UNIVERSITY OF PELOPONNESE
SCHOOL OF ECONOMICS, BUSINESS AND INFORMATION
DEPARTMENT OF ECONOMICS

COMBINING FORECASTS IN VARIOUS FINANCIAL APPLICATIONS

PhD Dissertation

By

Fameliti Stavroula

B.A., University of Peloponnese, 2010
M.Sc., University of Peloponnese, 2013

Committee in charge:
Assistant Professor Vasiliki Skintzi
Associate Professor Timotheos Angelidis
Professor Dimitrios Thomakos

TRIPOLIS
MAY 2019

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ

ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΑΣ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

COMBINING FORECASTS IN VARIOUS FINANCIAL APPLICATIONS

Stavroula P. Fameliti

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Υπεβλήθη στο Τμήμα Οικονομικών Επιστημών του Πανεπιστημίου Πελοποννήσου ως μέρος των απαιτήσεων για την απόκτηση Διδακτορικού Διπλώματος.

Η εκπόνηση της διατριβής συνετελέσθη με την καθοδήγηση τριμελούς Συμβουλευτικής Επιτροπής υπό την Εποπτεία της Επίκ. Καθηγήτριας Σκίντζη Βασιλικής.

ΤΡΙΠΟΛΗ, Μάιος 2019

Η ΕΠΙΒΛΕΠΟΥΣΑ,

Επικ. Καθηγήτρια Βασιλική Σκίντζη

Combining Forecasts in Various Financial Applications

Approved by

Date: May 8, 2019

…………………………	…………………………	…………………………
Skintzi Vasiliki Assistant Professor University of Peloponnese	**Angelidis Timotheos** Associate Professor University of Peloponnese	**Thomakos Dimitrios** Professor University of Peloponnese
…………………………	…………………………	…………………………
Alexopoulos Thomas Assistant Professor University of Peloponnese	**Anastasiou Athanasios** Assistant Professor University of Peloponnese	**Giotopoulos Ioannis** Assistant Professor University of Peloponnese
	…………………………	
	Andrikopoulos Andreas Associate Professor University of Aegean	

Acknowledgments

First and foremost, I offer my sincerest gratitude to my supervior, Assistant Professor Vasiliki Skintzi, who has supported me throughout my PhD with her patience and knowledge. Her kind help and warm-hearted encouragement were well beyond the call of duty. I would like to thank, also, Associate Professor Timotheos Angelidis and Professor Dimitrios Thomakos for their useful comments during my dissertation.

I am grateful to University of Peloponnese for providing me teaching opportunities. I would like to thank participants at the 4th International Conference on Applied Theory, Macro and Empirical Finance (AMEF) in Thessaloniki, the 14th Biennial Athenian Policy Forum (APF) Conference on Contemporary Economic, Financial, Business and Policy Issues in Piraeus and the 9th National Conference of the Financial Engineering and Banking Society (FEBS) in Athens for useful feedback and comments.

I would like to express my very great appreciation to Xu Gong and Boqiang Lin for providing their dataset from their paper: “The incremental information content of investor fear gauge for volatility forecasting in the crude oil futures market”, to complete my research.

Nobody has been more important to me in the pursuit of this project than the members of my family. I would like to thank my parents and my brother, whose love and guidance are with me in whatever I pursue. Without their support this thesis would not have come into existence.

Stavroula Fameliti

Tripoli 2019

Dedications

To my family

Conferences and Publications

Conferences

1. "Predictive Ability and Economic Gains from Volatility Forecast Combinations", 4th International Conference on Applied Theory, Macro and Empirical Finance (AMEF), University of Macedonia, Thessaloniki, 2-3 April 2018.
2. "Predictive Ability and Economic Gains from Volatility Forecast Combinations", 14th Biennial Conference of the Athenian Policy Forum on Contemporary, Economic, Financial, Business and Policy Issues, University of Piraeus, 6-8 July 2018.
3. "Combining Energy Price Volatility Forecasts", 9th National Conference of the Financial Engineering and Banking Society (FEBS), Athens University of Economics and Business, 21-22 December 2018.

Publications

1. "Predictive Ability and Economic Gains from Volatility Forecast Combinations", under review to the Journal of Foorecasting (co-authored with V. Skintzi).

Table of Contents

List of Tables

COMBINING FORECASTS IN VARIOUS FINANCIAL APPLICATIONS

STAVROULA P. FAMELITI

University of Peloponnese, 2019

Supervisor: Vasiliki D. Skintzi, Assistant Professor in Econometrics

Abstract

The availability of numerous modeling approaches for volatility forecasting leads to model uncertainty for both researchers and practitioners. Accurate forecasts of volatility are required across most applications in finance such as risk management, portfolio allocation and option pricing. A large number of studies provide evidence in favor of combination methods for forecasting a variety of financial variables, but most of them are implemented on returns' forecasting. Surprisingly, combinations of volatility forecasts have not received significant attention in the finance literature. This thesis is focused on evaluating the predictive ability of simple and complex combination techniques as well as on developing and investigating innovative methods for combining volatility forecasts with applications in the stock and oil markets.

Firstly, combinations of various volatility forecasts based on different combination schemes of S&P500 index are provided. We add to the literature by combining volatility forecasts from models based on daily, intraday and implied volatility data. Moreover, an exhaustive variety of combination methods to forecast volatility ranging

from simple techniques to time-varying techniques based on the past performance of the single models and regression techniques is used. The evaluation procedure is based on both statistical and economic loss functions indicating the superior performance of combination techniques. Although combination forecasts based on more complex regression methods perform better than simple combinations and single models, there is no dominant combination technique that outperforms the rest in both statistical and economic terms, implying that different combination schemes are preferable based on the economic application to be used.

Secondly, we propose new combination techniques based on portfolio and risk management loss functions to forecast crude oil price volatility. The forecasting performance of three types of volatility forecast combination is evaluated: forecast combinations involving high-frequency models, forecast combinations involving daily models and forecast combinations involving both high-frequency and daily models. By considering combination techniques based on portfolio and risk management loss functions, new evidence may be drawn regarding the combination forecasts techniques. Firstly, the results show that most combination forecasts produce more accurate volatility forecasts in both statistical and economic terms than single volatility models. Secondly, daily data generate higher economic gains when they are combined through portfolio loss functions especially in 1-step and 22-step ahead forecast horizons, while two single models indicate superior forecasting performance for the 5- step ahead forecasts. Thirdly, statistical combination forecasts from high-frequency models are more accurate according to statistical and economic loss functions when they are compared with the economic combinations suggesting that the information contained in these data can adequately predict economic gains even through statistical combinations. Finally, the two information channels lead to higher

economic gains when they are combined through portfolio loss functions for the 22-step ahead forecasting horizon.

Keywords: volatility forecasting, combination methods, combining volatility forecasts, forecasting performance, statistical evaluation, economic evaluation, energy markets, economic combinations

COMBINING FORECASTS IN VARIOUS FINANCIAL APPLICATIONS

STAVROULA P. FAMELITI

University of Peloponnese, 2019

Supervisor: Vasiliki D. Skintzi, Assistant Professor in Econometrics

Abstract in Greek

Η διαθεσιμότητα ενός μεγάλου αριθμού υποδειγμάτων για την πρόβλεψη της μεταβλητότητας οδηγεί σε αυξημένη αβεβαιότητα για την αξιοπιστία του μοντέλου τους ερευνητές και τους επαγγελματίες. Οι ακριβείς προβλέψεις για την μεταβλητότητα απαιτούνται στις περισσότερες χρηματοοικονομικές εφαρμογές, όπως η διαχείριση κινδύνων, η διαχείριση χαρτοφυλακίου και η τιμολόγηση των δικαιωμάτων προαίρεσης. Ένας μεγάλος αριθμός μελετών υποστηρίζει την χρήση συνδυαστικών μεθοδολογιών για την πρόβλεψη ενός μεγάλου αριθμού οικονομικών μεταβλητών, όμως οι περισσότερες από αυτές εφαρμόζονται στην πρόβλεψη των αποδόσεων. Παραδόξως, οι συνδυαστικές μεθοδολογίες για τη μεταβλητότητα δεν έχουν λάβει την δέουσα προσοχή στη βιβλιογραφία. Η έρευνα αυτή επικεντρώνεται στην αξιολόγηση της προβλεπτικής ικανότητας απλών και πιο σύνθετων συνδυαστικών μεθοδολογιών, όπως και στην ανάπτυξη και διερεύνηση καινοτόμων μεθοδολογιών πρόβλεψης της μεταβλητότητας με εφαρμογές στις χρηματιστηριακές αγορές και τις αγορές πετρελαίου.

Πρώτον, παρέχονται διάφορες συνδυαστικές προβλέψεις για τη μεταβλητότητα του δείκτη S&P500. Ενισχύουμε την βιβλιογραφία συνδυάζοντας προβλέψεις για την

μεταβλητότητα από υποδείγματα που βασίζονται σε ημερήσια, ενδοημερήσια και τεκμαρτής μεταβλητότητας δεδομένα. Συγκεκριμένα, χρησιμοποιείται μια μεγάλη ποικιλία συνδυαστικών μεθοδολογιών για την πρόβλεψη της μεταβλητότητας, στην οποία περιλαμβάνονται απλές τεχνικές, τεχνικές που βασίζονται στις προηγούμενες επιδόσεις των μεμονωμένων μοντέλων και τεχνικές παλινδρόμησης. Η διαδικασία αξιολόγησης βασίζεται τόσο σε στατιστικά όσο και σε οικονομικά μέτρα αποδεικνύοντας την υπεροχή των συνδυαστικών μεθοδολογιών. Αν και οι συνδυαστικές προβλέψεις που βασίζονται σε πιο πολύπλοκες μεθόδους έχουν καλύτερες επιδόσεις από τις απλές συνδυαστικές μεθοδολογίες και τα μεμονωμένα μοντέλα, δεν υπάρχει κάποια μεθοδολογία που να κυριαρχεί τις υπόλοιπες τόσο από στατιστική όσο και από οικονομική άποψη. Αυτό μας οδηγεί στο συμπέρασμα ότι οι διαφορετικές συνδυαστικές μεθοδολογίες επιλέγονται με βάση την οικονομική εφαρμογή.

Δεύτερον, προτείνουμε νέες συνδυαστικές μεθοδολογίες που βασίζονται σε μέτρα αξιολόγησης χαρτοφυλακίων και διαχείρισης κινδύνου στην πρόβλεψη της μεταβλητότητας των τιμών των συμβολαίων μελλοντικής εκπλήρωσης αργού πετρελαίου. Αξιολογείται η προβλεπτική ικανότητα τριών τύπων συνδυαστικών μεθοδολογιών μεταβλητότητας: συνδυαστικές προβλέψεις που βασίζονται σε δεδομένα υψηλής συχνότητας, συνδυαστικές προβλέψεις που βασίζονται σε ημερήσια δεδομένα και συνδυαστικές προβλέψεις που περιλαμβάνουν και τα δύο είδη δεδομένων. Με την εξέταση των συνδυαστικών μεθοδολογιών που βασίζονται σε μέτρα αξιολόγησης χαρτοφυλακίου και διαχείρισης κινδύνου, μπορούν να αντληθούν νέα στοιχεία σχετικά με τις συνδυαστικές προβλέψεις. Πρώτον, τα αποτελέσματα δείχνουν ότι οι συνδυαστικές μεθοδολογίες οδηγούν σε πιο ακριβείς προβλέψεις, σε στατιστικούς και οικονομικούς όρους, για τη μεταβλητότητα από τα μεμονωμένα

μοντέλα. Δεύτερον, τα μοντέλα που βασίζονται σε ημερήσια δεδομένα οδηγούν σε μεγαλύτερα οικονομικά οφέλη όταν συδυάζονται μέσω μέτρων αξιολόγησης χαρτοφυλακίου ιδίως στους χρονικούς ορίζοντες της μίας και των είκοσι δύο ημερών, ενώ δύο μεμονωμένα μοντέλα επιδεικνούουν καλύτερη παρουσία για τον χρονικό ορίζοντα των πέντε ημερών. Τρίτον, οι συνδυαστικές πρόβλεψεις που βασίζονται σε στατιστικά κριτήρια και προέρχονται από μοντέλα υψηλής συχνότητας είναι πιο ακριβείς σύμφωνα με στατιστικά και οικονομικά μέτρα υποδεικνύοντας ότι η πληροφορία που εμπεριέχεται σε αυτά τα δεδομένα μπορεί να οδηγήσει σε σωστή πρόβλεψη για τα οικονομικά οφέλη ακόμη και μέσω συνδυαστικών προβλέψεων που προέρχονται από στατιστικά μέτρα. Τέλος, ο συνδυασμών των δύο πηγών δεδομένων οδηγεί σε υψηλότερα οικονομικά κέρδη, όταν τα δεδομένα αυτά συνδυάζονται μέσω μέτρων αξιολόγησης χαρτοφυλακίου, για τον χρονικό ορίζοντα των 22 ημερών.

Λέξεις-κλειδιά: πρόβλεψη μεταβλητότητας, συνδυαστικές μέθοδοι, συνδυαστικές προβλέψεις μεταβλητότητας, αξιολόγηση προβλέψεων, στατιστική αξιολόγηση, οικονομική αξιολόγηση, αγορές ενέργειας, οικονομικές συνδυαστικές προβλέψεις

List of abbreviations

AFTER	Aggregated Forecast through Exponential Re-Weighting
APARCH	Asymmetric Power Autoregressive Conditional Heteroscedasticity
CER	Certainly Equivalent Return
DMSFE	Discounted Mean Squared Forecasting Error
EGARCH	Exponential Generalized Autoregressive Conditional Heteroscedasticity
FIGARCH	Fractionally Integrated Generalized Autoregressive Conditional Heteroscedasticity
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
GARCH-MIDAS	Generalized Autoregressive Conditional Heteroscedasticity – Mixed Data Sampling
GJR-GARCH	Glosten-Jagganathan-Runkle Generalized Autoregressive Conditional Heteroscedasticity
HAR-RV	Heterogeneous Autoregressive Realized Volatility
HAR-RV-J	Heterogeneous Autoregressive Realized Volatility with discontinuous Jump variation
HAR-RV-CJ	Heterogeneous Autoregressive Realized Volatility with continuous sample variation
HAR-RSV	Heterogeneous Autoregressive Realized Semi-Variances
HRLF	Homogeneous Robust Loss Function
HYGARCH	Hyperbolic Generalized Autoregressive Conditional Heteroscedasticity
IMSFE	Inverse Mean Squared Forecasting Error
LHAR-RV	Leverage Heterogeneous Autoregressive Realized Volatility
MAE	Mean Absolute Error
MCS	Model Confidence Set
MRC	Market Risk Capital
MSE	Mean Squared Error
MSFE	Mean Squared Forecasting Error

LS	Least Squares Regression
RV	Realized Volatility
SPA	Superior Predictive Ability
SR	Sharpe Ratio
TW	Triangular Weighting
VaR	Value-at-Risk
VIX	CBOE Volatility Index

Chapter 1

Introduction

1.1 Motivation for the Thesis

Over the last decades forecasting the second moments of asset returns has been one of the most active areas in financial econometrics. A vast literature on methods and models for volatility forecasting has been developed. Though, there is rarely any consensus on which model is most appropriate in providing accurate forecasts. The main problem is that volatility, unlike returns, is unobserved even ex post. Therefore, it has to be proxied. Accurate forecasts of volatility are required in a number of applications such as asset pricing, capital allocation, risk management and option pricing. While a strand of the literature has attempted to identify the single best forecasting model in the context of financial applications, a number of studies in financial forecasting have applied combination techniques to aggregate numerous individual forecasts into a pooled model. Despite the large number of combination forecasting techniques developed in the literature, little attention has been paid in the combination of volatility forecasts. This thesis focuses on investigating the forecasting performance of combinations of individual volatility models.

The literature on volatility forecasting models is vast. The seminal papers of Engle (1982) and Bollerslev (1986) introduced the class of Generalized Autoregressive Conditional Heteroscedasticity (GARCH) volatility models that have been proved to improve forecasting performance while several extensions have been proposed e.g. EGARCH, GJR-GARCH, FIGARCH, APARCH, HYGARCH models among others. Similarly, Engle et al. (2013) propose a new class of GARCH models, the so-called GARCH-MIDAS models that decompose volatility into a short run and a long run component. The long-run volatility component is a slowly-decaying function either of realized volatility or macroeconomic variables. Furthermore, a considerable amount of studies have explored the availability of high-frequency data and several studies have shown that using realized volatility measures based on intraday data improve forecast performance (Andersen et al., 2007). Under a similar perspective, Corsi (2009) proposes the Heterogeneous Autoregressive (HAR) models of Realized Volatility (HAR-RV) considering different volatility components realized over different time horizons that lead to good forecasting performance. Andersen et al. (2007) proposed the HAR-RV-J models by adding the daily discontinuous jump variation to the HAR-RV model, while the HAR-RV-CJ model that decomposes realized volatility into continuous sample path variation and discontinuous jump variation was introduced. Patton and Sheppard (2011) developed the HAR-RSV model which assumes that positive and negative realized semivariances can have different predictive abilities for different time horizons.

Almost five decades of extensive research and promising applications starting from the seminal work of Bates and Granger (1969) provide theoretical support and empirical evidence on the benefits of forecast combinations. Clemen (1989) summarizes the literature on forecast combinations and concludes that combining forecasts of various economic and financial variables leads to increased forecast accuracy. Similar conclusions are reached by Aksu and Gunter (1992) based on macroeconomic variables and firm specific series, Makridakis and Hibon (2003) based on the so-called M3 competition, Stock and Watson (2003, 2004) across various economic and financial variables, Swanson and Zeng (2001) using US macroeconomic variables, Marsellino (2004) on a large set of European macroeconomic variables, Rapach et al. (2010) on equity premium prediction and Benavides and Capistrán (2012) based on Mexican peso-US dollar exchange rate. From a more theoretical point of view, Timmerman (2006) provides a theoretical justification for the success of combination methods. Following the success of combination methods on forecasting the first moments of economic or financial time series, the research question of whether combinations methods can also improve the forecasts of the second moments arises. Despite the importance of volatility forecasting and the wide variety of combination models developed, the earliest study in combining various volatility forecasts dates back to 2008 by Becker and Clements. Becker and Clements (2008), investigate the forecasting performance of combination forecasts on S&P500 index volatility, indicating the superior forecasting performance of combination techniques. More, Liu and Maheu (2009) based on high-frequency data; use the Bayesian model averaging technique to construct realized volatility and density forecasts concluding that Bayesian model averaging provides adequate density forecasts and modest improvements in volatility forecasting. In a similar framework, Patton and Sheppard (2009) combine individual realized volatility estimators through various loss functions concluding that none of the combined estimators can be out-performed by any individual estimator. Optimal combination procedures have been also applied to improve the accuracy of individual quantile forecasts. For instance, McAleer et al. (2011) examine simple deterministic Value-at-Risk (VaR) forecasts while Halbleib and Pohlmeier (2012) combine VaR forecasts based on the maximization of conditional coverage rates and the minimization of the distance between the population quantiles and VaR's combinations and conclude that optimal combinations improve VaR performance during turbulent periods. Alternative combination procedures for optimally combining individual VaR models have been developed by Tsiotas (2015), Opschoor et al. (2017), amongst others, through density combination forecasting. Following the large variety of single models and combination techniques, two questions arise. Firstly, are there any economic gains from combining volatility forecasts. Secondly, are there any economic gains from combining volatility forecasts through economic loss functions.

