utilise Baek and Pipiras (2014) method (henceforth, BaPi) for the null of $I\left(0\right)$ with multiple structural breaks against the alternative of stationary long memory, $I\left(d\right)$. Depending on the test result, we expect that long memory and structural breaks may be more clearly discriminated. Moreover, it might help rule out any other possibilities of an alternative process such as a smooth trend or other omitted non-linearity properties. The tests employed can play a role to distinguish between long memory and structural breaks.

The test procedure of BaPi is based on the local Whittle estimation of the long memory parameters from the residual series obtained by sequentially removing changes in mean. Then, the proposed procedure evaluates and compares differences in the number of breaks estimated from $\sup F$, least squares and CUSUM-based approaches with the local Whittle method. They suggest that the size and significance of a difference in the number of breaks across the break estimation approaches lead to a consistent and updated stopping rule which helps estimate the changes in mean under the null and distinguish the structural changes and the long memory. For the sake of simplicity, we account for the stationary bootstrap procedure proposed in BaPi to obtain the p-value in a comparison of stopping rules between the CUSUM-based method and the LW method. The selection of the bandwidth of the LW estimator follows the techniques suggested in Baek and Pipiras (2014)[13].

Table 24: Distinguishing long memory and structural breaks

$\mathbf{H}_0$:	$I(d)$		$I(0) + breaks$
	Qu ($T^{0.60}$)	Qu ($T^{0.70}$)	BaPi
$\log r_t^2$	**1.513**	**1.144**	0.318
$\log RV_t$	0.357	0.607	**0.000**
$\log RK_t$	0.383	0.627	**0.000**

Note: The reported value for Qu is the test statistic, but BaPi is the p-value, obtained by stationary block bootstrap. The bandwidth for long memory estimator is indicated in the parenthesis. Significance at 10% level or less is in bold.

We conduct these tests to distinguish long memory and spurious long memory (structural breaks) in the level of log-volatility processes. Table 24 states the results of the tests, applied to log-variance series of S&P 500 index return in terms of squared return, realised variance and realised kernel. We report the test statistic of Qu[14] and the bootstrap p-value of BaPi. First, it confirms that both log-realised volatility series make consistent results in favour of long memory over the test employed, rather than spurious long memory, including structural breaks. In brief, we conclude that the log-realised volatility series may exhibit true long memory property in its level, or at least, the long memory is powerful enough to dominate any non-linearity. Martens and Zein (2004) and Koopman et al. (2005) documented that long memory is generally present in the logarithmic processes of the realised measures of volatility of the US stock market index return.

On the other hand, for the daily log-squared return series, the BaPi test does not reject the null hypothesis of structural breaks for the same series. The stationary long memory hypothesis of Qu is rejected at 5% level for $T^{0.60}$ and at 10% level for $T^{0.70}$, respectively. Such a test result for structural breaks in daily return-based volatility process of the US stock market index is evidently supported by a large body of literature. For example, the test results of Perron and Qu (2010) showed the statistical evidences of short memory with structural breaks for the S&P 500 stock index return volatility against stationary long memory in the period from 1990 to 2002. Zhang, Gabrys and Kokoszka (2007) also found significant structural breaks against long memory for the daily squared return series of the Dow-Jones

[13] Refer to detailed description at p.943 in Baek and Pipiras (2014).
[14] The critical value used is reported in Table 1 of Qu (2011, p.428).

Industrial Average Index (DJIA) in the sample period from July 1998 through to June 2006. A similar conclusion was reached by McCloskey and Perron (2013) for the log-absolute daily return of the S&P 500 from July 1962 to March 2004 and of the DJIA from March 1957 to October 2002. In the case of the recent sample period, Baek and Pipiras (2014) concluded that the log-absolute daily return series of DJIA and S&P 500 from January 2004 to December 2013 are subject to multiple structural breaks, rather than a pure stationary long memory. Based on such test results, we are willing to find the evidence of structural breaks and verify the effect of structural breaks in $\log r_t^2$ on to its persistence level.

4.4 Testing for the Presence of Structural Breaks

First, we carry out the procedure of Bai and Perron (1998, 2003) (henceforth, BP) within short memory-based break models to detect structural changes in the mean level of the log-squared return process. The break model of BP is defined as

$$y_t = c_j + u_t,$$

for $t = T_{j-1} + 1, T_{j-1} + 2, \ldots, T$, where $j = 1, 2, \ldots, k+1$. y_t is the logarithmic squared return, and c_j is the mean of y_t. The break points $(T_1, \ldots, T_k)$ are assumed to be unknown. u_t is serially correlated and heteroskedastic. Let $\mathrm{UD}_{\max}$ be the double maximum statistic, defined as $\mathrm{UD}_{\max} = \max_{1 \leq l \leq K} \left[\sup F_T(l)\right]$, where $\sup F_T(l)$ denote the F statistic for the null of no structural breaks against arbitrary number of breaks, and K is the maximum number of breaks allowed. In this study, we set to $K = 5$. Denote $\mathrm{WD}_{\max}$ to be the weighted double maximum statistic, which is given by $\mathrm{WD}_{\max} = \max_{1 \leq l \leq K} \left[w_l \sup F_T(l)\right]$, where w_l is the marginal p-values which are across values of l. $\sup F_T(l+1|\, l)$ is the test statistic that sequentially tests the null of l changes against the alternative of $l+1$ breaks. The trimming value for $K = 5$ is set to 0.15. See Bai and Perron (1998, 2003) for more details of the test procedures.

Since BP is supposed to test for $I(0)$ with structural breaks in levels against stationary short memory, we additionally implement the test of Berkes, Horváth, Kokoszka and Shao (2006) (henceforth, BHKS) which tests the null of weak dependence with a single change in mean at an unknown point against the alternative of long memory. The test statistic is based on the CUSUM-type statistics, where we calculate $P_{t,1}$ based on the observations up to the time of the estimated single break point and $P_{t,2}$ based on the observations after the time of the estimated single break point. Assuming a single break process, the test statistic is defined as $M_{t,2} = \max\left[P_{t,1}, P_{t,2}\right]$ which converges to a well-known distribution in the presence of structural breaks, but diverges to infinity under the alternative. In order to detect multiple structural breaks, we adopt a sequential procedure as proposed in BaPi until the null hypothesis of k breaks is not rejected against $k+1$ breaks alternative. Saying, if $M_{t,k+1} = \max\left[P_{t,1}, \ldots, P_{t,k+1}\right]$ is less than its critical value at $k+1$ break points through the sequential procedure, then we may confirm the existence of k structural breaks against a pure stationary long memory. Refer to Berkes et al. (2006) and Baek and Pipiras (2014) for further details.

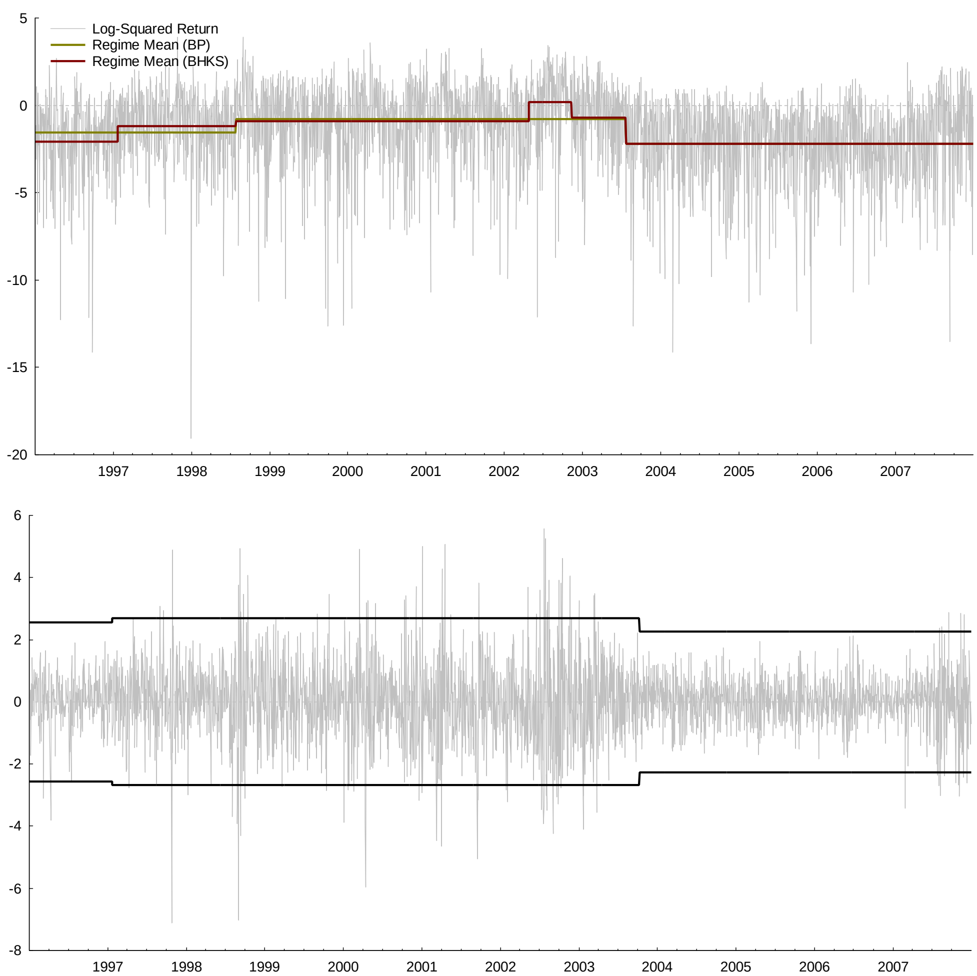
Table 25: Tests of structural changes for $\log r_t^2$ and r_t^2

BP		Break date		Regime mean	
$\mathrm{UD}_{\max}$	154.54*	$\hat{T}_1$	27/07/1998	$\hat{c}_1$	-1.567
$\mathrm{WD}_{\max}$	154.54*	$\hat{T}_2$	22/07/2003	$\hat{c}_2$	-0.801
$\sup F_T(2\|1)$	46.20*			$\hat{c}_3$	-2.210
$\sup F_T(3\|2)$	7.46*				
$\sup F_T(4\|3)$	0.78				
$\sup F_T(5\|4)$	0.00				
BHKS		Break date		Regime mean	
$\hat{M}_1$	4.959*	$\hat{T}_1$	20/01/1997	$\hat{c}_1$	-2.094
$\hat{M}_2$	3.188*	$\hat{T}_2$	27/07/1998	$\hat{c}_2$	-1.197
$\hat{M}_3$	2.271*	$\hat{T}_3$	25/04/2002	$\hat{c}_3$	-0.920
$\hat{M}_4$	1.692**	$\hat{T}_4$	11/11/2002	$\hat{c}_4$	0.194
$\hat{M}_5$	1.868*	$\hat{T}_5$	22/07/2003	$\hat{c}_5$	-0.720
$\hat{M}_6$	1.548			$\hat{c}_6$	-2.184
ICSS		Break date			
$\hat{C}_1$	1.387**	$\hat{T}_1$	20/01/1997		
$\hat{C}_2$	3.593**	$\hat{T}_2$	09/10/2003		

Note: * and ** indicate 1% and 5% significance, respectively. Date format is dd/mm/yyyy.

In addition, we consider potential structural breaks in the raw squared return series for the comparison purpose with the log-squared return. In doing so, the test statistic of Inclan and Tiao (1994) is adopted for the detection of multiple structural breaks in unconditional variance of the return, $E\left(r_t^2\right)$, against the stationary short memory process. In fact, the asymptotic distribution of the test statistic is obtained under the assumption that a given series of random variables follows Gaussian *i.i.d.* processes. However, the test statistic may suffer from upward size distortions as sample size increases, when a given sequence of observations is dependent on a process such as a GARCH process, as shown in Andreou and Ghysels (2002) and Sansó et al. (2004). To complement this drawback, we utilise the non-parametric-adjusted CUSUM statistic using the Bartlett kernel estimator to test the null of a homogenous unconditional variance against the alternative that the structural breaks present in unconditional variance, based on settings of the asymptotically valid test statistic of Sansó et al. (2004), denoting as C_k. In order to estimate multiple structural breaks, the modified iterated cumulative sum of squares algorithm (ICSS) of Inclan and Tiao (1994) is applied to the test statistic at 5% significance level.

Figure 5: Structural breaks for log-squared return and unconditional variance of return



Note: The plot above is the log-squared return series and regime means by BP and BHKS. The plot below is the return series and its two-standard-deviation bands for the regimes by the modified ICSS algorithm.

Table 25 reports the estimated break points, the break dates and the regime mean depending on each of the break points. The applied test statistics provide strong evidence of multiple structural breaks. The sequential procedure for the $\sup F$-based test of BP and CUSUM-based test of BHKS confirms two and five structural breaks over the full-sample period, respectively. The break locations detected by BP are exactly matched with some of the locations by BHKS, one in July 1998 and another in July 2003. Such mean-level shifts coincide with past stock market crises. The first break would be taken place by the Russian Financial crisis. The largest size of downward regime change is observed in July 2003 which might be around the end of a highly-volatile period, driven by the Dot-com bubble and the stock market crash. Related to the Asian financial crisis, it seems that BHKS reacts more sensitively than BP by additionally detecting the break at the beginning of 1997. BHKS has kept updating the break information and catching up the mean-levels by July 1998 which is the first break point located by BP. The modified ICSS has detected two structural changes in the unconditional variance process of r_t. The estimated break dates are 20 January 1997 and 9 October 2003. The first break point corresponds to the first break point of BHKS, and the second is slightly after the last break detected by BP and BHKS. Figure 5 displays the log-squared return series with regime changes by BP and BHKS, and the demeaned return series with regime changes in unconditional standard deviation by the modified ICSS. Further, it seems that we are less likely to find any existence of a strong time trend, rather than structural breaks, as seen by the pattern of the mean shifts in the $\log r_t^2$ process.

Table 26: Long memory tests and parameters for the break-adjusted $\log r_t^2$

	$\mathbf{H}_0$: $I(0)$		$\hat{d}_{dm}$				
	V/S	HML	GPH	LW	ELW	2ELW	2ELW$_d$
BP	0.221	0.121	0.179	0.171	0.170	0.173	0.172
BHKS	0.477	0.242	0.130	0.122	0.121	0.124	0.124

Note: We here report p-values of the V/S and HML tests. The bandwidth for the semiparametric long memory estimators is set to $T^{0.70}$.

Based on the estimated regime means by BP and BHKS for log-squared return series, we generate the break-adjusted series, subtracting the $\widehat{c}_j$ from $y_t = \log r_t^2$. In turn, we now conduct the V/S and HML tests for the null of stationary short memory against long memory, and also estimate the long memory parameters for the break-adjusted series to compare with $\widehat{d}$ from the original series. Let $\widehat{d}_{dm}$ denote the estimate of the semiparametric long memory estimator for the demeaned series. As reported in Table 26, once the effect of structural breaks is eliminated, the estimated persistence levels for the every long memory estimators are generally lying down between 0.12 and 0.15, decreasing by 0.1 point compared to $\widehat{d}$. Although $\widehat{d}_{dm}$ are still bounded in the stationary long memory condition, it seems that the level of persistence is considerably lower than a persistence level which is observed in a stationary financial asset return volatility process, in general. Further, the V/S and HML tests suggest that the break-adjusted series of $\log r_t^2$ is stationary short memory process, rather than long memory. None of the considerable difference can be seen in the estimated parameter values between 2ELW and 2ELW with *detrending* (2ELW$_d$). Therefore, we may conclude that the $\log r_t^2$ series is contaminated by a spurious long memory which is mainly induced by the structural breaks, rather than a smooth time trend.

4.5 Forecasting Methodology

4.5.1 Volatility Models

We compare the relative predictive ability of a variant of econometric models which generates conditional volatility forecasts. For realised measures of volatility, the ARFIMA$(1, d, 1)$ model is utilised to estimate and predict realised variance and realised kernel using their logarithmic transformed series as proposed by Andersen et al. (2003). Moreover, Martens, van Dijk and de Pooter (2009) and Hillebrand and Medeiros (2015) suggest that the out-of-sample fit of ARFIMA-type models can be improved by modelling long memory with non-linearity when forecasting the realised volatility of US stock market index returns. In this respect, we also adopt the Markov Switching-ARFIMA$(0, d, 0)$ (MSFI) model which particularly takes possible regime switching of the underlying volatility process into account. For the sake of simplicity and accuracy in estimation, we allow for one regime switching in an intercept term for MS-FI. In a discrete time process of the daily return volatility, we employ a variant of GARCH-type models, which is designed to capture memory properties and/or non-linearity in conditional variance process: GARCH$(1, 1)$ of Bollerslev (1986), exponential GARCH$(1, 1)$ (EGARCH) of Nelson (1991), fractionally integrated GARCH$(1, d, 1)$ (FIGARCH) of Baillie et al. (1996), fractionally integrated exponential GARCH$(1, d, 1)$ (FIEGARCH) of Bollerslev and Mikkelsen (1996) and Adaptive-FIGARCH$(1, d, 1, 1)$ (A-FIGARCH) of Baillie and

Morana (2009). In brief, the short memory EGARCH model can capture asymmetric behaviour of financial asset volatility. FIGARCH and FIEGARCH processes may exhibit longer memory properties than other competing short memory models. The A-FIGARCH specification is designed to allow for a fractional integrated parameter to capture long memory and a deterministic time-varying intercept that allows for structural breaks. We also consider a class of joint models of daily return and a realised measure of volatility, where the realised measure is additionally included as an exogenous variable to the standard GARCH specification: GARCHX$(1,1)$ of Engle (2002) and Realised GARCH$(1,1)$ (RGARCH) of Hansen et al. (2012).[15] In a GARCHX framework, we treat the realised measure of volatility as a strong exogenous variable, which is not dependent of the squared error process of the return. The RGARCH model allows for joint dependence between the realised measure of volatility and the return by specifying the realised measure equation in addition to the GARCH equation. All the GARCH-type conditional variance models are estimated by Quasi-Maximum Likelihood method under Gaussianity assumption, with conventional parameter restrictions for positiveness and stationarity of conditional variance process. The specification of the estimation models are described in Table 27. On the other hand, we additionally utilise a class of exponentially weighted moving average (EWMA) models additionally for the comparison purpose of out-of-sample forecasts, which are RiskMetrics EWMA (EWMA) of J.P. Morgan (1994) and Long Memory EWMA (LM-EWMA) of Zumbach (2006).

4.5.2 Estimation Window

On out-of-sample forecasting, we firstly divide the generated synthetic error series into the in-sample period and the out-of-sample period. Denote that p is the size of out-of-sample. $T-p$ is the length of the in-sample period. For the expanding window forecasting models, the in-sample observations are used to generate the first out-of-sample forecast. Namely, the initial set of observations spans from the first realisation up to $(T-p)$th observation. Once we obtain a new forecast, then, we expand the estimation window by one observation to forecast conditional variance for the next period, say the first observation through observation $T-p+1$. By repeating this procedure up to the end of the available out-of-sample period, we can finally obtain p numbers of out-of-sample forecasts for every single expanding window-

[15] We use Log-Realised GARCH model in this study, rather than the simple linear specification in empirical study. It automatically ensures positivity of conditional variance.

