Journal of Economic Theory and Econometrics, Vol. 26, No. 2, Jun. 2015, 36–56

The Effect of Level Shift in the Unconditional Variance on Predicting Conditional Volatility

Hojin Lee*

Abstract We evaluate out-of-sample forecasting performance of different prediction models using different estimation windows to account for a variety of statistical characteristics such as the long range dependence and the structural breaks of the process. We identify the timing of the deterministic shifts in the unconditional variance and evaluate the impact of accounting for the level shifts in the unconditional variance on out-of-sample volatility forecasting. The modified iterated cumulative sums of squares algorithm identifies two shifts in the unconditional variance for the KOSPI (Korea Composite Stock Price Index) returns. For the KOSPI returns process, the full sample performance of the recursive GARCH(1,1) model is worse than the competing models, which is unsurprising given two structural breaks in the process. The superiority of the competing models in forecasting performance can be attributed to the capability of the model which accommodates both the long range dependence by giving a slow hyperbolic rate of decaying weights on the past observations in forming the likelihood and the structural changes in the variance by discarding observations beyond a rolling window length distance in the past which may have come from a different regime. Although we try to improve the forecasting performance by incorporating statistical characteristics of the process into a prediction model, the out-of-sample performance of the prediction model can be tainted with uncertainties related to statistical tests and estimation methodologies.

Keywords Structural Break, Long Memory, Long Range Dependence, Out-of-Sample Forecast, Volatility Break, ICSS Algorithm

JEL Classification C12, C22

*Associate Professor, Department of Business Administration, Myongji University, 34 Geobukgol-ro, Seodaemun-Gu, Seoul 120-728, Korea, e-mail: hlee07@mju.ac.kr. We thank referees for very insightful comments on earlier drafts. The usual disclaimer applies.

Received October 20, 2014, Revised August 5, 2014, Accepted June 25, 2015

1. INTRODUCTION

The findings that the level shifts in the unconditional variance can generate the long range dependence (LRD) in volatility or the integrated GARCH behavior have received much attention in the extant literature. Lobato and Savin (1998) assert that the long memory property of the market index returns may be ascribed to the nonstationarity of the series. They suggest resolving the nonstationarity of the conditional volatility of the stock market index returns by disaggregation or by splitting into stationary subperiods of returns. Although it is hard to differentiate between the long memory process and the occasional-break model in terms of statistical test, Granger and Hyung (2004) suggest combining the fractionally integrated model with the break model to improve volatility forecasting performance. Diebold and Inoue (2001) show that a variety of stochastic regime switching models produce long memory property process. Mikosch and Stărică (2004) show that structural breaks in the variance produce the LRD in the long span of high frequency data. They argue that the statistical procedures to test for the LRD in volatility process cannot differentiate the seemingly IGARCH process.

Whereas the extant literature focuses on model misspecification under nonlinearities, Amado and Teräsvirta (2012, 2013, 2014) model the unconditional variance as time-varying by specifying the conditional variance as a GJR-GARCH and the unconditional variance as a linear combination of logistic transition functions. The out-of-sample forecasting performance evaluation in their work shows that the forecasting performance of the proposed model is dramatic when the subperiod covers a period of turmoil such as the Great Depression and the oil crisis in 1973. Their approach improves the performance of out-of-sample forecast due to the characterization of the time-varying nature of the unconditional volatility, however, the specification of the subperiod is not based on the statistical testing procedures.

This paper aims to model nonlinearities in stock market volatility. Specifically, we address the issue of long range dependence and level shifts in volatility in relation to out-of-sample forecasting performance of volatility. To that end, we evaluate out-of-sample forecasting performance of different prediction models using different estimation windows to account for a variety of statistical characteristics such as the LRD and the structural breaks of the processes. We use the ICSS (integrated cumulative sums of squares) algorithm to identify sudden shifts in the variance and the duration of such shifts. In the process of estimating the prediction model, we use the expanding estimation window or the rolling estimation window depending on the model specification.

The results of this paper show that the most distinct structural breaks in the unconditional volatility are associated with the global or regional financial crises. In terms of model specification, our estimation results show that the long range dependence in volatility may be due to the level shifts in the unconditional volatility. We also find that the out-of-sample forecasting performance of the model with the structural breaks is superior to that of the plain GARCH model.

The main contributions of the paper are two-fold. First, we show that it is important to take structural breaks into account in producing out-of-sample forecasts. Second, we provide evidence on the benefits from using pre-break data in estimating forecasting models by trading the bias off against forecast error variance.

The organization of the article is as follows. Section 2 presents statistical tests for long range dependence and for level shifts in variance. Then, we introduce a variety of prediction models and estimation windows for conditional volatility forecasts and the methodology to compare forecast performance of different models. Section 3 contains the empirical results for sudden changes in variance, along with a comparison of the forecasting performance of the models. Section 4 concludes the discussion.