While a number of studies provide evidence in favor of combination forecasts, most of the existing literature evaluates the performance of combination forecasts based solely on statistical evaluation criteria. Only a limited number of studies have used economic loss functions to evaluate the forecasting performance of individual volatility forecasts. In the volatility forecasting literature, González-Rivera et al. (2004) compute volatility under several models, using closing prices of call options of the S&P 500 index and indicate the importance of economic loss functions in the evaluation procedure. Pierdzioch et al. (2008) evaluate the forecasting performance of several individual volatility models using economic loss functions such as the utility criterion and an option-based criterion. The first objective of this thesis is to apply both statistical and economic loss functions in the evaluation process of combination forecasts. The fact that volatility is not observable and thereby necessitating the use of a volatility proxy as an input in statistical loss functions may affect the ability of these loss functions in discriminating between forecasts. In contrast, economic loss functions overcome the problem of the volatility proxy bias and are related to the actual economic use of volatility forecasts. More specifically, Chapter 3 investigates the economic gains derived from combining volatility forecasts through different economic loss functions. In contrast to previous studies, we evaluate the economic importance of our results based on loss functions involving Value-at-Risk (VaR), VaR-based market risk capital (MRC), option pricing and utility gains. Moreover, we adopt the Superior Predictive Ability (SPA) test of Hansen (2005) and the Model Confidence Set (MCS, Hansen et al., 2011) to test whether the models with the smallest loss values significantly outperform alternative models.

The second research question is approximated through an application on energy price volatility, one of the most important inputs into macroeconometric, option pricing, and portfolio selection models. Although a large number of methods and models has been developed, it is not clear which model is most appropriate in providing accurate oil price volatility forecasts. Numerous surveys indicate the use of GARCH-class models (Sadorsky, 2006; Kang et al., 2009; Mohammadi and Su, 2010; Arouri et al., 2012; Hou and Suardi, 2012; Wei et al., 2014; Lux et al., 2016; Kleian and Wather, 2016; Charles and Darné, 2017), but there is no model that consistently dominates the others. Sadorsky (2006), for example, examines the volatility of WTI crude oil futures and finds that GARCH-class models beat the random walk model. Kang et al. (2009) found that the Component GARCH and the Fractionally Integrated GARCH models yield superior forecasting performance, while Wei et al. (2014) concluded that no model outperforms all the others across different loss functions. Mohammadi and Su (2010) examining several GARCH models, conclude that even the APARCH model outperforms the others, in most cases, there is no clear winner among the models.

Although GARCH-class models exhibit adequate forecasting performance for energy price volatility; they do not incorporate the whole-day volatility information. The seminal work of Andersen and Bollerslev (1998) led to a new method of measuring volatility, i.e. the Realized Volatility (RV), based on high-frequency data develops rapidly the models based on them. A growing body of literature is based on the HAR-RV model proposed by Corsi (2009) which is based on high-frequency datasets. Ma et al. (2017) argue that GARCH-class models are constructed from low frequency data that lead to a substantial loss of intraday trading information. They use Markov regime switching models to HAR-RV models indicating their superior forecasting ability. Degiannakis and Filis (2017) extend the HAR-RV models incorporating information channels from other assets that improve predictive ability across various forecasting horizons. Haugom et al. (2014) include the CBOE Crude Oil Volatility Index (OVX) and other market variables into HAR-RV models to forecast crude oil volatility, pointing out the significantly improvements. The first models improve the forecasting accuracy in shorter forecasting horizons (i.e. daily and weekly basis) and the second in all forecasting horizons (i.e. daily, weekly and monthly basis). Wen et al. (2016) extend the HAR-RV models that control for structural breaks, finding that different models perform good across different forecasting horizons. Sévi (2014) applies nine extensions of the HAR model in forecasting the oil futures realized volatility, pointing out that the simple HAR-RV model outperforms often more complicated models. Ma et al. (2018a, 2018b) found that models incorporating both large and small jumps significantly perform better than the single models.

While a number of studies provide evidence in favor of HAR-type and GARCH-type models, a considerable amount of literature points out that volatility forecasts from individual models are very unstable and change over time (e.g. Corsi et al., 2010; Ma et al., 2017 amongst others). However, few studies investigate the use of combination forecasts on the prediction of the oil price volatility. Particularly, only five studies exist to date. Lux et al. (2016) combined oil volatility forecasts computing the optimal weights through a forecast encompassing test, while Wei et al. (2017) and Ma et al. (2018c) used the Dynamic Model Averaging (DMA) approach that takes into account the historical performance of the competing models. Zhang and Ma (2018) based on high frequency data on oil futures combined several models pointing out their superior performance. Zhang et al. (2019) combined several HAR models using standard combination techniques and compared them with the forecasts derived from shrinkage methods of the elastic net and lasso regressions indicating the superior forecasting performance of the shrinkage methods. While a number of studies provide evidence in favor of combination forecasts, all the existing studies combine forecasts based solely on statistical criteria. The superior forecasting performance in both statistical and economic terms of combination models has been corroborated in Chapter 3. In Chapter 4, we introduce new combination techniques based on economic and risk management loss functions and evaluate their forecasting performance using statistical and economic evaluation measures. We expect to improve the forecasting accuracy in economic terms by combining volatility forecasts based on them. To the best of our knowledge there is no study that combines forecasts based on economic criteria to date.

To compute the combination weights, we use standard economic and risk management loss functions widely used in volatility forecasting literature. Ma et al. (2017), for example compute the volatility of WTI futures contract based on 5-minute returns and evaluate its economic performance through the *Sharpe* ratio. Zhang and Ma (2018) and Ma et al. (2018a, 2018b) compute the volatility of WTI oil futures and spot prices respectively using the *Certainly Equivalent Return* criterion due to Rapach et al. (2010) indicating the out-of-sample economic gains. We expect to use these economic loss functions and a risk management loss function proposed by González-Rivera et al. (2004) (i.e. the Smoothed-*Q* loss function) to optimally combine volatility forecasts of crude oil futures prices.

1.2 Contribution of the Thesis

This thesis is primarily concerned with the investigation of the forecasting performance of simple and more complex combination techniques as weel as on the development and use of new, more accurate combination volatility methods in financial applications. While the idea of combining volatility forecasts is not new, it still has been exploited little in the existing literature. To the best of our knowledge, there are a few papers that make reference to this topic (Li et al., 2013; Fuertes et al., 2009 and Wang et al., 2015 amongst others). In this thesis, two important aspects of combination volatility forecasts are examined. Firstly, a large variety of simple and more complex combination techniques in the context of volatility forecasting and the related economic gains of these techniques are investigated. Secondly, combination techniques based on economic and risk management loss functions are proposed and examined using statistical and economic measures.

Concerning the first contribution, various volatility forecasts based on different combination schemes are investigated (i.e. simple combinations, combinations based on the relative performace of the single models and combinations based on regression approaches). Forecasting the volatility of asset returns has received significant attention in the literature. The common volatility models seem to provide accurate forecasts during tranquil periods but not in turbulent periods. The instability in financial markets during the global financial crisis of 2007-2009 was characterized by extreme asset price movements and high volatility revealing the insufficiency of the existing single volatility models. Investors faced with extreme losses, while these losses underlined the need for accurate volatility forecasting.

A number of studies compare widely common used models (e.g. GARCH-class models due to Engle (1982) and Bollerslev (1986), the GARCH-MIDAS models by Engle et al. (2013) and the so-called HAR models of Corsi (2009)) yielding to a different best volatility forecasting model each time. However, the evidence about which is the best volatility forecasting model is far from being conclusive. The first part of this study examines the following question: "Does it worth combining volatility forecasts in both statistical and economic terms?". We address this question by examining whether combination forecasts based on three different class models can lead to higher volatility forecasting accuracy for the S&P500 index in Chapter 3. The three different class models include models based on low-frequency (daily) data, models based on high-frequency (5 minutes) data and option-based data (i.e. the VIX index). We expect the combination of different information channels, i.e. different datasets, to be more efficient than combining models based on the same dataset. Moreover, several statistical and economic loss functions widely used in the forecasting literature are used to evaluate their forecasting accuracy. The empirical evidence indicates that combination forecasts based on regression and more complex schemes lead to higher forecasting accuracy and economic gains than the single volatility models and the simple combination techniques. Our results are consistent with the literature on combining volatility forecasts as many studies suggest regression techniques to optimally combine volatility forecasts (Fuertes et al., 2009; Li et al., 2013; Yang et al., 2015 amongst others). The superior forecasting performance of such combinations can be attributed to three stylized facts. Firstly, different volatility models capture different market microstructures. Secondly, regression-based forecasts correct more properly any bias compared with the simple average that gives the same weight even to mispecified models. Thirdly, different information channels come from three different datasets, improve the pooled model. Though the main drawback of our study is that there is no clear winner across all loss functions suggesting that different combination techniques are preferable based on the economic application to be used. These results are important for investors or financial institutions. For instance, taking into consideration the instability in the performance of single forecasts and the fact that different loss functions are relevant for different decision makers, they are encouraged to use the combination forecasting methodology to improve forecasting accuracy and economic gains. On the other hand, if no such methodology exists, an investor or financial institution can use the existing volatility models that may lead to substantial losses due to the miscalculated volatility.

The second contribution of the thesis is the development and application of new combination methodologies based on economic and risk management loss functions. The main research question is: “Does it worth combining volatility forecasts through economic loss functions across different model-types and forecasting horizons?”. While most combination approaches take into consideration statistical measures or are based on simple combination techniques that have been found to work well and sometimes outperform more complex techniques (see Stock and Watson, 2004; Smith and Wallis, 2009 amongst others), Chapter 4 of this thesis proposes combination techniques based on economic and risk management loss functions in the context of crude oil price futures volatility. Since each model incorporates different aspects of volatility, they arguably contain useful information not included in a single volatility model. The proposed methodology has an important advantage compared to alternative combinations. It is based on economic and risk management measures that can lead to substantial economic gains as the optimal weights are computed through the relative performance of each model according to the economic loss functions based on a training period. To the best of our knowledge there is no other study in the literature that combines volatility forecasts using economic and risk management loss functions.

The economic combination techniques are based on the *Certainly Equivalent Return (CER)* measure, the *Sharpe* ratio and the Smoothed-*Q* loss function. More specifically, we extend the triangular weighting methodology due to Timmermann (2006) by substituting basic statistical loss functions with the aforementioned economic and risk management loss functions. We also include a new trimmed mean combination technique, where the combination forecast is calculated through the equal weighted average of all forecasts after trimming the one with the worst out-of-sample performance according to the economic and risk management measures. The last economic combination technique is based on an alternative combination strategy of Stock and Watson (2004) that places all the weight on the individual forecast that has the best post performance during the last period. We implement this strategy based on economic loss functions rather than statistical loss functions.

To examine the relative accuracy of the proposed methodologies, we implement three types of combinations: combinations based on GARCH models (i.e. low-frequency data), combinations based on HAR models (i.e. high-frequency data); and combinations based on both model types. The empirical results of this chapter propose that the best performing model is different for each combination type and forecasting horizon. Considering the GARCH combinations, our results suggest that combination techniques often outperform the single models in both statistical and economic terms. Interestingly, two of the proposed combination schemes (i.e. the combinations derived from the *CER* measure and the *Sharpe* ratio), increase the economic gains for an investor for the 1-step and 22-step ahead forecasts. The superior performance of these combination techniques is attributed to the fact that different volatility models capture different market microstructures (for example the FIGARCH model accounts for long memory and asymmetries while the EGARCH model responds in a different way to positive and negative returns). As a result, a combination based on an economic measure can increase the economic gains even if daily data are used. Turning to the HAR combinations, the combination techniques based on statistical loss functions indicate superior forecasting performance for both statistical and economic loss functions. This result is consistent for all the forecasting horizons examined. Although in some cases the economic combinations are ranked amongst the best performers, they seldomly outperform the statistical combinations, suggesting that the considerable amount of information contained in high-frequency data can increase the forecasting accuracy of several models when they are combined even in a statistical measure that corrects better for bias (e.g. regression techniques). When two information channels (i.e. the low-frequency and the high-frequency models) are used, the economic gains for an investor are maximized for longer forecasting horizons. Our analysis, however, suggests that a single model, the EGARCH model indicates superior forecasting accuracy under statistical and risk management evaluation in some cases. Although this result is not expected, a limited number of studies (Arouri et al., 2011; Wei et al., 2017) indicate that in some cases single models outperform combination techniques.

1.3 Overview of the Thesis

In this section a brief overview of the thesis is provided. Chapter 2 reviews the existing studies on the development and application of combination forecasts on economics. Firstly, simple combination techniques (i.e. Mean, Median amongst others) are presented. Secondly, a number of studies that include more complex combination techniques are presented. These methodologies include linear combinations based on least squares regression approaches as well as non-parametric techniques that take into consideration the ranking of the models over the last periods (for example the triangular weighting method due to Timmermann ,2006). Moreover, based on Yang's (2004) argument that linear combinations can lead sometimes to worse forecasting performance, nonlinear combinations are examined. Furthermore, methods based on more complex techniques such as the Bayesian Model Averaging and the Dynamic Model Averaging are discussed. Finally, the shrinkage and the iterated combinations are considered.

Chapter 3 of this thesis examines the benefits of combination techniques under an economic perspective. Firstly, numerous forecasting volatility models are combined using an exhaustive set of simple and statistical combination techniques. Secondly, we combine various volatility forecasts to forecast the volatility of S&P500 index based on different combination schemes and different model-types (i.e. we include GARCH models computed from daily observations, HAR models computed from intraday 5-minutes returns, and an implied volatility index). Thirdly, their forecasting accuracy is evaluated in both statistical and economic context. The methodology consists of simple combination methods such as the mean, the geometric mean, the harmonic mean and a transformed trimmed mean method and more complex combination methods. More specifically, the discounted Mean Squared Forecasting

Error (DMSFE) combination proposed by Diebold and Pauly (1987) and Stock and Watson (2004) is used. From a regression approach, we consider the linear least squares combinations proposed by Aksu and Gunter (1992). We also use Yang's (2004) nonlinear and AFTER combination model (Yang, 2004; Zou and Yang, 2004). Finally, we include nonparametric combinations such as the Kernel regression approach (Härdle, 1990), the triangular weighting (TW) method due to Timmermann (2006), and the Shrinkage Combination scheme based on Stock and Watson (2004). In contrast to previous studies, we evaluate the economic importance of our results based on loss functions involving Value-at-Risk (VaR), VaR-based market risk capital (MRC), option pricing and utility gains. Moreover, the Superior Predictive Ability (SPA) test of Hansen (2005) and the Model Confidence Set (MCS, Hansen et al., 2011) are adapted to test whether the models with the smallest loss values significantly outperform alternative models.

Chapter 4 introduces new combination techniques based on economic and risk management loss functions in the context of crude oil futures price volatility. The volatility forecasting models considered include several GARCH and HAR models. To compute these two type models, we use both daily and high-frequency data. Furthermore, we include to our calculations combinations derived from GARCH models (daily data), combinations derived from HAR models (high-frequency data) and combinations derived from both GARCH and HAR models. We include to our calculations simple combination methods (i.e. the mean, the median, the geometric mean, the harmonic mean and a transformed trimmed mean method), and more complex combination schemes based on statistical loss functions. Apart from a vast variety of methodologies based on statistical measures, we include three new methodologies based on economic and risk management loss functions. Particularly, the economic combinations are analyzed as: the TW-*CER*, the TW-*Sharpe*, the TW-*Q*, the Trimmed-*CER*, the Trimmed-*Sharpe*, the Trimmed-*Q*, the Best-*CER*, the Best-*Sharpe* and the Best-*Q*. Finally, we evaluate the economic significance of our results based on loss functions involving Value-at-Risk (VaR) and portfolio gains. Moreover, the Superior Predictive Ability (SPA) test of Hansen (2005) and the Model Confidence Set (MCS, Hansen et al., 2011), as well as a modified version of the SPA test proposed by Hsu et al. (2010) are adapted.

Finally, Chapter 5 presents the major conclusions of this thesis and suggests topics and areas of further research.

Chapter 2

Combination Forecasts - A Review

Forecasting the volatility of asset returns has become one of the most active areas in financial econometrics. Although, a vast literature on models and methods has been developed, it is not clear which model is most appropriate in providing accurate volatility forecasts. Instability in financial markets during the global financial crisis of 2007-2009 was characterized by extreme asset price movements and high volatility revealing the insufficiency of the existing single volatility methods. Investors faced with extreme losses, while these losses underlined the need for accurate volatility forecasting. While a strand of literature has attempted to identify the single best forecasting model in the context of financial applications, a limited number of studies in financial forecasting have applied combination techniques to aggregate numerous individual forecasts into a pooled model. Combining forecasts can reduce uncertainty risk with a single predictive model, while forecasts combinations are more robust to unknown instabilities (i.e. structural breaks). This chapter presents the existing studies to date concerning combination forecasts. Furthermore, stylized facts of combination forecasts are reported, while the existing methodologies of combination forecasts are presented.