Table 27: ARFIMA and GARCH volatility model specifications

ARFIMA$(1,d,1)$	$(1-\phi L)(1-L)^d(x_t-\mu)=(1+\theta L)\varepsilon_t$
MS-ARFIMA$(0,d,0)$	$(1-L)^d(x_t-\mu_i)=\varepsilon_t$
GARCH$(1,1)$	$\sigma_t^2=\omega+\alpha r_{t-1}^2+\beta\sigma_{t-1}^2$
EGARCH$(1,1)$	$\log\sigma_t^2=\omega+(1-\beta L)^{-1}(1-\alpha L)\{\gamma z_t+\lvert z_t\rvert-\sqrt{2/\pi}\}$
FIGARCH$(1,d,1)$	$\sigma_t^2=\omega+[1-(1-\beta L)^{-1}(1-\alpha L)(1-L)^d]r_t^2$
FIEGARCH$(1,d,1)$	$\log\sigma_t^2=\omega+(1-\beta L)^{-1}(1-L)^{-d}(1-\alpha L)\{\gamma z_t+\lvert z_t\rvert-\sqrt{2/\pi}\}$
Adaptive-FIGARCH$(1,d,1,1)$	$\sigma_t^2=\omega_t+[1-(1-\beta L)^{-1}(1-\alpha L)(1-L)^d]r_t^2$ $\omega_t=\omega+\gamma\sin(2\pi t/T)+\delta\cos(2\pi t/T)$
GARCH-X$(1,1)$	$\sigma_t^2=\omega+\alpha r_{t-1}^2+\beta\sigma_{t-1}^2+\gamma x_{t-1}$
Log-Realised GARCH$(1,1)$	$\log\sigma_t^2=\omega+\beta\log\sigma_{t-1}^2+\gamma\log x_{t-1}$ $\log x_t=\xi+\varphi\log\sigma_t^2+\tau_1 z_t+\tau_2(z_t^2-1)+u_t$

Note: x_t represents a series of realised measure of volatility. Particularly, the ARFIMA models utilise logarithmic realised measures of volatility to estimate and forecast conditional variance.

based model. Basically, the forecasts of the one-day-ahead conditional volatility are generated by the framework of expanding window forecasting.

In addition to the expanding window, we consider two rolling window forecasts of the standard GARCH model. The model is estimated with two different rolling window sizes that are one-half and one-quarter lengths of the in-sample period. Let r denote the rolling window size. Thus, $r = 0.50$ and 0.25 in our case. Then the initial sample size used for the estimation is from $round\,[(1-r)\times(T-p)]+1$ to $T-p$. Once we obtain a new forecast, we roll over the estimation window by one observation to forecast conditional variance for the next period. Specifically, the new estimation window covers the observations from $round\,[(1-r)\times(T-p)]+2$ to $T-p+1$. Then, we repeat this procedure up to the end of the available out-of-sample period. Corresponding to given window sizes, these forecasting models are denoted as GARCH(0.50) and GARCH(0.25), respectively. In using a shorter estimation window, the forecast model has a relatively smaller number of observations available to estimate GARCH parameters, but it is more likely to reduce an overlapping part in data between different regimes.

Also, we take a post-break sample estimation window into account for the GARCH forecast. The detection of structural breaks is based on the modified ICSS framework, which was discussed in the previous section. The last break point among all of the estimated break points is used to determine the estimation window size for the GARCH forecasts. Specifically, the GARCH model could be estimated using part of the in-sample observations from $k_f + 1$ up to $T - p$, where k_f is the final structural break point detected. We can then obtain the first out-of-sample forecast. If no break is detected, the generated forecasts must be equivalent to the forecast from the GARCH with expanding window. After that, the second out-of-sample forecast can be generated using the observations from the new break point by the modified ICSS to $T - p + 1$. By repeating the described procedure up to the end of the full-sample, we finally obtain p number of out-of-sample forecasts which may account for the potential (final) structural breaks throughout the entire sample. We denote the forecasts produced by means of this framework as GARCH(break). However, GARCH(break) is likely to suffer from an issue, related to the number of observations to be used for reasonably reliable estimates of the GARCH parameters. Namely, if the detected break point is located too close around the forecast date, then a short sample would be available for estimation.

4.5.3 Loss Functions and Evaluation Criteria

We evaluate the forecast performance across employed conditional volatility models using a range of statistical and econometric measurements. We consider realised kernel as the proxy of actual volatility. As noted earlier, it is generally known that the realised measures are more efficient proxies than the squared return series. Moreover, the realised kernel is known as more robust to the market microstructure noise than the realised variance. It is however noted that the realised kernel is still a noisy proxy for the true underlying variance process. We compare three different loss functions, instead of utilising the unique best criterion only in terms of loss evaluation. For measuring forecasting accuracy, we adopt mean squared error (MSE), mean absolute error (MAE) and QLIKE loss functions to evaluate forecasting errors across the individual forecasts. The loss functions for one-step ahead forecast evaluation are given by

$$
\begin{aligned}
\text{MSE}_1 &= \frac{1}{p}\sum_{t=T-p+1}^{T}\left(\tilde{\sigma}_t^2-\hat{\sigma}_{t|t-1}^2\right)^2, \\
\text{MSE}_2 &= \frac{1}{p}\sum_{t=T-p+1}^{T}\left(\tilde{\sigma}_t-\hat{\sigma}_{t|t-1}\right)^2, \\
\text{MAE}_1 &= \frac{1}{p}\sum_{t=T-p+1}^{T}\left|\tilde{\sigma}_t^2-\hat{\sigma}_{t|t-1}^2\right|, \\
\text{MAE}_2 &= \frac{1}{p}\sum_{t=T-p+1}^{T}\left|\tilde{\sigma}_t-\hat{\sigma}_{t|t-1}\right|, \\
\text{QLIKE} &= \frac{1}{p}\sum_{t=T-p+1}^{T}\left(\log\hat{\sigma}_{t|t-1}^2+\frac{\tilde{\sigma}_t^2}{\hat{\sigma}_{t|t-1}^2}\right),
\end{aligned}
$$

where $\tilde{\sigma}_t^2$ is the proxy of volatility, realised kernel, and $\hat{\sigma}_{t|t-1}^2$ is the one-day ahead out-of-sample forecast of volatility. Unlike MSE or MAE, QLIKE may be informative for the loss implied by the Gaussian log-likelihood. Patton (2011) shows that the MSE_1 and QLIKE are robust when using an unbiased but noisy (imperfect) volatility proxy, and a forecasting model ranking based on MSE_1 and QLIKE would be preserved in use of various conditional volatility forecasts. The MAE loss functions may not be robust to the noise of the proxy when using the unobservable conditional volatility. In addition to the loss functions, we measure the forecast bias and information content using the Mincer-Zarnowitz (MZ) regression, by regressing realised kernel on the conditional volatility forecast, which is given by

$$
\tilde{\sigma}_t^2=\beta_0+\beta_1\hat{\sigma}_{t|t-1}^2+e_t.
$$

An accurate forecast is conditionally unbiased if and only if $\beta_0 = 0$ and $\beta_1 = 1$. Although a forecasting model ranking based on a loss comparison could be informative in a selection of the best-performing (the smallest loss) forecasting model, it does not mean that the loss difference between forecasting models is statistically significant. Moreover, it does not indicate whether such a rank order is consistently robust in a different sample. In this sense, the significance of the difference in the forecast losses is evaluated by the equal predictive ability test of Diebold and Mariano (1995) and West (1996) (DMW), the superior predictive ability (SPA) of Hansen (2005) and the model confidence set (MCS) approach of Hansen et al. (2011). The DMW test is applied to the pairwise comparison of the MSE_1 losses between the ARFIMA-type models and some GARCH-based models which have low MSE_1 relative to the others as well as the entire GARCH-based models. The SPA and the MCS approaches are utilised to select the best forecasting model against a multiple number of competing models over MSE_1, MAE_1 and QLIKE losses.

The null hypothesis of the DMW is to test for the equal predictive accuracy of different forecasting models, given by $\mathbf{H}_0 : E\left[d_{t,ij}\right] = 0$, where $d_{t,ij} = L\left(\hat{\sigma}^2_{t,i}, \tilde{\sigma}^2_t\right) - L\left(\hat{\sigma}^2_{t,j}, \tilde{\sigma}^2_t\right)$, for $i, j = 1, \ldots, J$, where J is the number of forecasting models. $d_{t,ij}$ is a relative loss differential between model i and j for $i \neq j$. that is the loss differential between model i and j for $i \neq j$. The DMW test statistic is given by

$$T^{DMW} = \frac{\sqrt{p}\overline{d}_{ij}}{\widehat{\sigma}_{ij}},$$

where $\overline{d}_{ij}$ is the sample mean of $d_{t,ij}$, that is $p^{-1}\sum_{t=T-p+1}^{T} d_{t,ij}$. $\widehat{\sigma}^2_{ij}$ is the asymptotic long-run variance of $\sqrt{p}\overline{d}_{ij}$ from a weighted sum of sample autocovariances. The DMW statistic is asymptotically distributed standard normal for non-nested model comparison as shown in West (1996). The rejection of the null hypothesis indicates that the i model outperforms against the j model when $T^{DMW} < 0$, and vice versa. However, the DMW has some drawbacks when it applies to a large set of competing forecasting models. For example, White (2000) points out a data snooping problem, and Hansen et al. (2011) concerns a high-dimensionality issue in the estimation of a covariance matrix with a large number of competing forecasting models, amongst others. Moreover, the SPA test and the MCS method are valid to both nested and non-nested models in evaluation of relative forecasting performance.

The SPA test is conducted for the multiple comparison of competing forecasts. We are interested in testing the null hypothesis that one particular model fixed as the benchmark is not worse than any of the competing forecasts in terms of expected loss. The null hypothesis is given by $\mathbf{H}_0 : E\left[\mathbf{d}_t\right] \leq \mathbf{0}$,where $\mathbf{d}_t = \left(d_{t,1}, \ldots, d_{t,J}\right)$, in which $d_{t,i} = L\left(\hat{\sigma}^2_{t,0}, \tilde{\sigma}^2_t\right) - L\left(\hat{\sigma}^2_{t,i}, \tilde{\sigma}^2_t\right)$. $\hat{\sigma}^2_{t,0}$ is a conditional variance forecast of a benchmark model, and $\hat{\sigma}^2_{t,i}$ is a conditional variance forecast of i competing model. $d_{t,i}$ denotes the loss differential of forecast i relative to the benchmark forecast. The SPA test statistic for a forecast sample size p is given by

$$T^{SPA} = \max\left[\max_i \frac{\sqrt{p}\overline{d}_i}{\widehat{\sigma}_i}, 0\right]$$

respectively, where $\overline{d}_i = p^{-1}\sum_{t=T-p+1}^{T} d_{t,i}$ and $\hat{\sigma}_i^2$ is a consistent estimator of the asymptotic variance of $\sqrt{p}\overline{d}_i$. In our applications, we set each of individual forecasts as the benchmark and the rest of the others as an alternative for the comparison. The p-value of T^{SPA} is computed using the stationary bootstrap of Politis and Romano (1994) with 1000 bootstrap replications[16]. A high p-value implies that we cannot reject that a benchmark model does not outperform competing models. However, the SPA test would have somewhat limiting attributes if a large number of competing models are considered. In brief, the SPA test needs for a pre-specified benchmark model, and its information is restrictive to a benchmark model. It requires a composite hypothesis to test. See more relevant details in Hansen et al. (2011).

As improvement of the DMW test and the SPA test, the MCS approach is not only available for a large set of competing models, but it also does not require a benchmark model and a composite hypothesis. Hansen et al. (2011) address that the MCS is a confidence interval which contains the best forecast at a confidence level α. Let $\mathcal{M}^0$ be the initial set of one-step-ahead forecast of volatility. The MCS is a subset of $\mathcal{M}^0$, denoted as $\mathcal{M}_\alpha$. $\mathcal{M}_\alpha^*$ can be found by an 4-stage iterative testing framework by testing for the null hypothesis, $\mathbf{H}_0 : E\left[d_{t,ij}\right] = 0$ for all i and j, where $d_{t,ij} = L\left(\hat{\sigma}_{t,i}^2, \tilde{\sigma}_t^2\right) - L\left(\hat{\sigma}_{t,j}^2, \tilde{\sigma}_t^2\right)$. In the first stage, the initial hypothesis being tested is that all forecasts in $\mathcal{M}^0$ exhibit equal predictive ability. We calculate the loss differentials of the forecast models from the empirical data and the bootstrapped loss differentials. Based on $d_{t,ij}$, the bootstrapped loss differentials $d_{t,b,ij}$ are generated for $b = 1, \ldots, B$, where b is the number of bootstrap replication. Next, define the average forecast losses of model i relative to model j from the empirical data and the bootstrap samples as $\overline{d}_{ij} = p^{-1}\sum_{t=T-p+1}^{T} d_{t,ij}$ and $\overline{d}_{b,ij} = p^{-1}\sum_{t=T-p+1}^{T} d_{t,b,ij}$, respectively. Also, define the average losses of model i relative to all other competing models as $\overline{d}_i = J^{-1}\sum_{j\in I} d_{ij}$ and $\overline{d}_{b,i} = J^{-1}\sum_{j\in I} d_{b,ij}$. In addition, we estimate $\widehat{\sigma}_{b,ij}^2 = B^{-1}\sum_b \left(\overline{d}_{b,ij} - \overline{d}_{ij}\right)^2$ and $\widehat{\sigma}_{b,i}^2 = B^{-1}\sum_b \left(\overline{d}_{b,i} - \overline{d}_i\right)^2$. In the third stage of the MCS method, we test the null hypothesis by applying the empirical and the bootstrapped range test statistics, given by

[16]For SPA, we report the consistent p-value of T_r^{SPA}.

$$T^{MCS} = \max_{i,j} \frac{\left|\overline{d}_{ij}\right|}{\widehat{\sigma}_{b,ij}},$$
$$T_b^{MCS} = \max_{i,j} \frac{\left|\overline{d}_{b,ij} - \overline{d}_{ij}\right|}{\widehat{\sigma}_{b,ij}}.$$

The p-value of the test statistics is computed as

$$p^{MCS} = \frac{1}{B} \sum I \left\{ T_b^{MCS} > T^{MCS} \right\},$$

where I is an indicator function, which equals to unity if $T_b^{MCS} > T^{MCS}$, otherwise zero. If p^{MCS} is greater than α, the null cannot be rejected and the MCS procedure stops. It implies that all the models constitute the MCS, $\mathcal{M}_\alpha^0 = \mathcal{M}_\alpha^*$, holding equal predictive ability. Otherwise, if p^{MCS} is smaller than α, the null of equal predictive ability is rejected, then we move onto the final stage. The fourth stage determines a ranking of the competing forecasts, by removing the worst forecast from $\mathcal{M}_\alpha$. The worst performing model can be identified as

$$i^{(-)} = \arg\max_i \frac{\overline{d}_i}{\widehat{\sigma}_i}.$$

After a trimming of $i^{(-)}$ model, the MCS procedure goes back to the second stage and the procedure is repeated until the null is not rejected or only one model remains in $\mathcal{M}_\alpha$. Then, the forecast models in $\mathcal{M}_\alpha$ at the point where the procedure stops consist of $\mathcal{M}_\alpha^*$. In the applications to our study, following Hansen et al. (2011), we set the confidence interval and the bootstrap sample size to $\alpha = 0.25$ and $B = 1,000$, respectively. A block bootstrapping method is used to generate the bootstrap samples with 2 block length.

4.6 Empirical Results

For the forecasting analysis, we evaluate the relative forecasting performance of conditional volatility models using realised measures of volatility and daily squared return volatility. We have found statistical evidence that a source of the persistence of the (log-) squared return-based volatility process for the S&P 500 stock index price neglects structural breaks, whereas the realised variance and the realised kernel processes exhibit true long memory properties.

Based on these findings, the evaluation focuses on the predictive accuracy comparisons between a set of the realised measurement-based ARFIMA-type models and a set of historical measurement-based GARCH-type models. In addition, it is of particular interest to identify whether the GARCH-type models, which account for structural changes in estimation and forecasting, can generate more accurate forecasts than long memory-based GARCH models, even though the proxy of volatility (RK_t) follows a true long memory. For the generation of the forecasts, the out-of-sample period is set to $p = 502$, so that the in-sample period is from 3 January 1996 to 30 December 2005 (2469 observations). A number of observations used to estimate rolling window GARCH(0.50) and GARCH(0.25) is 1235 and 617, respectively. Those number of observations are fairly bigger than the suggested number of observations by Hwang and Valls Pereira (2006), to reliably estimate the GARCH models.[17] In the loss evaluation, the log-volatility forecasts are transformed to the nominal value by taking exponent.

4.6.1 In-Sample Analysis

Since this study is mainly focusing on evaluating the relative predictive accuracy of the forecast models, the MSE_1 and the QLIKE loss functions are applied for a direct comparison of the in-sample performance of all of the volatility models. The in-sample estimation results are presented in Table 28 and 29, with the MSE_1 and QLIKE goodness-of-fit statistics. We utilise realised kernel as the proxy for true volatility. The fitted values are the series of the in-sample forecasts of the conditional volatility models.

In general, the ARFIMA and MSFI models show better goodness-of-fit than the GARCH-based models. This result is natural because the daily squared return might be a more noisy estimator than the realised variance and realised kernel. In the models of the realised measures of volatility, the regime switching model slightly outperforms in terms of MSE_1 and QLIKE for both $\log RV_t$ and $\log RK_t$ series, which are ascertained to be long memory by the Qu and BaPi tests employed. The estimated persistence estimates of MSFI-RV and -RK are lower than those of ARFIMA-RV and -RK, but the persistence levels of the conditional variance in both models are quite high. As tested by V/S and HML in the earlier section, $\log RV_t$ and $\log RK_t$ are not covariance stationary over the full-sample period, and their semiparametric

[17] Hwang and Valls Pereira (2006) propose that at least 250 observations are needed for ARCH(1) models and 500 observations for GARCH$(1,1)$ models under the consideration of the size of biases and convergence errors.

Table 28: In-sample estimation for daily return volatility I

	ARFIMA$(1,d,1)$		MS-ARFIMA$(0,d,0)$	
	$\log RV_t$	$\log RK_t$	$\log RV_t$	$\log RK_t$
μ_1	-1.118	-1.075	-7.761	-7.971
	(0.633)	(0.630)	(0.545)	(0.511)
μ_2			-0.780	-0.754
			(0.456)	(0.472)
ϕ	0.383	0.384		
	(0.103)	(0.105)		
θ	0.559	0.562		
	(0.132)	(0.137)		
d	0.581	0.581	0.444	0.443
	(0.064)	(0.066)	(0.016)	(0.016)
MSE_1	0.255	0.252	0.247	0.244
QLIKE	0.221	0.231	0.187	0.194

Note: Standard errors in parentheses.

long memory estimates are outside the stationary long memory bound, slightly greater than 0.5. Therefore, it can be said that the in-sample ARFIMA and MSFI estimates for $\widehat{d}$ are consistent with the full-sample property of the logarithmic series of the realised measures of volatility. In the case of the GARCH-type model estimation, it can be seen that the models which combine with the realised measure of volatility generally outperform, even if they do not take either regime changes or long memory into account in their specification. The in-sample forecast of GARCHX-RV has the smallest MSE_1. RGARCH-RV and -RK exhibit the most accurate fit to others in terms of QLIKE. The long memory FIEGARCH model follows those of the combined models. Interestingly, we can see that the persistence size of FIEGARCH, $\widehat{d} = 0.068$ is quite low relative to that of FIGARCH, $\widehat{d} = 0.440$ or GARCH, $\widehat{\alpha} + \widehat{\beta} = 0.993$. In this sense, we conjecture that the effects of structural changes driven by large negative shocks to return might be better explained by those model specifications, rather than GARCH or FIGARCH. Further, it could mean that the estimated persistence of GARCH and FIGARCH models might be spurious. The A-FIGARCH model, which can simultaneously capture long memory and structural change, does not seem to fit well relative to any other models, except GARCH. The standard short memory GARCH model is chosen as the worst in its in-sample fit for both loss measures.