2. METHODOLOGY

2.1. CHANGE-POINT TEST

We investigate the possibility that the level shifts in the unconditional variance process can produce the long range dependence (LRD) in volatility or the integrated GARCH behavior. Mikosch and Stărică (2004) assert that the long memory process could be spurious due to the fact that the statistics to test for a long memory process against a nonstationary process cannot differentiate between the two stochastic processes. In addition, in-sample fit of a GARCH model to a stochastic process with level shifts generates the IGARCH effect. Lamoureux and Lastrapes (1990) show that the unaccounted structural breaks in the unconditional variance induce upward biases in GARCH estimates of persistence in variance.

In this paper, we address the issue of the unaccounted structural breaks. We identify the timing of the deterministic shifts in the unconditional variance and evaluate the impact of accounting for the level shifts in the unconditional variance on out-of-sample volatility forecasting. The level shifts in the unconditional variance are identified by applying the iterated cumulative sums of squares (ICSS) methodology due to Inclan and Tiao (1994). What we expect from the analysis is that whether the modeling strategy to take the level shifts into account reduces the persistence of volatility and enhances the out-of-sample forecasting performance of the model.

We use the Inclan and Tiao (1994) statistic to identify the timing of deterministic shifts in the unconditional variance with the ICSS algorithm. Define the process $X_t = |r_t|^\delta$ for $\delta = 1, 2$, where r_t is the stock market index returns process. Inclan and Tiao (1994) statistic to test the null hypothesis of a constant unconditional variance against the alternative of a deterministic level shift in the unconditional variance is $\sqrt{T/2}\max_k |D_k|$, $D_k = \frac{\sum_{j=1}^{k} X_j}{\sum_{j=1}^{k} X_j} - \frac{k}{T}$ and $D_0 = D_T = 0$. We identify the timing of the structural break in the unconditional variance where the statistic is maximized at a certain value of k in the sample.

The problem of the Inclan and Tiao statistic lies in the assumption that the process is independently and identically distributed, despite the fact that the series analyzed in this study are strongly dependent. Andreou and Ghysels (2002) examine the performance of the Inclan and Tiao test with non-independent process and suggest the Kokoszka and Leipus (2000) test for breaks in an ARCH process due to good power properties. Sanso et al. (2004) suggest the application of non-parametrically modified Inclan and Tiao test for multiple breaks to a variety of strongly dependent processes such as heteroskedastic conditional variance processes. The modified Inclan and Tiao statistic for constant unconditional variance is used to correct for size distortions of the Inclan and Tiao statistic when the processes are conditionally heteroskedastic and iid non-mesokurtic.

2.2. FORECASTING MODELS AND ESTIMATION WINDOW

The success of the GARCH models and their extensions in forecasting volatility can be attributed to the capability of characterizing the long range dependence in the absolute and squared returns and the non iid mesokurtic behavior of the process. However, as previously noted, these characteristics can also be obtained as a result of the level shifts in the unconditional variance process. Although the in-sample fit of a GARCH volatility model to a long range dependent volatility process can be good, the predictive power of the model may not be accurate when the variance process has structural breaks.

We use the out-of-sample forecasts evaluation procedures which have been extensively investigated in Rapach and Strauss (2008) and use the GARCH(1,1) model as the benchmark. From a variety of estimation windows, we use an expanding window with the GARCH(1,1) specification under the assumption of no structural breaks as the benchmark. We reserve the last 500 observations for the out-of-sample forecasts evaluation. The models are estimated with the first 3,487 observations. For a given model, we forecast the volatilities for the 500 out-of-sample forecast period. We evaluate the out-of-sample forecast performance based on the mean squared forecast error (MSFE) criterion.

We compare the performance of the GARCH(1,1) with that of the following models. Firstly, we choose the FIGARCH(1,d,1) model used in Baillie et al. (1996). The use of the FIGARCH model is motivated by the presence of long memory property in the autocorrelations of squared returns of the stock market index. The inclusion of the FIGARCH model as the competing model is intended to confirm whether the parameters of the misspecified GARCH models are very persistent and show the IGARCH effect in the conditional variance. To this end, a slow hyperbolic rate of decaying but mean-reverting behavior of the FIGARCH model may be more appropriate than an extreme behavior of the IGARCH process for volatility forecasting purposes, where the effect of a shock to the IGARCH volatility process persists for good.

As Pesaran and Timmermann (2007) present evidence on the performance of a forecasting model, the parameter estimates of a prediction model using the data from the pre-break sample period are informative even after the break. Also shown by Pesaran and Timmermann (2007) is that the potential gains in prediction accuracy from using pre-break data are dependent on locating the timing of break points. The out-of-sample forecasting performance is determined not only by the selection of the model but by the choice of the optimal estimation window.

If the unconditional volatility process has no breaks over the sample, the parameter estimates of the prediction models with an expanding window estimation procedure can be made efficient. In this regard, we choose the GARCH(1,1) model with an expanding window as the benchmark model in the forecast competition. As an alternative, a rolling window estimator which uses a fixed number of the most recent data can be used to estimate prediction models in cases where we suspect parameter constancy of the models due to structural breaks over the sample.