2.1 Simple Combinations

Several studies have examined the benefits derived from combining different forecasts (Aksu and Gunter, 1992; Makridakis and Hibon, 2003; Stock and Watson, 2004; Swanson and Zeng, 2001; Marsellino, 2004; Benavides and Capistrán, 2012; Timmerman, 2006). In general, the combination forecast of s_t derived from a set of n individual volatility forecasts based on the information available at time $t\text{-}1$, f_{it}, is given by

$$f_t^c = g\left(f_{1t},..., f_{nt}; \mathbf{w} \right) \tag{1}$$

where g is a function (linear or non-linear) and $\boldsymbol{w}$ is a vector of parameters/weights. In the following chapters, we include every single model presented to all combination schemes, as each model is expected to add significant information to the combined model. With the exception of the first class of simple combination forecasts, all of the combination schemes allow for time-varying weights.

At first, simple combination forecasts including the mean, the harmonic mean, the geometric mean, the trimmed mean, and the median are considered. Smith and Wallis (2009) found that simple combinations of point forecasts outperform sophisticated weighted combinations using several models to predict US GDP and the industrial production index during the 2001 recession. Similarly, Stock and Watson (2004) using a large variety of combination approaches, found that the more sophisticated models perform worse than the simple combinations. Their results indicate that among the simple combination forecasts there seems to be little difference between the mean and the trimmed mean while the median combination produces the higher MSFE.

Recently, Wang et al. (2017) proposed a transformed trimmed mean combination computed as the equal-weighted average from single forecasts after trimming the one with the worst past performance. Similarly, Jose and Winkler (2008) proposed the winsorized mean, computed by averaging the forecasts of all models, winsorizing the p highest and the p lowest forecasts.

Palm and Zellner (1992) pointed out the advantages of the simple average forecasts. They argue that the combination weights do not have to be estimated that is an important advantage in cases that there is little evidence on the performance of single forecasts or if the parameters of the model generating the forecasts are time-varying. More, they conclude that a simple average of the forecasts can achieve significant reduction in variance and bias and finally that it often dominates forecasts based on more complex schemes in terms of MSE. Although the simple combination approaches seem to provide good forecasting performance, Genre et al. (2010) pointed out that there is no single combination approach which appears to dominate across either variable or at different horizons. Based on this argument, several more sophisticated combination approaches are examined.

2.2 Linear Time Varying Parameter Combinations

The importance of combination forecasts was firstly made apparent in Bates and Granger (1969) pioneering work on combination forecasts. The core problem for an investor is which the superior forecast to use is. The main assumption is that the minimized error variance is no greater than the smaller of the n individual forecast error variances. The second class of combination methods considered combines all the individual models to a linear combination of the form:

$$f_t^c = w_0 + \sum_{i=1}^{n} w_{it} f_{it} \quad (2)$$

where f_t^c is the one-step ahead combination forecast at time t, w_0 is the constant term if required, w_{it} are the combination weights and f_{it} the individual one-step ahead forecasts. In these combination schemes a holdout period for estimating the combination weights is required. The last q observations of the in-sample period are used as the initial holdout period. In the following methodologies, time-varying weights are considered in order to capture the dynamics of the examined series and weights depend on the historical performance of the single forecasts. The idea to use a regression approach for combining forecasts was firstly proposed by Granger and Ramanathan (1984). They argued that an advantage of the ordinary least squares (OLS) forecast combinations is that a combined forecast including an intercept is unbiased, even if one of the single forecasts is biased. However, a disadvantage of the proposed methodology is that it places no restriction on the combination weights, which complicates their interpretation. Aksu and Gunter (1992) propose a large variety of combination techniques based on a regression approach. More specifically, the parameters (i.e. weights) are recursively updated and computed through the least squares (LS) regression considering different combinations depending on the parameters' constraints and a pseudo-out-of-sample period. The proposed models are:

- OLS_c: Unrestricted LS regression with a constant term;
- OLS_{nc}: LS regression with no constant term;
- $ERLS_c$: LS regression with a constant term and the sum of weights equals unity;

- $ERLS_{nc}$: LS regression without a constant term and the sum of weights equals unity;
- $NRLS_{c}$: LS regression with a constant term, while the weights are positive;
- $NRLS_{pc}$: LS regression with a positive constant term, while the weights are positive;
- $NERLS_{c}$: LS regression with a constant term, while the sum of positive weights equals unity;
- $NERLS_{pc}$: LS regression with a positive constant term, while the sum of the positive weights equals unity;
- $NRLS_{c}$: LS regression without a constant term and positive weights;
- $NERLS_{nc}$: LS regression without a constant term, while the sum of the positive weights equals unity.

The constrained least squares combinations is sub-optimal compared to the OLS model of Granger and Ramanathan (1984) as they lack the asymptotic properties admitted by OLS. However, the weights are more easily interpretable while it often indicates better forecasting performance (Clemen, 1986; Weiss et al., 2018). Following Aksu and Gunter (1992); Becker and Clements (2008) combined volatility forecasts for the S&P500 index using both a mean and a regression approach concluding that a combination of model based forecasts is the dominant approach and that the implied volatility index is an inferior forecast of the S&P500 volatility. Next, Chan et al. (1999) and Stock and Watson (2004) considered the principal component analysis combination to estimate the static common factors from the panel of forecasts and regress a subset of these on the target variable. The combination forecast is based on the fitted k principal components of the uncentered second order matrix of the single model forecasts. The weights are estimated through the OLS regression. In this combination methodology, the number of the fitted *k* principal components is estimated through an information criterion (e.g. the AIC or the BIC). The projection on the mean combination was proposed as a superior forecasting methodology by Capistrán and Timmermann (2009). In this method the combination forecast is computed through a linear projection of the targer variable on the equally weighted forecast.

Ma et al. (2018c) following Tibshirani (1996) combined forecasts through the LASSO regression. In this regression approach, the coefficients for the variables that are not significantly important take values of zero, making a simpler model by focusing on variables that are strong predictors without increasing bias. More, the variance of the OLS estimates is higher when the number of observations is small and the predictors' number is large. Granger and Jeon (2004) proposed the "thick modeling approach" that removes the models with the worst past performance. In this methodology, the poorly performing models are removed in a step that precedes the calculation of combination weights, while the optimal weights are calculated through a least squares regression.

Similarly, Hansen (2008) proposed a forecast combination based on the method of Mallows Model Averaging (MMA), where the combination weights are calculated through the minimization of the unbiased estimate of both the in-sample MSE and the out-of-sample one-step-ahead MSFE. Amendola and Storti (2008) combined volatility forecasts using several GARCH-class models for the S&P500 daily returns using the Generalized Method of Moments (GMM) to estimate the combination weights, concluding that the proposed combination algorithm can be considered as a useful tool for risk management applications and financial modeling. More, they propose to use their application to multi-step ahead forecasts. Gurung et al. (2017) combined forecasts using a linear regression approach. However, they estimated the coefficients through the Kalman Filter[1] that ensures the minimization of the forecasts error variances. Differently, Patton and Sheppard (2009) combined forecasts to predict realized volatility estimators concluding that none of the single models outperform the forecasting accuracy of the combinations.

One disadvantage of regression approaches is that many models do not guarantee the positiveness of the estimated weights that is required for volatility estimation. In order to deal with cases of negative volatility Amendola and Storti (2016) proposed two transformations to models that are not restricted to provide positive volatility forecasts[2]. To guarantee the positive definiteness of the variance, an exponential transformation is used computed as:

$$f_t^c = \exp\left(w_{0t} + \sum_{i=1}^{n} w_{it} f_{it} \right) \quad (3)$$

Alternatively, the square root transformation on the estimations derived from the OLS models is computed as:

$$f_t^c = w_0 + \sum_{i=1}^{n} w_{it} \sqrt{f_{it}} \quad (4)$$

Wei et al. (2017) and Ma et al. (2018c) combined volatility forecasts for oil futures through the Dynamic Model Averaging (DMA) approach proposed by Raftery et al. (2010). The DMA approach allows both set of predictors (i.e. the forecasting models

[1] The Kalman Filter is a set of mathematical equations that provides a powerful way to estimate the unknown state of a process, while it is recursively used to estimate the forecasts and the forecasts error variances. In particular, the next observation is estimated using the previous real observation and the estimate of the previous observation.

[2] The transformations are implemented to the following models: OLS_c, OLS_{nc}, $ERLS_c$, $ERLS_{nc}$, $NERLS_c$.

used in a combination method) and their coefficients to vary over time. The DMA combination is described as:

$$f^{c}_{t} = \sum_{i=1}^{n} \pi_{(t,t-1,i)} f^{(i)}_{t-1} w^{(i)}_{t-1} \tag{5}$$

where $f^{(i)}_{t-1} \subseteq f_{t-1}$ for $i = 1,2,...,n$ denotes the specific predictor set and $\pi_{(t,t-1,i)} = \Pr\left(L_t = i \mid f^{t-1}\right)$. The equation $L_t = i$ indicates the model that is selected at time *t*. The DMA approach computes its forecast by taking the average of all the *i* models in terms of their historical forecasting performance (i.e. $\pi_{(t,t-1,i)}$).

An alternative combination approach consists of Bayesian Model Averaging. Let's consider *M* forecasting models with a prior probability of each model equal to $p\left(M_i\right)$ the prior distribution of the parameters in each model equal to $p\left(\theta_i \mid M_i\right)$ and the likelihood function $L\left(f_i \mid \theta_i, M_i\right)$. The posterior probabilities of the models are based on Bayes rule:

$$p\left(M_i \mid f_i\right) = \frac{m\left(f_i \mid M_i\right) p\left(M_i\right)}{\sum_{i=1}^{M} m\left(f_i \mid M_j\right) p\left(M_j\right)} \tag{6}$$

where the marginal likelihood of model M_i is $m\left(f_i \mid M_i\right) = \int L\left(f_i \mid \theta_i, M_i\right) p\left(\theta_i \mid M_i\right) d\theta_i$.

The expected forecast based on the posterior probabilities of the model is equal to:

$$\hat{f}^{c}_{t} = E\left(f^{c}_{t} \mid f_i\right) = \sum_{j=1}^{M} E\left(f^{c}_{t} \mid f_t, M_j\right) p\left(M_j \mid f_t\right) \tag{7}$$

Although the methodology leads to optimal forecasts, conditional on the true model being included in the set of models, it faces the problem that the forecast combination is adversely affected by in-sample overfitting of the data.

2.3 Relative Performance Combinations

Apart from the linear time varying parameter combinations, several combination techniques based on basic statistical measures and the ranking of a model during a pseudo-out-of-sample period have been proposed. Diebold and Pauly (1987) and Stock and Watson (2004) proposed the discounted Mean Squared Forecasting Error (DMSFE) combination forecast. The recursively updated weights are computed through:

$$w_{it} = \frac{m_{it}^{-1}}{\sum_{j=1}^{n} m_{jt}^{-1}} \tag{8}$$

where $m_{it} = \sum_{s=T_0}^{t-h} \delta^{t-h-s} \left(RV_{s+h} - f_{i,s+h} \right)^2$ and δ is the discount factor, assumed to equal 0.9 in this case, and h is the forecast horizon. Recently, the inverse MSFE (IMSFE) combination scheme by simply imposing $\delta = 1$ was proposed by Baumeister and Kilian (2015).

A related combination approach is the most recently best which places all the weight on the single forecast that has the lowest average squared forecast error over the previous four periods (Stock and Watson, 2004).

Kolassa (2011) proposed two combination approaches based on typical statistical information criteria, the Akaike Information Criterion and the Bayes Information

Criterion. He found that the proposed combination methodologies lead to higher forecasting performance for short-range forecasts. The combination weights are inversely proportional to the models' information criterion:

$$w_{it} = \frac{\exp\left(-\frac{1}{2}\Delta_{IC(i)}\right)}{\sum_{i=1}^{n}\exp\left(-\frac{1}{2}\Delta_{IC(i)}\right)} \tag{9}$$

where $\Delta_{IC(i)} = IC(i) - IC(k)$, $IC(i)$ is the information criterion (either the AIC or the BIC) of the model *I* and $IC(k)$ is the minimum information criterion of model *k*.

Swanson and Zeng (2001) proposed combination forecasts based on an Heteroscedasticity and Autocorrelation-Consistent (HAC) *t*-statistics as well as combinations based on AIC, BIC and MSE. In particular, the best combination is pared down by eliminating the forecasts whose weights are insignificant based on the HAC statistic concluding that these combinations dominate forecasts from either individual or other combinations. Tsangari (2007) following Lupoletti and Webb (1986) combined forecasts based on the minimization of the MSE. She argued that since the errors have zero means and are uncorrelated, the best choice for the weights is a ratio of the sums of squared errors.

From a different point of view, Granger and Pesaran (2000) and Patton and Timmermann (2007) concluded that combination forecasts based on symmetric loss functions (i.e. the combinations where the weights are computed through the minimization of the MSE) can lead to serial correlation in the forecast error at the single-period forecast horizon and to the increase of the forecast error variance as the horizon grows. Based on these arguments Timmermann (2006) considered the LINEX loss function as an alternative asymmetric loss function that washes out bias in individual forecasts.

2.4 Non-linear Combinations

The forth class of combination models builds on Yang's (2004) argument that linear combinations can sometimes lead to worse out-of-sample performance compared to single models. He argues that this happens if and when the forecasts are strongly collinear. Thus, Yang (2004) proposed a nonlinear combination where the weights are computed as a function of the mean squared error as follows:

$$w_{it} = \frac{\pi_i \exp\left(-\lambda \sum_{k-1}^{t-1} \left(RV_k - f_{ik} \right)^2 \right)}{\sum_{j=1}^{n} \left(\pi_j \exp\left(-\lambda \sum_{k-1}^{t-1} \left(RV_k - f_{jk} \right)^2 \right) \right)} \tag{10}$$

For simplicity, he supposes $\pi_i = \lambda = 1$. This forecasting procedure ensures that combinations achieve a performance similar to that of the best single forecasting model up to a constantly penalty and a proportionality factor.

Moreover, Yang (2004) and Zou and Yang (2004) proposed the Aggregated Forecast Through Exponential Re-weighting (AFTER) model, as a modified version of the Yang's model. The weights of each single forecast are recursively updated and obtained according to:

$$w_{it} = \frac{v_{t-1,i}^{-1/2} \exp\left(-\frac{\left(RV_t - f_{it} \right)^2}{2 v_{t-1,i}} \right) w_{t-1,i}}{\sum_{i=1}^{n} v_{t-1,i}^{-1/2} \exp\left(-\frac{\left(RV_t - f_{it} \right)^2}{2 v_{t-1,i}} \right) w_{t-1,i}} \tag{11}$$

where $w_{1,i} = 1/n$, $v_{t-1,i} = \frac{1}{t-1}\sum_{s=1}^{t-1}\left(RV_t - f_{it}\right)^2$, $0 \le w_{it} \le 1$, $\sum_{i=1}^{n} w_{it} = 1$. A modified version of the AFTER combination, including a forgetting factor in order to adapt the combination more quickly to different situations was proposed by Sánchez (2008). Cheng and Yang (2015) proposed also another modified version of the AFTER combination that ensures for outlier protection through a synthetic loss function indicating the advantages of the new method by providing combined forecasts with fewer large forecast errors.

In this combination class is also included the basic combination approach of Bates and Granger (1969). This technique assigns weights depending on the inverse of means squares error prediction errors. The optimal weight is:

$$w_{it} = \frac{\left(\sum_{t=T_0}^{t-h}\left(RV_{s+h} - f_{i,s+h}\right)^2\right)^{-1}}{\sum_{j=1}^{n}\left(\sum_{t=T_0}^{t-h}\left(RV_{s+h} - f_{j,s+h}\right)^2\right)^{-1}} \qquad (12)$$

where $0 \le w_{it} \le 1$ and $\sum_{i=1}^{n} w_{it} = 1$.

2.5 Non-parametric Combinations

Timmermann (2006) argued that linear and nonlinear combinations require stationarity at least for the time involved in the estimation, as well as a large data sample in order to be robust to outliers. However, the non-parametric combination methods take into account the ranking of each model based on its forecasting performance (often measured by the MSFE) up to time *t*. More specifically, Timmermann's (2006) triangular weighting (TW) method is less sensitive to outliers, while the weights are expected to be more robust than the weights derived from the

DMSFE methodology. The combination weights are inversely proportional to the model's rank:

$$w_{it} = \frac{R_{i,t}^{-1}}{\sum_{i=1}^{n} R_{i,t}^{-1}} \tag{13}$$

where $R_{i,t}$ is the ranking of model *i* based on its MSFE up to time *t*. One advantage of the methodology is that any correlations across forecasts are ignored. Moreover, Timmermann (2006) argues that this combination forecast generally leads to increased forecasting accuracy as the information contained in each single model is combined; it averages across differences in the way single forecasts are affected by structural breaks; and it is less sensitive to possible misspecification of single forecasting models.

Another combination technique is the Spread Combination proposed by Aiolfi and Timmermann (2006) that assumes weights of the form:

$$w_{it} = \begin{cases} \frac{1+\bar{w}}{aN}, \; if \; R_{it} \leq aN \\ 0, if \; aN < R_{it} < (1-a)N \\ \frac{-\bar{w}}{aN}, if \; R_{it} \leq (1-a)N \end{cases} \tag{14}$$

where α is the proportion of the top models that gets a weight equal to $\frac{1+\bar{w}}{aN}$ based on the performance up to time *t*. Similarly a proportion of the models gets a weight equal to $\frac{-\bar{w}}{aN}$. The larger the value of α, the wider the set of the top and bottom models used in the combination. The larger the value of $\bar{w}$, the bigger the difference in weights on top and bottom models.

Similarly, Härdle (1990) introduced the non-parametric Kernel regression approach with time-varying parameters, while Tsangari (2007) indicates the superior performance of the Kernel regression approach through an application on exchange rates. The Nadaraya-Watson kernel weights are given by:

$$w_{it} = \frac{K_l\left(RV_t - f_{it}\right)}{T^{-1}\sum_{t=1}^{T} K_l\left(RV_t - f_{it}\right)} \tag{15}$$

where $K_l\left(u\right) = l^{-2}K\left(u/l\right)$, $u = \left(u_1, u_2\right)$ and $K\left(u\right) = K\left(u_1\right)K\left(u_2\right)$ is the product of each individual univariate kernel *K*. The shape of the Kernel weights depends on the kernel *K* and the size of the weights is parameterized by *l*, the bandwidth.