Table 29: In-sample estimation for daily return volatility II

	GAR	EGA	FIG	FIEG	A-FIG	GX-RV	GX-RK	RG-RV	RG-RK
ω	0.013	0.005	0.093	0.002	0.181	0.105	0.110	0.226	0.210
	(0.005)	(0.003)	(0.025)	(0.001)	(0.004)	(0.029)	(0.031)	(0.017)	(0.016)
α	0.077	0.108	0.109	0.086	0.077	0.019	0.010		
	(0.012)	(0.019)	(0.070)	(0.015)	(0.041)	(0.013)	(0.013)		
β	0.916	0.980	0.487	0.989	0.323	0.693	0.668	0.637	0.640
	(0.013)	(0.006)	(0.115)	(0.003)	(0.062)	(0.069)	(0.076)	(0.021)	(0.021)
d			0.440	0.068	0.287				
			(0.071)	(0.022)	(0.005)				
γ		-1.024		-1.189	0.191	0.289	0.323	0.296	0.292
		(0.146)		(0.169)	(0.003)	(0.075)	(0.082)	(0.018)	(0.018)
δ					0.035				
					(0.002)				
ξ								-0.761	-0.714
								(0.032)	(0.033)
φ								1.125	1.132
								(0.043)	(0.044)
τ_1								-0.161	-0.160
								(0.010)	(0.010)
τ_2								0.047	0.047
								(0.005)	(0.005)
MSE_1	0.845	0.534	0.836	0.428	0.710	0.376	0.377	0.413	0.413
QLIKE	1.580	1.307	1.311	1.190	1.340	1.081	1.053	0.934	0.934

Note: GAR (GARCH), EGA (EGARCH), FIG (FIGARCH), FIEG (FIEGARCH), A-FIG (A-FIGARCH), GX-RV, RK (GARCHX with RV_t, RK_t), RG-RV, RK (Log-RGARCH with RV_t, RK_t). Standard errors in parentheses.

4.6.2 Out-of-Sample Analysis

One-step-ahead conditional variance forecasts are constructed for realised measures of volatility and daily return-based volatility models, as described in Section 4.5. Table 30 presents the forecast accuracy with respect to five loss functions. In terms of the out-of-sample predictive accuracy for all the loss functions, the ARFIMA and the MSFI models using the realised measures of volatility outperforms relative to the GARCH-type models which are based on the conditional variance of the daily return. It is also difficult to find any big difference in the losses from the ARFIMA and the MSFI models between the one-day-ahead forecasts of RV_t and RK_t. In this respect, our finding may support the fact that the daily return conditional variance is even noisier than realised variance or realised kernel, even though realised kernel is a noisy proxy of a true volatility. In terms of MSE_1 and QLIKE, the simple long memory-based ARFIMA models present slightly better out-of-sample fit than the regime switching MSFI models. In contrast, it is reported by two MAE losses that the MSFI models produce slightly smaller mean losses than ARFIMA. In a comparison between a variance-based forecast loss and a standard deviation-based forecast loss, the mean losses from MSE_2 and MAE_2 look to be quite comparable across the employed models, so that the standard deviation-based losses is less likely to order the rank of the models than the variance-based loss functions. Moreover, the out-of-sample fit between RV_t and RK_t is not quite distinguishable.

Table 30: Out-of-sample evaluation of the one-day-ahead forecasts

	Loss functions					MZ regression		
	MSE_1	MSE_2	MAE_1	MAE_2	QLIKE	β_0	β_1	p-value
ARFIMA-RV	0.154	0.039	0.194	0.135	-0.049	-0.010 (0.039)	1.251 (0.140)	0.000
ARFIMA-RK	0.154	0.039	0.194	0.135	-0.049	-0.011 (0.039)	1.252 (0.141)	0.000
MSFI-RV	0.156	0.039	0.193	0.134	-0.052	-0.030 (0.040)	1.323 (0.145)	0.000
MSFI-RK	0.157	0.039	0.193	0.134	-0.052	-0.031 (0.040)	1.321 (0.145)	0.000
GARCH	0.327	0.113	0.414	0.276	0.103	-0.009 (0.038)	0.580 (0.067)	0.000
GARCH(0.50)	0.249	0.088	0.351	0.240	0.058	-0.013 (0.040)	0.677 (0.081)	0.000
GARCH(0.25)	0.213	0.074	0.316	0.223	0.036	-0.033 (0.045)	0.777 (0.095)	0.000
GARCH(break)	0.237	0.081	0.333	0.232	0.048	0.011 (0.041)	0.680 (0.086)	0.000
EGARCH	0.318	0.107	0.414	0.271	0.079	-0.034 (0.038)	0.598 (0.064)	0.000
FIGARCH	0.272	0.096	0.376	0.256	0.074	-0.036 (0.038)	0.651 (0.070)	0.000
FIEGARCH	0.289	0.098	0.393	0.260	0.064	-0.044 (0.037)	0.625 (0.063)	0.000
A-FIGARCH	0.233	0.088	0.354	0.249	0.071	-0.090 (0.043)	0.765 (0.080)	0.000
GARCHX-RV	0.250	0.103	0.407	0.288	0.129	-0.210 (0.059)	0.852 (0.094)	0.000
GARCHX-RK	0.241	0.098	0.394	0.280	0.117	-0.195 (0.059	0.852 (0.095)	0.000
RGARCH-RV	0.330	0.109	0.428	0.285	0.102	-0.039 (0.041)	0.593 (0.065)	0.000
RGARCH-RV	0.301	0.101	0.408	0.274	0.088	-0.043 (0.041)	0.619 (0.068)	0.000
EWMA	0.301	0.095	0.357	0.233	0.047	0.075 (0.037)	0.531(0.077)	0.000
LM-EWMA	0.233	0.081	0.330	0.227	0.038	-0.014 (0.037)	0.691 (0.077)	0.000

Note: p-values are for the F-test of the joint hypothesis of $\mathbf{H}_0$: $\beta_0 = 0$ and $\beta_1 = 1$. The standard errors of the MZ regression coefficients are in parentheses.

Next, the predictive ability of the GARCH-type models varies overall, but some models consistently keep their relative superiority in terms of loss rank. It can be seen that the GARCH rolling window model with the size of 0.25 consistently generates the most accurate forecasts in the entire loss functions. The LM-EWMA model also produces more accurate one-day-ahead forecasts than other GARCH models, following the best-performed GARCH(0.25) in terms of the model rank. However, the GARCH(break) and the A-FIGARCH forecasting models, which can account for the effect of the structural breaks in their own specifications perform better than the long memory-based GARCH models such as FIGARCH and FIE-GARCH. As discussed in Section 4.4, we have verified that the $\log r_t^2$ process exhibits spurious long memory, and is contaminated by structural changes. Therefore, it can be said that the presence of structural changes is an issue when forecasting GARCH-type volatility by means of the daily return volatility, even if the actual volatility is a true long memory. Further, when we compare the out-of-sample fit between two different rolling window sizes for GARCH(0.25) and GARCH(0.50), we realise that a proper choice of the size of rolling window is critical in prediction accuracy under possible presence of structural breaks. Unlike the in-sample fit, the recursive forecasts of the GARCH with realised measures cannot fit well to the true volatility proxy.

Table 30 summarises the results of the Mincer-Zarnowitz regression of RK_t on each of individual forecasts. We report the estimate of the MZ coefficients and the p-values of the joint test for the null of unbiasedness hypothesis. It can be ensured that all the generated forecasts are statistically biased from the volatility proxy RK_t, rejecting the null at any significance level. On the other hand, it is also worthy to note that we have yet any clues whether such loss differences by given loss functions are statistically significant to rank the forecasts.

Table 31: DMW test results for the MSE_1 loss I

	ARFIMA-RV	ARFIMA-RK	MSFI-RV	MSFI-RV	GARCH(0.25)
ARFIMA-RK	-1.931				
	(0.054)				
MSFI-RV	-1.125	-0.922			
	(0.261)	(0.356)			
MSFI-RK	-1.242	-1.071	-1.416		
	(0.214)	(0.284)	(0.157)		
GARCH(0.25)	-3.725	-3.684	-3.387	-3.359	
	(0.000)	(0.000)	(0.001)	(0.001)	
LM-EWMA	-3.884	-3.849	-3.591	-3.567	-3.108
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)

Note: p-values in parentheses. A minus (plus) sign indicates that the model in the column has smaller (greater) MSE_1 than the model in the row.

The equal predictive ability test for the MSE_1 losses is conducted utilising DMW. Since the MSE_1 for the ARFIMA-type models are considerably smaller than the one for the GARCH models, we firstly carry out pairwise comparisons for the ARFIMA-type models with GARCH(0.25) and LM-EWMA. The pairwise comparison result of the test is described in Table 31. If the sign of the DMW test statistic is negative, it implies that a forecast model in the column produces more accurate forecasts than a model in the row, and vice versa. The DMW results in that the MSE_1 of the ARFIMA-RV forecasts is smaller than that of ARFIMA-RK with 5.4% significance. On the other hand, the MSE_1 losses of GARCH(0.25) and LM-EWMA are significantly larger than all of the ARFIMA and the MSFI models. However, it is still difficult to determine the superiority of the relative predictive accuracy among other models which fail to reject the null of the equality of the loss. On the other hand, we separately evaluate the equal predictive accuracy only for the GARCH-type models, and the DMW outcomes are displayed in Table 32. In the evaluation for the GARCH-type models, the pattern of forecast model superiority by MSE_1 keeps holding in the pairwise comparison by the DMW test. The GARCH(0.25) forecasts are significantly less biased than all of the other competing forecasts. It can be also seen that GARCH(break), A-FIGARCH and LM-EWMA are preferred relative to the long memory-based models and the GARCHX and RGARCH models with the realised measures. However, the significance of equal predictive accuracy are not always ensured for some cases of comparison.

In turn, we now deliver further comparisons for a large set of multiple competing forecasts. The results of forecasting have clearly addressed that the ARFIMA-type models with realised measures of volatility produce better predictive accuracy than a class of GARCH forecasts. And also, it has been shown that a class of GARCH forecasts, which may take structural changes into account, can outperform the long memory-based model and the GARCHX and the RGARCH models with realised measures of volatility. Nonetheless, it is obvious that these results are obtained from the selected empirical data, particularly adopted for this study. In this sense, we apply the SPA test and the MCS approach to assess whether we can obtain similar outcomes for ensuring a consistent decision in terms of forecast superiority among the employed models. As noted earlier, it can be seen that there is a large difference in the losses between the ARFIMA and GARCH-type models. And also, the DMW test has shown that the most accurate model selected among a class of the GARCH-type models still performs worse than every ARFIMA-type model. Therefore, we deliver the SPA and the MCS evaluations for these two types of models separately. In particular, for the SPA test, we set each of the individual forecasting models as a benchmark, and set all the others in a given class of models as a competing set. A high p-value in the SPA implies non-rejection of the null that is a benchmark not outperformed. Thus, the model with the highest p-value can be chosen as the best of the competing models. An implication of the high p-value for the MCS

Table 32: DMW test results for the MSE_1 loss II

	GAR	GA50	GA25	GwB	EGA	FIG	FIEG	A-FIG	GX-RV	GX-RK	RG-RV	RG-RK	EWMA
GARCH(0.50)	7.028												
	(0.000)												
GARCH(0.25)	6.937	5.183											
	(0.000)	(0.000)											
GARCH(break)	6.008	1.431	-5.485										
	(0.000)	(0.152)	(0.000)										
EGARCH	0.603	-3.397	-4.284	-3.414									
	(0.547)	(0.001)	(0.000)	(0.001)									
FIGARCH	7.074	-2.397	-4.555	-2.820	2.831								
	(0.000)	(0.017)	(0.000)	(0.005)	(0.005)								
FIEGARCH	2.094	-1.902	-3.116	-2.161	3.427	-0.946							
	(0.036)	(0.057)	(0.002)	(0.031)	(0.001)	(0.344)							
A-FIGARCH	6.838	2.248	-2.901	0.497	3.852	4.405	2.521						
	(0.000)	(0.025)	(0.004)	(0.619)	(0.000)	(0.000)	(0.012)						
GARCHX-RV	3.971	-0.081	-2.233	-0.682	3.386	1.350	1.850	-1.108					
	(0.000)	(0.936)	(0.026)	(0.496)	(0.001)	(0.177)	(0.064)	(0.268)					
GARCHX-RK	4.219	0.489	-1.710	-0.211	3.648	1.812	2.189	-0.515	5.796				
	(0.000)	(0.625)	(0.087)	(0.833)	(0.000)	(0.070)	(0.029)	(0.606)	(0.000)				
RGARCH-RV	-0.087	-2.554	-3.460	-2.680	-0.467	-2.020	-1.491	-3.029	-3.911	-4.310			
	(0.931)	(0.011)	(0.001)	(0.007)	(0.640)	(0.043)	(0.136)	(0.002)	(0.000)	(0.000)			
RGARCH-RK	0.965	-1.836	-2.911	-2.038	0.785	-1.132	-0.461	-2.382	-3.062	-3.571	7.000		
	(0.335)	(0.066)	(0.004)	(0.042)	(0.432)	(0.258)	(0.645)	(0.017)	(0.002)	(0.000)	(0.000)		
EWMA	5.792	-5.438	-6.064	-4.983	0.978	-3.118	-0.612	-5.173	-2.500	-2.829	0.882	-0.023	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.328)	(0.002)	(0.541)	(0.000)	(0.012)	(0.005)	(0.378)	(0.982)	
LM-EWMA	8.601	4.200	-3.108	0.723	4.226	5.137	2.712	0.177	1.143	0.552	3.085	2.431	7.049
	(0.000)	(0.000)	(0.002)	(0.470)	(0.000)	(0.000)	(0.007)	(0.860)	(0.253)	(0.581)	(0.002)	(0.015)	(0.000)

Note: GAR (GARCH), GA50 (GARCH 0.50 rolling window), GA25 (GARCH 0.25 rolling window), GwB (GARCH with break), EGA (EGARCH), FIG (FIGARCH), FIEG (FIEGARCH), A-FIG (A-FIGARCH), GX-RV, RK (GARCHX with RV_t, RK_t), RG-RV, RK (Log-RGARCH with RV_t, RK_t). p-values in parentheses. A minus (plus) sign indicates that the model in the column has smaller (greater) MSE_1 than the model in the row.

is similar with that of the SPA. Although the MCS approach does not require a benchmark, the highest p-valued model is the best-performing forecast relative to the competing models within a given model confidence set.

Table 33 presents the SPA and the MCS results for the MSE_1, MAE_1 and QLIKE losses. For the ARFIMA-type models, the best-performing model varies between the SPA and the MCS, being mismatched with the first ranked model with respect to each of the loss functions. By the SPA loss evaluation criterion, ARFIMA-RV, MSFI-RV and MSFI-RK are chosen as the best models with respect to MSE_1, MAE_1 and QLIKE, respectively. The MCS approach chooses the MSFI-RK as the best in terms of MSE_1 and MAE_1. ARFIMA-RK outperforms for QLIKE in the MCS criterion. Recalling forecasting results for loss functions, the ARFIMA models have shown predictive accuracy relative to the MSFI models in terms of MSE_1 and QLIKE. In contrast, the MSFI models have had lower MAE_1 than the ARFIMA models. Comparing all these results above, it would be difficult to find certain consistent patterns of forecast superiority across a class of ARFIMA-type models. Moreover, it is hard to identify any particular preference between two different realised measurements either. Finally, these results may indicate that each of the individual models does not produce significantly different forecasts when an actual volatility exhibits a true long memory. On the other hand, for a class of the GARCH-type models, the SPA and the MCS consistently report that GARCH(0.25) is the most preferred model over the entire loss functions, followed by LM-EWMA and A-FIGARCH. Unlike the loss function-based and DMW predictive ability, however, the GARCH forecasting model with structural changes has no significant preference in terms of relative predictive ability by the evaluation based on these loss criteria. The long memory-based FIGARCH and FIEGARCH forecasts are inferior against the proxy process which is a long memory. Finally, we may conclude that the GARCH rolling window forecast with a proper estimation size can dominate the long memory GARCH forecasts, even if the return volatility is spurious long memory which is mainly driven by the structural breaks, and also the proxy of volatility exhibits a true long memory property.

4.6.3 Sensitivity Analysis of Rolling Window Size for GARCH Model

Based on the forecast evaluation, it has been found that the GARCH 0.25 rolling window forecast is selected as the best-performing model among the class of GARCH-type models.

Table 33: Results of SPA and MCS tests

	SPA			MCS		
	MSE_1	MAE_1	QLIKE	MSE_1	MAE_1	QLIKE
ARFIMA-RV	1.000	0.189	0.094	0.201	0.180	0.820
ARFIMA-RK	0.046	0.058	0.102	0.201	0.180	1.000
MSFI-RV	0.240	1.000	0.686	0.201	0.180	0.297
MSFI-RK	0.154	0.032	1.000	1.000	1.000	0.297

	SPA			MCS		
	MSE_1	MAE_1	QLIKE	MSE_1	MAE_1	QLIKE
GARCH	0.000	0.000	0.000	0.000	0.000	0.000
GARCH(0.50)	0.000	0.000	0.000	0.000	0.000	0.000
GARCH(0.25)	1.000	1.000	1.000	1.000	1.000	1.000
GARCH(break)	0.000	0.000	0.000	0.000	0.000	0.000
EGARCH	0.000	0.000	0.000	0.000	0.000	0.000
FIGARCH	0.000	0.000	0.000	0.000	0.000	0.000
FIEGARCH	0.006	0.000	0.000	0.000	0.000	0.000
A-FIGARCH	0.007	0.000	0.000	0.054	0.062	0.000
GARCHX-RV	0.000	0.000	0.000	0.000	0.000	0.000
GARCHX-RK	0.110	0.000	0.000	0.054	0.062	0.000
RGARCH-RV	0.000	0.000	0.000	0.000	0.000	0.000
RGARCH-RV	0.002	0.000	0.000	0.000	0.000	0.000
EWMA	0.000	0.000	0.146	0.000	0.000	0.355
LM-EWMA	0.004	0.002	0.767	0.054	0.062	0.756

Note: We reprot p-values of the SPA and the MCS. For the SPA test, the benchmark corresponds to the model in the first column.

However, the GARCH 0.50 rolling window model cannot produce comparable forecasts against the best model, even though GARCH(0.50) outperforms the long memory-based models. In this respect, in the presence of structural breaks, we suspect here that the performance of a rolling window GARCH forecast is quite sensitive to the size of rolling window, and the forecasting performance can be improved when the rolling window size is about to reflect the effect of change in return volatility. To assess this argument, we would further like to demonstrate how the relative performance of the GARCH rolling window forecasting depends on a certain rolling window size. In addition to 0.25 and 0.50, we consider three more estimation sizes for rolling window forecasting which are 0.40, 0.60 and 0.75.

As displayed in Table 34, the first observation in an initial set of sub-sample data used to produce the first out-of-sample forecast (denoted as 'Starting date') is relatively close to the last break point detected by BP, BHKS, and the modified ICSS. The GARCH(0.75)-starting date is located around the first break point of BP and the modified ICSS. Overall, the starting date determined by rolling window size is much closer to the break dates of log-squared return

Table 34: Rolling window and structural changes

	GARCH(0.25)	GARCH(0.40)	GARCH(0.50)	GARCH(0.60)	GARCH(0.75)
Rolling window size	617	988	1235	1481	1852
Starting date	16/07/2003	14/01/2002	03/01/2001	06/01/2000	13/07/1998
Close break (r_t^2)	09/10/2003				20/01/1997
Close break ($\log r_t^2$)	22/07/2003	25/04/2002			27/07/1998

Note: Starting date indicates the date of the first observation in the initial estimation window which is used to generate the first out-of-sample forecast. Refer to Table 4 for the detected break dates of r_t^2 and $\log r_t^2$. Date format is dd/mm/yyyy.

series. The GARCH(0.50) and GARCH(0.60) cases have no corresponding structural change points.

Next, we generate forecasts of those five rolling window models and calculate their mean losses using the MSE_1 and the QLIKE functions. The results are presented in Table 35. When we consider 0.40 rolling window size, the mean loss is slightly smaller than the 0.25 rolling window size in both MSE_1 and QLIKE. It can be found that other rolling window forecasts tend to get inferior in loss evaluation as the rolling window size increases. These outcomes may imply that the last break point detected is more likely to be effective in order to improve the predictive ability of the GARCH rolling window model, and an inclusion of the small number of pre-break observations is also helpful to obtain more accurate forecasts, as pointed out by Pesaran and Timmermann (2007) and Clark and McCracken (2009), amongst others.