We partition a sample of 3,987 observations into an in-sample of 3,487 observations and an out-of-sample of 500 observations. The competing models except for the GARCH with breaks estimate prediction models based on data from different regimes. There are trade-offs between the bias and the forecast error variance, dependently of whether we include pre-break data. That is, we may trade off an increased persistence in the GARCH estimates by including pre-break data against a decreased forecast error variance from obtaining more observations in the estimation window. We evaluate the forecasting performance of the models by setting the size of the estimation window at 50% (long rolling estimation window) and 25% (short rolling estimation window) of the in-sample. The prediction models are estimated using the in-sample observations and the out-of-sample forecast exercises are repeated. The statistical tests on the forecast errors from the competing models for predictive accuracy are exercised to determine the appropriate forecasting model.

The fifth model for potential changes in variance is based on weighted maximum likelihood (WML) estimation procedure proposed in Mittnik and Paollela (2000) and also used for out-of-sample forecasting evaluation in Rapach and Strauss (2008). The rolling window WML procedure accommodates both the long range dependence by giving a slow hyperbolic rate of decaying weights on the past observations in forming the weighted likelihood and the level shifts in the variance by discarding observations beyond a rolling window length distance in the past which may have come from a different regime.

On the contrary to the approaches so far, we may use a GARCH model with variance breaks, so that we use only observations over the post-break period. We estimate the time of breaks using the modified ICSS algorithm by Inclan and Tiao (1994). The prediction model is estimated using the subsample observations identified by the most recent structural break to produce the out-of-sample forecast for the conditional variance. Next, we add one more observation in the expanding window to test for structural breaks. If we identify structural breaks, we estimate the post-break prediction model. If we find no evidence on structural instability in the variance process, the prediction model is estimated using the entire sample at the time of the forecast formation. A word of caution is in order. As investigated in Hwang and Pereira (2006), small sample properties of the ML estimates of GARCH(1,1) model may affect the parameter estimates of the prediction model and the forecasting performance of the post-break GARCH models.

3. EMPIRICAL RESULTS

3.1. DATA

The data used in this study is the daily KOSPI and the S&P 500 index during the period from January 2, 1998 to February 28, 2014. The first difference of the logarithmic price is taken to calculate the continuously compounded returns. Table 1 reports the descriptive statistics of the index returns. The KOSPI return is more volatile with the standard deviation of 1.88%. The two returns processes are leptokurtic and negatively skewed, meaning that they are more likely to have extreme negative returns than positive ones. From the modified Ljung-Box Q-statistics in the squared index returns, we reject the null hypothesis of no serial correlation in the squared returns and confirm that the conditional volatilities are time-varying. Engle (1982) Lagrange multiplier test shows strong evidence of ARCH effects.

Figure 1 illustrates the return and the conditional standard deviation of the S&P 500 index in the left panel and those of the KOSPI in the right panel. The graphs in the right panel identify significant changes in the conditional variance of the KOSPI returns, which indicates that the variance of the process conditional on its past history may change over time, or equivalently, the process may exhibit time varying conditional heteroskedasticity or volatility clustering. While the conditional volatility of the KOSPI returns shows a downward trend, the conditional volatility of the S&P 500 index returns is nearly constant.

Table 1: Descriptive statistics and test statistics

	S&P500 return	KOSPI return
A. Descriptive statistics		
Mean	0.02(0.02)	0.04(0.03)
Standard Deviation	1.30(0.04)	1.88(0.04)
Skewness	-0.19(0.24)	-0.16(0.16)
Kurtosis	10.22(1.46)	7.30(0.61)
Minimum	-9.47	-12.80
Maximum	10.96	11.28
B. Test statistics for serial correlation		
Q(20)	28.51(0.10)	15.74(0.73)
Q(20) on the squared returns	5327.05(0.00)	1632.76(0.00)
C. Test statistics for ARCH effects		
LM(1)	154.96(0.00)	114.60(0.00)
LM(6)	930.87(0.00)	434.78(0.00)
LM(12)	1096.03(0.00)	476.61(0.00)

Note: The entries in panel A are reported in percentage unit except the skewness and kurtosis. The entries in the parentheses in panels A are the heteroskedasticity and autocorrelation consistent standard errors. The entries in the parentheses in panels B and C are the p-values. The Q(20) stands for the modified Ljung-Box Q-statistic for serial correlation for 20 lags. The LM(p) stands for the Engle (1982) Lagrange multiplier test for ARCH(p) effects for p lags