2.6 Shrinkage and Iterated Combinations

The next combination approach is based in a portfolio application by Ledoit and Wolf (2003) who propose to shrink the weights towards a point implied by a single factor structure. Diebold and Pauly (1990) and Stock and Watson (2004) proposed the shrinkage combination where the weights shrunk the combination weights to the equal weight solution so the combination gives a convex combination of the least-squares and equal weights. The weights are shrunken linearly toward the equal weight solution and computed as follows:

$$w_{it} = \lambda\hat{\beta}_{it} + (1-\lambda)(1/n) \tag{16}$$

where $\lambda = \max\{0, 1-\kappa[n/(t-h-T_0-n)]\}$, κ is a constant that controls the amount of shrinkage toward equal weighting and $\hat{\beta}_{it}$ is the *i*th estimated recursive coefficient from the combination[3].

Finally, Lin et al. (2017) and Zhang et al. (2018) proposed an advanced combination approach, the iterated combination, which is based on standard combination forecasts and a simple benchmark forecast. They argued that the iterated combination forecasts can generate a smaller forecast error relative to the standard combination forecasts. The specification for the iterated combination is:

$$f_t^{c'} = (1-\lambda) f_{it,B} + \lambda f_t^c + \varepsilon_t \qquad (17)$$

where λ is the restricted regression coefficient estimated through the restricted least-squares regression, $f_{it,B}$ is the benchmark forecast on day *t*, and f_t^c is a standard combination forecast on day *t*.

Concluding the major advantage of combining volatility forecasts is that it has been found to be a successful alternative to using just a single model. However, there is no combination technique that dominates the others. In Chapter 3, we examine a large variety of the presented combination techniques through an application on the volatility of the S&P500 index. More, in Chapter 4, new combination methodologies based on economic and risk management loss function are proposed and evaluated through an application to crude oil futures price volatility.

---

[3] The Shrinkage technique is applied to all the combination models that do not contain a constant term, i.e. the DMSFE, the OLS_{nc}, the $ERLS_{nc}$, the $NRLS_{nc}$, the $NERLS_{nc}$, the IMSFE, the Nonlinear and the Triangular Weighting, since the weights are computed as an average of the estimated weights and equal weighting.

Chapter 3

Predictive Ability and Economic Gains from Volatility Forecast Combinations

The availability of numerous modeling approaches for volatility forecasting leads to model uncertainty for both researchers and practitioners. A large number of studies provide evidence in favor of combination methods for forecasting a variety of financial variables, but most of them are implemented on returns' forecasting and evaluate their performance based solely on statistical evaluation criteria. In this chapter, we combine various volatility forecasts based on different combination schemes and evaluate their performance in forecasting the volatility of S&P500 index. We use an exhaustive variety of combination methods to forecast the volatility ranging from simple techniques to time-varying techniques based on the past performance of the single models and regression techniques. Then, we evaluate the forecasting performance of single and combination volatility forecasts based on both statistical and economic loss functions. The empirical analysis yields an important conclusion. Although combination forecasts based on more complex methods perform better than the simple combinations and single models, there is no dominant combination technique that outperforms the rest in both statistical and economic terms.

This chapter is organized as follows. Section 3.1 reviews the use of combination forecasts on stock market volatility. Section 3.2 describes the single volatility models used in this application. Section 3.3 presents the combination techniques used in this chapter. Section 3.4 describes the statistical evaluation process of the out-of-sample performance of volatility forecasts, while Section 3.5 presents the economic evaluation process. In Section 3.6 the tests of forecasting performance (i.e. the Superior Predictive Ability (SPA) and the Model Confidence Set (MCS)) are presented. Section 3.7 presents the data and the results of the current study. Finally, Section 3.8 presents some concluding remarks.

3.1 Combination forecasts and Stock Market Volatility

Forecasting stock market volatility is an important and challenging task for both academics and practitioners. Volatility prediction is the key variable in forecasting the prices of stocks and, in general, the risk that investors face. However, a single forecasting model is likely to result in worse forecasting performance because of model misspecification. A recent trend to improve forecasting accuracy is to combine single forecasts under various statistical loss functions or simple combination techniques. Given the importance of volatility forecasting, a limited number of studies have examined the benefits derived from forecast combination in terms of stock market volatility as it is noticed by Poon and Granger (2003).

Amendola and Storti (2008) combined volatility forecasts for S&P500 index using the Generalized Method of Moments (GMM) imposing conditions on the standardized residuals implied by a given set of combination weights. Their methodology has been found to provide adequate results for the one-step-ahead prediction of volatility as it minimizes the forecasting error. Furthermore, Amendola and Storti (2016) extend their research by applying exponential and square root transformations to combination volatility forecasting, concluding that the results are very sensitive to the choice of model and the combination strategy.

On the other hand, Becker and Clements (2008) combine volatility forecasts for the S&P500 index and compare the forecasting ability among several single models and the so-called VIX index. Using two combination approaches (i.e. the mean combination and a regression approach), they indicate their dominance as the combinations capture different dynamics in volatility. Claessen and Mittnik (2002) combined volatility forecasts using several GARCH models and an implied volatility index for the DAX index concluding that the combination of two sources of information improves the forecasting accuracy.

Jing-Rong et al. (2011) combined stock market volatility forecasts using least squares regressions under different parameter constraints and a regime switching approach driven by a latent variable, concluding that the regime switching combination approach has a better forecasting accuracy than the single models. In a similar framework, Fuertes et al. (2009) and Li et al. (2013) concluded that combination forecasts derived from a regression approach is better than an equally weighted approach as time-varying weights take into consideration the market's bias and reduce model uncertainty.

Furthermore, Yang et al. (2015) implemented several combination techniques to forecast the volatility of Shanghai Stock Exchange Composite Index and five sectoral indices. The results indicated that a non-parametric Kernel regression and a non-negative restricted least squares regression are the best performing models for forecasting realized volatility under structural breaks. A different approach was adapted from Wang et al. (2015). Using a large number of HAR-RV models, they combined forecasts to predict the volatility of the S&P500 index under the Dynamic Model Averaging (DMA) approach and extended methods based on DMA. More, they evaluated their forecasting accuracy based on statistical measures and portfolio measures, concluding that combinations lead to more accurate forecasts in both statistical and economic terms.

Ma et al. (2018d) combined volatility forecasts derived from high-frequency and low-frequency datasets. In total, they implement three types of combinations, combinations based on high-frequency data, combinations based on low-frequency data and combinations based on both datasets. The results show that combinations based on GARCH models lead to higher forecasting accuracy, while the combinations based on HAR models do not surpass the forecasting ability of the single models. However, incorporating both GARCH and HAR models to combinations lead to higher forecasting accuracy.

Analyzing the properties of combination forecasts in Chapter 2, and motivated by their adequate forecasting performance in stock market volatility, we employ several combination schemes based on both daily and high-frequency datasets.

3.2 Single Volatility Models

Several models have been proposed to forecast volatility, but there is no agreement regarding which model is superior in terms of out-of-sample forecasting accuracy. Firstly, we analyze several volatility models to capture a number of stylized facts in volatility behavior such as asymmetry, long-memory and persistence, and then numerous combinations based on these models are considered.

The return process is usually represented as:

$$r_t = \mu + \varepsilon_t \tag{18}$$

where μ is a constant mean and $\varepsilon_t = s_t z_t$ is the innovation term with $z_t \sim N(0,1)$ and s_t is the conditional volatility.

We divide the total sample of T observations into an in-sample portion composed of the first T_0 observations and an out-of sample portion of T_1=T-T_0 observations. We generate out-of sample volatility forecasts from each of the following described models using a rolling estimation window.

The first specification consists of the so-called GARCH (*p*,*q*) model due to Bollerslev (1986) considered as sufficient state of the art in volatility forecasting. In its general form a GARCH model is presented as:

$$s_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-1}^2 + \sum_{i=1}^{p} \beta_i s_{t-1}^2 \tag{19}$$

where s_t^2 is the conditional variance, α_0, α_i and β_i are parameters that have to be positive, while $a_i + \beta_i \leq 1$, *p*, *q* are the lags of the ARCH and GARCH terms respectively. In our application, we use the GARCH (*1,1*) model. To account for potential asymmetric response of volatility on positive and negative innovations, Nelson's (1991) exponential GARCH (EGARCH) model is considered. The EGARCH model adopts the following volatility process:

$$\ln s_t^2 = \alpha_0 + \sum_{i=1}^{q} \left(\alpha_i \left| \frac{\varepsilon_{t-i}}{s_{t-i}} \right| + \gamma_i \frac{\varepsilon_{t-i}}{s_{t-i}} \right) + \sum_{i=1}^{p} \beta_j \ln s_{t-j}^2 \qquad (20)$$

where the leverage effect is captured if $\gamma_i < 0$.

Another extension of the GARCH model that captures the long memory effect is the Fractionally Integrated Model (FIGARCH model) proposed by Baillie et al. (1996) and described as:

$$s_t^2 = \alpha_0 + \beta(L) s_{t-1}^2 + \left[1-\beta(L)\right]\varepsilon_{t-1}^2 - \varphi(L)(1-L)^d \varepsilon_{t-1}^2 = \alpha_0 \left[1-L\right]^{-1} + \lambda(L)\varepsilon_{t-1}^2$$

(21)

where L is the lag operator and $(1-L)^d$ is the fractional differencing operator.

Another strand of the literature assumes that the information contained in macroeconomic variables provides useful information for estimating and forecasting volatility. Engle et al. (2013) propose a new class of volatility models, the GARCH-MIDAS models that incorporate economic fundamentals into a volatility model. In particular, a daily GARCH process and a MIDAS polynomial applied to monthly, quarterly or bi-annual macroeconomic or financial variables are used. We consider two versions of the model. In the first specification, the long term component is based on realized volatility over a different time basis (i.e. monthly), while in the second version the long term component links to a macroeconomic variable (i.e. the inflation rate and the industrial production index)[4]. A GARCH-MIDAS model with a rolling window of the Realized Volatility or a macroeconomic variable is described as:

$$s_t = g_t m_\tau \tag{22}$$

$$g_t = (1-\alpha-\beta) + \alpha \frac{(r_{t-1}-\mu_{t-1})^2}{m_\tau} + \beta g_{t-1} \tag{23}$$

$$\log(m_\tau) = m + \theta_i \sum_{k=1}^{K} \phi_k(\omega) X_{\tau-k} \tag{24}$$

where g_t is the short-run and m_τ the long-run volatility component, X_τ is either the Realized Volatility or a macroeconomic variable, and φ_k is the weighting scheme that can be either Beta or Exponentially weighting.

In our application the monthly realized volatility, the inflation rate and the industrial production index growth rate are used and a restricted beta weighting scheme is adopted where the weights are computed as follows.

$$\phi_k(\omega) = \frac{(1-k/K)^{\omega-1}}{\sum_{l=1}^{K}(1-l/K)^{\omega-1}}, \quad k = 1,\ldots,K \tag{25}$$

Numerous studies indicate that models based on high frequency data achieve better forecasting performance compared to conventional models based on daily returns (Martens, 2002; Martens and Zein, 2004). First, we compute the Realized Volatility (RV) as the sum of squared intraday returns, i.e.:

[4] Macroeconomic variables can be expressed in level or volatility basis, or in both. In our case, we use the level basis.

$$RV_t = \sum_{t=1}^{m} r_{t,m}^2 \tag{26}$$

Then, we employ three frequently used models directly applied on the realized volatility series. The first model is Corsi's (2009) Heterogeneous AutoRegressive model of Realized Volatility (HAR-RV) that encompasses different time aggregations of RV (i.e. the RV of the previous day, the weekly RV and the monthly RV). The HAR model is represented from the following equation:

$$RV_t^2 = \beta_0 + \beta_d RV_{t-1,d} + \beta_w RV_{t-1,w} + \beta_m RV_{t-1,m} + u_t \tag{27}$$

where $RV_{t-1,d}$ is the daily realized volatility, $RV_{t-1,w}$ is the weekly realized volatility and $RV_{t-1,m}$ is the monthly realized volatility. An autoregressive moving average, ARIMA(p,d,q) model is also employed on the realized volatility series as follows:

$$\Phi(L)\Delta^d (RV_t - \mu) = \Theta(L) u_t \tag{28}$$

where Φ and Θ are polynomials of orders p and q, Δ is the difference operator and d is the order of integration.

Taking into consideration the observed strong serial dependence of realized volatility, we finally consider the autoregressive fractionally integrated moving average ARFIMA (p,d,q) model given by:

$$\Phi(L)(1-L)^d (RV_t - \mu) = \Theta(L) u_t \tag{29}$$

where $(1-L)^d$ is the fractional integration operator that captures the long-term dependence of the series. If $d \in (0, 0.5)$, the process exhibits long memory, while if

$d \in (-0.5, 0)$, it exhibits antipersistence. The ARFIMA model achieves long memory in a parsimonious way by imposing a set of infinite-dimensional restrictions on the infinite variable lags.

Finally, we consider the VIX index, the core index for US equities that reflects the market view on the expected volatility by averaging the weighted prices of put and call options on the S&P500 index over a wide range of strike prices (CBOE, 2014). As the VIX index is annualized, we turn it to a daily basis, according to Kambouroudis et al. (2016):

$$VIX_d = \left[\left(\frac{VIX}{\sqrt{365}} \right) * 100 \right]^2 \qquad (30)$$

3.3 Combination Techniques

Stock and Watson (2004) argued that the predictability of a single model is very unstable and changes over time. Consequently, we use a large variety of combination techniques presented in Chapter 2. In particular, we include to our calculations the following combinations; the Mean, the Geometric Mean, the Harmonic Mean, the linear least regressions and their transformations depicted from the equations (2), (3) and (4). Also we consider the Nonlinear combination (equation (10)) by Yang (2004) and the AFTER combination (equation (11)). More, the DMSFE and the IMSFE based on equation (8) are implemented. Finally, the non-parametric Kernel regression, the TW and the Shrinkage combinations based on equations (15), (13) and (16) are included to our analysis.

3.4 Statistical Evaluation Measures

To evaluate the relative accuracy of the proposed methodologies, we rank the models using several statistical loss functions. We consider realized volatility (RV) as a proxy for the "true" but unobserved volatility. From a statistical point of view, we use, at first, statistical loss functions widely used in the existing forecasting literature such as the mean absolute error (MAE) and the mean squared error (MSE). The loss functions are defined as:

$$MAE^i = (\mathrm{T}-\mathrm{T}_0)^{-1}\sum_{t=\mathrm{T}_0+1}^{T}\left|f_{it}-RV_t\right| \tag{31}$$

$$MSE^i = \left(T-\mathrm{T}_0\right)^{-1}\sum_{t=\mathrm{T}_0+1}^{T}\left(f_{it}-RV_t\right)^2 \tag{32}$$

A challenging issue in volatility forecasting is that the over-prediction of volatility does not have the same practical implications with the under-prediction of volatility in various financial applications and asymmetric loss functions are required. For example in risk management applications under-prediction of volatility is far more severe that over-prediction. Moreover, in option pricing, a buyer (seller) of an option might want to penalize more heavily volatility over (under)-prediction. In the class of asymmetric loss functions, the linear exponential (LINEX) loss function proposed by Varian (1975) and Zellner (1986) is considered. The LINEX measure is computed as:

$$L^i_{LINEX} = \left(RV_t, f_{it};\alpha\right) = \left(T-T_0\right)^{-1}\sum_{t=T_0+1}^{T}\left(\exp\left[a\left(f_{it}-RV_t\right)\right]-\alpha\left(f_{it}-RV_t\right)-1\right) \tag{33}$$

where α is a scalar parameter that controls for the degree of asymmetry. We consider only positive values for the scalar parameter that penalize more under-prediction than over-prediction. In the same context, we include the Homogeneous Robust Loss Function (HRLF) proposed by Patton (2011) and given by:

$$L_{\mathrm{HR}}^{t}\left(RV, f_i; b\right)=\begin{cases}\left(T-T_0\right)^{-1}\sum_{t=T_0+1}^{T}\left[\frac{1}{(b+1)(b+2)}\left(RV_t^{2b+4}-f_{it}^{b+2}\right)-\frac{1}{b+1}f_i^{b+1}\left(RV_t-f_{it}\right)\right], for\ b\notin\{-1,-2\}\\ \left(T-T_0\right)^{-1}\sum_{t=T_0+1}^{T}\left(f_{it}-RV_t+RV_t\log\frac{RV_t}{f_{it}}\right), for\ b=-1\\ \left(T-T_0\right)^{-1}\sum_{t=T_0+1}^{T}\left(\frac{RV_t}{f_{it}}-\log\frac{RV_t}{f_{it}}-1\right), for\ b=-2\end{cases} \quad (34)$$

where the parameter b controls the shape of the function. This loss function nests two other functions, the MSE when *b=0*, and the QLIKE when *b=-2*. Similar to the LINEX loss function, in the empirical application we focus on negative values of the parameter b that penalize more volatility under-prediction than over-prediction.

3.5 Economic Evaluation Measures

In order to assess the performance of volatility forecasts in economic terms, we consider a number of economic loss functions based on various financial applications and compute the predictive gains derived from the used methods. Firstly, following González-Rivera et al. (2004), a utility loss function is considered to compare the predictive performance of the volatility models. Based on the fact that a risk averse agent has lower expected utility when the conditional variance is underestimated than overestimated, the loss function is defined as:

$$U^i=-\left(T-T_0\right)^{-1}\sum_{t=T_0+1}^{T}\left(c_t(\gamma)+d_t(\gamma)x\left(\left(r_t-r_{ft}\right)^2, f_{i,t}\right)\right) \quad (35)$$

where r_{ft} is the risk-free asset return, γ is the risk aversion parameter, $c_t(\gamma)\equiv r_{f,t}-0.5\gamma r_{f,t}^2$, $d_t(\gamma)\equiv \mu_t^2\frac{\left(1-\gamma r_{f,t}\right)^2}{\gamma}$ and $x\left(\left(r_t-r_{f,t}\right)^2, f_{i,t}\right)\equiv\frac{1}{\left(\mu_t^2+f_{i,t}\right)}-0.5\frac{\left(\mu_t^2+\left(r_t-r_{f,t}\right)^2\right)}{\left(\mu_t^2+f_{i,t}\right)^2}$.