Table 35: MSE_1 and QLIKE losses and loss ratios, relative to GARCH(0.25)

	GARCH(0.25)	GARCH(0.40)	GARCH(0.50)	GARCH(0.60)	GARCH(0.75)
MSE_1	0.213	0.212	0.249	0.283	0.303
QLIKE	0.036	0.036	0.058	0.073	0.080
MSE_1 ratio	1.000	0.995	1.172	1.330	1.427
QLIKE ratio	1.000	0.989	1.607	2.006	2.216

Note: Loss ratio is calculated by dividing loss of GARCH(0.25) into the loss of a competing rolling window model.

To complete a comparison among the GARCH rolling window forecasting model, we apply the conditional predictive ability test of Giacomini and White (2006), which is available to check the significance of equal predictive ability for nested and non-nested rolling window-based forecasting models. Giacomini and White (2006) propose a two-step procedure to determine the best forecast at a certain point in the future, using current information. The test is designed to provide a forecast decision rule in a case where equal (conditional) predictive ability is rejected. The null hypothesis is $\mathbf{H}_0 : E\left[d_{t,ij}|\,\mathcal{F}_{t-1}\right] = 0$ for all i and j, where $d_{t,ij} = L\left(\hat{\sigma}^2_{t,i}, \tilde{\sigma}^2_t\right) - L\left(\hat{\sigma}^2_{t,j}, \tilde{\sigma}^2_t\right)$. The conditional moment condition for the expectation of loss differential can be derived as $E\left[f_t d_{t+1,ij}\right] = 0$, where f_t is a $q \times 1$ vector of any $\mathcal{F}_{t-1}$ measurable function. We use the test function $f_t = \left(1, d_{t,ij}\right)'$. As the first step of the CPA procedure, we regress $d_{t+1,ij}$ on f_t over the out-of-sample period and let ψ be the vector of the regression coefficient.

$$d_{t+1,ij} = \psi' f_t + e_t.$$

In the context, $\widehat{\psi}' f_t \approx E\left[d_{t,ij}|\,\mathcal{F}_{t-1}\right]$, so that the original null can be tested by testing for $\psi = 0$. The test statistic of the CPA is formed as a Wald-type test statistic, given by

$$T^{CPA} = (p-1)\left(\frac{1}{p-1}\sum_{t=T-p+1}^{T-1} f_t d_{t+1,ij}\right)' \widehat{\Omega}_p^{-1}\left(\frac{1}{p-1}\sum_{t=T-p+1}^{T-1} f_t d_{t+1,ij}\right),$$

where $\widehat{\Omega}_p$ is defined as $p^{-1}\sum_{t=T-p+1}^{T-1}\left(f_t d_{t+1,ij}\right)'\left(f_t d_{t+1,ij}\right)$, which is a consistent estimate of variance of $f_t d_{t+1,ij}$. The test statistic asymptotically converges to χ^2_q with q degrees of freedom. If we cannot reject the null hypothesis of equal predictive ability, both models would have no statistically significant difference in terms of forecasting performance, conditioning on the information used in the procedure. In contrast, a rejection of the null implies that f_t is statistically informative to identify which forecasting model is superior relative to the other in a pairwise comparison, for the future date of interest. In case of a rejection, we proceed to the second step. The decision rule in this step is to select $\hat{\sigma}^2_{t,i}$ if $\widehat{\psi}' f_t > 0$ as superior one on interpreting the test results, otherwise select $\hat{\sigma}^2_{t,j}$ if $\widehat{\psi}' f_t < 0$.

The CPA test results are reported in Table 36. The critical value is obtained from χ^2 distribution at 5% significance level. The upper triangular part of Table 36 is based on the QLIKE loss series of a class of GARCH rolling window models. The lower triangular part contains the test result based on the MSE_1 losses. It is evident that the model superiority by the loss functions is consistent with the model superiority by the CPA test. In effect, it can be said that the loss differences across the GARCH rolling window models are statistically significant. Therefore, it has been empirically showed that the appropriate choice of the rolling window size for the GARCH model is of importance to achieve relatively better predictive ability in the structural breaks even when the proxy of an actual volatility exhibits

Table 36: CPA test results for MSE_1 and QLIKE losses

	GARCH(0.25)	GARCH(0.40)	GARCH(0.50)	GARCH(0.60)	GARCH(0.75)
GARCH(0.25)		-83.92	78.20	77.77	87.76
		(0.000)	(0.000)	(0.000)	(0.000)
GARCH(0.40)	31.72		60.42	72.82	91.15
	(0.000)		(0.000)	(0.000)	(0.000)
GARCH(0.50)	-41.19	-35.42		104.45	85.05
	(0.000)	(0.000)		(0.000)	(0.000)
GARCH(0.60)	-54.04	-46.74	-47.26		78.63
	(0.000)	(0.000)	(0.000)		(0.000)
GARCH(0.75)	-47.68	-40.06	-32.63	-17.27	
	(0.000)	(0.000)	(0.000)	(0.000)	

Note: The CPA test outcomes are presented in the lower triangular part for MSE_1, and in the upper triangular part for QLIKE. *p*-values in parentheses. A minus (plus) sign indicates that the model in the column outperforms (is outperformed by) the model in the row more than 50% of the time, in terms of either MSE_1 or QLIKE.

long memory. And also, our empirical result recommends to use some of pre-break data when forecasting volatility by means of rolling window GARCH, as suggested in the relevant literature.

4.7 Concluding Remarks

The memory properties of the historical and realised measures of volatility series have been investigated. The baseline task of this study has been to identify whether the daily S&P 500 index return volatility process is long memory or spurious long memory, subject to structural changes. Moreover, we have evaluated the relative performance of forecasting models and have analysed the predictive content based on identified memory properties of the given data generating volatility processes. Prior to the out-of-sample forecast analysis, we found that two different unbiased volatility measures exhibit inconsistent memory properties for the same underlying S&P 5000 stock index return series. As shown by the testing frameworks employed, both realised variance and realised kernel volatility processes show long memory, while the squared daily return process is stationary short memory contaminated by structural changes. Through forecast performance evaluations, we found that a class of ARFIMA models with realised measures of volatility have significantly outperformed the historical volatility-based GARCH class models. The applied loss functions have made the consistent results in favour of superiority of the realised measurement-based models for the realised kernel proxy. In a class of GARCH models, the rolling GARCH with 0.25 estimation window model apparently generates more accurate forecasts relative to other historical volatility-based models. All the statistical tests applied for evaluating forecast losses support those findings as well, in terms of the pairwise comparison for equal predictive ability (DMW) and the multiple model comparison for superior predictive ability (SPA) and model confidence set approach (MCS). Moreover, it has been found that long memory-based FIGARCH and FIEGARCH models are outperformed by some GARCH forecasts which can reflect the effects of the structural breaks. In the context of GARCH volatility, it implies that the presence of structural breaks in estimated volatility processes does matter, so that its effect should be accounted for when forecasting volatility, even if the true volatility shows long memory properties. It has also been revealed that the rolling window size is quite a sensitive factor in generating more accurate forecasts of conditional volatility under the structural breaks to produce accurate forecast.

Our empirical findings may contribute to practitioners in a variety of field of financial applications when forecasting models need to be chosen amidst in a confusion regarding structural breaks and long memory. Along the lines of the existing literature, the availability of high-frequency-based realised measures of volatility still guarantees apparently relative superiority of realised volatility measurement by means of long memory-based forecasting models. In the level of lower-frequency daily squared return volatility, the presence of structural breaks is critical to determine short-term forecasting model superiority even when the true volatility generation is long memory. Our results also suggest the importance of choosing an proper rolling window size when estimating and forecasting conditional volatility by means of the GARCH model under structural breaks. However, the results illustrated in this chapter are based on one-day-ahead forecast case only, so that our study can be extended by evaluating the predictive content of even longer-term forecasts such as weekly or monthly forecasts. Further, it would also be interesting to attempt a theoretical demonstration of possible causes and effects of different memory properties across different volatility measures for the same underlying return series.

Chapter 5 Conclusion

The aim of this thesis is to contribute to better understanding of topics in financial time series. We have introduced a new misspecification testing framework for the GARCH-based parametric model under the joint dependence between the realised measure of volatility and the squared error process. Moreover, this thesis has evaluated the relative predictive ability and financial economic benefits of a variety of financial volatility models with particular emphasis on structural breaks and long memory properties.

In the second chapter, we provide a unifying and generic class of misspecification testing frameworks for the Realised GARCH(p, q) model. The proposed test statistics are constructed based on the conditional moment principle, having an asymptotic chi-square distribution under the null. The misspecification test procedures can be simply applicable in practice without further bootstrapping procedures etc., and help reduce the cost of time and computation load. In addition, our analysis of the conditional mean effect in estimation and testing has provided theoretical soundness for the test statistic to make it robust to the conditional heteroskedasticity of the return process. We have shown that the conditional mean parameter effects in the estimation and testing for the Realised GARCH are effectively negligible in the absence of the squared error process in the GARCHX specification. Therefore, the Realised GARCH model with various types of mean specifications can be tested using the proposed generic framework without any loss of generality. Our Monte Carlo experiment reveals that the proposed test statistics have good finite sample size properties and high degrees of power against alternative data generation process. In particular, the test statistic that accounts for the recursive nature of the conditional variance appears to be a powerful tool in the detection of the potential misspecification of the null model arising from asymmetry behaviour in financial asset returns. The empirical application also supports that the test statistic with the recursive nature of the processes works very well when the size of the asymmetry in the leverage effect is large enough. Specifically, the asymmetry test rejects the null at any significance level for the stock returns with a higher degree of asymmetric volatility. The further discussions would be of interest, for example, to investigate the asymptotic properties of the proposed misspecification tests of the Realised GARCH model. Moreover, the misspecification testing framework can be extended to dealing with a non-stationary covariate case such as long memory.

In the third chapter, a Monte Carlo experiment is conducted to investigate the relative out-of-sample predictive ability of a class of parsimonious conditional variance models when either a structural break or long-run dependence is allowed for a conditional variance process. The results of our experiment reveal some supporting evidence of the discussions of the existing relevant literature. If the conditional variance process is stationary short or long memory in the absence of a structural break, the forecasting models which are able to capture the properties of the true process are more favourable than any other misspecified models. When the true short memory process is contaminated by a structural break, the detection of the break may play an important role in choosing a proper window size for the short-run forecasting. Further we have found that spurious long memory may strongly dominate the true structural break in long-run forecasting when the true short memory process is highly persistent. However, it has not been easy to justify any consistent features or patterns in forecast superiority among the individual forecasting models when the structural break is located around the end of the in-sample period. It might be due to relatively small number of observations used for estimation. Nevertheless, it can be seen that the long memory-based forecasts are generally better off than the short memory-based competing forecasts in the presence of the most recent break. On the other hand, two forecast combinations are very favourable in the presence of a structural break, regardless of the forecast horizon and the level of persistence. A number of extensions would be possible, based on the limitations of this study. For example, our simulation design can be naturally extended to accounting for more general non-stationary volatility processes which are subject to structural break and long memory simultaneously or other non-linearities. Moreover, it would be more informative in a general sense if we additionally take more various conditional volatility models such as stochastic volatility models and markov-switching models into account.

In the fourth chapter, the memory properties of the historical and realised measures of volatility series have been investigated. Moreover, we have evaluated the relative performance of forecasting models and have analysed the predictive content based on identified memory properties of given data generating volatility processes. Prior to the out-of-sample forecast analysis, we have found that two different unbiased volatility measures exhibit inconsistent memory properties for the same underlying S&P 5000 stock index return series. As shown by the several of the testing frameworks employed, both realised variance and realised kernel volatility processes show long memory, while the squared daily return process is stationary short memory contaminated by structural changes. Through forecast performance evaluations, we found that a class of ARFIMA models with realised measures of volatility have significantly outperformed the historical volatility-based GARCH class models. The applied loss functions have made consistent results in favour of superiority of the realised measurement-based models for the realised kernel proxy. In a class of GARCH models, the rolling GARCH with 0.25 estimation window model apparently generates more accurate forecasts than the other historical volatility-based models. In the context of GARCH volatility, this implies that the presence of structural breaks in estimated volatility processes does matter, so that their effects should be accounted for when forecasting volatility, even if the true volatility shows long memory properties. And also, it has been revealed that the rolling window size is quite a sensitive factor in generating more accurate forecasts of conditional volatility under structural breaks to produce more accurate forecasts. The results illustrated in this chapter are based on one-day-ahead forecast case only, so that our study can be extended by evaluating the predictive content of even longer-term forecasts such as weekly or monthly forecasts. Further, it would also be interesting to attempt a theoretical demonstration of possible causes and effects of different memory properties across different volatility measures for the same underlying return series.

References

Andersen, Torben G., Bollerslev, Tim 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* **39**(4): 885–905.

Andersen, Torben G., Bollerslev, Tim, Diebold, Francis X., Ebens, Heiko 2001a. The distribution of stock return volatility. *Journal of Financial Economics* **61**: 43–76.

Andersen, Torben G., Bollerslev, Tim, Diebold, Francis X., Labys, Paul 2001b. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* **96**(453): 42–55.

_ 2003. Modeling and forecasting realized volatility. *Econometrica* **71**(2): 579–625.

Andreou, Elena, Ghysels, Eric 2002. Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics* **17**(5): 579–600.

_ 2009. Structural breaks in financial time series. In *Handbook of financial time series,* ed. Torben G. Andersen, Richard a. Davis, Jens-Peter Kreiss, Thomas Mikosch. Springer Berlin Heidelberg pp. 839–870.

Awartani, Basel M., Corradi, Valentina 2005. Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries. *International Journal of Forecasting* **21**(1): 167–183.

Baek, Changryong, Pipiras, Vladas 2014. On distinguishing multiple changes in mean and long-range dependence using local Whittle estimation. *Electronic Journal of Statistics* **8**: 931–964.

Bai, Jushan, Perron, Pierre 1998. Estimating and testing linear models with multiple structural changes. *Econometrica.*

_ 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* **18**(1): 1–22.

Baillie, Richard T. 1996. Long memory processes and fractional integration in econometrics. *Journal of Econometrics* **73**(1): 5–59.

Baillie, Richard T., Bollerslev, Tim, Mikkelsen, Hans O. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **74**(1): 3–30.

Baillie, Richard T., Morana, Claudio 2009. Modelling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach. *Journal of Economic Dynamics and Control* **33**(8): 1577–1592.

Banerjee, Anindya, Urga, Giovanni 2005. Modelling structural breaks, long memory and stock market volatility: an overview. *Journal of Econometrics* **129**(1-2): 1–34.

Barndorff-Nielsen, Ole E., Hansen, Peter R., Lunde, Asger, Shephard, Neil 2008. Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* **76**(6): 1481–1536.

_ 2009. Realized kernels in practice: trades and quotes. *Econometrics Journal* **12**(3): C1–C32.

Barndorff-Nielsen, Ole E., Shephard, Neil 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* **64**(2): 253–280.

_ 2007. Variation, jumps and high frequency data in financial econometrics. In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Econometric Society Monographs,* ed. Richard Blundell, Torsten Persson, Whitney K. Newey. Cambridge: Cambridge University Press pp. 328–372.

Berkes, István, Horváth, Lajos, Kokoszka, Piotr 2003. GARCH processes: Structural and estimation. *Bernoulli* **9**(2): 201–227.

Berkes, István, Horváth, Lajos, Kokoszka, Piotr, Shao, Qi-Man 2006. On discriminating between long-range dependence and changes in mean. *The Annals of Statistics* **34**(3): 1140–1165.

Bollerslev, Tim 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3): 307–327.

Bollerslev, Tim, Engle, Robert F. 1993. Common persistence in conditional variances. *Econometrica* **61**(1): 167.

Bollerslev, Tim, Mikkelsen, Hans O. 1996. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* **73**(1): 151–184.

Choi, Kyongwook, Yu, Wei-Choun, Zivot, Eric 2010. Long memory versus structural breaks in modeling and forecasting realized volatility. *Journal of International Money and Finance* **29**(5): 857–875.

Choi, Kyongwook, Zivot, Eric 2007. Long memory and structural changes in the forward discount: An empirical investigation. *Journal of International Money and Finance* **26**(3): 342–363.

Clark, Todd E., McCracken, Michael W. 2009. Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review* **50**(2): 363–395.

Davidson, James E.H. 2004. Moment and Memory Properties of Linear Conditional Heteroscedasticity Models, and a New Model. *Journal of Business & Economic Statistics* **22**(1): 16–29.

Diebold, Francis X. 1986. Modeling the persistence of conditional variances: A comment. *Econometric Reviews* **5**(1): 51–56.

Diebold, Francis X., Inoue, Atsushi 2001. Long memory and regime switching. *Journal of Econometrics* **105**(1): 131–159.

Diebold, Francis X., Mariano, Roberto S. 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* **13**(3): 134–144.

Engle, Robert F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**(4): 987–1008.

_ 2002. New frontiers for ARCH models. *Journal of Applied Econometrics* **17**(5): 425–446.

Engle, Robert F., Gallo, Giampiero M. 2006. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* **131**(1-2): 3–27.

Engle, Robert F., Ng, Victor K. 1993. Measuring and testing the impact of news on volatility. *The Journal of Finance* **48**(5): 1749–1778.

Fleming, Jeff, Kirby, Chris, Ostdiek, Barbara 2008. The specification of GARCH models with stochastic covariates. *Journal of Futures Markets* **28**(10): 911–934.

Forsberg, Lars, Bollerslev, Tim 2002. Bridging the gap between the distribution of realized (ECU) volatility and ARCH modelling (of the Euro): the GARCH-NIG model. *Journal of Applied Econometrics* **17**(5): 535–548.

Francq, Christian, Zakoian, Jean-Michel 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**(4): 605–637.

Geweke, John, Porter-Hudak, Susan 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4**(4): 221–238.

Giacomini, Raffaella, White, Halbert 2006. Tests of conditional predictive ability. *Econometrica* **74**(6): 1545–1578.

Giraitis, Liudas, Kokoszka, Piotr, Leipus, Remigijus, Teyssière, Gilles 2003. Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics* **112**(2): 265–294.

González-Rivera, Gloria, Lee, Tae-Hwy, Mishra, Santosh 2004. Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting* **20**(4): 629–645.

Granger, Clive W. J. 1980. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* **14**(2): 227–238.

_ 1981. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* **16**(1): 121–130.

Granger, Clive W. J., Ding, Zhuanxin 1996. Varieties of long memory models. *Journal of Econometrics* **73**(1): 61–77.

Granger, Clive W. J., Hyung, Namwon 2004. Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance* **11**(3): 399–421.

Granger, Clive W. J., Joyeux, Roselyne 1980. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* **1**(1): 15–29.

Halunga, Andreea G., Orme, Chris D. 2009. First-order asymptotic theory for parametric misspecification tests of GARCH models. *Econometric Theory* **25**(02): 364.

Hamilton, James D, Susmel, Raul 1994. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* **64**(1-2): 307–333.

Han, Heejoon, Kristensen, Dennis 2014. Asymptotic theory for the QMLE in GARCH-X models with stationary and non-stationary covariates. *Journal of Business & Economic Statistics* **32**(3): 416–429.

Han, Heejoon, Park, Myung D. 2013. Comparison of realized measure and implied volatility in forecasting volatility. *Journal of Forecasting* **32**: 522–533.

Hansen, Peter R. 2005. A test for superior predictive ability. *Journal of Business & Economic Statistics* **23**(4): 365–380.

Hansen, Peter R., Huang, Zhuo 2012. Exponential GARCH modeling with realized measures of volatility. *Working paper.*

Hansen, Peter R., Huang, Zhuo, Shek, Howard H. 2012. Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* **27**(6): 877–906.

Hansen, Peter R., Lunde, Asger 2005. A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* **20**(7): 873–889.

_ 2011. Forecasting volatility using high frequency data. In *Oxford Handbook of Economic Forecasting,* ed. Michael P. Clements and David F. Hendry. Oxford: Oxford University Press chapter 19, pp. 525–556.