Figure 1: Daily log returns and the conditional standard deviation of the index

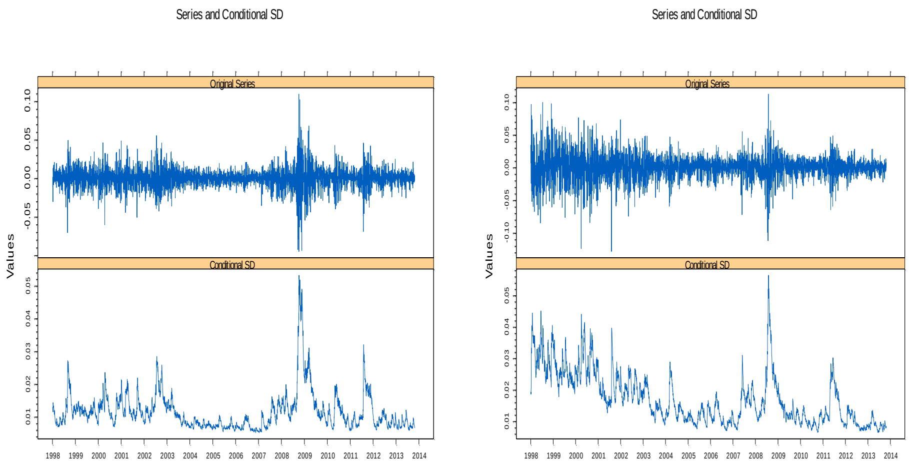
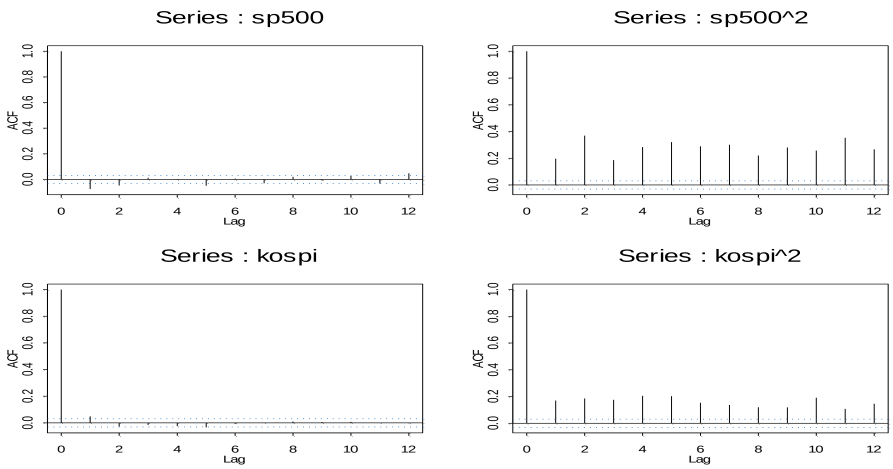


Figure 2 shows the sample ACF of the returns and of the squared returns. The significant sample autocorrelations at long lags in the squared returns can be interpreted as the IGARCH effect or as the long range dependence. As Sanso et al. (2004), Stărică and Granger (2005) and Wang and Moore (2009) suggest, the existence of significant sample autocorrelation at long lags in the squared returns may be caused by deterministic level shifts in the unconditional variance.

Figure 2: Sample ACF of the S&P 500 index return and the KOSPI return



The empirical findings in this section confirm that the daily index return process exhibits the IGARCH effect or at least the long range dependence property in variance. However, the persistence of the process may be explained by the presence of deterministic level shifts in the unconditional volatility. We perform tests for the existence of long range dependence or structural breaks in the unconditional variance and proceed with the application of the results to evaluating the out-of-sample forecasting performance of a variety of models in the next sections.

3.2. LONG RANGE DEPENDENCE

Firstly, we test for long range dependence or long memory in the stock market index returns processes using the modified range over standard deviation (R/S) statistic due to Lo (1991). The modified R/S statistic is corrected for the bias of the classical R/S statistic which is sensitive to short range dependence of the process. As we report in Table 1, averages of the index return processes are 0.02 and 0.04 %, respectively. Thus, the absolute return of each process can be used as a volatility measure. We test for long memory in the absolute returns of the index. From both the modified R/S statistic and the classical R/S statistic reported in Table 2. The modified R/S statistic is different from the classical R/S

Table 2: Tests for long memory

Statistics	S&P500 return	KOSPI return
Modified R/S	3.22**	5.29**
Classical R/S	6.29**	9.52**
GPH	5.53**	5.88**

Note: The modified R/S statistic is different from the classical R/S statistic in its denominator which is the square root of the partial sums variance. In calculating the long run variance, we use the Newey and Wests procedure in obtaining the bandwidth parameter. If we set the bandwidth parameter equals to 0, then the classical R/S statistic is computed. ** denotes the significance at the 1% level.

statistic in its denominator which is the square root of the partial sums variance. In calculating the long run variance, we use the Newey and Wests procedure in obtaining the bandwidth parameter. If we set the bandwidth parameter equals to 0, then the classical R/S statistic is computed. The null hypothesis of no long range dependence is rejected at the 1% significance level for the two absolute return processes. The estimated fractional difference parameters are 0.440 and 0.323 using the R/S method, respectively. The semi-nonparametric GPH test for long memory due to Geweke and Porter-Hudak (1983) can also be used. The estimated fractional difference parameters are 0.498 and 0.530, and the GPH tests for long memory in the S&P 500 index absolute returns and the KOSPI absolute returns reject the null hypotheses of no long memory at the 1% significance level.