Considering a simulated option-pricing framework, we also evaluate the volatility forecast methods based on the option criterion of Engle's et al. (1997), by finding the model that generates the highest profits for an investor. Options are priced according to the Black and Scholes (1972) model. We simulate an options market comprising *n* agents who trade straddles based on their volatility forecasts. We assume that a trader with a higher forecast price for the option buys a straddle on one US$ of the S&P 500 index from the remaining traders with lower forecasts. The agent's average daily profit is:

$$\pi^{i} = \left(T - T_0\right)^{-1} \sum_{t=T_0+1}^{T} \sum_{j=1}^{n} \pi_t^{(i,j)} \tag{36}$$

where $\pi_t^{(i,j)}$ is the relative profit of agents *(i)* and *(j).* The trader's *(i)* profit is either $\pi_{t+1}^{(i,j)} = \pi_{t+1} - \left(S_{t+1|t}^{(i)} + S_{t+1|t}^{(j)}\right)$ if $S_{t+1|t}^{(i)} > S_{t+1|t}^{(j)}$ or $\pi_{t+1}^{(i,j)} = \left(S_{t+1|t}^{(i)} + S_{t+1|t}^{(j)}\right) - \pi_{t+1}$ if $S_{t+1|t}^{(i)} < S_{t+1|t}^{(j)}$ where $S_{i,t+1|t} = 4N\left(0.5 f_{i,t+1|t}\right) - 2$ is the straddle's price and $N(\cdot)$ denotes the cumulative normal distribution function. While the daily profit of each agent holding the straddle is computed through $\pi_{t+1} = \max\left(\exp\left(r_{t+1}\right) - \exp\left(r_{f|t+1}\right), \exp\left(r_{f|t+1}\right) - \exp\left(r_{t+1}\right)\right)$ where r_{t+1} is the daily return of S&P500 index and $r_{f|t+1}$ is the daily risk free rate.

In the last set of loss functions we examine includes value-at-risk (VaR) based loss functions. The conditional value-at-risk, denoted as $VaR_{t+1}^{i,a}$ is estimated from

$$VaR_{t+1}^{i,a} = \mu_t + \Phi_t^{-1}\left(a\right) f_{i,t} \tag{37}$$

where $\mu_{i,t}$ is the conditional mean and Φ_t is the cumulative distribution function (assumed normal in our case). Firstly, a loss function based on quantile estimation (Koenker and Bassett, 1978) that penalizes more heavily observations with a violation is used, i.e.:

$$Q^i \equiv \left(T - T_0\right)^{-1} \sum_{t=T_0+1}^{T} \left(a - d_{t+1}^{i,a}\right)\left(r_{t+1} - VaR_{t+1}^{i,a}\right) \tag{38}$$

where $d_{t+1}^{i,a} \equiv 1\left(r_{t+1} < VaR_{t+1}^{i,a}\right)$. A smaller value for Q indicates a better fit for the model. As the Q loss function faces the problem of nondifferentiability, we also employ a smoothed version of the Q loss function proposed by González-Rivera et al. (2004) where the indicator function is replaced with a continuous differentiable function. The smoothed Q loss function is derived from:

$$\tilde{Q}^i \equiv \left(T - T_0\right)^{-1} \sum_{t=T_0+1}^{T} \left(\alpha - m_\delta\left(r_{t+1}, VaR_{t+1}^{i,\alpha}\right)\right)\left(r_{t+1} - VaR_{t+1}^{i,\alpha}\right) \tag{39}$$

where $m_\delta\left(a,b\right) = \left[1 + \exp\left\{\delta\left(a-b\right)\right\}\right]^{-1}$ and the parameter $\delta>0$ controls for the smoothness. The $\tilde{Q}$ loss function is closer to the Q loss function when higher values of the smoothness parameter are used.

Lastly, the VaR-based market risk capital requirement is used as an additional economic loss function. Under the current framework of the Basel II Accord (Basel Committee on Banking Supervision 1996, 2004) financial institutions determine their market risk minimum capital requirement (MRC) according to their internal VaR estimates. The daily risk capital charges must be set at the higher of the average VaR over the previous 60 business days or the previous day's VaR, multiplied by a penalty factor based on the three-zone approach. Thus, for a long position are formulated by:

$$MRC_{t+1}^i = \left(T - T_0\right)^{-1} \sum_{t=T_0+1}^{T} \left(k \times \max\left(VaR_t^{i,0.99}, \frac{1}{60}\sum_{s=0}^{59} VaR_{t-s}^{i,0.99}\right)\right) \tag{40}$$

where $VaR_t^{i,0.99}$ is the VaR estimate generated on day t based on model i and a 10-day holding period and k is penalty factor ranging from 3 to 4 depending on the number of exemptions over the past 250 days.[5] Models that produce lower MRC are preferred since they suggest lower capital charges and allow for higher bank profits.

3.6 Tests of Forecasting Performance

The loss functions described above allow forecasts to be ranked according to their out-of sample forecasting performance. However, they give no indication whether the forecasting losses across the various models are significantly different. Since our aim is to compare a large number of volatility forecasts based on various models and investigate the statistical significance of their comparative forecasting performance based on multiple comparisons we adopt two testing procedures that seem to be the most suitable in our analysis. First, we employ the superior predictive ability (SPA) test of Hansen (2005) that allows for comparing the forecasting performance of two or more models at a time. Consider M+1 different models, i.e. the single and combination models discussed in the previous sections. Forecasts are compared against a benchmark model M_0 and based on a predefined loss function i.e. the performance of model i relative to a benchmark model M_0 is:

$$d_{i,t} = L_{0,t} - L_{i,t}, i = 1,...,M \qquad (41)$$

where $L_{i,t}$ is a statistic or economic loss functions.

Under the assumption of stationarity for $d_{i,t}$, the expected performance of model i relative to the benchmark is defined as μ_i=E[$d_{i,t}$] for i=1,…,M. The null hypothesis is that the benchmark model M_0 is not outperformed by any of the other M competitive

[5] For 0 to 4 violations (green zone) the penalty factor is 3; for 5 to 9 violations (yellow zone) the penalty factor is 3.4, 3.5, 3,65, 3.75 and 3.85, respectively; for more than 10 violations (red zone) the penalty factor is 4.

models and can be expressed as $H_0 : \max_{i=1,\ldots,M} \mu_i \le 0$. The associated test statistic is $\max_{i=1,\ldots,M} M\sqrt{T}\,\bar{d}_i / \hat{\omega}_{ii}$, where $\bar{d}_i = T^{-1}\sum_{t=1}^{T} d_{i,t}$ and $\hat{\omega}_{ii}$ is a consistent estimate of the asymptotic variance of d_i. An estimator of ω_{ii} and the p-values of the test are obtained using the stationary bootstrap procedure of Politis and Romano (1994).

As a way to determine the best performing model(s) in a model-rich environment, we employ the MCS proposed by Hansen et al. (2011). Given a universe of model based forecasts, the test actually picks the best model based on in and out-of-sample evaluations under a specific loss function. The MCS applies sequential trimming to the set of candidate models, M_0. At each step the null hypothesis of equal predictive ability (EPA) is $H_0 : E\left(d_{ij,t}\right) = 0 \forall i, j \in M$ where $d_{ij,t}$ is the loss function differential between models i and j. In each step the worst performing model is eliminated from the model set, and the set of surviving models is the model confidence set at the α level of significance. In order to test the null hypothesis the semi-quadratic statistic:

$$T_{SQ} = \sum_{i<j}\left(\left(\bar{d}_{ij}\right)^2 / \operatorname{var}\left(\bar{d}_i\right)\right) \tag{42}$$

is employed where $\bar{d}_{ij} = T^{-1}\sum_{t=1}^{T} d_{ij,t}$ measures the relative performance between models i and j and $\bar{d}_i \equiv T^{-1}\sum_{j=1}^{M} \bar{d}_{ij}$ measures the model's i performance relative to the averages of the models in the model set. If the null hypothesis is rejected, the worst model is excluded from the model set. A block bootstrap with two bootstrap samples and 10,000 replications is used to obtain the distribution under the null and the confidence level is set to 10%.

The SPA and MCS tests are employed on all statistic and economic loss functions defined in equations (41) – (42). Both tests have been used in the literature mostly on statistical loss functions such as the MSE and more rarely on economic loss functions (González-Rivera et al, 2004; Jiang et al., 2014; Dark, 2015; Tian and Hamori, 2015; Wang et al., 2016). As noted by Laurent et al. (2012) they are both tests of conditional predictive ability and, thus, suitable for both nested and non-nested models while they also account for the estimation method, parameter uncertainty, the estimation and evaluation sample, and data heterogeneity.

3.7 Data and Empirical Results

For the purposes of our study, daily closing prices and intraday 5 minute quotes of the S&P 500 index are used. The sample extends from January 3, 2006 to December 30, 2016 including 2769 daily observations. Daily data are collected from Datastream, while the intraday dataset is obtained from Olsen and Associates. The intraday returns are used to compute the realized volatility measure and the associated models as they contain more information than the simple daily squared returns computed through two arbitrary points in time (Chou et al., 2010). The index value is quoted from 9:30 AM to 4:00 PM, Monday to Friday. As global financial markets tend to integrate and new information arrives during non-trading hours, we adjust the high-frequency dataset to capture the overnight return, excluding the first five minutes of each trading day.. According to Ahoniemi and Lanne (2013) the incorporation of overnight return leads to more accurate forecasts for index realized volatility. Moreover, we use the VIX index, constructed by the Chicago Board of Options Exchange, as a forward-looking indicator of the expected S&P500 index volatility[6].

To evaluate the performance of the competing models we divide the sample into three periods. The period from January 3, 2006 to December 31, 2010 is considered as the in-sample period, while the period from January 3, 2011 to December 30, 2016 is used as the evaluation period. For the combination schemes that require a holdout period for estimating the pseudo-out-of sample forecasts we use the period from June 1, 2009 to December 31, 2010. A total of 20 models are considered using rolling one-step ahead out-of-sample forecasts to evaluate the predictive ability and the economic significance of the examined models[7], while estimates of the actual volatility are obtained using the Realized Volatility (RV) measure by aggregating the intra-day squared returns. The results for the S&P500 index are shown in Tables 1 to 4. The SPA and the MCS tests are used to assess the statistical significance of loss differences among the models presented in section 2.2.3. The models are also ranked according to their forecasting performance. However, the rankings given by the used loss functions differ, since the loss functions penalize differently the forecast errors.

In Table 3.1, results based on the symmetric statistical loss functions are presented. Under the MAE loss function, the two transformed unrestricted OLS models provide the smallest losses compared to all models, while implied volatility is the best performing single model. We verify the superior performance of the OLS-based

[6] For technical details on the index construction, see CBOE (2003).

[7] Although we initially explore a large number of combination models as presented in the previous section, for reasons of brevity we include in the empirical results selected models (i.e. 10 single models and 10 combination models) from each combination category based on their performance. From the simple combinations, we present the results for the Harmonic Mean since it exhibits the best forecasting performance amongst the simple combinations. From the OLS-based models we include the square root transformed OLS_c model ($OLS_{C\text{-}SQRT}$), the exponential transformed OLS_{nc} model ($OLS_{NC\text{-}EXP}$),and the $NRLS_{NC}$ model. More extended results including 20 combination models are available on the Appendix.Results from excluded models are available from the authors upon request.

models against all the single models and the majority of the combination techniques under the SPA and the MCS test using a 10% significance level. Considering the MSE results, the best performing model is the TW model of Timmermann (2006) that is not affected by outliers, structural breaks or shocks, while its form allows for maximum flexibility; followed by the Shrinkage and the $NRLS_{NC}$ model. Regarding the statistical significance, all models except for the FIGARCH model pass the SPA test, while the p-values for most of the combinations are significantly higher compared to the single models. Under the MSE criterion more models are included to the 10% MCS compared to the MAE since the two loss functions penalize differently the losses. .Similarly to Becker and Clements (2008), we find that combination schemes significantly outperform single forecasting models across symmetric statistical evaluation measures. We argue that regression and non-parametric combinations indicate superior forecasting performance as more accurate models are weighted more heavily suggesting that different volatility models capture different market microstructures.

Table 3.1: Statistical Evaluation under symmetric loss functions

	MSE			MAE		
	Loss function	SPA	Rank	Loss function	SPA	Rank
GARCH	2.2469*	(0.1691)	18	0.6242	(0.0000)	18
EGARCH	2.0737*	(0.6102)	12	0.6925	(0.0000)	20
FIGARCH	2.5466	(0.0344)	20	0.6500	(0.0000)	19
MIDAS-RV	2.1044*	(0.5626)	13	0.5976	(0.0000)	16
MIDAS-CPI	2.0631*	(0.6812)	10	0.5965	(0.0000)	15
MIDAS-IP	2.0442*	(0.7980)	8	0.5885	(0.0000)	14
ARMA	2.1565*	(0.3564)	16	0.5590	(0.0000)	13
HAR	2.1490*	(0.3792)	15	0.5429	(0.0000)	9
ARFIMA	2.2846*	(0.1745)	19	0.5415	(0.0000)	8
VIX	2.0006*	(0.9528)	4	0.5349	(0.0000)	5
Harmonic Mean	2.0620*	(0.7811)	9	0.5398	(0.0000)	7
MSFE	2.0369*	(0.9238)	7	0.5552	(0.0000)	11
$OLS_{C\text{-}SQRT}$	2.0710*	(0.7029)	11	0.4330*	(0.2691)	2
$OLS_{NC\text{-}EXP}$	2.2203*	(0.2993)	17	**0.4283***	(1.0000)	1
$NRLS_{NC}$	1.9909*	(0.9818)	3	0.4978	(0.0000)	3
TW	**1.9831***	(1.0000)	1	0.5374	(0.0000)	6
Trimmed MSPE	2.0251*	(0.9631)	5	0.5553	(0.0000)	12
AFTER	2.1240*	(0.4910)	14	0.5540	(0.0000)	10
Shrinkage $OLS_{NC\text{-}EXP}$	2.0312*	(0.8890)	6	0.6191	(0.0000)	17
Shrinkage $NRLS_{NC}$	1.9906*	(0.9791)	2	0.4984	(0.0000)	4

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. The Shrinkage factor is set equal to *d=0.50*. However, we included to our calculations the values 0.25 and 1 and all gave us similar results and ranking. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Results from the asymmetric statistical loss functions are reported in Table 2. The HRLF results suggest that the simple and Shrinkage $NRLS_{NC}$ models followed by the TW model attain the lowest values for the corresponding loss function. It is interesting to note that all the single models are ranked amongst the worst performers apart from the VIX index. The superior performance of the combination models is further corroborated by both the SPA and the MCS tests..The asymmetric QLIKE criterion also confirms the superior performance of OLS models, whilst three of them are included to the MCS. Similarly, Becker and Clements (2008) found that combination forecasts clearly dominate the single models for the asymmetric QLIKE loss function that penalizes more the under prediction of the volatility. The superior performance of the transformed OLS models is further confirmed under the LINEX loss function, presented on the last columns of Table 2. Considering the MCS test, our results indicate that when the scalar parameter is set equal to 0.5, only the two best performing combinations are included to the optimal set, while when a value of 1 is considered, only the best performer i.e. the $OLS_{NC\text{-}EXP}$ model is included. On the contrary the results for the SPA test indicate that all the combinations pass the test for superior predictive ability in both cases. The dominance of regression techniques is evident across all asymmetric loss functions, as they may account more properly for bias correction than simple techniques. In other words, by updating daily the optimal combination weights, regression-based combinations may accommodate more efficiently the “model uncertainty” problem. This result is in accordance with evidence in previous studies, supporting the superior forecasting performance of regression-based combinations for forecasting stock market volatility, i.e. Li et al., 2013, Yang et al., 2015.

Table 3.2: Statistical Evaluation under asymmetric loss functions

	HRLF *(b=-1)*			QLIKE			LINEX *(a=0.5)*			LINEX *(a=1)*		
	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank
GARCH	0.3410	(0.0452)	16	0.3598	(0.0001)	17	0.2572	(0.0269)	19	6.7130	(0.6531)	16
EGARCH	0.3398	(0.1619)	15	0.3755	(0.0000)	18	0.2170	(0.2004)	15	2.4278	(0.6688)	13
FIGARCH	0.3984	(0.0214)	20	0.4022	(0.0006)	20	0.4404	(0.0434)	20	33.5479	(0.0988)	20
MIDAS-RV	0.3064*	(0.3801)	12	0.3267	(0.0296)	12	0.2496	(0.1309)	18	7.4104	(0.5137)	17
MIDAS-CPI	0.3045*	(0.4110)	9	0.3306	(0.0535)	13	0.2486	(0.0822)	17	8.9119	(0.4676)	18
MIDAS-IP	0.3025*	(0.4298)	8	0.3264	(0.0345)	11	0.2068	(0.2048)	14	5.1104	(0.7087)	15
ARMA	0.3466	(0.0707)	18	0.3568	(0.0015)	16	0.1370	(0.5078)	10	1.2870	(0.7553)	10
HAR	0.3432	(0.0836)	17	0.3471	(0.0100)	15	0.1212	(0.5826)	6	0.7029	(0.7128)	4
ARFIMA	0.3795	(0.0469)	19	0.3896	(0.0042)	19	0.1566	(0.4212)	12	3.0749	(0.7105)	14
VIX	0.2960*	(0.5994)	4	0.3234	(0.0566)	8	0.0881	(0.7014)	3	0.3026	(0.7920)	2
Harmonic Mean	0.3063*	(0.2983)	11	0.3207	(0.0393)	7	0.1147	(0.6126)	4	0.7588	(0.7738)	6
MSFE	0.2979*	(0.5119)	7	0.3161	(0.0836)	5	0.1594	(0.3824)	13	2.2119	(0.7543)	12
$OLS_{C\text{-}SQRT}$	0.2974*	(0.3554)	6	0.3035*	(0.2760)	3	0.0879*	(0.7548)	2	0.3355	(0.8454)	3
$OLS_{NC\text{-}EXP}$	0.3338*	(0.1294)	14	0.3248	(0.1194)	10	**0.0801***	(1.0000)	1	**0.2149***	(1.0000)	1
$NRLS_{NC}$	0.2809*	(0.9313)	2	**0.2855***	(1.0000)	1	0.1210	(0.6023)	5	1.2244	(0.7809)	8
TW	0.2898*	(0.7082)	3	0.3059	(0.2005)	4	0.1302	(0.5364)	9	1.1236	(0.7377)	7
Trimmed MSPE	0.2971*	(0.5241)	5	0.3162	(0.0842)	6	0.1541	(0.4160)	11	1.9477	(0.7559)	11
AFTER	0.3047*	(0.4127)	10	0.3235	(0.0494)	9	0.2453	(0.1296)	16	12.6978	(0.4001)	19
Shrinkage $OLS_{NC\text{-}EXP}$	0.3129*	(0.3158)	13	0.3441	(0.0085)	14	0.1301	(0.5485)	8	0.7267	(0.7022)	5
Shrinkage $NRLS_{NC}$	**0.2809***	(1.0000)	1	0.2857*	(0.6022)	2	0.1212	(0.6001)	7	1.2271	(0.7748)	9

Note. 1. Values in bold denote that the corresponding model has the lowest loss function under the specific criterion. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. The Shrinkage factor is set equal to *d=0.50*. However, we included to our calculations the values 0.25 and 1 and all gave us similar results and ranking. 5. EXP and SQRT denote the exponential and the square root transformation respectively.