Hansen, Peter R., Lunde, Asger, Nason, James M. 2011. The Model Confidence Set. *Econometrica* **79**(2): 453–497.

Hansen, Peter R., Lunde, Asger, Voev, Valeri 2014. Realized beta GARCH: a multivariate GARCH model with realized measures of volatility. *Journal of Applied Econometrics.*

Hansen, Peter Reinhard, Lunde, Asger 2006. Consistent ranking of volatility models. *Journal of Econometrics* **131**(1-2): 97–121.

Harris, David, McCabe, Brendan, Leybourne, Stephen 2008. Testing for long memory. *Econometric Theory* **24**(01): 143–175.

Harris, Richard D.F., Nguyen, Anh 2013. Long memory conditional volatility and asset allocation. *International Journal of Forecasting* **29**(2): 258–273.

Harvey, David, Leybourne, Stephen, Newbold, Paul 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting* **13**(2): 281–291.

Heber, Gerd, Lunde, Asger, Shephard, Neil, Sheppard, Kevin 2009. OMI's realized library.

Hillebrand, Eric 2005. Neglecting parameter changes in GARCH models. *Journal of Econometrics* **129**(1): 121–138.

Hillebrand, Eric, Medeiros, Marcelo C. 2015. Nonlinearity, breaks, and long-range dependence in time-series models. *Journal of Business & Economic Statistics* p. Forthcoming.

Hwang, Soosung, Valls Pereira, Pedro L. 2006. Small sample properties of GARCH estimates and persistence. *The European Journal of Finance* **12**(6-7): 473–494.

Inclan, Carla, Tiao, George C. 1994. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association* **89**(427): 913–923.

Jorion, Philippe 1995. Predicting volatility in the foreign exchange market. *The Journal of Finance* **50**(2): 507–528.

J.P. Morgan 1994 *RiskMetrics.* Technical Report. JP Morgan.

Kim, Tae-Hwan, Leybourne, Stephen J., Newbold, Paul 2000. Spurious rejections by Perron tests in the presence of a break. *Oxford Bulletin of Economics and Statistics* **62**(3): 433–444.

Koopman, Siem Jan, Jungbacker, Borus, Hol, Eugenie 2005. Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* **12**(3): 445–475.

Kunsch, Hans 1987. Statistical aspects of self-similar processes. In *Proceedings of the First World Congress of the Bernoulli Society,* ed. Y. Prohorov and V.V. Sazanov, vol. 1. Utrecht: VNU Science Press pp. 67–74.

Lamoureux, Christopher G., Lastrapes, William D. 1990. Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics* **8**(2): 225–234.

Leybourne, Stephen J., Mills, Terence C., Newbold, Paul 1998. Spurious rejections by Dickey-Fuller tests in the presence of a break under the null. *Journal of Econometrics* **87**(1): 191–203.

Li, W. K., Mak, T. K. 1994. On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis* **15**(6): 627–636.

Ljung, Greta M., Box, George E. P. 1978. On a measure of lack of fit in time series models. *Biometrika* **65**(2): 297–303.

Lu, Yang K., Perron, Pierre 2010. Modeling and forecasting stock return volatility using a random level shift model. *Journal of Empirical Finance* **17**(1): 138–156.

Lundbergh, Stefan, Teräsvirta, Timo 2002. Evaluating GARCH models. *Journal of Econometrics* **110**(2): 417–435.

Lux, Thomas, Morales-Arias, Leonardo 2013. Relative forecasting performance of volatility models: Monte Carlo evidence. *Quantitative Finance* **13**(9): 1375–1394.

Martens, Martin, van Dijk, Dick, de Pooter, Michiel 2009. Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of Forecasting* **25**(2): 282–303.

Martens, Martin, Zein, Jason 2004. Predicting financial volatility: High-frequency time-series forecasts vis-a-vis implied volatility. *Journal of Futures Markets* **24**(11): 1005–1028.

McCloskey, Adam, Perron, Pierre 2013. Memory Parameter Estimation in the Presence of Level Shifts and Deterministic Trends. *Econometric Theory* **29**(6): 1196–1237.

Mikosch, Thomas, Starica, Catalin 2004. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics* **86**(1): 378–390.

Morana, Claudio, Beltratti, Andrea 2004. Structural change and long-range dependence in volatility of exchange rates: either, neither or both? *Journal of Empirical Finance* **11**(5): 629–658.

Nelson, Daniel B. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59**(2): 347–370.

Newey, Whitney K. 1985. Maximum likelihood specification testing and conditional moment tests. *Econometrica* **53**(5): 1047–1070.

Newey, Whitney K., West, Kenneth D. 1994. Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies* **61**(4): 631–653.

Patton, Andrew J. 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* **160**(1): 246–256.

Perron, Pierre 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* **57**(6): 1361–1401.

_ 2006. Dealing with structural breaks. In *Palgrave Handbook of Econometrics,* ed. Terence C. Mills and Kerry D. Patterson, vol. 1. London: Palgrave Macmillan pp. 278–352.

Perron, Pierre, Qu, Zhongjun 2010. Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business & Economic Statistics* **28**(2): 275–290.

Pesaran, M. Hashem, Timmermann, Allan 2007. Selection of estimation window in the presence of breaks. *Journal of Econometrics* **137**(1): 134–161.

Politis, Dimitris N., Romano, Joseph P. 1994. The stationary bootstrap. *Journal of the American Statistical Association* **89**(428): 1303–1313.

Qu, Zhongjun 2011. A test against spurious long memory. *Journal of Business & Economic Statistics* **29**(3): 423–438.

Rapach, David E., Strauss, Jack K. 2008. Structural breaks and GARCH models of exchange rate volatility. *Journal of Applied Econometrics* **23**(1): 65–90.

Rappoport, Peter, Reichlin, Lucrezia 1989. Segmented trends and non-stationary time series. *The Economic Journal* **99**(395): 168–177.

Robinson, Peter M. 1995. Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* **23**(5): 1630–1661.

Sansó, Andreu, Aragó, Vincent, Carrion, Josep L. 2004. Testing for changes in the unconditional variance of financial time series. *Revista de Economía financiera* **4**: 32–53.

Shephard, Neil, Sheppard, Kevin 2010. Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* **25**(2): 197–231.

Shimotsu, Katsumi 2010. Exact Local Whittle Estimation of Fractional Integration With Unknown Mean and Time Trend. *Econometric Theory* **26**(02): 501.

Shimotsu, Katsumi, Phillips, Peter C. B. 2005. Exact local Whittle estimation of fractional integration. *The Annals of Statistics* **33**(4): 1890–1933.

Starica, Catalin, Granger, Clive W. J. 2005. Nonstationarities in stock returns. *Review of economics and statistics* **87**(3): 503–522.

Straumann, Daniel, Mikosch, Thomas 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics* **34**(5): 2449–2495.

Tauchen, George 1985. Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* **30**: 415–443.

West, Kenneth D. 1996. Asymptotic Inference about Predictive Ability. *Econometrica* **64**(5): 1067–1084.

White, Halbert 2000. A reality check for data snooping. *Econometrica* **68**(5): 1097–1126.

Zhang, Aonan, Gabrys, Robertas, Kokoszka, Piotr 2007. Discriminating between long memory and volatility shifts. *Austrian Journal of Statistics* **36**(4): 253–275.

Zumbach, Gilles 2006 *The RiskMetrics 2006 methodology.* Technical Report. RiskMetrics Group.

_ 2011. The statistical properties of the innovations in multivariate ARCH processes in high dimensions. *Quantitative Finance* **13**(1): 29–44.

Codes, Programmes, and Sources

Chapter 2

For all the results, the codes are newly written by Author using GAUSS, based on the reference sources for Realised GARCH of Hansen et al. (2012), which are written by OX and available at Journal of Applied Econometrics Data Archive. The graphic is drawn using Excel.

Chapter 3

Based on the original work of Rapach and Strauss (2008), the main codes are considerably revised by Author using GAUSS for the simulation experiment. The original source includes all of the GARCH-based estimation and forecasting, VaR, MSFE, ICSS, DMW, RC and SPA calculation. Original GAUSS procedure files are available at David Rapach's Homepage: http://sites.slu.edu/rapachde/home/research. For some other works, Author wrote some new GAUSS codes for EWMA and LM-EWMA forecast. The graphic is drawn using MATLAB.

Chapter 4

Table 22: Time Series Modelling by James Davidson, University of Exeter.

Table 23, 26: V/S, HML, GPH, LW - Time Series Modelling; ELW, 2ELW, 2ELWd - MATLAB code available at Katsumi Shimotsu's Homepage: http://shimotsu.web.fc2.com/Site/Matlab_Codes.html.

Table 24: Qu test - R code available at Zhongjun Qu's Webpage: http://people.bu.edu/qu/code.htm; BaPi test – R code available at Changryong Baek's Homepage: http://web.skku.edu/~crbaek/code.htm.

Table 25: BP test – GAUSS code available at Pierre Perron's Homepage: http://people.bu.edu/perron/; BHKS test – R code available at Changryong Baek's Homepage: http://web.skku.edu/~crbaek/code.htm; ICSS test – GAUSS code available at David Rapach's Homepage: http://sites.slu.edu/rapachde/home/research.

Table 28: Time Series Modelling.

Table 29: GARCH, FIGARCH, GARCHX-RV(RK), RGARCH-RV(RK) – GAUSS procedure written by Author; EGARCH, FIEGARCH – Time Series Modelling; Adaptive-FIGARCH – GAUSS code slightly revised by Author, the original GAUSS procedure is provided by Claudio Morana.

Table 30: MSE, MAE, QLIKE – Calculated using GAUSS by Author; MZ regression - GAUSS code available at David Rapach's Homepage: http://sites.slu.edu/rapachde/home/research.

Table 32: DMW test - GAUSS code available at David Rapach's Homepage: http://sites.slu.edu/rapachde/home/research.

Table 33: SPA - GAUSS code available at David Rapach's Homepage: http://sites.slu.edu/rapachde/home/research; MCS – MATLAB code available at MFE Toolbox by Kevin Sheppard: https://www.kevinsheppard.com/MFE_Toolbox.

Table 36: CPA – MATLAB code available at Runmycode: http://www.runmycode.org/companion/view/88.

All the graphics in Chapter 4 are drawn using Time Series Modelling.

* All of the codes are available from Author upon request. Any remaining errors are my own. The followings are the sample GAUSS codes for the finite empirical size of the test (Chapter 2) and the Monte Carlo simulation experiment (Chapter 3).

```
/**************************************************************************/
/***************  Chapter 2: Finite empirical size of test  ***************/
/**************************************************************************/

new;
output file=C:\gauss12\procs\phd1\results\size_asy\SPY.out reset;
library cml;
cmlset;
_cml_MaxIters=3000;
_cml_Algorithm=4;
tim=hsec;

/* general settings */
dist=1;   @ 1: standard normal, otherwise: standardised student t @
nu=12;    @ degrees of freedoms for t distribution                @
lags=20;  @ Ljung-Box and Engle's LM test lags                    @

/* true parameter values for Realized GARCH(1,1) */

h0=-0.333;
om= 0.061;
be= 0.550;
ga= 0.410;
xi=-0.183;
ph= 1.036;
t1=-0.067;
t2= 0.073;
vu= 0.146;

v_param=om|be|ga|xi|ph|t1|t2|vu;

/* DGP and simulation settings */
t=3000;
t_0=2000;

seed1=1492037;
seed2=7518391;
replic=1000;

/* output initialisation */

v_par=zeros(replic,10);
testat1=zeros(replic,1);
pvalue1=zeros(replic,1);
testat2=zeros(replic,1);
pvalue2=zeros(replic,1);
testat3=zeros(replic,1);
pvalue3=zeros(replic,1);
testat4=zeros(replic,1);
pvalue4=zeros(replic,1);

counter=1;
do until counter>replic;

{et,xt}=genrgarch(h0,v_param,t_0,t,dist,nu);
d_et=et[1:rows(et),1];
s_et=d_et.^2;
//h1=meanc(s_et);           @ unconditional variance of squared return @
d_xt=xt[1:rows(xt),1];

/* initialising the starting parameter values */

startvalues=h0|v_param;

{vp,fout,gout,vcv,ret,d_ht,d_zt}=cmlrgarch(d_et,d_xt,startvalues);