As shown in Hosking (1981), when the fractional difference parameter d, $0 < d < 1/2$, the absolute returns process is stationary and long range dependent and when $|d| > 1/2$, the process is nonstationary. Thus, we investigate the fractional difference parameter using a fractionally integrated GARCH (FIGARCH) model to find an appropriate prediction model for the forecast competition.

We choose the FIGARCH(1,d,1) model used in Baillie et al. (1996) for a prediction model. The use of the FIGARCH model is motivated by the presence of long memory property in the realized volatility of the stock market index. The inclusion of the FIGARCH model as the competing model is intended to confirm whether the parameters of the misspecified GARCH models are very persistent and show the IGARCH effect in the conditional variance. To this end, a slow hyperbolic rate of decaying but mean-reverting behavior of the FIGARCH model

Table 3: FIGARCH(1,d,1) Model $h_t = \omega + (1-\alpha L)(1-L)^d e_t^2 + \beta h_{t-1}$ estimation results with the full sample

	S&P500 return	KOSPI return
A. Estimation		
$\hat{\omega}$	0.03(0.00)	0.03(0.01)
$\hat{\alpha}$	0.01(0.02)	0.16(0.03)
$\hat{\beta}$	0.70(0.04)	0.57(0.05)
$\hat{d}$	0.71(0.06)	0.46(0.04)
B. Diagnostic statistics		
Q(12) on the standardized residuals	26.58(0.01)	19.27(0.08)
Q(12) on the sq. standardized residuals	13.48(0.34)	9.72(0.64)
TR^2	13.19(0.36)	9.06(0.70)

Note: The table shows the FIGARCH(1,d,1) model fit to the full sample of data. The entries in parentheses are standard errors. The entries in panel A are the estimation results of the FIGARCH(1,d,1) model. The entries in the parentheses in panels B are the p-values. The Q(12) stands for the modified Ljung-Box Q-statistic for up to twelfth-order serial correlation in the standardized residuals and the squared standardized residuals. The diagnostic statistics TR^2 are reported in panel B

may be more appropriate than an extreme behavior of the IGARCH process for volatility forecasting purposes.

The fit of the FIGARCH(1,d,1) model to the full sample of data is reported in Table 3. For the S&P 500 index returns, the fractional difference parameter estimate of the FIGARCH model is 0.71, which indicates the nonstationarity of the process. For the KOSPI returns, however, the fractional difference parameter estimate of the FIGARCH model is 0.46, meaning that the process has the long range dependence. From the estimation results, we confirm that the FIGARCH model provides a better fit than the GARCH model for the S&P 500 index returns. The FIGARCH model produces lower AIC and BIC values than the GARCH model. For the KOSPI returns, the FIGARCH model has lower AIC and BIC values than the GARCH model. Also, the GARCH estimates of persistence in variance of the FIGARCH(1,d,1) model are stationary and mean-reverting

Table 4: GARCH(1,1) Model $h_t = \omega + \alpha e_{t-1}^2 + \beta h_{t-1}$ estimation results with the full sample

	S&P500 return	KOSPI return
A. Estimation		
$\hat{\omega}$	0.02(0.00)	0.01(0.00)
$\hat{\alpha}$	0.09(0.01)	0.07(0.01)
$\hat{\beta}$	0.90(0.01)	0.92(0.01)
$\hat{\sigma}^2$	1.69(0.20)	3.55(0.28)
B. Diagnostic statistics		
Q(12) on the standardized residuals	26.66(0.01)	18.99(0.09)
Q(12) on the sq. standardized residuals	30.45(0.00)	13.02(0.37)
TR^2	29.56(0.00)	11.09(0.52)

Note: The table shows the GARCH(1,1) model fit to the full sample of data. The entries in parentheses are standard errors. The entries in panel A are the estimation results of the GARCH(1,1) model. The entries in the parentheses in panels B are the p-values. The Q(12) stands for the modified Ljung-Box Q-statistic for up to twelfth-order serial correlation in the standardized residuals and the squared standardized residuals. The diagnostic statistics TR^2 are reported in panel B.

3.3. LEVEL SHIFTS IN VARIANCE

The QMLE fit of the GARCH(1,1) model to the full sample of data uses the GAUSS Constrained Maximum Likelihood 2.0 module and is reported in Table 4. Consistent with the prior literature, the estimated GARCH parameters are both significant and the sum of the GARCH coefficients is close to one, meaning that the processes in both markets show persistence in volatility. The half life of a volatility shock implied by the mean reverting rate for the S&P 500 index process is 67 days and that for the KOSPI series is 668 days, respectively.

Also included are the diagnostic statistics such as the Ljung-Box Q-statistics applied to the standardized residuals and the TR^2. From the statistically significant diagnostic results, we conclude that the GARCH(1,1) model does not fit to the S&P 500 index returns well. However, for the KOSPI, neither the standardized residuals nor the squared standardized residuals of the estimated model show further autocorrelation left. The TR^2 statistic with the standardized residuals fails to reject the null hypothesis.