In addition to comparing the models' statistical performance, we also evaluate their forecasting performance in an economic context. Table 3 reports the risk management performance of volatility forecasts applied to the S&P500 index. The results include the out-of-sample VaR evaluation and daily capital charges based on VaR forecasts derived from single and combination volatility models under the 1% confidence level. The first column of Table 3 reports the empirical percentage of violations during the out-of sample period for 1% VaR. We note that the percentage of violations for all models is always higher than 1% suggesting that all models underforecast VaR and none of the models is adequately reliable as an internal VaR model. The problem is less severe for the EGARCH, the MIDAS and the Shrinkage $OLS_{NC\text{-}EXP}$models where the empirical percentage is close to 1%. These findings are further corroborated by the results for the Q and smoothed Q loss functions, as the Shrinkage $OLS_{NC\text{-}EXP}$ model produces the smallest errors for both loss functions. The results are noteworthy because contrary to the previous section, the transformed OLS schemes perform worse also in terms of the SPA test. We attribute the superior performance of the single forecasts and only a few of the combination forecasts to the fact that they tend to overestimate the daily volatility. We also assess the economic cost related to VaR models using a loss function based on the MRC requirements. The MRC requirements are determined by a 99% VaR over a 10-day holding period. The $OLS_{NC\text{-}EXP}$ model reduces the economic costs significantly by requiring the lowest MRC followed by the rest OLS models. On the contrary, all the single models tend to maintain large capital, suffering from big opportunity costs. Moreover, the SPA and the MCS tests reveal superior forecasting performance solely for the $OLS_{NC\text{-}EXP}$ model. It should be pointed out that contrary to previous findings; the $OLS_{NC\text{-}EXP}$ model is the best performer based on MRC calculations. In other words, the best performing models according to Q and Smoothed-Q loss functions tend to overestimate variance during periods of higher volatility, resulting in larger MRC values.

Table 3.3: Economic Evaluation using VaR-based loss functions

		Q-loss (α=0.01)			Smoothed-Q (α=0.01)			MRC		
	Percentage of violations (α=1%)	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank
GARCH	1.79%	0.0325*	(0.1274)	14	0.0324*	(0.1264)	14	22.48	(0.0000)	18
EGARCH	1.13%	0.0302*	(0.6510)	5	0.0301*	(0.6638)	5	23.51	(0.0000)	20
FIGARCH	2.19%	0.0352	(0.0292)	17	0.0351	(0.0296)	17	23.40	(0.0000)	19
MIDAS-RV	1.39%	0.0301*	(0.7317)	3	0.0300*	(0.7408)	3	21.30	(0.0000)	12
MIDAS-CPI	1.26%	0.0301*	(0.7329)	2	0.0300*	(0.7459)	2	21.24	(0.0000)	11
MIDAS-IP	1.33%	0.0302*	(0.7022)	4	0.0301*	(0.7179)	4	21.49	(0.0000)	13
ARMA	1.72%	0.0328*	(0.1285)	15	0.0327*	(0.1398)	15	21.85	(0.0000)	14
HAR	2.12%	0.0337*	(0.0782)	16	0.0336*	(0.0826)	16	22.17	(0.0000)	16
ARFIMA	2.45%	0.0360	(0.0302)	18	0.0359	(0.0315)	18	22.28	(0.0000)	17
VIX	1.59%	0.0313*	(0.3065)	9	0.0312*	(0.3142)	9	20.69	(0.0000)	6
Harmonic Mean	1.59%	0.0318*	(0.1462)	13	0.0317*	(0.1488)	13	20.86	(0.0000)	10
MSFE	1.59%	0.0309*	(0.3938)	8	0.0308*	(0.4220)	8	20.81	(0.0000)	9
$OLS_{C\text{-}SQRT}$	3.45%	0.0367	(0.0061)	19	0.0365	(0.0067)	19	20.04	(0.0021)	2
$OLS_{NC\text{-}EXP}$	3.78%	0.0379	(0.0054)	20	0.0377	(0.0069)	20	**19.80***	(1.0000)	1
$NRLS_{NC}$	1.72%	0.0314*	(0.2253)	11	0.0313*	(0.2551)	11	20.31	(0.0000)	4
TW	1.66%	0.0304*	(0.6402)	6	0.0303*	(0.6695)	6	20.73	(0.0000)	8
Trimmed MSPE	1.52%	0.0306*	(0.5127)	7	0.0305*	(0.5253)	7	20.73	(0.0000)	7
AFTER	1.59%	0.0316*	(0.1559)	12	0.0316*	(0.1556)	12	20.57	(0.0000)	5
Shrinkage $OLS_{NC\text{-}EXP}$	1.33%	**0.0295***	(1.0000)	1	**0.0295***	(1.0000)	1	22.15	(0.0000)	15
Shrinkage $NRLS_{NC}$	1.66%	0.0314*	(0.2278)	10	0.0313*	(0.2333)	10	20.10	(0.0005)	3

Note. 1. Values in bold denote that the corresponding model has the lowest loss function under the specific criterion. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. The Shrinkage factor is set equal to $d=0.50$. However, we included to our calculations the values 0.25 and 1 and all gave us similar results and ranking. 5. The Q-loss function and the Smoothed-Q loss function are calculated for VaR level $\alpha=0.01$ and α=0.05. For the Smoothed-Q calculation we set the smoothness parameter $\delta=25$. 6. EXP and SQRT denote the exponential and the square root transformation respectively.

The results for the utility-based criterion of González-Rivera et al. (2004) are reported in Table 4. A lower value for this function can be interpreted as a higher utility gain for an investor that uses a specific volatility forecast since the loss function has to be minimized. There is a most preferred model that clearly dominates all the rest; this is the $OLS_{C\text{-}SQRT}$ that attains the highest utility gains, passes the SPA test and is included to the 10% model confidence set. The ranking of the competing models indicates that most of the combination schemes can lead to higher utility gains compared to the single models except from the ARFIMA and the HAR models. Turning to the option based loss function, we find that the Shrinkage $NRLS_{NC}$ model is the dominant model, as it provides the highest mean daily profit for an investor, while most of the single models generate losses. Considering the MCS test, the dominance of the combination techniques is obvious against the single models as only one combination model is excluded. In total, regression, non-parametric and shrinkage combinations produce both higher utility gains and daily profits suggesting that they are preferable for volatility forecasting. This result reveals that the combination of three sources of information, along with the incorporation of different market microstructure of each volatility model may improve the combination techniques in the context of stock market volatility. This result is consistent with the aforementioned results of statistical evaluations.

While a significant number of studies (e.g. Rapach et al., 2010; Jordan et al., 2014) find evidence in favor of simple combination techniques as better predictors of stock market returns and equity premium, our study is consistent with the literature in the context of combination forecasts for volatility (e.g. Li et al., 2013; Yang et al., 2015) suggesting that more complex combination techniques, including regression-based, non-parametric and shrinkage methods, lead to higher gains. The main issue is that volatility cannot be directly observed and a more accurate approximation is needed. As a result, simple combinations perform poor as they ignore any information about the relative quality of the single forecasts, while the time-varying weights examined in this paper allow for maximum flexibility among the forecasts.

Table 3.4: Economic Evaluation using utility-based and option-based loss functions

	Utility (gamma=3)			Option criterion		
	Loss function	SPA	Rank	Loss function	SPA	Rank
GARCH	-0.0469	(0.0000)	19	-2.6080	(0.0273)	17
EGARCH	-0.0466	(0.0000)	20	-3.1119*	(0.0382)	19
FIGARCH	-0.0469	(0.0000)	18	-4.4414	(0.0170)	20
MIDAS-RV	-0.0476	(0.0000)	11	-0.0562*	(0.3017)	13
MIDAS-CPI	-0.0476	(0.0000)	14	0.3497*	(0.4100)	10
MIDAS-IP	-0.0476	(0.0000)	12	-0.0806*	(0.2893)	14
ARMA	-0.0474	(0.0000)	16	-0.0308*	(0.3147)	11
HAR	-0.0478	(0.0000)	8	0.6350*	(0.4619)	8
ARFIMA	-0.0481	(0.0000)	6	0.4287*	(0.4082)	9
VIX	-0.0475	(0.0000)	15	-0.5834*	(0.2141)	15
Harmonic Mean	-0.0479	(0.0000)	7	-0.0509*	(0.2634)	12
MSFE	-0.0478	(0.0000)	9	1.5643*	(0.7804)	5
$OLS_{C\text{-}SQRT}$	**-0.0511***	(1.0000)	1	1.0040*	(0.4777)	7
$OLS_{NC\text{-}EXP}$	-0.0509	(0.0066)	2	-0.8213*	(0.2458)	16
$NRLS_{NC}$	-0.0488	(0.0000)	3	2.7404*	(0.9913)	2
TW	-0.0481	(0.0000)	5	2.5149*	(0.9571)	3
Trimmed MSPE	-0.0478	(0.0000)	10	1.9650*	(0.8541)	4
AFTER	-0.0476	(0.0000)	13	1.2483*	(0.6314)	6
Shrinkage $OLS_{NC\text{-}EXP}$	-0.0471	(0.0000)	17	-2.7803	(0.0510)	18
Shrinkage $NRLS_{NC}$	-0.0488	(0.0000)	4	**2.8067***	(1.0000)	1

Note. 1. Values in bold denote that the corresponding model has the lowest loss function under the specific criterion. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. The Shrinkage factor is set equal to *d=0.50*. However, we included to our calculations the values 0.25 and 1 and all gave us similar results and ranking. 5. The risk aversion parameter for the Utility function is set equal to 3. However, we included values ranging from 1 to 5 leading to similar results. 6. EXP and SQRT denote the exponential and the square root transformation respectively. 7. To calculate the economic gains, the risk free rate return is derived from the 3 month T-bill.

3.8 Concluding Remarks

One of the most frequently asked question in the finance literature is which model should to be used to forecast the volatility of asset returns. Given the uncertainty of financial markets the issue of selecting a single model in all cases of volatility forecasting is quite complicated. Following an extensive literature in forecasting, composite forecasts have been shown to be a good way to improve forecasting accuracy.

In this chapter we investigate the benefits of combining forecasts using simple and more complex combination techniques. We compare an exhaustive set of forecast combination methods and single models in both statistical and economic terms to determine the best performing models. From a statistical point, we assess the volatility models using both symmetric and asymmetric loss functions and conclude that the OLS-based schemes outperform the single and other combination techniques. Testing for the statistical significance of our findings, the SPA and the MCS tests support their superior performance. The examination of the economic loss functions results yields several useful conclusions. For the VaR-based loss functions, which focus on the tails of the density, no model is proved adequately reliable, apart from the Shrinkage $OLS_{NC\text{-}EXP}$. Taking into consideration the market risk capital requirements, the $OLS_{NC\text{-}EXP}$ scheme is the best performing model and the only model that passes both tests indicating significant superior performance. Under the utility-based loss function and the simulated option pricing framework, the OLS-based combination models provide the highest gains, while the SPA test verifies their superior performance. These results are in accordance with the literature, as several studies point out the dominance of regression techniques in the context of stock market volatility prediction (Fuertes et al., 2009; Li et al., 2013; Yang et al., 2015 amongst others), while our resputs further support the use of non-parametric and shrinkage techniques to optimally combine volatility forecasts. The superior forecasting performance of more complex combination techniques is attributed to three factors: (i) the combination of three sources of information, i.e. the combination of different information channels is more efficient than combining models based on the same dataset; (ii) regression-based, non-parametric and shrinkage forecasts account for bias correction more properly than the simple average as the optimal combination weights are updated daily; (iii) volatility models incorporate different market microstructures that vary over time. As a result, accounting for the importance of each volatility model in a time-varying framework reduces the "model uncertainty" problem and improves the combination forecast in both statistical and economic terms.

Concluding, although statistical and economic loss functions support the superior performance of combination forecasts based on more regression-based, non-parametric and shrinkage methods, there is no clear winner across all loss functions suggesting that different combination schemes are preferable based on the economic application to be used. Taking into consideration the instability in the performance of single forecasts and the fact that different loss functions are relevant for different decision makers, our results encourage investors or financial institutions to use the combination forecasting methodology to improve forecasting accuracy and economic gains. A logical next step could be to optimally combine forecasts based on economic loss functions expecting these combination forecasts to lead to higher forecasting performance.

Chapter 4

Combining Energy Price Volatility Forecasts through Economic Loss Functions

Forecasting energy price volatility is an important input in macroeconomic, option pricing, and portfolio selection models. Although a vast literature on models has been developed, there is rarely any consensus on which model is most appropriate in providing accurate volatility forecasts for energy price. Numerous studies provide evidence in favor of GARCH-class and HAR-class models (for example see Sadorsky, 2006; Kang et al., 2009; Wei et al., 2014; Lux et al., 2016; Kleian and Wather, 2016; Charles and Darné, 2017; Degiannakis and Filis, 2017; Wen et al., 2016; Ma et al., 2018 among others) but a limited number examines the benefits of combination forecasts.

The objective of this chapter is threefold. Firstly, we introduce new combination methodologies based on economic and risk management loss functions where the optimal weights are computed through these loss functions. We use two economic loss functions widely used in portfolio evaluation literature, i.e. the Certainly Equivalent Return (*CER*) and the *Sharpe* ratio (Ma et al., 2018a; Rapach et al., 2010); and a risk management loss function, the Smoothed-*Q* loss function due to González-

Rivera et al. (2004). We expect to optimally combine forecasts that generate higher economic gains compared to other combinations and single models. To test the predictive ability of the new combination techniques, we use a large variety of simple combinations and techniques based on statistical measures to predict the volatility of WTI crude oil futures. Secondly, we use both low-frequency and high-frequency datasets to our combinations. Specifically, we combine forecasts based on: (i) GARCH models (daily data), (ii) HAR models (intraday data), and (iii) both GARCH and HAR models. Thirdly, we evaluate the proposed methodologies with the rest combinations and the single models using both statistical and economic evaluation criteria.

The empirical analysis yields some important conclusions. At first, the results indicate an improvement to the forecasting performance from the combinations based on GARCH or all models and economic loss functions, while there is no improvement according to HAR combinations. More specifically, economic combination forecasts based on either GARCH or all models are more profitable for 1-step and 22-step ahead forecast horizons for a mean-variance investor, while economic combination forecasts based on risk management and standard statistical loss functions are more profitable for one-step-ahead forecasts.

The remainder of the chapter is organized as follows. Section 4.1 reviews the literature on combination forecasts and energy price volatility. Section 4.2 presents the single volatility models used in this application. Section 4.3 presents the combination models used in this chapter, while Section 4.4 describes the new combinations based on economic and risk management loss functions. Section 4.5 presents the evaluation measures, while in Section 4.6 the tests for forecasting performance are described. Section 4.7 describes the dataset and reports the empirical results, and Section 4.8 concludes.

4.1 Combination Forecasts and Energy Price Volatility

Over the last decade, energy products have become a popular asset class for investors and financial institutions. Accurate forecasts of oil price are important to firms, consumers and governments. However, as real-time data account for delays and revisions in data releases and crude oil price is assigned to the highest weight in constructing some popular commodity prices indices, combination forecasts can be seen as an alternative solution. Combination forecasts have been found to work well for forecasting macroeconomic variables (Timmermann, 2006), but there is little evidence considering energy products. The question arises is: Do they work well in crude oil price and volatility forecasting?

Although the no-change forecast has been documented as the best forecast of future oil prices (e.g. Davies, 2007; Hamilton, 2009), the recent literature provides several econometric forecasting models that outperform at least at some horizons the no-change forecast (see, e.g. Baumeister and Kilian, 2012). However, a limited number of studies examines the benefits from combination forecasts. Baumeister and Kilian (2015) using an equal average combination scheme and a DMSPE approach, conclude that forecast combinations are superior to the no-change forecast as they generate lower MSPE ratios and have higher directional accuracy.

Similarly, Baumeister et al. (2014) combine several oil price forecasting models under three combination techniques; a simple average, a simple average of the models after dropping out the models with the largest MSPE, and the best single model according to the lowest recursive MSPE for the previous period. They find that the two first combination schemes lead to smaller MSPE for all forecasting periods while the third model indicates unstable forecasting ability over time. Furthermore, Wang et al. (2017) combine forecasts using several parametric and non-parametric techniques to predict oil prices and the annotated density. Their results indicate that a rank-based model (i.e. the TW model due to Timmermann, 2006) generates more accurate point and density forecasts. Recently, Zhang et al. (2018) use the so-called iterated combination approach (due to Lin et al., 2017), which is based on both the standard combination forecasts and simple benchmark forecasts, as an alternative method to improve the predictability of the returns of oil prices. The results indicate that the iterated combination outperforms standard combination approaches in both statistical and economic measures.

While a significant amount of literature provide evidence for the use of combination techniques to oil price forecasting, there is little evidence on the benefits derived from combination forecasts to crude oil price volatility. First, Lux et al. (2016) combined several GARCH and Markov-Switching Models through a weighted linerar combination, where the optimal weights are computed through a forecast encompassing test due to Harvey et al. (1998) for non-nested models that is based on least squares regression. Pointing out the benefits derived from DMA approach, Wei et al. (2017) and Ma et al. (2018c) used the Dynamic Model Averaging (DMA) approach that takes into account the historical performance of the competing models. Their results indicated the superior predictive ability of the combination techniques compared to other single models.