```

```
/* testing */

{stat1,pval1,stat2,pval2}=mtest(d_ht,d_xt,d_zt,vp);

v_par[counter,1]=vp[1];
v_par[counter,2]=vp[2];
v_par[counter,3]=vp[3];
v_par[counter,4]=vp[4];
v_par[counter,5]=vp[5];
v_par[counter,6]=vp[6];
v_par[counter,7]=vp[7];
v_par[counter,8]=vp[8];
v_par[counter,9]=vp[9];
v_par[counter,10]=ret;

s_zt=d_zt.^2;

/* Ljung-Box Test */
{stat3,pval3}=lbqtest(s_zt,lags); @ squared standardised residuals @
/* Engle's LM Test */
{stat4,pval4}=lmtest(d_zt,lags); @ squared standardised residuals @

testat1[counter,1]=stat1;
pvalue1[counter,1]=pval1;
testat2[counter,1]=stat2;
pvalue2[counter,1]=pval2;
testat3[counter,1]=stat3;
pvalue3[counter,1]=pval3;
testat4[counter,1]=stat4;
pvalue4[counter,1]=pval4;

counter=counter+1;
endo;

ind1=(pvalue1).<0.05;
ind2=(pvalue2).<0.05;
ind3=(pvalue3).<0.05;
ind4=(pvalue4).<0.05;

afr1=meanc(ind1);
afr2=meanc(ind2);
afr3=meanc(ind3);
afr4=meanc(ind4);

format 10,4;
"==========================================";
"True and Estimated Parameters";
"==========================================";
"Empirical Size: SPY";
"------------------------------------------";
if dist==1;
"Error Distribution  : Normal";
else;
"Error Distribution  : t(" nu ")";
endif;
"No. of Replications : " replic;
tim=(hsec-tim)/6000;
"Running Time        : " tim;
"------------------------------------------";
"omega  = " om~meanc(v_par[.,2]);
"beta   = " be~meanc(v_par[.,3]);
"gamma  = " ga~meanc(v_par[.,4]);
"xi     = " xi~meanc(v_par[.,5]);
"phi    = " ph~meanc(v_par[.,6]);
"tau1   = " t1~meanc(v_par[.,7]);
"tau2   = " t2~meanc(v_par[.,8]);
```

```
"var(u) = " vu~meanc(v_par[.,9]);
"========================================";
"Actual Frequency of Rejection";
"Lags for LB and LM tests:" lags;
"========================================";
"New test without recursion  = " 100*afr1;
"New test with recursion     = " 100*afr2;
"LB test                     = " 100*afr3;
"LM test                     = " 100*afr4;
"========================================";
save path=c:\gauss12\procs\phd1\results\size_asy\SPY v_par;
save path=c:\gauss12\procs\phd1\results\size_asy\SPY testat1;
save path=c:\gauss12\procs\phd1\results\size_asy\SPY pvalue1;
save path=c:\gauss12\procs\phd1\results\size_asy\SPY testat2;
save path=c:\gauss12\procs\phd1\results\size_asy\SPY pvalue2;
save path=c:\gauss12\procs\phd1\results\size_asy\SPY testat3;
save path=c:\gauss12\procs\phd1\results\size_asy\SPY pvalue3;
save path=c:\gauss12\procs\phd1\results\size_asy\SPY testat4;
save path=c:\gauss12\procs\phd1\results\size_asy\SPY pvalue4;

/************ Procedure Definition ************/

/* data generation using Realized GARCH(1,1) model */

proc(2)=genrgarch(h0,param,t_s,t_t,dist,dof);
    local j,e,h,x,u,z,lnh,lnx,us;
    local co0,co1,co2,me0,me1,me2,me3,vau;

/* dimensions of variables */
e=zeros(t_s+t_t,1);
lnh=zeros(t_s+t_t,1);
lnx=zeros(t_s+t_t,1);
h=zeros(t_s+t_t,1);
x=zeros(t_s+t_t,1);
u=zeros(t_s+t_t,1);
z=rndns(t_s+t_t,1,seed1);
us=rndns(t_s+t_t,1,seed2);

co0=param[1];  @ omega @
co1=param[2];  @ beta  @
co2=param[3];  @ gamma @
me0=param[4];  @ xi    @
me1=param[5];  @ phi   @
me2=param[6];  @ tau1  @
me3=param[7];  @ tau2  @
vau=param[8];  @ var(u)@

//lnh[1]=(co0+me0*co2)/(1-co1-me1*co2); @ initial logarithm conditional
variance @
//h[1]=exp(lnh[1]);

h[1]=exp(h0);

if dist==1;
    e[1]=sqrt(h[1])*z[1];
    else;
    e[1]=(sqrt((dof-2)/dof).*sqrt(h[1]).*rndns(1,1,seed1)./sqrt(sumc(rndns(dof
,1,seed1).^2)'./dof));
    z[1]=e[1]./sqrt(h[1]);
endif;

u[1]=sqrt(vau)*us[1];
lnx[1]=me0+me1*lnh[1]+me2*z[1]+me3*((z[1].^2)-1)+u[1];
x[1]=exp(lnx[1]);

j=2;
```

```
do until j>t_s+t_t;

lnh[j]=co0+co1*lnh[j-1]+co2*lnx[j-1];
h[j]=exp(lnh[j]);

if dist==1;
    e[j]=sqrt(h[j])*z[j];
    else;
    e[j]=(sqrt((dof-2)/dof).*sqrt(h[j]).*rndns(1,1,seed1)./sqrt(sumc(rndns(dof,1,seed1).^2)'./dof));
    z[j]=e[j]./sqrt(h[j]);
endif;

u[j]=sqrt(vau)*us[j];
lnx[j]=me0+me1*lnh[j]+me2*z[j]+me3*((z[j].^2)-1)+u[j];
x[j]=exp(lnx[j]);

j=j+1;
endo;

et=e[t_s+1:t_s+t_t];
xt=x[t_s+1:t_s+t_t];

retp(et,xt);
endp;

@===========================================================@
/* constrained MLE for Realized GARCH(1,1) */

proc(7)=cmlrgarch(dr,dm,startvalues);
    local p,q,start,x,rt,rt_s,rt_m,xt,xt_m,xt1,ut,zt,lht,ht;
    local h0l,h0h,oml,omh,bel,beh,gal,gah,xil,xih,phl,phh,t1l,t1h,t2l,t2h,vul,vuh;
    local parl,parh,parcons,b,f,g,vcv,ret;
    local h0,c0,c1,c2;

cmlset;

x=dr~dm;

/* parameter constraints */
h0l=-10;   h0h=3;
oml=-3;    omh=3;
bel=-1;    beh=2;
gal=-1;    gah=2;
xil=-10;   xih=3;
phl=-0.5;  phh=5;
t1l=-0.5;  t1h=0.5;
t2l=-0.5;  t2h=0.5;
vul=0.01;  vuh=4;

parl=h0l|oml|bel|gal|xil|phl|t1l|t2l|vul;
parh=h0h|omh|beh|gah|xih|phh|t1h|t2h|vuh;
parcons=parl~parh;

_cml_ParNames="h0"|"omega"|"beta"|"gamma"|"xi"|"phi"|"tau1"|"tau2"|"var_u";
_cml_IneqProc=&cfunc;
_cml_Bounds=parcons;
__output=0;

start=startvalues;

{b,f,g,vcv,ret}=cml(x,0,&rgarchll,start);

rt=x[.,1];
rt_s=rt.^2;
```

```
rt_m=meanc(rt_s);

xt=x[.,2];
xt_m=meanc(xt);
xt1=shiftr(xt',1,xt_m)';

h0=b[1];
c0=b[2];
c1=b[3];
c2=b[4];

lht=recserar((c0+c2*ln(xt1)),h0,c1);
ht=exp(lht);
zt=rt./sqrt(ht);

retp(b,f,g,vcv,ret,ht,zt);
endp;

@===========================================================@
/* Log-Likelihood function of RGARCH(1,1) */

proc rgarchll(b,x);
    local inh,ome,bet,gam,xii,phi,ta1,ta2,vau;
    local r_t,r_s,rs_m,x_t,xt_m,xt_1,z_t,lnh_t,h_t,u_t,lik;

    r_t=x[.,1];
    r_s=r_t.^2;
    rs_m=meanc(r_s);

    x_t=x[.,2];
    xt_m=meanc(x_t);
    xt_1=shiftr(x_t',1,xt_m)';

    inh=b[1];
    ome=b[2];
    bet=b[3];
    gam=b[4];
    xii=b[5];
    phi=b[6];
    ta1=b[7];
    ta2=b[8];
    vau=b[9];

    lnh_t=recserar((ome+gam*ln(xt_1)),inh,bet);
    h_t=exp(lnh_t);
    z_t=r_t./sqrt(h_t);
    u_t=ln(x_t)-xii-phi*ln(h_t)-ta1*z_t-ta2*((z_t.^2)-1);

    lik=(-0.5)*(ln(h_t)+(r_s./h_t)+ln(vau)+((u_t.^2)/vau));

    retp(lik);
endp;

@===========================================================@
/* persistent parameter constraint: nonlinear */

proc cfunc(b);
    local bc,gc,pc; /* bc: beta, gc: gamma, pc: phi */

    bc=0; gc=0; pc=1;
    bc=b[3,1];
    gc=b[4,1];
    pc=b[6,1];

    retp(0.999999-abs(bc+pc*gc)); /* constraint: b+g*p is greater than or
equal to 0. */
```

```
endp;
@==========================================================@
/* Testing */

proc(4)=mtest(ht,xt,zt,par);
    local cv1,cv2,cv3,rv1,rv2,rv3,rv4,vou;
    local n,m,lh,lx,lhm,lxm,lh1,lx1,ut,ct,ict,ctd,ictd,hd,mt,ud,udd,dls;
    local d_lam,s_lam,s_psi,s_sig,score,v_cov;
    local md,e1,e2,e3,e4,e5,e6,e7,e8,e9,e10,e11,e12,e13;
    local it11,it12,it13,it21,it22,it23,it31,it32,it33,it,it_inv;
    local jt11,jt12,jt13,jt21,jt22,jt23,jt31,jt32,jt33,jt;
    local kt,kt11,kt12,kt13,mt11,vt1,pt11,pt21,pt31,pt1,at1;
    local k,k11,k12,k13,m11,vt2,p11,p21,p31,pt2,at2,dl2;
    local zt_m,zt1,zt_s,zts_m,zts_1,zt2,itv;
    local diag_ut,diag_ud,diag_udd,diag_dls,vcv1,vcv2,inv_v1,inv_v2;
    local dt,tv1,tv2,wt1,wt2,ts1,ts2,pv1,pv2;

    cv1=par[2,1];
    cv2=par[3,1];
    cv3=par[4,1];
    rv1=par[5,1];
    rv2=par[6,1];
    rv3=par[7,1];
    rv4=par[8,1];
    vou=par[9,1];

    lh=ln(ht);
    lhm=ln(meanc(ht));
    lx=ln(xt);
    lxm=ln(meanc(xt));
    lh1=shiftr(lh',1,lhm)';                     /* log ht-1 */
    lx1=shiftr(lx',1,lxm)';                     /* log xt-1 */

    n=rows(zt);
    zt_s=zt.^2;
    mt=(ones(n,1)~lh~zt~(zt_s-1));              /* n by 4 */
    ut=lx-rv1-rv2*lh-rv3*zt-rv4*(zt_s-1);
    ud=-rv2+0.5*rv3*zt+rv4*zt_s;

    /*** first deriv. of log ht wrt lamda ***/
    ct=(ones(n,1)~lh1~lx1);                     /* n by 3 */
    ict=1~lhm~lxm;
    hd=recserar(ct,ict,cv2.*ones(1,3));

    d_lam=(1-(zt.^2)+(2/vou)*(ut.*ud));

    /*** analytical score ***/

//  s_lam=(-0.5)*(1/n)*(hd'd_lam);
//  s_psi=(1/vou)*(1/n)*(mt'ut);
//  s_sig=(-0.5)*((1/vou)^2)*meanc(vou-(ut.^2));

//  score=(s_lam|s_psi|s_sig);
//  v_cov=(score'score)';

    zt_m=meanc(zt);
    zt1=shiftr(zt',1,zt_m)';                    /* z_t-1 */
    zts_m=meanc(zt_s);
    zts_1=shiftr(zt_s',1,zts_m)';               /* square of z_t-1 */
    zt2=zts_1-1;                                /* (z_t-1)^2-1 */

    dt=-d_lam;
    tv1=(zt1~zt2);                              /* test variable without recursion*/
    wt1=(1/n)*tv1'dt;                           /* test indicator for tv1 : 2 by 1 */
```

```
    m=cols(tv1);
    itv=(meanc(zt1)~meanc(zt2));
    tv2=recserar(tv1,itv,cv2.*ones(1,2));     /* test variable with recursion*/
    wt2=(1/n)*tv2'dt;                         /* test indicator for tv2 : 2 by 1 */

    /*** calculating variance matrix ***/

    /* I matrix */
    md=(zeros(n,1)~ones(n,1)~(-0.5*zt)~(-zt_s));
    e1=(1/n)*(ud'ud);                         /* E(ud^2) */
    e2=(1/n)*(hd'hd);                         /* E(hdhd') */
    diag_ud=diagrv(zeros(n,n),ud);
    diag_ut=diagrv(zeros(n,n),ut);
    e3=(1/n)*((mt'diag_ud+md'diag_ut)*hd);    /* E((ud*mt+ut*md)*hd) */
    e4=(1/n)*(mt'mt);                         /* E(mtmt')*/

    it11=(0.5+(e1/vou))*e2;                   /* 3 by 3 matrix */
    it21=(-1/vou)*e3;                         /* 4 by 3 matrix */
    it12=it21';
    it22=(1/vou)*e4;                          /* 4 by 4 matrix */
    it31=zeros(1,3);
    it13=it31';
    it32=zeros(1,4);
    it23=it32';
    it33=0.5/(vou^2);

    it=(it11~it12~it13)|(it21~it22~it23)|(it31~it32~it33);  /* 8 by 8 matrix */
    it_inv=invpd(it);

    /* J matrix */
    e5=meanc(d_lam.^2);                           /* E(d_lam) */
    e6=(1/n)*(mt'diag_ud*hd);                     /* E((ud*mt)'hd) */
    e7=meanc(ut.^3);
    e8=meanc(ud);
    e9=((1/n)*sumc(hd))';
    e10=((1/n)*sumc(mt))';                        /* 1 by 4 */
    e11=meanc((1-((ut.^2)/vou)).^2);

    jt11=0.25*e5*e2;                              /* 3 by 3 */
    jt21=(-1/vou)*e6;                             /* 4 by 3 */
    jt12=jt21';
    jt22=it22;                                    /* 4 by 4 */
    jt31=(-0.5/(vou^3))*e7*e8*e9;
    jt13=jt31';
    jt32=(0.5/(vou^3))*e7*e10;
    jt23=jt32';
    jt33=(0.25/(vou^2))*e11;

    jt=(jt11~jt12~jt13)|(jt21~jt22~jt23)|(jt31~jt32~jt33);  /* 8 by 8 matrix */

    /* Iphi matrix without recursion*/
    pt11=zeros(3,m);
    pt21=zeros(4,m);
    pt31=zeros(1,m);

    pt1=(pt11|pt21|pt31);

    /* Iphi matrix with recursion*/
    p11=zeros(3,m);
    p21=zeros(4,m);
    p31=zeros(1,m);

    pt2=(p11|p21|p31);

```

```
    /* K matrix without recursion */
    dls=d_lam.^2;
    diag_dls=diagrv(zeros(n,n),dls);
    e12=(1/n)*(tv1'ud);
    kt11=(1/n)*(0.5)*tv1'diag_dls*hd;
    kt12=(1/n)*(-2/vou)*tv1'diag_ud*mt;
    kt13=(-1/(vou^3))*e7*e12;

    kt=(kt11~kt12~kt13);

    /* K matrix with recursion */
    e13=(1/n)*(tv2'ud);
    k11=(1/n)*(0.5)*tv2'diag_dls*hd;
    k12=(1/n)*(-2/vou)*tv2'diag_ud*mt;
    k13=(-1/(vou^3))*e7*e13;

    k=(k11~k12~k13);

    /* M matrix without recursion */
    mt11=(1/n)*tv1'diag_dls*tv1;

    /* M matrix with recursion */
    m11=(1/n)*tv2'diag_dls*tv2;

    /* V matrix without recursion */
    vt1=((jt~kt')|(kt~mt11));

    /* V matrix with recursion */
    vt2=((jt~k')|(k~m11));

    /* A matrix without recursion */
    at1=(((-pt1)'it_inv)~eye(m));

    /* A matrix with recursion */
    at2=(((-pt2)'it_inv)~eye(m));

    /* Variance estimator without recursion */
    vcv1=at1*vt1*at1';
    inv_v1=invpd(vcv1);

    /* Variance estimator with recursion */
    vcv2=at2*vt2*at2';
    inv_v2=invpd(vcv2);

    ts1=n*wt1'inv_v1*wt1;  @ test statistic without recursion @
    pv1=cdfchic(ts1,m);    @ p-value without recursion        @

    ts2=n*wt2'inv_v2*wt2;  @ test statistic with recursion @
    pv2=cdfchic(ts2,m);    @ p-value with recursion        @

    retp(ts1,pv1,ts2,pv2);
endp;
@============================================================@
/* Ljung Box Test */
proc(2)=lbqtest(data,lag);
    local ac,sac,t,idx,q,stat,pval;

    ac=acf(data,lag,0);
    sac=ac.^2;
    t=rows(data);
    idx=t-seqa(1,1,lag);
    q=sac./idx;
    stat=t*(t+2)*cumsumc(q);

    stat=stat[lag];
    pval=cdfChic(stat,lag);
```

```
    retp(stat,pval);
endp;
@==========================================================@
/* Engle's LM test */
proc(2)=lmtest(data,lag);
    local y,ylags,Tnew,b,sigma,rsq,stat,pval;

    y=(data-meanc(data)).^2;
    ylags=shiftr((ones(1,lag).*.y)',seqa(1,1,lag),-exp(20))';
    ylags=delif(ylags,ylags[.,cols(ylags)].==-exp(20));
    Tnew=rows(ylags);

    y=y[rows(y)-Tnew+1:rows(y)];
    ylags=ones(rows(ylags),1)~ylags;
    b=inv(ylags'ylags)*ylags'*y;
    sigma=(y-ylags*b)'(y-ylags*b)/rows(y);
    rsq=1-sigma/((y-meanc(y))'(y-meanc(y))/rows(y));

    stat=rows(ylags)*rsq;
    pval=cdfchic(stat,lag);
retp(stat,pval);
endp;
```

```
/*************************************************************************/
/***************  Chapter 3: Monte Carlo Simulation Sample  ***************/
/*************************************************************************/

new;
format 12,6;
output file=C:\gauss12\procs\phd2\results_70\break_low_s22\break_low.out reset;
library cml;
cmlset;
_cml_MaxIters=3000;
_cml_Algorithm=4;
tim=hsec;

#include c:\gauss12\src\icss.src;
#include c:\gauss12\src\variance.src;

/* Simulation and Forecasting Settings */
numb=1;          @ number of breaks (artificial) : 0 or 1 @
b_point=0.7;     @ artificial break point in sample @
t=2600;          @ total number of samples @
t0=3000;         @ to remove initial effect of DGP @
p=100;           @ number of out-of-sample observations @
s=22;            @ forecast horizon @
cri=0|1|4;       @ settings for break test @
replic=1000;     @ number of replications  @
nsim=2000;       @ number of simulations for VAR forecast @
ploss=0.05;      @ probability of loss for VAR forecast @
tl=2000;         @ truncation lag for FIGARCH @

/* GARCH Parameters: Setting single break in GARCH uncoditional variance */
omega=0.04; alpha=0.05; beta=0.75; shock_variance=5.0; @ shock to uncondition variance (intercept) @

/* FIGARCH Parameters (Starting values for estimation)*/
fi_omega=0.10; fi_phi=0.25; fi_beta=0.60; fi_d=0.45;

/* GARCH estimation starting value */
start1=omega|alpha|beta;
/* FIGARCH estimation starting value */
start2=fi_omega|fi_phi|fi_beta|fi_d;

seed1=1492038;
seed2=3941072;

/***** Vector Initialisation *****/

v_dgp=zeros(t,replic);          @ DGP @
v_frt=zeros(p,replic);          @ Returns for out-of-samples @
v_garch=zeros(replic,5);        @ GARCH parameters estimated: omega, alpha, beta, alpha+beta @
v_figarch=zeros(replic,7);      @ FIGARCH parameters estimated: omega, phi, beta, d, phi+d, 1-2*phi @

/* for storing first ob used to estimate */
exga=zeros(p,1); exfi=zeros(p,1); ga50=zeros(p,1); ga25=zeros(p,1); gawb=zeros(p,1);

/* for storing parameters estimates */
p_exga=zeros(p,3); p_exfi=zeros(p,4); p_ga50=zeros(p,3); p_ga25=zeros(p,3); p_gawb=zeros(p,3);

/* for storing most recent h estimate */
h_exga=zeros(p,1); h_exfi=zeros(p,1); h_ga50=zeros(p,1); h_ga25=zeros(p,1); h_gawb=zeros(p,1);

/* for storing number of breaks detected */
```

```
nbr_detected=zeros(p,1); @ for out of forecasting @
v_nbr=zeros(replic,1);   @ breaks detected for each replications @

/* for storing return codes for GARCH estimations */
ret_exga=zeros(replic,p); ret_exfi=zeros(replic,p); ret_ga50=zeros(replic,p);
ret_ga25=zeros(replic,p); ret_gawb=zeros(replic,p);

/* for storing single model forecasts */
f_exga=zeros(p-(s-1),s); f_exfi=zeros(p-(s-1),s); f_ga50=zeros(p-(s-1),s);
f_ga25=zeros(p-(s-1),s); f_gawb=zeros(p-(s-1),s); f_smrm=zeros(p-(s-1),s);
f_lmrm=zeros(p-(s-1),s); f_mean=zeros(p-(s-1),s); f_trim=zeros(p-(s-1),s);

fc_exga=zeros(p-(s-1),s*replic); fc_exfi=zeros(p-(s-1),s*replic); fc_ga50=
zeros(p-(s-1),s*replic);
fc_ga25=zeros(p-(s-1),s*replic); fc_gawb=zeros(p-(s-1),s*replic); fc_smrm=
zeros(p-(s-1),s*replic);
fc_lmrm=zeros(p-(s-1),s*replic); fc_mean=zeros(p-(s-1),s*replic); fc_trim=
zeros(p-(s-1),s*replic);

/* for storing single model VAR forecasts */
fv_exga=zeros(p-(s-1),1); fv_exfi=zeros(p-(s-1),1); fv_ga50=zeros(p-(s-1),1);
fv_ga25=zeros(p-(s-1),1); fv_gawb=zeros(p-(s-1),1); fv_smrm=zeros(p-(s-1),1);
fv_lmrm=zeros(p-(s-1),1); fv_mean=zeros(p-(s-1),1); fv_trim=zeros(p-(s-1),1);

fcv_exga=zeros(p-(s-1),replic); fcv_exfi=zeros(p-(s-1),replic); fcv_ga50=zeros
(p-(s-1),replic);
fcv_ga25=zeros(p-(s-1),replic); fcv_gawb=zeros(p-(s-1),replic); fcv_smrm=zeros
(p-(s-1),replic);
fcv_lmrm=zeros(p-(s-1),replic); fcv_mean=zeros(p-(s-1),replic); fcv_trim=zeros
(p-(s-1),replic);

/* average squared error loss function (loss series, mse and ratio) */
sim_loss_exga=zeros(p-(s-1),replic); sim_loss_exfi=zeros(p-(s-1),replic);
sim_loss_ga50=zeros(p-(s-1),replic);
sim_loss_ga25=zeros(p-(s-1),replic); sim_loss_gawb=zeros(p-(s-1),replic);
sim_loss_smrm=zeros(p-(s-1),replic);
sim_loss_lmrm=zeros(p-(s-1),replic); sim_loss_mean=zeros(p-(s-1),replic);
sim_loss_trim=zeros(p-(s-1),replic);

sim_mse_exga=zeros(replic,1); sim_mse_exfi=zeros(replic,1); sim_mse_ga50=zeros
(replic,1);
sim_mse_ga25=zeros(replic,1); sim_mse_gawb=zeros(replic,1); sim_mse_smrm=zeros
(replic,1);
sim_mse_lmrm=zeros(replic,1); sim_mse_mean=zeros(replic,1); sim_mse_trim=zeros
(replic,1);

sim_ratio_exga=zeros(replic,1); sim_ratio_exfi=zeros(replic,1); sim_ratio_ga50
=zeros(replic,1);
sim_ratio_ga25=zeros(replic,1); sim_ratio_gawb=zeros(replic,1); sim_ratio_smrm
=zeros(replic,1);
sim_ratio_lmrm=zeros(replic,1); sim_ratio_mean=zeros(replic,1); sim_ratio_trim
=zeros(replic,1);

/* 5% VAR (average quantile, percentage) */
sim_avg_exga=zeros(replic,1); sim_avg_exfi=zeros(replic,1); sim_avg_ga50=zeros
(replic,1);
sim_avg_ga25=zeros(replic,1); sim_avg_gawb=zeros(replic,1); sim_avg_smrm=zeros
(replic,1);
sim_avg_lmrm=zeros(replic,1); sim_avg_mean=zeros(replic,1); sim_avg_trim=zeros
(replic,1);

sim_per_exga=zeros(replic,1); sim_per_exfi=zeros(replic,1); sim_per_ga50=zeros
(replic,1);
sim_per_ga25=zeros(replic,1); sim_per_gawb=zeros(replic,1); sim_per_smrm=zeros
(replic,1);
sim_per_lmrm=zeros(replic,1); sim_per_mean=zeros(replic,1); sim_per_trim=zeros
```

```
(replic,1);

/* 5% VAR loss (loss seires, mean and ratio) */
sim_vloss_exga=zeros(p-(s-1),replic); sim_vloss_exfi=zeros(p-(s-1),replic);
sim_vloss_ga50=zeros(p-(s-1),replic);
sim_vloss_ga25=zeros(p-(s-1),replic); sim_vloss_gawb=zeros(p-(s-1),replic);
sim_vloss_smrm=zeros(p-(s-1),replic);
sim_vloss_lmrm=zeros(p-(s-1),replic); sim_vloss_mean=zeros(p-(s-1),replic);
sim_vloss_trim=zeros(p-(s-1),replic);

sim_mvar_exga=zeros(replic,1); sim_mvar_exfi=zeros(replic,1); sim_mvar_ga50=
zeros(replic,1);
sim_mvar_ga25=zeros(replic,1); sim_mvar_gawb=zeros(replic,1); sim_mvar_smrm=
zeros(replic,1);
sim_mvar_lmrm=zeros(replic,1); sim_mvar_mean=zeros(replic,1); sim_mvar_trim=
zeros(replic,1);

sim_vratio_exga=zeros(replic,1); sim_vratio_exfi=zeros(replic,1);
sim_vratio_ga50=zeros(replic,1);
sim_vratio_ga25=zeros(replic,1); sim_vratio_gawb=zeros(replic,1);
sim_vratio_smrm=zeros(replic,1);
sim_vratio_lmrm=zeros(replic,1); sim_vratio_mean=zeros(replic,1);
sim_vratio_trim=zeros(replic,1);


/***** Starting simulation *****/
counter=1;
do until counter>replic;

/* Generation of GARCH(1,1) DGP */
{rt,bp}=gengarch(t0,t,numb,p,b_point,omega,alpha,beta,shock_variance);

{garch_par,xx,xx,xx,ret_garch,ht_garch}=garch(rt,start1);

v_garch[counter,1]=garch_par[1];
v_garch[counter,2]=garch_par[2];
v_garch[counter,3]=garch_par[3];
v_garch[counter,4]=garch_par[2]+garch_par[3];
v_garch[counter,5]=ret_garch;

{figarch_par,xx,xx,xx,ret_figarch,ht_figarch}=figarch(rt,start2);

v_figarch[counter,1]=figarch_par[1]; @ omega @
v_figarch[counter,2]=figarch_par[2]; @ phi   @
v_figarch[counter,3]=figarch_par[3]; @ beta  @
v_figarch[counter,4]=figarch_par[4]; @ d     @
v_figarch[counter,5]=figarch_par[2]+figarch_par[4]; @ phi+d   @
v_figarch[counter,6]=1-2*figarch_par[2];            @ 1-2*phi @
v_figarch[counter,7]=ret_figarch;

v_dgp[.,counter]=rt; /* vector of DGPs */

numr=rows(rt);
r=numr-p;  @ in-sample period @

/* Parameter Estimation for Forecasting */
iter=0;
do until iter>p-1;

  /* Expanding GARCH */
  {par1,xx,xx,xx,ret1,ht1}=garch(rt[1:r+iter],start1);
  ret_exga[counter,iter+1]=ret1;
  exga[iter+1]=1;
  p_exga[iter+1,.]=par1';
  h_exga[iter+1,.]=ht1[rows(ht1)];