For the KOSPI returns process, the diagnostic test results to the standardized

Table 5: GARCH(1,1) model fit to each KOSPI subsample

	Subsample 1 (1998.1.2-2000.10.31)	Subsample 2 (2000.11.1-2009.12.3)	Subsample 3 (2009.12.4-2014.2.28)
$\hat{\omega}$	1.93(0.79)	0.03(0.01)	0.03(0.01)
$\hat{\alpha}$	0.09(0.04)	0.07(0.01)	0.08(0.02)
$\hat{\beta}$	0.69(0.11)	0.92(0.01)	0.89(0.02)
$\hat{\sigma}^2$	8.73(0.67)	3.02(0.32)	1.26(0.20)

Note: The table shows the GARCH(1,1) model fit to each subsample identified by the modified ICSS algorithm. The entries in parentheses are standard errors

residuals of the estimated GARCH(1,1) model rationalize the GARCH modeling of the serial volatility dependencies. However, as Lamoureux and Lastrapes (1990) and Mikosch and Stărică (2004) have shown, in-sample fit of GARCH models to stochastic processes with unaccounted level shifts in variance may generate upward biases in GARCH estimates of persistence in variance or the IGARCH effect.

Since the index returns we analyze are leptokurtic and conditionally heteroskedastic, we need to employ the ICSS algorithm with the modified Inclan and Tiao statistic to detect structural changes in the unconditional variance. While the modified ICSS algorithm identifies no variance break for the S&P 500 index returns, two shifts in the unconditional variance are identified with the KOSPI returns. Table 5 reports the dates of structural changes in the unconditional variance using the ICSS algorithm. The dates of the deterministic shifts in the unconditional variance are November 1, 2000 and December 4, 2009. The subsample period 1 includes to the period of the Asian financial crisis in 1997 and records the highest level of the unconditional variance. The test for the deterministic shifts also reveals that the unconditional variance process experiences another structural break during the global financial crisis of the late 2000s. After the shift to a relatively stable period in the unconditional variance on December 4, 2009, the level of volatility in the KOSPI return process becomes subdued. The estimation results of the GARCH(1,1) model across each subsample are reported in Table 5. The values of the GARCH coefficients are decreased when the model is fit to the subsample 1, ranging from January 5, 1998 to October 31, 2000.

3.4. OUT-OF-SAMPLE FORECAST

In choosing prediction models and estimation windows for the forecast competition, we have the following considerations. Firstly, there are trade-offs between the bias and the forecast error variance, dependently of whether we include pre-break data. That is, we may trade off an increased persistence in the GARCH estimates by including pre-break data against a decreased forecast error variance from obtaining more observations in the estimation window. Secondly, the rolling window WML procedure accommodates both the long range dependence by giving a slow hyperbolic rate of decaying weights on the past observations in forming the weighted likelihood and the level shifts in the variance by discarding observations beyond a rolling window length distance in the past which may have come from a different regime. Thirdly, small sample properties of the ML estimates of GARCH(1,1) model may affect the parameter estimates of the prediction model and the forecasting performance of the post-break GARCH models. Fourthly, the inclusion of the FIGARCH model as the competing model is intended to confirm whether the parameters of the misspecified GARCH models are very persistent and show the IGARCH effect in the conditional variance. To this end, a slow hyperbolic rate of decaying but mean-reverting behavior of the FIGARCH model is more appropriate than an extreme behavior of the IGARCH process for volatility forecasting purposes.

Table 6 reports the results of s=1, 20, 60 and 120-step ahead forecasts of the prediction models. The out-of-sample covers the period from February 21, 2012 to February 28, 2014. The entries reported are the MSFE of the recursive GARCH(1,1) model and the ratios of the MSFE of the competing models to that of the benchmark recursive GARCH(1,1) model.[1] Thus, the ratios above unity mean worse forecasting performance and ratios below unity better forecasting performance than the benchmark.

For the KOSPI returns process, the full sample performance of the recursive GARCH(1,1) model is worse than the competing models, which is unsurprising given two structural breaks in the process. We observe that the recursive GARCH(1,1) model has the larger MSFE than the competing models. We also find that the predictive gains of the competing models range from 2 % to 73 % depending on the out-of-sample forecasting horizon. Along with the short rolling window GARCH(1,1) model, the predictive gains of the rolling window WML GARCH(1,1) model are distinct. The superiority of the model in forecasting performance can be attributed to the capability of the model which ac-

[1]We are grateful to Dave E. Rapach for providing his GAUSS code used in this forecast performance evaluation.

commodates both the long range dependence by giving a slow hyperbolic rate of decaying weights on the past observations in forming the weighted likelihood and the structural changes in the variance by discarding observations beyond a rolling window length distance in the past which may have come from a different regime.