Zhang and Ma (2018) based on high frequency data on oil futures combined several models, ranging from simple combinations, such as the mean and the median, to more complex combination schemes, such as the DMSPE combination, the Shrinkage and the iterated combination. The out-of-sample evaluation indicated that the iterated combination forecasts surpasses the standard ones and the single models in both statistical and economic terms. On the contrary, Zhang et al. (2019) combined eight HAR models using several combination techniques and two shrinkage methods (i.e. the elastic net and lasso), concluding that the two shrinkage methods outperform not only the single models but the combination models in both statistical and economic perspectives.In this chapter, we introduce new combination techniques based on economic and risk management loss functions, while we expect to improve the forecasting accuracy in both statistical and economic terms.

4.2 Single Volatility Forecasting Models

4.2.1 GARCH-type Volatility Models

In this chapter we use several (linear and nonlinear) GARCH-type and HAR-type models to forecast volatility of crude oil futures prices. The models capture a number of stylized facts in volatility behavior such as asymmetry, long-memory and persistence. Then, numerous combinations based on these models are considered, as well as the out-of-sample evaluation procedure. The total sample of T observations is divided into an in-sample portion composed of the first T_0 observations and an out-of sample portion of $T_1=T-T_0$ observations. We generate out-of sample volatility forecasts from each of the following described models using a rolling estimation window.

Considering a return process equal to (18) presented in the previous section, we include several common GARCH models, such as the GARCH (*p*,*q*), the EGARCH and the FIGARCH models presented by the equations (19), (20) and (21) respectively. However, GARCH and EGARCH models assume that the conditional variance is a linear function of lagged squared returns. As a result, we also use the asymmetric power GARCH (APARCH) model of Ding et al. (1993). The APARCH model is more flexible in modeling the conditional variance and is written as:

$$s_t^{\delta} = a_0 + \sum_{i=1}^{p} a_i \left(|\varepsilon_{t-i}| - \gamma \varepsilon_{t-i} \right)^{\delta} + \sum_{j=1}^{q} \beta s_{t-j}^{\delta} \tag{43}$$

where $\delta > 0$ and $-1 < \gamma < 1$. To capture asymmetric volatility, the GJR-GARCH model due to Glosten et al. (1993) is considered. The conditional variance specification for this model is:

$$s_t^2 = a_0 + \sum_{i=1}^{q} \left(a_i + \gamma I_{t-i} \right) \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j s_{t-j}^2 \tag{44}$$

where $I_{t-i} = 1$ if $\varepsilon_{t-i} < 0$, and 0 otherwise. The parameters are restricted to $a_0 > 0$, $a_i, \gamma_i, \beta_j \geq 0$ and $a_i + \gamma_i \geq 0$ in order to guarantee the positive variance and the process is stationary if $a_i + \beta_j + \left(\gamma_i / 2 \right) < 1$. If the asymmetry coefficient γ is greater than zero, the volatility rises more after large negative shocks than after large positive shocks. The GJR-GARCH model nests to the GARCH model when $\gamma = 0$.

Another long-memory model which generalizes the FIGARCH model is the hyperbolic GARCH (HYGARCH) model of Davidson (2004). In this model the existence of second moments at more large widths compared to FIGARCH model is permitted. The HYGARCH(*p*,*d*,*q*) model is described as:

$$s_t^{\ 2} = a_0\left[1-\beta(L)\right]^{-1} + \left\{1-\left[1-\beta(L)\right]^{-1}\varphi(L)(1+k)\left[(1-L)^d - 1\right]\right\}\varepsilon_t^{\ 2} \qquad (45)$$

where $k \geq 0$, $d \geq 0$. For $0 < k < 1$ the process is stationary, while the model nests the FIGARCH and GARCH models when $k = 1$ and $k = 0$.

4.2.2 HAR-type Volatility Models

Numerous studies indicate that models based on high frequency data achieve better forecasting performance compared to conventional models based on daily returns (e.g. Andersen et al., 2007; Corsi et al., 2010; Patton and Sheppard, 2015, Wen et al., 2016, Ma et al., 2017). First, we compute the Realized Volatility (RV) as the sum of squared intraday returns, i.e.:

$$RV_t = \sum_{t=1}^{m} r_{t,m}^2 \qquad (46)$$

Then, we employ several frequently used models directly applied on the realized volatility series. The first model is Corsi’s (2009) Heterogeneous AutoRegressive model of Realized Volatility (HAR-RV) that encompasses different time aggregations of RV represented from (27). Andersen (2007) proposed the HAR-RV-J models by adding the daily discontinuous jump variation to the HAR-RV model. The model is considered to improve the forecasting performance of the simple HAR-RV model through the jump component. The HAR-RV-J models is:

$$RV_t^2 = \beta_0 + \beta_d RV_{t-1,d} + \beta_w RV_{t-1,w} + \beta_m RV_{t-1,m} + \gamma_d J_{t-1,d} + u_t \qquad (47)$$

where $J_{d,t-1}$ is the daily discontinuous jump variation. Following Andersen et al. (2007), we employ the HAR-RV-CJ model by decomposing realized volatility into continuous sample path variation and discontinuous jump variation. The model is represented as:

$$RV_t^2 = \beta_0 + \beta_d C_{t-1,d} + \beta_w C_{t-1,w} + \beta_m C_{t-1,m} + \gamma_d J_{t-1,d} + \gamma_w J_{t-1,w} + \gamma_m J_{t-1,m} + u_t \quad (48)$$

where $C_{t-1,d}$ is the daily continuous sample path variation, $C_{t-1,w}$ is the weekly continuous sample path variation, $C_{t-1,m}$ is the monthly continuous sample path variation, $J_{t-1,w}$ is the weekly discontinuous jump variation; and $J_{t-1,m}$ is the monthly discontinuous jump variation. Patton and Sheppard (2011) developed the HAR-RSV model which assumes that positive and negative realized semivariances can have different predictive abilities for different time horizons. The HAR-RSV model is equal to:

$$RV_t^2 = \beta_0 + \beta_d RSV_{t-1,d^+} + \beta_w RSV_{t-1,w^+} + \beta_m RSV_{t-1,m^+} + \gamma_d RSV_{t-1,d^-} + \gamma_w RSV_{t-1,w^-} + \gamma_m RSV_{t-1,m^-} + u_t \quad (49)$$

where RSV_{t-1,d^+} is the daily positive realized semivariance, RSV_{t-1,w^+} is the weekly positive realized semivariance, RSV_{t-1,m^+} is the monthly positive realized semivariance, RSV_{t-1,d^-} is the daily negative realized semivariance, RSV_{t-1,w^-} is the weekly negative realized semivariance and RSV_{t-1,m^-} is the monthly negative realized semivariance.

Finally, the LHAR-RV model is included to our calculations. Asai et al. (2012) introduced the LHAR-RV model that takes into consideration the leverage effects. The model is:

$$RV_t^2 = \beta_0 + \beta_d RSV_{t-1,d^+} + \beta_w RSV_{t-1,w^+} + \beta_m RSV_{t-1,m^+} + \lambda_d r_{t-1,d^-} + \lambda_w r_{t-1,w^-} + \lambda_m r_{t-1,m^-} + u_t \quad (50)$$

where r_{t-1,d^-} is the daily negative returns (i.e. $r_{t-1,d^-} = r_t I\{r_t < 0\}$), r_{t-1,w^-} is the weekly negative returns and r_{t-1,m^-} is the monthly negative returns.

4.3 Combination Models

As Bates and Granger (1969) stated combining forecasts is generally considered an useful forecasting method that leads to improved forecasting accuracy. Furthermore, Timmermann (2006) provided a theoretical justification for combining forecasts. In this chapter, a large variety of combination techniques is used. In our calculations, the following combinations are included. At first, all the simple combinations (i.e. the Mean, the Geometric Mean, the Harmonic Mean, the Trimmed Mean and the Median) are included. More, the linear least regressions and their transformations represented by equations (2), (3) and (4). Also we consider the nonlinear AFTER combination in equation (11). More, the DMSFE and the IMSFE based on equation (8) are used. Finally, the non-parametric TW and the Kernel regression based on equations (13) and (15) are included.

4.4 Combinations based on economic and risk management loss functions

We extend the existing forecasting literature by providing combination forecasts based on economic loss functions. We include to our calculations three economic loss functions. The former is based on value-at-risk while the second and the third are based on Certainly Equivalent Return and Sharpe ratio, respectively.

In the first combination scheme we incorporate the three loss functions to the TW combination technique described at equation (13). In this case the combination weights are inversely proportional to the model's rank according to the specific loss function. Firstly, we use the value-at-risk (VaR) loss function, where the conditional value-at-risk, denoted as $VaR_{t+1}^{i,a}$ is estimated from

$$VaR_{t+1}^{i,a} = \mu_t + \Phi_t^{-1}(a)\, f_{i,t} \tag{51}$$

where $\mu_{i,t}$ is the conditional mean and Φ_t is the cumulative distribution function (assumed normal in our case). Using VaR we employ a loss function (i.e. the *Q* smoothed loss function) proposed by González-Rivera et al. (2004), described as:

$$\tilde{Q}^i = \left(\alpha - m_\delta\left(r_{t+1}, VaR_{t+1}^{i,\alpha}\right)\right)\left(r_{t+1} - VaR_{t+1}^{i,\alpha}\right) \tag{52}$$

where $m_\delta(a,b) = \left[1 + \exp\{\delta(a-b)\}\right]^{-1}$ and the parameter $\delta>0$ controls for the smoothness.

The second loss function is based on portfolio performance as the volatility is one of the key determinants for portfolio allocation following Ma et al. (2018a). We assume that a mean-variance allocates his/her assets between a risky asset (i.e. oil price) and a risk free asset (e.g. Rapach et al., 2010). The portfolio return is equal to:

$$R_{pt}{}^{i} = w_t r_t^* + r_{ft} \tag{53}$$

where r_t^* is the excess return of the risky asset, i.e. $r_t^* = r_t - r_{ft}$ and r_{ft} is the risk free return. Following Cambell and Thompson (2008), the investor decides the proportion to the risky asset at time *t* as:

$$w_t = \frac{1}{\gamma}\left(\frac{r_{t-1}}{s_{t-1}{}^2}\right) \tag{54}$$

where r_{t-1} is the historical moving average returns of the in-sample data and $s_{t-1}{}^{2}$ is a rolling window forecasting variance and γ is the risk aversion parameter. We restrict the weights to 0 and 1.5. The mean-variance investor realizes a Certainly Equivalent Return (*CER*) equal to:

$$CER_p{}^{i} = \hat{\mu}_p - \frac{\gamma}{2}\hat{\sigma}_p{}^{2} \quad (55)$$

where $\hat{\mu}_p$ is the out-of-sample mean portfolio returns and $\hat{\sigma}_p{}^{2}$ is its variance respectively. The *CER* ratio is the difference between the average utility of the models and the benchmark. Finally, we use the *Sharpe* ratio to combine optimally forecasts derived by:

$$S_p{}^{i} = \frac{\hat{\mu}_p}{\hat{\sigma}_p{}^{2}} \quad (56)$$

We denote these combinations as TW-*Q* and TW-*CER* and TW-*Sharpe* respectively.

The second combination technique is based on Wang et al. (2017) who computed optimal forecasts after trimming the one with the worst past performance according to MSFE. In this case, the excluded model is the one with the worst past performance in terms of the economic loss functions described on equations (52), (55) and (56). We denote these combinations as Trimmed-*Q*, Trimmed-*CER* and Trimmed-*Sharpe* respectively.

Finally we include an alternative combination strategy following Stock and Watson (2004) that places all the weight on the individual forecast that has the best average post performance during the last four periods. This strategy was also followed by Zhang and Ma (2018) but in contrast to them we implement this strategy based on economic loss functions (i.e. the smoothed-*Q*, the *CER* and the *Sharpe* loss functions) rather than statistical loss functions. We denote these combinations as Best-*Q* and Best-*CER* and Best-*Sharpe* respectively.

4.5 Evaluation Measures

To evaluate the relative accuracy of the proposed methodologies, we rank the models using several statistical loss functions. We consider realized volatility (RV) as a proxy for the "true" volatility. From a statistical point of view, we use, at first, statistical loss functions widely used in the existing forecasting literature such as the mean absolute error (MAE) and the mean squared error (MSE) defined in equations (31) and (32). In various financial applications, volatility over-prediction has different impacts than volatility under-prediction. Consequently, we use two asymmetric loss functions, the HRLF and the QLIKE represented by equation (34).

In order to assess the performance of volatility forecasts in economic terms, we consider a number of economic loss functions based on various financial applications and compute the predictive gains derived from the used methods. Specifically, in this chapter, we use the economic loss functions described in equations (52), (55) and (56) to evaluate the economic gains derived from the combinations and single models.

4.6 Tests of forecasting performance

The loss functions described above allow forecasts to be ranked according to their out-of sample forecasting performance. However, they give no indication whether the forecasting losses across the various models are significantly different. Since our aim is to compare a large number of volatility forecasts based on various models and investigate the statistical significance of their comparative forecasting performance based on multiple comparisons we adopt three testing procedures that seem to be the most suitable in our analysis. First, we employ the superior predictive ability (SPA) test of Hansen (2005) described in the previous section by equation (41) and the MCS proposed by Hansen et al. (2011) represented by equation (42). The SPA and MCS tests are employed on all statistic loss functions defined in equations (31), (32) and (34).

Although the SPA test introduced by Hansen (2005) is considered as a powerful test, it is designed to test whether the best model from a universe of alternative models beats the benchmark of not. Several extensions proposed by the literature lack power while many poor models are included in the test (e.g. the stepwise reality check test due to Romano and Wolf (2005)). Hsu et al. (2010) proposed a more powerful test, the stepwise SPA (StepSPA) test that controls for only one false rejection, which is a stringent criterion. In this application, we use this test to evaluate the out-of-sample forecasting performance of alternative models relative to the benchmark model according to two economic loss functions, the *CER* and the *Sharpe ratio*.

In the StepSPA test, all the significant models are identified when the null hypothesis is rejected. The SPA and the StepSPA test differ in the calculation of losses during the bootstrap process. The method to calculate the StepSPA by Hsu et al. (2010) is as follows:

Step 1 All the models are rearranged in a descending order according to $d_{i,t}$.

Step 2 The null hypothesis is rejected when there is at least one model with a statistic greater than the critical value, specified at the 1-α quantile of the empirical distribution that is obtained through a bootstrapping procedure.

Step 3 If the null hypothesis is rejected, the models with statistics greater than the critical value are eliminated and a new critical value is computed through the remaining set of models. This procedure is repeated until the hypothesis is not rejected. As a result, the models that are not eliminated are considered as superior.

4.7 Data and Empirical Results

To empirically test the performance of our single and combination models, the following two datasets are considered: daily prices and intraday 5 minute quotes of the front-month WTI crude oil futures. The sample extends from May 10, 2007 to June 30, 2016 including 2281 daily observations. Daily data are collected from the U.S. Energy Information Administration (EIA) while the intraday dataset is obtained from the paper of Gong and Lin (2018)[8]. Our sample data are divided into three groups: 1) in-sample data for volatility modeling from May 10, 2007 to June 10, 2011; 2) out-of-sample data for forecasting evaluation from June 13, 2011 to June 30, 2016; and 3) pseudo-out-of-sample data for training the combination weights from June 1, 2010 to June 10, 2011.

A total of 170 models are considered using rolling 1-step, 5-step and 22-step ahead out-of-sample forecasts to evaluate the predictive ability and the economic significance of the examined models[9]. The results for the WTI crude oil futures are shown in Tables 4.1 to 4.27. The models are also ranked according to their forecasting

[8] We would like to thank Gong and Lin (2018) for providing us the realized volatility series used in their paper: "The incremental information content of investor fear gauge for volatility forecasting in the crude oil futures market"

[9] Although a large number of combination models is presented in the previous section, for reason of brevity we include in this section the empirical results for selected models (i.e. the single models and 26 combination models). We exclude from the evaluation process models in the same class that give similar results, such as the Shrinkage models. From the OLS-based models we include to the evaluation process only the square root transformed and exponential transformed $ERLS_c$ model and the $NRLS_{nc}$ model that indicated superior performance against the other OLS models. Results from excluded models are available from the authors upon request.

performance. We note that the rankings given by the used loss functions differ, since the loss functions penalize differently the forecast errors. We evaluate all models using the SPA test due to Hansen (2005), the MCS test due to Hansen et al. (2011) and the StepSPA test due to Hsu et al. (2010), which is implemented on the evaluation according to economic criteria.

4.7.1 Out-of-Sample Forecasting Performance for GARCH combinations

Statistical Evaluation

We use the rolling forecasting methodology to generate the 1-, 5- and 22-step ahead volatility forecasts of six competing GARCH-type models and combinations based on them. We compare their forecasting performance based on the symmetric MAE and MSE loss functions and the asymmetric HR and QLIKE loss functions. The superior predictive ability (SPA) proposed by Hansen (2005) and the model confidence set (MCS) proposed by Hansen et al. (2011) are used to assess the statistical significance of the single models and the proposed combinations. Based on Laurent et al. (2012), we set the confidence level α equal to 0.10, which means that if the *p*-value obtained from either the SPA or the MCS test is smaller than 0.10, we can exclude the corresponding model. The *p*-values are obtained from 10,000 bootstraps with 2 block length. Table 4.1 reports the obtained results, while the bold face numbers indicate the best model in terms of volatility forecasting. Under the MAE loss function, the results for the 1-step ahead forecasts indicate the superior forecasting performance of an OLS transformed model, while the same results hold for the MSE. However, the SPA test indicates that under the MSE all models pass the test, and a big number of combinations are included to the MCS.

Considering the asymmetric loss functions, HR and QLIKE, the OLS_{SQRT} and the TW combinations are ranked first for each loss function. The Kernel regression combination is proved inadequate according to SPA test, while five and four single models pass the test respectively. From the rest combinations, most of them are included to the MCS, while only some combinations based on Smoothed-Q are included to the MCS.