```

```
    /* Expanding FIGARCH */
    {par2,xx,xx,xx,ret2,ht2}=figarch(rt[1:r+iter],start2);
    ret_exfi[counter,iter+1]=ret2;
    exfi[iter+1]=1;
    p_exfi[iter+1,.]=par2';
    h_exfi[iter+1,.]=ht2[rows(ht2)];

    /* 0.50 Rolling GARCH */
    ro50=round(0.50*r)+1; @ starting obs for 0.50 rolling window @
    {par3,xx,xx,xx,ret3,ht3}=garch(rt[ro50+iter:r+iter],start1);
    ret_ga50[counter,iter+1]=ret3;
    ga50[iter+1]=ro50+iter;
    p_ga50[iter+1,.]=par3';
    h_ga50[iter+1,.]=ht3[rows(ht3)];

    /* 0.25 Rolling GARCH */
    ro25=round(0.75*r)+1; @ starting obs for 0.25 rolling window @
    {par4,xx,xx,xx,ret4,ht4}=garch(rt[ro25+iter:r+iter],start1);
    ret_ga25[counter,iter+1]=ret4;
    ga25[iter+1]=ro25+iter;
    p_ga25[iter+1,.]=par4';
    h_ga25[iter+1,.]=ht4[rows(ht4)];

    /* GARCH with Break */

   {cpr,nbr}=icss(rt[1:r+iter],2,cri);
      if nbr==0;
          nbr_detected[iter+1]=nbr;
          p_gawb[iter+1,.]=p_exga[iter+1,.];
          h_gawb[iter+1,.]=h_exga[iter+1];
      else;
          nbr_detected[iter+1]=nbr;
          first_br=1+cpr[rows(cpr)-1]; /* The first observation after the last break point detected */
          {par5,xx,xx,xx,ret5,ht5}=garch(rt[first_br:r+iter],start1);
          ret_gawb[counter,iter+1]=ret5;
          if par5[2]<0.00000000001;
              p_gawb[iter+1,.]=meanc(rt[first_br:r+iter]^2)~0~0;
              h_gawb[iter+1,.]=meanc(rt[first_br:r+iter]^2);
          else;
              p_gawb[iter+1,.]=par5';
              h_gawb[iter+1,.]=ht5[rows(ht5)];
          endif;
      endif;

  iter=iter+1;
  endo;

  /* Forecasting */

  it=0;
  do until it>p-s;

     /* Expanding GARCH forecast */
     p_garch_ex=p_exga[it+1,.]';
     fc1=garch_fc(p_garch_ex,rt[r+it],h_exga[it+1],s);
     f_exga[it+1,.]=fc1';
     fv_exga[it+1]=garch_varfc(p_garch_ex,fc1[1],s,ploss,nsim);

     /* Expanding FIGARCH forecast */
     p_garch_fi=p_exfi[it+1,.]';
     fc2=figarch_fc(p_garch_fi,rt[1:(r+it)],h_exfi[it+1],s);
     f_exfi[it+1,.]=fc2';
     fv_exfi[it+1]=figarch_varfc(p_garch_fi,fc2[1],rt[exfi[it+1]:r+it],s,ploss,nsim);