It appears to be appropriate to estimate the parameters of the forecasting model using the post-break sample data to address the IGARCH effect or the long range dependence in variance process when there are structural breaks in the unconditional variance process. The use of post-break observations in estimating the prediction model produces better forecasting performance than the benchmark recursive GARCH(1,1) model in most of the forecasting horizons. However, the forecasting performance of the post-break GARCH(1,1) model is not impressive among the competing models. The out-of-sample forecasting performance evaluation results suggest that the forecasting performance may be affected by the power of the test for structural breaks. That is, there are uncertainties related to test and estimate the time of the break in the unconditional variance process. Further, the most recent break in the unconditional variance is on December 4, 2009, and we have 546 observations in the post-break subsample. As Hwang and Pereira (2006) suggest, it is desirable to have at least 500 observations to correct for the biases in GARCH estimates considering the convergence errors. So, we are on the verge of having the convergence errors in estimating the prediction models which may have affected the forecasting performance of the GARCH model with structural breaks.

For the S&P 500 index returns process, the recursive GARCH(1,1) model generates the lowest MSFE values except the weighted maximum likelihood (WML) GARCH(1,1) model due to Mittnik and Paollela (2000) and Rapach and Strauss (2008). However, the predictive gains of the rolling window WML GARCH(1,1) model are not dramatic compared to those of the competing models for the KOSPI returns process. The predictive gains of the WML GARCH(1,1) model range from 0 % to 47 % depending on the out-of-sample forecasting horizon. We can interpret that a recursive estimation window procedure as in the benchmark recursive GARCH(1,1) model increases efficiency of a prediction model when there are no structural breaks in the unconditional variance process.

4. CONCLUSION

The success of the GARCH models and their extensions in forecasting volatility can be attributed to the capability of characterizing the long range dependence

Table 6: Out-of-sample KOSPI return volatility forecasting results

Model	s=1	s=20	s=60	s=120
Recursive GARCH(1,1)	1.85	135.92	1995.46	19137.60
Recursive FIGARCH(1,d,1)	0.99	0.84	0.53	0.35
GARCH(1,1) long rolling window	0.99	0.90	0.75	0.58
GARCH(1,1) short rolling window	**0.98**	**0.75**	**0.47**	**0.29**
WML GARCH(1,1) rolling window	0.99	0.79	**0.46**	**0.27**
Post-break GARCH(1,1)	1.00	0.99	0.95	0.90

Note: The entries for the recursive GARCH(1,1) model are the MSFEs. The entries for the competing models, we report the ratio of the MSFE of the corresponding model to the MSFE of the benchmark recursive GARCH(1,1) model. Therefore, the ratios above unity mean worse forecasting performance and ratios below unity better forecasting performance than the benchmark recursive GARCH(1,1) model

Table 7: Out-of-sample S&P 500 return volatility forecasting results

Model	s=1	s=20	s=60	s=120
Recursive GARCH(1,1)	1.08	63.77	706.47	4809.17
Recursive FIGARCH(1,d,1)	1.01	1.29	1.72	2.24
GARCH(1,1) long rolling window	1.01	1.14	1.30	1.25
GARCH(1,1) short rolling window	1.01	1.16	1.34	1.38
WML GARCH(1,1) rolling window	**1.00**	**0.89**	**0.63**	**0.53**
Post-break GARCH(1,1)	1.00	1.00	1.00	1.00

Note: The entries for the recursive GARCH(1,1) model are the MSFEs. The entries for the competing models, we report the ratio of the MSFE of the corresponding model to the MSFE of the benchmark recursive GARCH(1,1) model. Therefore, the ratios above unity mean worse forecasting performance and ratios below unity better forecasting performance than the benchmark recursive GARCH(1,1) model in the variance and the non iid mesokurtic behavior of the process. However, these characteristics can also be obtained as a result of the level shifts in the unconditional variance process. Although the in-sample fit of a GARCH volatility model to a long range dependent volatility process can be good, the predictive power of the model may not be accurate when the variance process has structural breaks.

This paper aims to clarify the assertion that the out-of-sample forecasting performance is determined not only by the selection of the model but by the choice of the optimal estimation window. If the unconditional volatility process has no breaks over the sample, the parameter estimates of the prediction models with an expanding window estimation procedure can be made efficient. In this regard, we choose the GARCH(1,1) model with an expanding window as the benchmark model in the forecast competition. As an alternative, the GARCH(1,1) model with a rolling window estimator which uses a fixed number of the most recent data can be used to estimate prediction models in cases where we suspect parameter constancy of the models due to structural breaks over the sample.

We address the issue of long range dependence and level shifts in variance in relation to out-of-sample volatility forecasting performance. To that end, we evaluate out-of-sample forecasting performance of the GARCH(1,1) prediction models using different estimation windows to account for a variety of statistical characteristics such as the LRD and the structural breaks of the process. We identify the timing of the deterministic shifts in the unconditional variance and evaluate the impact of accounting for the level shifts in the unconditional variance on out-of-sample volatility forecasting. The level shifts in the unconditional variance are identified by applying the ICSS methodology. While the modified ICSS algorithm identifies no variance break for the S&P 500 index returns, two shifts in the unconditional variance are identified with the KOSPI returns.