```

```
    /* 0.50 rolling GARCH forecast */
    p_garch_50=p_ga50[it+1,.]';
    fc3=garch_fc(p_garch_50,rt[r+it],h_ga50[it+1],s);
    f_ga50[it+1,.]=fc3';
    fv_ga50[it+1]=garch_varfc(p_garch_50,fc3[1],s,ploss,nsim);

    /* 0.25 rolling GARCH forecast */
    p_garch_25=p_ga25[it+1,.]';
    fc4=garch_fc(p_garch_25,rt[r+it],h_ga25[it+1],s);
    f_ga25[it+1,.]=fc4';
    fv_ga25[it+1]=garch_varfc(p_garch_25,fc4[1],s,ploss,nsim);

    /* GARCH w/breaks forecast */
    p_garch_br=p_gawb[it+1,.]';
    if p_garch_br[2]==0;
       f_gawb[it+1,.]=p_garch_br[1]*ones(1,s);
       fv_gawb[it+1]=cons_varfc(p_garch_br[1],s,ploss,nsim);
    else;
       fc5=garch_fc(p_garch_br,rt[r+it],h_gawb[it+1],s);
       f_gawb[it+1,.]=fc5';
       fv_gawb[it+1]=garch_varfc(p_garch_br,fc5[1],s,ploss,nsim);
    endif;

    /* EWMA forecast */
    lam=0.94;
    rs=rt[1:r+it].^2;
    rsr=rev(rs);
    tau=seqa(0,1,rows(rs));
    lamtau=lam^tau;
    fc6=(1-lam)*lamtau'rsr*ones(s,1);
    f_smrm[it+1,.]=fc6';
    fv_smrm[it+1]=cons_varfc(fc6[1],s,ploss,nsim);

    /* Long Memory EWMA forecast */
    tau0=1560;
    tau1=4;
    taumax=512;
    rho=sqrt(2);
    kmax=round(1+((ln(taumax)-ln(tau1))/ln(rho)));
    kseq=seqa(1,1,kmax);                           /* k by 1 */
    tauk=tau1*(rho^(kseq-ones(rows(kseq),1)));     /* k by 1 */
    muk=exp(-1/tauk);                              /* k by 1 */

    muktau=zeros(kmax,rows(rs));                   /* k by T */

    i=1;
    do until i>kmax;
    muktau[i,.]=(muk[i]^tau)';                     /* 1 by T */
    i=i+1;
    endo;

    hk=(1-muk).*(muktau*rsr);                      /* k by 1 */
    c=kmax-sumc(ln(tauk)./ln(tau0));
    wk=(1/c)*(1-(ln(tauk)./ln(tau0)));             /* k by 1 */
    fc7=wk'hk*ones(s,1);
    f_lmrm[it+1,.]=fc7';
    fv_lmrm[it+1]=cons_varfc(fc7[1],s,ploss,nsim);

    /* Combination forecasting: mean and trimmed mean*/
    f_all=f_exga[it+1,.]|f_ga50[it+1,.]|f_ga25[it+1,.]|f_gawb[it+1,.];
    f_mean[it+1,.]=(meanc(f_all))';
    ftrim=(1/(rows(f_all)-2))*(sumc(f_all)-(minc(f_all)+maxc(f_all)));
    f_trim[it+1,.]=ftrim';

    fv_all=fv_exga[it+1,.]|fv_ga50[it+1,.]|fv_ga25[it+1,.]|fv_gawb[it+1,.];
    fv_mean[it+1,.]=(meanc(fv_all))';
```

```
    fvtrim=(1/(rows(fv_all)-2))*(sumc(fv_all)-(minc(fv_all)+maxc(fv_all)));
    fv_trim[it+1,.]=fvtrim';

    it=it+1;
 endo;

 f_rt=rt[r+1:r+p];
 frt_s=rt[r+1:r+p].^2;
 v_frt[.,counter]=frt_s;

 /* MSE averaged loss, ratios */
 {loss_exga,mse_exga}=loss_se_avg(f_exga,frt_s); {loss_exfi,mse_exfi}=loss_se_avg(f_exfi,frt_s);
 {loss_ga50,mse_ga50}=loss_se_avg(f_ga50,frt_s); {loss_ga25,mse_ga25}=loss_se_avg(f_ga25,frt_s);
 {loss_gawb,mse_gawb}=loss_se_avg(f_gawb,frt_s); {loss_smrm,mse_smrm}=loss_se_avg(f_smrm,frt_s);
 {loss_lmrm,mse_lmrm}=loss_se_avg(f_lmrm,frt_s); {loss_mean,mse_mean}=loss_se_avg(f_mean,frt_s);
 {loss_trim,mse_trim}=loss_se_avg(f_trim,frt_s);

 /************* 5% VAR ***************/

 /* average quantile, percentage, LR statistics and p-values */
 {avg_exga,per_exga}=var_stats(fv_exga,f_rt,s);
 {avg_exfi,per_exfi}=var_stats(fv_exfi,f_rt,s);
 {avg_ga50,per_ga50}=var_stats(fv_ga50,f_rt,s);
 {avg_ga25,per_ga25}=var_stats(fv_ga25,f_rt,s);
 {avg_gawb,per_gawb}=var_stats(fv_gawb,f_rt,s);
 {avg_smrm,per_smrm}=var_stats(fv_smrm,f_rt,s);
 {avg_lmrm,per_lmrm}=var_stats(fv_lmrm,f_rt,s);
 {avg_mean,per_mean}=var_stats(fv_mean,f_rt,s);
 {avg_trim,per_trim}=var_stats(fv_trim,f_rt,s);

 {vloss_exga,mvar_exga}=loss_var(fv_exga,f_rt,s,ploss);
 {vloss_exfi,mvar_exfi}=loss_var(fv_exfi,f_rt,s,ploss);
 {vloss_ga50,mvar_ga50}=loss_var(fv_ga50,f_rt,s,ploss);
 {vloss_ga25,mvar_ga25}=loss_var(fv_ga25,f_rt,s,ploss);
 {vloss_gawb,mvar_gawb}=loss_var(fv_gawb,f_rt,s,ploss);
 {vloss_smrm,mvar_smrm}=loss_var(fv_smrm,f_rt,s,ploss);
 {vloss_lmrm,mvar_lmrm}=loss_var(fv_lmrm,f_rt,s,ploss);
 {vloss_mean,mvar_mean}=loss_var(fv_mean,f_rt,s,ploss);
 {vloss_trim,mvar_trim}=loss_var(fv_trim,f_rt,s,ploss);

 if nbr<1;
     cpr="n/a";?;
 else;
     cpr=trimr(cpr,1,1);
 endif;

 format 12,6;
 "=====================================================";
 "Counter No.         = " counter;
 "Artificial break    = " bp;
 "Break points        = " cpr';
 "Number of breaks    = " nbr;
 "=====================================================";
 " Average Squared Loss Function (MSFE and Ratio)";
 "=====================================================";
 "Expanding GARCH     = " mse_exga~(mse_exga/mse_exga);
 "0.50 Rolling GARCH = " mse_ga50~(mse_ga50/mse_exga);
 "0.25 Rolling GARCH = " mse_ga25~(mse_ga25/mse_exga);
 "GARCH w/breaks      = " mse_gawb~(mse_gawb/mse_exga);
 "-----------------------------------------------------";
 "Expanding FIGARCH  = " mse_exfi~(mse_exfi/mse_exga);
 "Short Memory EWMA  = " mse_smrm~(mse_smrm/mse_exga);
```

```
"Long Memory EWMA    = " mse_lmrm~(mse_lmrm/mse_exga);
"-----------------------------------------------------";
"MSE GARCH Mean      = " mse_mean~(mse_mean/mse_exga);
"MSE GARCH Trimmed   = " mse_trim~(mse_trim/mse_exga);
"=====================================================";
" 5% VaR Loss Function (mean, ratio)";
"=====================================================";
"Expanding GARCH     = " mvar_exga~(mvar_exga/mvar_exga);
"0.50 Rolling GARCH = " mvar_ga50~(mvar_ga50/mvar_exga);
"0.25 Rolling GARCH = " mvar_ga25~(mvar_ga25/mvar_exga);
"GARCH w/breaks      = " mvar_gawb~(mvar_gawb/mvar_exga);
"-----------------------------------------------------";
"Expanding FIGARCH   = " mvar_exfi~(mvar_exfi/mvar_exga);
"Short Memory EWMA   = " mvar_smrm~(mvar_smrm/mvar_exga);
"Long Memory EWMA    = " mvar_lmrm~(mvar_lmrm/mvar_exga);
"-----------------------------------------------------";
"VaR GARCH Mean      = " mvar_mean~(mvar_mean/mvar_exga);
"VaR GARCH Trimmed   = " mvar_trim~(mvar_trim/mvar_exga);
"=====================================================";?;

sr=s*(counter-1)+1;
er=s*counter;
fc_exga[.,sr:er]=f_exga; fc_exfi[.,sr:er]=f_exfi; fc_ga50[.,sr:er]=f_ga50;
fc_ga25[.,sr:er]=f_ga25; fc_gawb[.,sr:er]=f_gawb; fc_smrm[.,sr:er]=f_smrm;
fc_lmrm[.,sr:er]=f_lmrm; fc_mean[.,sr:er]=f_mean; fc_trim[.,sr:er]=f_trim;

fcv_exga[.,counter]=fv_exga; fcv_exfi[.,counter]=fv_exfi; fcv_ga50[.,counter]=
fv_ga50;
fcv_ga25[.,counter]=fv_ga25; fcv_gawb[.,counter]=fv_gawb; fcv_smrm[.,counter]=
fv_smrm;
fcv_lmrm[.,counter]=fv_lmrm; fcv_mean[.,counter]=fv_mean; fcv_trim[.,counter]=
fv_trim;

sim_loss_exga[.,counter]=loss_exga; sim_loss_exfi[.,counter]=loss_exfi;
sim_loss_ga50[.,counter]=loss_ga50;
sim_loss_ga25[.,counter]=loss_ga25; sim_loss_gawb[.,counter]=loss_gawb;
sim_loss_smrm[.,counter]=loss_smrm;
sim_loss_lmrm[.,counter]=loss_lmrm; sim_loss_mean[.,counter]=loss_mean;
sim_loss_trim[.,counter]=loss_trim;

sim_mse_exga[counter,1]=mse_exga; sim_mse_exfi[counter,1]=mse_exfi;
sim_mse_ga50[counter,1]=mse_ga50;
sim_mse_ga25[counter,1]=mse_ga25; sim_mse_gawb[counter,1]=mse_gawb;
sim_mse_smrm[counter,1]=mse_smrm;
sim_mse_lmrm[counter,1]=mse_lmrm; sim_mse_mean[counter,1]=mse_mean;
sim_mse_trim[counter,1]=mse_trim;

sim_ratio_exga[counter,1]=mse_exga/mse_exga; sim_ratio_exfi[counter,1]=
mse_exfi/mse_exga;
sim_ratio_ga50[counter,1]=mse_ga50/mse_exga; sim_ratio_ga25[counter,1]=
mse_ga25/mse_exga;
sim_ratio_gawb[counter,1]=mse_gawb/mse_exga; sim_ratio_smrm[counter,1]=
mse_smrm/mse_exga;
sim_ratio_lmrm[counter,1]=mse_lmrm/mse_exga; sim_ratio_mean[counter,1]=
mse_mean/mse_exga;
sim_ratio_trim[counter,1]=mse_trim/mse_exga;

sim_avg_exga[counter,1]=avg_exga; sim_avg_exfi[counter,1]=avg_exfi;
sim_avg_ga50[counter,1]=avg_ga50;
sim_avg_ga25[counter,1]=avg_ga25; sim_avg_gawb[counter,1]=avg_gawb;
sim_avg_smrm[counter,1]=avg_smrm;
sim_avg_lmrm[counter,1]=avg_lmrm; sim_avg_mean[counter,1]=avg_mean;
sim_avg_trim[counter,1]=avg_trim;

sim_per_exga[counter,1]=per_exga; sim_per_exfi[counter,1]=per_exfi;
sim_per_ga50[counter,1]=per_ga50;
```

```
sim_per_ga25[counter,1]=per_ga25; sim_per_gawb[counter,1]=per_gawb;
sim_per_smrm[counter,1]=per_smrm;
sim_per_lmrm[counter,1]=per_lmrm; sim_per_mean[counter,1]=per_mean;
sim_per_trim[counter,1]=per_trim;

sim_vloss_exga[.,counter]=vloss_exga; sim_vloss_exfi[.,counter]=vloss_exfi;
sim_vloss_ga50[.,counter]=vloss_ga50;
sim_vloss_ga25[.,counter]=vloss_ga25; sim_vloss_gawb[.,counter]=vloss_gawb;
sim_vloss_smrm[.,counter]=vloss_smrm;
sim_vloss_lmrm[.,counter]=vloss_lmrm; sim_vloss_mean[.,counter]=vloss_mean;
sim_vloss_trim[.,counter]=vloss_trim;

sim_mvar_exga[counter,1]=mvar_exga; sim_mvar_exfi[counter,1]=mvar_exfi;
sim_mvar_ga50[counter,1]=mvar_ga50;
sim_mvar_ga25[counter,1]=mvar_ga25; sim_mvar_gawb[counter,1]=mvar_gawb;
sim_mvar_smrm[counter,1]=mvar_smrm;
sim_mvar_lmrm[counter,1]=mvar_lmrm; sim_mvar_mean[counter,1]=mvar_mean;
sim_mvar_trim[counter,1]=mvar_trim;

sim_vratio_exga[counter,1]=mvar_exga/mvar_exga; sim_vratio_exfi[counter,1]=
mvar_exfi/mvar_exga;
sim_vratio_ga50[counter,1]=mvar_ga50/mvar_exga; sim_vratio_ga25[counter,1]=
mvar_ga25/mvar_exga;
sim_vratio_gawb[counter,1]=mvar_gawb/mvar_exga; sim_vratio_smrm[counter,1]=
mvar_smrm/mvar_exga;
sim_vratio_lmrm[counter,1]=mvar_lmrm/mvar_exga; sim_vratio_mean[counter,1]=
mvar_mean/mvar_exga;
sim_vratio_trim[counter,1]=mvar_trim/mvar_exga;

v_nbr[counter,1]=nbr;

counter=counter+1;
endo;

freq_fc=rows(selif(v_nbr,v_nbr[.,1] .gt 0));

format 12,6;
"=====================================================";
"Simulation Result Summary";
"Number of Replications    : " replic;
"Frequency: break          : " freq_fc;
"-----------------------------------------------------";
"Total number of sample    : " numr;
"In-sample observations    : " r;
"Out-of-sample forecasts   : " p-(s-1);
"Forecast horizon          : " s;
"=====================================================";
" Average Squared Loss Function";
"=====================================================";
"Expanding GARCH    = " meanc(sim_mse_exga);
"0.50 Rolling GARCH = " meanc(sim_ratio_ga50);
"0.25 Rolling GARCH = " meanc(sim_ratio_ga25);
"GARCH w/breaks     = " meanc(sim_ratio_gawb);
"-----------------------------------------------------";
"Expanding FIGARCH  = " meanc(sim_ratio_exfi);
"Short Memory EWMA  = " meanc(sim_ratio_smrm);
"Long Memory EWMA   = " meanc(sim_ratio_lmrm);
"-----------------------------------------------------";
"MSE GARCH Mean     = " meanc(sim_ratio_mean);
"MSE GARCH Trimmed  = " meanc(sim_ratio_trim);
"=====================================================";
" 5% VAR Loss Function";
"=====================================================";
"Expanding GARCH    = " meanc(sim_mvar_exga);
"0.50 Rolling GARCH = " meanc(sim_vratio_ga50);
"0.25 Rolling GARCH = " meanc(sim_vratio_ga25);
```

```
"GARCH w/breaks      = " meanc(sim_vratio_gawb);
"-------------------------------------------------------";
"Expanding FIGARCH   = " meanc(sim_vratio_exfi);
"Short Memory EWMA   = " meanc(sim_vratio_smrm);
"Long Memory EWMA    = " meanc(sim_vratio_lmrm);
"-------------------------------------------------------";
"VAR GARCH Mean      = " meanc(sim_vratio_mean);
"VAR GARCH Trimmed   = " meanc(sim_vratio_trim);
"=======================================================";?;

mse=sim_mse_exga~sim_mse_exfi~sim_mse_ga50~sim_mse_ga25~sim_mse_gawb
   ~sim_mse_smrm~sim_mse_lmrm~sim_mse_mean~sim_mse_trim;
mse_ratio=sim_ratio_exga~sim_ratio_exfi~sim_ratio_ga50~sim_ratio_ga25~sim_ratio_gawb
         ~sim_ratio_smrm~sim_ratio_lmrm~sim_ratio_mean~sim_ratio_trim;
loss_series=sim_loss_exga|sim_loss_exfi|sim_loss_ga50|sim_loss_ga25|sim_loss_gawb
           |sim_loss_smrm|sim_loss_lmrm|sim_loss_mean|sim_loss_trim;
mvar=sim_mvar_exga~sim_mvar_exfi~sim_mvar_ga50~sim_mvar_ga25~sim_mvar_gawb
    ~sim_mvar_smrm~sim_mvar_lmrm~sim_mvar_mean~sim_mvar_trim;
mvar_ratio=sim_vratio_exga~sim_vratio_exfi~sim_vratio_ga50~sim_vratio_ga25~sim_vratio_gawb
          ~sim_vratio_smrm~sim_vratio_lmrm~sim_vratio_mean~sim_vratio_trim;
vloss_series=sim_vloss_exga|sim_vloss_exfi|sim_vloss_ga50|sim_vloss_ga25|sim_vloss_gawb
            |sim_vloss_smrm|sim_vloss_lmrm|sim_vloss_mean|sim_vloss_trim;
avg_var=sim_avg_exga~sim_avg_exfi~sim_avg_ga50~sim_avg_ga25~sim_avg_gawb
   ~sim_avg_smrm~sim_avg_lmrm~sim_avg_mean~sim_avg_trim;
per_var=sim_per_exga~sim_per_exfi~sim_per_ga50~sim_per_ga25~sim_per_gawb
   ~sim_per_smrm~sim_per_lmrm~sim_per_mean~sim_per_trim;

/* Saving results */
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 v_dgp;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 v_frt;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 v_garch;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 v_figarch;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fc_exga;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fc_exfi;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fc_ga50;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fc_ga25;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fc_gawb;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fc_smrm;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fc_lmrm;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fc_mean;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fc_trim;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fcv_exga;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fcv_exfi;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fcv_ga50;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fcv_ga25;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fcv_gawb;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fcv_smrm;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fcv_lmrm;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fcv_mean;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 fcv_trim;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 mse;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 mse_ratio;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 loss_series;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 mvar;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 mvar_ratio;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 vloss_series;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 avg_var;
save path=C:\gauss12\procs\phd2\results_70\break_low_s22 per_var;

/******************************************************************/
/************************    Procedures    ************************/
/******************************************************************/
```

```
proc(2)=gengarch(t0,t,br,oos,bp,w,a,b,sv);
    local eps,i,j,k,k0,h,zt,w1;

    h=zeros(t0+t,1);
    eps=zeros(t0+t,1);
    zt=rndns(t0+t,1,seed1);
    w1=w*sv;
    h[1]=(w/(1-a-b));

    j=2;
    do until j>t0+t;

    if br<1;
        eps[1]=sqrt(h[1])*zt[1];
        h[j]=w+a*(eps[j-1].^2)+b*h[j-1];
        eps[j]=sqrt(h[j])*zt[j];
    else;
        k0=round(bp*(t-oos));       @ artificial break point for in-sample period @
        eps[1]=sqrt(h[1])*zt[1];
        h[j]=w*(j.<=t0+k0)+w1*(j.>t0+k0)+a*(eps[j-1].^2)+b*h[j-1];
        eps[j]=sqrt(h[j])*zt[j];
    endif;
    j=j+1;
endo;
eps=eps[t0+1:t0+t];

if br<1; k="n/a";
else; k=k0;
endif;

retp(eps,k);
endp;
/******************************************************************/
proc(6)=garch(e,startvalues);
   local b,f,g,vcv,ret,hhat,_ww_;

   cmlset;

   _ww_={-1e250 1e250};
   _cml_DirTol=0.001;
   _cml_Bounds=ones(3,2).*_ww_;
   _cml_Bounds[1,1]=0.0001;
   _cml_Bounds[2:3,1]=zeros(2,1);

   _cml_C=zeros(1,3);
   _cml_C[1,2:3]=-ones(1,2);
   _cml_D=-0.99999;

   __output=0;
   {b,f,g,vcv,ret}=cml(e,0,&garch11_loglike,startvalues);
   _cml_covPar=3;

   hhat=garch11_hhat(b,e);
   retp(b,f,g,vcv,ret,hhat);
endp;
/******************************************************************/
proc garch11_loglike(b,e);
   local es,esm,est1,omega,alpha,beta,h,lik;

   es=e.^2;
   esm=meanc(es);
   est1=shiftr(es',1,esm)';

   omega=b[1]; alpha=b[2]; beta=b[3];
```

```
      h=recserar((omega+alpha*est1),esm,beta);
      h=substute(h,h.<1e-6,1e-6);

      lik=-0.5*ln(2*pi*h)-0.5*(e.^2)./h;
      retp(lik);
   endp;
   /******************************************************************/
   proc garch11_hhat(b,e);
      local t,es,esm,est1,omega,alpha,beta,h,lik;

      t=rows(e);
      es=e.^2;
      esm=meanc(es);
      est1=shiftr(es',1,esm)';

      omega=b[1]; alpha=b[2]; beta=b[3];

      h=recserar((omega+alpha*est1),esm,beta);
      h=substute(h,h.<1e-6,1e-6);

      retp(h);
   endp;
   /******************************************************************/
   proc(6)=figarch(e,startvalues);
      local b,f,g,vcv,ret,hhat,_ww_;
      cmlset;

      _ww_={-1e250 1e250};
      _cml_DirTol=0.001;
      _cml_Bounds=ones(4,2).*_ww_;
      _cml_Bounds[1,1]=0.0001;
      _cml_Bounds[2:4,1]=zeros(3,1);

      _cml_C=zeros(2,4);
      _cml_C=(0~-2~0~-1)|(0~1~-1~1);
      _cml_D=-0.99999|0;

      __output=0;
      {b,f,g,vcv,ret}=cml(e,0,&figarch11_loglike,startvalues);
      hhat=figarch11_hhat(b,e);
      retp(b,f,g,vcv,ret,hhat);
   endp;
   /******************************************************************/
   proc figarch11_loglike(b,e);
      local i,j,t,es,esm,est1,bclength,bcweight,seq_bcwe,backcast,aug_es,bc_es;
      local omega,phi,beta,d,lambda,delta,tau,h,lik;

      t=rows(e);
      es=e.^2;
      esm=meanc(es);
      est1=shiftr(es',1,esm)';

      omega=b[1]; phi=b[2]; beta=b[3]; d=b[4];

      lambda=zeros(tl,1);
      delta=zeros(tl,1);
      lambda[1]=phi-beta+d;
      delta[1]=d;

      j=2;
      do until j>tl;
      delta[j]=((j-1-d)/j)*delta[j-1];
      lambda[j]=beta*lambda[j-1]+(delta[j]-phi*delta[j-1]);
      j=j+1;
      endo;
```

```
    bclength=maxv(floor(sqrt(t)),1);
    seq_bcwe=seqa(0,1,bclength+1)';
    bcweight=0.05*(0.9.^seq_bcwe);
    bcweight=bcweight/sumr(bcweight);
    bc_es=es[1:bclength+1];
    backcast=bcweight*bc_es;

    if backcast==0;
        backcast=vcx(e);
    endif;

    aug_es=zeros(tl,1)|es;
    aug_es[1:tl]=ones(tl,1)*backcast;

    h=zeros(rows(aug_es),1);

    i=tl+1;
    do until i>tl+t;
    h[i]=omega+lambda'aug_es[i-1:i-tl];
    i=i+1;
    endo;

    h=h[tl+1:tl+t];
    h=substute(h,h.<1e-6,1e-6);

    lik=-0.5*ln(2*pi*h)-0.5*(e.^2)./h;

    retp(lik);
 endp;
 /*******************************************************************/
 proc figarch11_hhat(b,e);
    local i,j,t,es,esm,est1,bclength,bcweight,seq_bcwe,backcast,aug_es,bc_es;
    local omega,phi,beta,d,lambda,delta,tau,h,lik;

    t=rows(e);
    es=e.^2;
    esm=meanc(es);
    est1=shiftr(es',1,esm)';

    omega=b[1]; phi=b[2]; beta=b[3]; d=b[4];

    lambda=zeros(tl,1);
    delta=zeros(tl,1);
    lambda[1]=phi-beta+d;
    delta[1]=d;

    j=2;
    do until j>tl;
    delta[j]=((j-1-d)/j)*delta[j-1];
    lambda[j]=beta*lambda[j-1]+(delta[j]-phi*delta[j-1]);
    j=j+1;
    endo;

    bclength=maxv(floor(sqrt(t)),1);
    seq_bcwe=seqa(0,1,bclength+1)';
    bcweight=0.05*(0.9.^seq_bcwe);
    bcweight=bcweight/sumr(bcweight);
    bc_es=es[1:bclength+1];
    backcast=bcweight*bc_es;

    if backcast==0;
        backcast=vcx(e);
    endif;

    aug_es=zeros(tl,1)|es;
    aug_es[1:tl]=ones(tl,1)*backcast;
```

```
   h=zeros(rows(aug_es),1);

   i=tl+1;
   do until i>tl+t;
   h[i]=omega+lambda'aug_es[i-1:i-tl];
   i=i+1;
   endo;

   h=h[tl+1:tl+t];
   h=substute(h,h.<1e-6,1e-6);

   retp(h);
endp;
/***************************************************************/
proc(1)=garch_fc(theta,e,hhat,s);
   local om,al,be,h_t1,h_ts,iter,sig2;
   om=theta[1];
   al=theta[2];
   be=theta[3];
   h_t1=om+al*e.^2+be*hhat;
   h_ts=h_t1;
   sig2=om/(1-al-be);
   if s>1;
      iter=2;
      do until iter>s;
         h_ts=h_ts|(sig2+(al+be)^(iter-1)*(h_t1-sig2));
         iter=iter+1;
      endo;
   endif;
   retp(h_ts);
endp;

/***************************************************************/
proc(1)=figarch_fc(params,e,hhat,s);
   local i,j,t,es,rs,r2,omega,phi,beta,d,lambda,delta,tau,h,fc;

   t=rows(e);
   es=e.^2;

   omega=params[1]; phi=params[2]; beta=params[3]; d=params[4];

   lambda=zeros(tl,1);
   delta=zeros(tl,1);
   lambda[1]=phi-beta+d;
   delta[1]=d;

   j=2;
   do until j>tl;
   delta[j]=(j-1-d)/j*delta[j-1];
   lambda[j]=beta*lambda[j-1]+(delta[j]-phi*delta[j-1]);
   j=j+1;
   endo;

   fc=zeros(s,1);
   h=hhat;
   tau=rows(lambda);

   i=1;
   do until i>s;
   rs=es[t-(tau-i-1):t];
   r2=rs|h;
   fc[i]=omega+lambda'rev(r2);
   h=h|fc[i];
   i=i+1;
   endo;
```

```

   retp(fc);
endp;
/**************************************************************/
proc(2)=loss_se_avg(fc,a);
   local n,p,fcavg,aavg,iter,l_series,l_mean;
   n=rows(fc);
   p=cols(fc);
   fcavg=sumc(fc');
   aavg=zeros(n,1);
   iter=1;
   do until iter>n;
      aavg[iter]=sumc(a[iter:iter+(p-1)]);
      iter=iter+1;
   endo;
   l_series=(fcavg-aavg)^2;
   l_mean=meanc(l_series);
   retp(l_series,l_mean);
endp;
/*************************************************************
PROC: GARCH11_VAR_FC

This procedure calculates a VaR quantile forecast for aggregate
returns for a GARCH(1,1) model at a horizon of s using a standard
normal distribution for the standardized residuals. The process is
assumed to have zero conditional and unconditional means.

Format: varq=garch11_var_fc(b_garch,fc,s,alpha,sims);

Intput:

b_garch = 3-vector of GARCH(1,1) estimates (omega|alpha|beta)
fc      = one-step-ahead conditional variance forecast
s       = forecast horizon
alpha   = probability of loss
sims    = number of simulations

Output:

varq = VaR quantile forecast
*************************************************************/
proc(1)=garch_varfc(b_garch,fc,s,alpha,sims);
   local tretstar,itersims,zstar,hstar,estar,iters,alphai,varq;
   tretstar=ones(sims,1);
   itersims=1;
   do until itersims>sims;
      zstar=rndns(s,1,seed2);
      if s>2;
         hstar=fc|zeros(s-1,1);
      else;
         hstar=fc;
      endif;
      estar=zeros(s,1);
      estar[1]=sqrt(hstar[1])*zstar[1];
      if s>2;
         iters=2;
         do until iters>s;
            hstar[iters]=b_garch[1]+b_garch[2]*estar[iters-1]^2
               +b_garch[3]*hstar[iters-1];
            estar[iters]=sqrt(hstar[iters])*zstar[iters];
            iters=iters+1;
         endo;
      endif;
      tretstar[itersims]=sumc(estar);
      itersims=itersims+1;
   endo;
```

```
   alphai=round(alpha*sims);
   tretstar=sortc(tretstar,1);
   varq=tretstar[alphai];
   retp(varq);
endp;

/*****************************************************************
PROC: CONSTANT_VAR_FC

This procedure calculates a VaR quantile forecast for aggregate
returns for a constant variance model at a horizon of s using a
standard normal distribution for the standardized residuals. The
process is assumed to have zero conditional and unconditional means.

Format: varq=constant_var_fc(fc,s,alpha,sims);

Intput:

fc      = variance forecast
s       = forecast horizon
alpha   = probability of loss
sims    = number of simulations

Output:

varq = VaR quantile forecast
*****************************************************************/
proc(1)=cons_varfc(fc,s,alpha,sims);
   local tretstar,itersims,estar,alphai,varq;
   tretstar=ones(sims,1);
   itersims=1;
   do until itersims>sims;
      estar=sqrt(fc)*rndns(s,1,seed2);
      tretstar[itersims]=sumc(estar);
      itersims=itersims+1;
   endo;
   alphai=round(alpha*sims);
   tretstar=sortc(tretstar,1);
   varq=tretstar[alphai];
   retp(varq);
endp;

/*****************************************************************
PROC: FIGARCH11_VAR_FC

This procedure calculates a VaR quantile forecast for aggregate
returns for a FIGARCH(1,d,1) model at a horizon of s using a
standard normal distribution for the standardized residuals. The
process is assumed to have zero conditional and unconditional means.

Format: varq=figarch11_var_fc(b_figarch,fc,e,s,alpha,sims);

Intput:

b_figarch = 4-vector of FIGARCH(1,d,1) estimates (omega|alpha|beta|d)
fc        = one-step-ahead conditional variance forecast
e         = T-vector of observations
s         = forecast horizon
alpha     = probability of loss
sims      = number of simulations

Output:

varq = VaR quantile forecast
*****************************************************************/
proc(1)=figarch_varfc(b_figarch,fc,e,s,alpha,sims);
```

```
    local i,j,t,es,rs,e2,r2,omega,phi,beta,d,lambda,delta,tau;
    local tretstar,itersims,zstar,estar,hstar,alphai,varq;

    t=rows(e);
    es=e.^2;
    omega=b_figarch[1]; phi=b_figarch[2]; beta=b_figarch[3]; d=b_figarch[4];

    lambda=zeros(tl,1);
    delta=zeros(tl,1);
    lambda[1]=phi-beta+d;
    delta[1]=d;

    j=2;
    do until j>tl;
    delta[j]=((j-1-d)/j)*delta[j-1];
    lambda[j]=beta*lambda[j-1]+(delta[j]-phi*delta[j-1]);
    j=j+1;
    endo;

    tau=rows(lambda);
    tretstar=ones(sims,1);

    itersims=1;
    do until itersims>sims;
    zstar=rndns(s,1,seed2);
    if s>2;
    hstar=fc|zeros(s-1,1);
    else;
    hstar=fc;
    endif;
    estar=zeros(s,1);
    estar[1]=sqrt(hstar[1])*zstar[1];
    if s>1;

    i=2;
    do until i>s;
    rs=es[t-(tau-i-1):t];
    e2=estar[1:i].^2;
    r2=rs|e2;
    hstar[i]=omega+lambda'rev(r2);
    estar[i]=sqrt(hstar[i])*zstar[i];
    i=i+1;
    endo;
    endif;
    tretstar[itersims]=sumc(estar);
    itersims=itersims+1;
    endo;

    alphai=round(alpha*sims);
    tretstar=sortc(tretstar,1);
    varq=tretstar[alphai];
    retp(varq);
endp;
/*****************************************************************
PROC: VAR_STATS

This procedure calculates the average Value at Risk quantile
forecasts and the percentage of times the actual return is less
than the Value at Risk quantile forecast.

Format: {varq_avg,varpercent}=var_stats(fc,a,s);

Intput:

fc = vector of VaR quantile forecasts
a  = vector of actual values for returns
```

```
s  = forecast horizon

Output:

varq_avg    = average VaR quantile forecast
varpercent = percent
*****************************************************************/
proc(2)=var_stats(fc,a,s);
   local n,varq_avg,aagg,iter,varpercent;
   n=rows(fc);
   varq_avg=meanc(fc);
   aagg=zeros(n,1);
   iter=1;
   do until iter>n;
      aagg[iter]=sumc(a[iter:iter+(s-1)]);
      iter=iter+1;
   endo;
   varpercent=sumc(aagg.<fc)/n;
   retp(varq_avg,varpercent);
endp;
/*****************************************************************
PROC: LOSS_VAR

This procedure calculates values for the VaR-based loss
function described in Section 4.3 of Gonzalez-Rivera et al.
(2004) for aggregate returns. It also calculates the mean
loss.

Format: {loss_series,loss_average}=loss_var(fc,a,s,alpha);

Intput:

fc     = vector of VaR quantile forecasts
a      = vector of actual values for one-period returns
s      = forecast horizon
alpha = probability of loss

Output:

loss_series  = vector of VaR-based loss function values
loss_average = mean loss

Reference

G. Gonzalez-Rivera, T.-H. Lee, and S. Mishra (2004),
"Forecasting Volatility: A Reality Check Based on Option
Prcing, Utility Function, Value-at-Risk, and Predictive
Likelihood," International Journal of Forecasting, 20(4),
629-645
*****************************************************************/
proc(2)=loss_var(fc,a,s,alpha);
   local n,aagg,iter,d_alpha,loss_series,loss_average;
   n=rows(fc);
   aagg=zeros(n,1);
   iter=1;
   do until iter>n;
      aagg[iter]=sumc(a[iter:iter+(s-1)]);
      iter=iter+1;
   endo;
   d_alpha=aagg.<fc;
   loss_series=(alpha-d_alpha).*(aagg-fc);
   loss_average=meanc(loss_series);
   retp(loss_series,loss_average);
endp;
```