For the KOSPI returns process, the full sample performance of the recursive GARCH(1,1) model is worse than the competing models, which is unsurprising given two structural breaks in the process. We observe that the recursive GARCH(1,1) model has the larger MSFE than the competing models. We also find that the predictive gains of the competing models range from 2 % to 73 % depending on the out-of-sample forecasting horizon and that the predictive gains increase as we extend the forecasting horizon. Along with the short rolling window GARCH(1,1) model, the predictive gains of the rolling window WML GARCH(1,1) model are conspicuous. The superiority of the model in forecasting performance can be attributed to the capability of the model which accommodates both the long range dependence by giving a slow hyperbolic rate of decaying weights on the past observations in forming the likelihood and the structural changes in the variance by discarding observations beyond a rolling window length distance in the past which may have come from a different regime.

It appears to be appropriate to estimate the parameters of the forecasting model using the post-break sample data to address the IGARCH effect or the long range dependence in variance process when there are structural breaks in the unconditional variance process. The use of post-break observations in estimating the prediction model produces better forecasting performance than the benchmark recursive GARCH(1,1) model in most of the forecasting horizons. However, the forecasting performance of the post-break GARCH(1,1) model is not impressive among the competing models. The out-of-sample forecasting performance evaluation results suggest that the forecasting performance may be affected by the power of the test for structural breaks. That is, there are uncertainties related to test and estimate the time of the break in the unconditional variance process. Further, the most recent break in the unconditional variance is on December 4, 2009, and we have 546 observations in the post-break subsample. As Hwang and Pereira (2006) suggest, it is desirable to have at least 500 observations to correct for the biases in GARCH estimates considering the convergence errors. So, we are on the verge of having the convergence errors in estimating the prediction models which may have affected the forecasting performance of the GARCH model with structural breaks.

Thus, although we try to improve the forecasting performance by incorporating statistical characteristics of the process such as long memory behavior or structural breaks into a prediction model, the out-of-sample performance of the prediction model can be tainted with uncertainties related to statistical tests and estimation methodologies.

The results of this paper show that the most distinct structural breaks in the unconditional volatility are associated with the global or regional financial crises. In terms of model specification, our estimation results show that the long range dependence in volatility may be due to the level shifts in the unconditional volatility. We also find that the out-of-sample forecasting performance of the GARCH model with structural breaks is superior to that of the expanding window GARCH model.

REFERENCES

Amado, C., and Teräsvirta, T. (2012). Specification and testing of multiplicative time-varying GARCH models with applications. Manuscript, Aarhus University.

Amado, C., and Teräsvirta, T. (2013). Modelling volatility by variance decomposition, *Econometrica*, 175, 142-153.

Amado, C., and Teräsvirta, T. (2014). Modelling changes in the unconditional variance of long stock return series, *Journal of Empirical Finance*, 25, 15-35.

Andreou, E., and E. Ghysels (2002). Detecting multiple breaks in financial market volatility dynamics, *Journal of Applied Econometrics*, 17, 579-600.

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

Diebold, F. X., and Inoue, A. (2001). Long memory and regime switching, *Journal of Econometrics*, 105, 131-159.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1007.

Geweke, J., and Porter-Hudak, S. (1983). The estimation and application of long memory time series models, *Journal of Time Series Analysis*, 4, 221-237.

Granger, C. W. J., and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P500 absolute stock returns, *Journal of Empirical Finance*, 11, 399-421.

Hosking, J. (1981). Fractional differencing, *Biometrika*, 68, 165-176.

Hwang, S., and Valls Pereira, P. S. (2006). Small sample properties of GARCH estimates and persistence, *European Journal of Finance*, 12, 473-494.

Inclan, C., and Tiao, G. C. (1994). Use of cumulative sum of squares for retrospective detection of change of variance, *Journal of the American Statistical Association*, 89, 913-923.

Kokoszka, P., and Leipus, R. (2000). Change-point estimation in ARCH models, *Bernoulli*, 6, 1-28.

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

Lo, A. W. (1991). Long-term memory in stock market prices, *Econometrica*, 59, 1279-1313.

Lobato, I. N., and Savin, N. E. (1998). Real and spurious long-memory properties of stock-market data, *Journal of Business and Economic Statistics*, 16, 261-268.

Mikosch, T., and Stărică, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects, *Review of Economics and Statistics*, 86, 378-390.

Mittnik, S., and Paolella, M. S. (2000). Conditional density and value-at-risk prediction of Asian currency exchange rates, *Journal of Forecasting*, 19, 313-333.

Pesaran, M. H., and Timmermann, A. (2007). Selection of estimation window in the presence of breaks, *Journal of Econometrics*, 137, 134-161.

Rapach, D. E., and Struass, J. K. (2008). Structural breaks and GARCH models of exchange rate volatility, *Journal of Applied Econometrics*, 23, 65-90.

Sanso, A., Arrago, V., and Carrion, J. L. (2004). Testing for change in the unconditional variance of financial time series, *Revista de Economia Financiera*, 4, 32-53.

Stărică, C., and Granger, C. W. J. (2005). Nonstationarities in stock returns, *Review of Economics and Statistics*, 87, 503-522.

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

Wang, P., and Moore, T. (2009). Sudden changes in volatility: The case of five central European stock markets, *Journal of International Financial Markets, Institutions and Money*, 19, 33-46.