Table 4.1 1-step ahead Statistical Forecasting Performance for GARCH combinations

	MAE			MSE			HRLF *(b=-1)*			QLIKE		
	Loss Function	SPA	Rank	Loss Function	SPA	Rank	Loss Function	SPA	Rank	Loss Function	SPA	Rank
GARCH	2.89	(0.0000)	30	33.34	(0.5757)	29	1.39	(0.2898)	29	0.31	(0.1914)	31
EGARCH	2.70	(0.0000)	10	32.05	(0.7242)	26	1.31*	(0.6620)	26	0.29*	(0.6721)	22
GJR	2.71	(0.0000)	11	31.77	(0.7838)	23	1.29*	(0.7589)	20	0.29*	(0.6465)	11
APARCH	2.75	(0.0000)	18	31.77*	(0.7523)	22	1.31*	(0.6412)	25	0.29	(0.6195)	23
FIGARCH	2.74	(0.0000)	17	30.93*	(0.8552)	8	1.28*	(0.7523)	14	0.29*	(0.6275)	21
HYGARCH	2.79	(0.0000)	24	31.83	(0.7488)	25	1.30*	(0.6630)	22	0.28*	(0.8235)	8
Mean	2.71	(0.0000)	12	30.78*	(0.8994)	4	1.26*	(0.9514)	5	0.28*	(0.8178)	7
Geometric Mean	2.69	(0.0000)	7	30.83*	(0.8886)	5	1.25*	(0.9747)	4	0.28*	(0.9479)	3
Harmonic Mean	2.67	(0.0000)	4	30.88*	(0.8729)	6	1.25*	(0.9801)	3	0.28*	(0.9571)	2
Trimmed Mean	2.70	(0.0000)	9	30.91*	(0.8716)	7	1.26*	(0.9471)	8	0.28*	(0.8656)	4
Median	2.69	(0.0000)	8	31.04*	(0.8465)	13	1.26*	(0.9093)	9	0.28*	(0.8470)	5
AFTER	2.77	(0.0000)	22	31.01*	(0.8346)	12	1.28*	(0.7102)	16	0.29*	(0.6909)	15
DMSFE	2.73	(0.0000)	16	30.96*	(0.9121)	9	1.27*	(0.8863)	10	0.28*	(0.7657)	9
IMSFE	2.76	(0.0000)	20	31.06*	(0.8837)	14	1.28*	(0.8180)	13	0.29	(0.7305)	14
Kernel	2.65	(0.0131)	3	45.53	(0.1115)	32	1.61	(0.0870)	32	0.34	(0.0555)	32
OLS_{EXP}	2.60	(0.0041)	2	**29.93***	(1.0000)	1	1.24*	(0.7441)	2	0.29*	(0.6606)	16
OLS	2.73	(0.0000)	14	31.28	(0.8483)	19	1.28*	(0.7968)	15	0.28*	(0.8477)	10
OLS_{SQRT}	**2.33***	(1.0000)	1	30.49*	(0.9520)	2	**1.24***	(1.0000)	1	0.30	(0.3189)	28
TW	2.68	(0.0000)	5	31.12*	(0.8877)	15	1.26*	(0.9547)	7	**0.28***	(1.0000)	1
Trimmed MSPE	2.72	(0.0000)	13	30.78*	(0.9369)	3	1.26*	(0.9486)	6	0.28*	(0.8486)	6
TW-*CER*	2.82	(0.0000)	28	31.78	(0.7323)	24	1.31	(0.6154)	24	0.29	(0.5602)	25
Trimmed-*CER*	2.79	(0.0000)	25	31.24	(0.8568)	17	1.29	(0.6977)	18	0.29	(0.6671)	18
Best-*CER*	2.93	(0.0000)	32	33.85	(0.5211)	31	1.40	(0.2862)	30	0.30	(0.3253)	29
TW-$Q_{0.05}$	2.73	(0.0000)	15	31.24*	(0.8339)	16	1.28*	(0.7805)	17	0.29	(0.6839)	17
TW-$Q_{0.01}$	2.78	(0.0000)	23	31.53	(0.7669)	21	1.29	(0.7023)	21	0.29	(0.6673)	20
Trimmed-$Q_{0.05}$	2.75	(0.0000)	19	30.98*	(0.9010)	10	1.27*	(0.8255)	11	0.29*	(0.7344)	12
Trimmed-$Q_{0.01}$	2.77	(0.0000)	21	31.00*	(0.8941)	11	1.27*	(0.8001)	12	0.29	(0.7289)	13
Best-$Q_{0.05}$	2.68	(0.0000)	6	32.21	(0.7044)	28	1.34	(0.5204)	27	0.30	(0.4570)	26
Best-$Q_{0.01}$	2.86	(0.0000)	29	33.65	(0.5573)	30	1.40	(0.2620)	31	0.30	(0.2731)	30
TW-*Sharpe*	2.81	(0.0000)	27	31.42*	(0.8031)	20	1.30	(0.6560)	23	0.29	(0.5518)	24
Trimmed-*Sharpe*	2.79	(0.0000)	25	31.24*	(0.8545)	17	1.29*	(0.7041)	18	0.29	(0.6474)	18
Best-*Sharpe*	2.91	(0.0000)	31	32.10	(0.6806)	27	1.35	(0.4139)	28	0.30	(0.3540)	27

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \le w_{it} \le 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Regarding the 5-step and 22-step ahead forecasting horizon (Table 4.2 and Table 4.3), we can see that for the MAE loss function, the OLS_{SQRT} model has still the best forecasting performance, as it passes the SPA test and is the only model included to the MCS. Under the MSE loss function, all combinations based on CER and two combinations based on Sharpe ratio reject the null hypothesis of superior forecasting performance. However, the Best-$Q_{0.05}$ combination produces the smallest losses, followed by the least squares combinations and two single models (i.e. the EGARCH and the APARCH models).

Under the HR and QLIKE loss functions, the EGARCH model that accounts for potential asymmetric response of volatility on positive and negative innovations is considered as the best performing model. The only single models that pass the SPA test are the EGARCH, the GJR, and the APARCH models. It is worth noting that only economic combinations based on Smoothed-Q loss function pass the SPA test while the rest economic combinations are proved inadequate.

Table 4.2 5-step ahead Statistical Forecasting Performance for GARCH combinations

	MAE			MSE			HRLF *(b=-1)*			QLIKE		
	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank
GARCH	3.21	(0.0000)	30	39.00	(0.0027)	30	1.76	(0.0000)	31	0.39	(0.0014)	32
EGARCH	2.81	(0.0000)	8	34.19*	(0.8391)	5	**1.48***	(1.0000)	1	**0.33***	(1.0000)	1
GJR	2.87	(0.0000)	10	34.95	(0.3930)	11	1.53*	(0.3875)	6	0.33	(0.6449)	2
APARCH	2.92	(0.0000)	16	34.68*	(0.4883)	8	1.53*	(0.3893)	7	0.34	(0.5063)	3
FIGARCH	3.03	(0.0000)	24	37.57	(0.0189)	27	1.67	(0.0006)	28	0.37	(0.0278)	28
HYGARCH	3.07	(0.0000)	28	38.24	(0.0024)	29	1.69	(0.0000)	29	0.36	(0.0971)	22
Mean	2.93	(0.0000)	21	35.44	(0.2692)	19	1.56	(0.1312)	17	0.35	(0.3094)	10
Geometric Mean	2.91	(0.0000)	14	35.25	(0.3110)	14	1.55*	(0.1764)	12	0.34	(0.3591)	7
Harmonic Mean	2.88	(0.0000)	11	35.08	(0.3616)	13	1.55*	(0.2428)	10	0.34	(0.4229)	4
Trimmed Mean	2.92	(0.0000)	17	35.44	(0.2704)	20	1.56	(0.1465)	15	0.34	(0.3637)	6
Median	2.92	(0.0000)	15	35.51	(0.2476)	21	1.56	(0.1353)	16	0.34	(0.3730)	5
AFTER	2.77	(0.0000)	5	33.98*	(0.9603)	3	1.52*	(0.4810)	5	0.35	(0.1694)	20
DMSFE	2.93	(0.0000)	19	35.42	(0.2864)	18	1.58	(0.0772)	20	0.35	(0.1961)	14
IMSFE	2.98	(0.0000)	23	35.72	(0.2250)	22	1.60	(0.0299)	22	0.36	(0.1199)	21
Kernel	2.61	(0.0001)	3	37.70	(0.1007)	28	1.62	(0.0540)	23	0.36	(0.0856)	23
OLS_{EXP}	2.56	(0.0000)	2	34.09*	(0.8617)	4	1.49*	(0.7570)	2	0.35	(0.2375)	16
OLS	2.77	(0.0000)	6	33.96*	(0.9877)	2	1.51*	(0.6479)	3	0.35	(0.3239)	9
OLS_{SQRT}	**2.36***	(1.0000)	1	34.47*	(0.7141)	7	1.56*	(0.2357)	13	0.38	(0.0493)	30
TW	2.86	(0.0000)	9	35.02	(0.3797)	12	1.54*	(0.2576)	9	0.34	(0.3676)	8
Trimmed MSPE	2.92	(0.0000)	18	35.42	(0.2808)	17	1.57	(0.0747)	19	0.35	(0.2073)	12
TW-*CER*	3.07	(0.0000)	27	37.17	(0.0494)	25	1.66	(0.0011)	26	0.37	(0.0406)	26
Trimmed-*CER*	3.04	(0.0000)	25	36.60	(0.0953)	23	1.64	(0.0039)	24	0.36	(0.0585)	24
Best-*CER*	3.22	(0.0000)	31	39.63	(0.0022)	31	1.75	(0.0000)	30	0.38	(0.0060)	29
TW-$Q_{0.05}$	2.88	(0.0000)	12	34.68*	(0.4899)	9	1.55*	(0.2052)	11	0.35	(0.2121)	11
TW-$Q_{0.01}$	2.89	(0.0000)	13	34.82*	(0.4321)	10	1.56	(0.1538)	14	0.35	(0.1728)	18
Trimmed-$Q_{0.05}$	2.93	(0.0000)	20	35.26	(0.3180)	15	1.57	(0.0973)	18	0.35	(0.2046)	13
Trimmed-$Q_{0.01}$	2.94	(0.0000)	22	35.29	(0.3107)	16	1.58	(0.0737)	21	0.35	(0.1770)	19
Best-$Q_{0.05}$	2.76	(0.0000)	4	**33.90***	(1.0000)	1	1.52*	(0.5726)	4	0.35	(0.2106)	15
Best-$Q_{0.01}$	2.79	(0.0000)	7	34.38*	(0.7112)	6	1.54*	(0.3302)	8	0.35	(0.1888)	17
TW-*Sharpe*	3.08	(0.0000)	29	37.18	(0.0449)	26	1.66	(0.0007)	27	0.37	(0.0279)	27
Trimmed-*Sharpe*	3.04	(0.0000)	25	36.60	(0.1002)	23	1.64	(0.0027)	24	0.36	(0.0578)	24
Best-*Sharpe*	3.24	(0.0000)	32	39.71	(0.0010)	32	1.77	(0.0000)	32	0.39	(0.0017)	31

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \le w_{it} \le 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Table 4.3 22-step ahead Statistical Forecasting Performance for GARCH combinations

	MAE			MSE			HRLF *(b=-1)*			QLIKE		
	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank
GARCH	3.88	(0.0000)	30	51.14	(0.0000)	30	2.34	(0.0000)	30	0.48	(0.0080)	30
EGARCH	3.06	(0.0000)	8	38.85*	(0.6010)	5	**1.75***	(1.0000)	1	**0.39***	(1.0000)	1
GJR	3.28	(0.0000)	12	40.74	(0.0544)	9	1.86	(0.1976)	7	0.40	(0.4888)	3
APARCH	3.34	(0.0000)	14	40.74	(0.0758)	8	1.87	(0.1726)	8	0.41	(0.3992)	5
FIGARCH	3.56	(0.0000)	24	46.55	(0.0001)	26	2.17	(0.0000)	26	0.46	(0.0455)	28
HYGARCH	3.63	(0.0000)	27	47.61	(0.0000)	27	2.18	(0.0000)	28	0.44	(0.1425)	23
Mean	3.41	(0.0000)	20	42.97	(0.0003)	21	1.97	(0.0089)	20	0.42	(0.2921)	14
Geometric Mean	3.37	(0.0000)	16	42.61	(0.0005)	14	1.96	(0.0115)	15	0.42	(0.3042)	12
Harmonic Mean	3.45	(0.0000)	23	43.85	(0.0003)	23	2.01	(0.0016)	23	0.43	(0.1782)	22
Trimmed Mean	3.40	(0.0000)	19	42.89	(0.0002)	20	1.97	(0.0099)	19	0.42	(0.2900)	13
Median	3.38	(0.0000)	18	42.70	(0.0004)	15	1.96	(0.0126)	16	0.42	(0.3109)	11
AFTER	3.02	(0.0000)	7	38.73*	(0.7341)	4	1.80*	(0.5205)	5	0.41	(0.1352)	6
DMSFE	3.34	(0.0000)	13	42.72	(0.0017)	16	1.96	(0.0073)	14	0.42	(0.2558)	16
IMSFE	3.45	(0.0000)	22	43.85	(0.0001)	22	2.01	(0.0012)	22	0.43	(0.1795)	21
Kernel	2.74	(0.0000)	3	41.18*	(0.0537)	10	1.89	(0.0713)	9	0.42	(0.1537)	9
OLS_{EXP}	2.68	(0.0003)	2	39.28*	(0.4367)	6	1.84*	(0.2404)	6	0.43	(0.1089)	19
OLS	3.01	(0.0000)	4	38.72*	(0.7469)	3	1.79*	(0.6034)	3	0.41	(0.2888)	4
OLS_{SQRT}	**2.50***	(1.0000)	1	39.32*	(0.2922)	7	1.93	(0.0849)	13	0.51	(0.0360)	32
TW	3.27	(0.0000)	9	41.83	(0.0049)	13	1.91	(0.0380)	11	0.42	(0.3464)	8
Trimmed MSPE	3.36	(0.0000)	15	42.87	(0.0010)	19	1.96	(0.0066)	17	0.42	(0.2528)	15
TW-*CER*	3.69	(0.0000)	29	47.97	(0.0000)	29	2.18	(0.0000)	27	0.46	(0.0530)	26
Trimmed-*CER*	3.58	(0.0000)	25	45.97	(0.0000)	24	2.10	(0.0000)	24	0.45	(0.1122)	24
Best-*CER*	3.96	(0.0000)	31	54.28	(0.0000)	31	2.38	(0.0000)	31	0.48	(0.0062)	29
TW-$Q_{0.05}$	3.27	(0.0000)	10	41.26	(0.0207)	11	1.90	(0.0615)	10	0.42	(0.3335)	10
TW-$Q_{0.01}$	3.28	(0.0000)	11	41.33	(0.0147)	12	1.92	(0.0270)	12	0.42	(0.2468)	18
Trimmed-$Q_{0.05}$	3.37	(0.0000)	17	42.74	(0.0012)	17	1.96	(0.0077)	18	0.42	(0.2652)	17
Trimmed-$Q_{0.01}$	3.41	(0.0000)	21	42.86	(0.0002)	18	1.99	(0.0028)	21	0.43	(0.1932)	20
Best-$Q_{0.05}$	3.02	(0.0000)	6	**38.49***	(1.0000)	1	1.77*	(0.7845)	2	0.40	(0.5101)	2
Best-$Q_{0.01}$	3.02	(0.0000)	5	38.71*	(0.7534)	2	1.80*	(0.5438)	4	0.41	(0.1498)	7
TW-*Sharpe*	3.68	(0.0000)	28	47.82	(0.0000)	28	2.18	(0.0000)	29	0.46	(0.0399)	27
Trimmed-*Sharpe*	3.58	(0.0000)	25	45.97	(0.0000)	24	2.10	(0.0000)	24	0.45	(0.1063)	24
Best-*Sharpe*	4.04	(0.0000)	32	54.57	(0.0000)	32	2.43	(0.0000)	32	0.50	(0.0006)	31

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Economic Evaluation

In this section, we examine the economic gains of introducing economic combinations in volatility forecasting. This is achieved through the calculation of *CER* and *Sharpe* ratios. To test whether the *CER* and the *Sharpe* ratio based on each forecasting model is significantly larger than the aforementioned loss functions of the rest models, we use the StepSPA test proposed by Hsu et al. (2010). The following tables report our findings for both *CER* and *Sharpe* ratio for the one-step ahead forecast horizon[10].Testing for the economic gains derived from the GARCH models and the combinations based on them, we find that the forecasting performance of each model varies greatly with the change in horizons. The results for the 1-step ahead forecasting horizon are presented in Table 4.4 and indicate the superior performance of an economic combination technique, the Best-*CER*, that produces the largest gains from a mean-variance investor under the *CER* measure and the *Sharpe* ratio. The StepSPA test indicates the adequate performance for most of the economic combination techniques, while the transformed least squares combination produce the smallest economic gains. It is interesting to note that the transformed least squares schemes that indicated the good forecasting performance according to statistical loss functions, are found to lead an investor to losses.

[10] In this section, we present the results for a risk aversion parameter equal to 3. We use several other values ranged from 1 to 5, but all lead to quite similar results.

Table 4.4 1-step ahead Economic Forecasting Performance for GARCH combinations

	CER			Sharpe		
	Loss function	StepSPA	Rank	Loss Function	StepSPA	Rank
GARCH	0.0307	(0.5090)	17	3.15	(0.5060)	15
EGARCH	0.0220	(0.3000)	30	2.32	(0.5740)	30
GJR	0.0278	(0.3270)	27	2.85	(0.4670)	27
APARCH	0.0236	(0.2940)	29	2.53	(0.3010)	29
FIGARCH	0.0364	(0.9150)	3	3.57	(0.9120)	5
HYGARCH	0.0318	(0.4700)	12	3.06	(0.6670)	19
Mean	0.0302	(0.4670)	19	3.03	(0.6090)	20
Geometric Mean	0.0295	(0.4670)	23	2.96	(0.6000)	25
Harmonic Mean	0.0291	(0.4570)	25	2.92	(0.5650)	26
Trimmed Mean	0.0295	(0.4380)	21	2.99	(0.4920)	24
Median	0.0295	(0.4120)	22	3.00	(0.5010)	22
AFTER	0.0329	(0.3290)	10	3.27	(0.3950)	11
DMSFE	0.0301	(0.0190)	20	3.10	(0.0220)	17
IMSFE	0.0316	(0.0350)	13	3.24	(0.0660)	12
Kernel	0.0336	(0.6090)	7	3.62	(0.4180)	3
OLS_{EXP}	-0.0239	(0.0910)	31	0.21	(0.0560)	31
OLS	0.0294	(0.1510)	24	2.99	(0.1950)	23
OLS_{SQRT}	-0.0367	(0.0970)	32	0.18	(0.0730)	32
TW	0.0311	(0.3240)	16	3.14	(0.3620)	16
Trimmed MSPE	0.0312	(0.2150)	14	3.18	(0.1700)	14
TW-*CER*	0.0354	(0.7240)	4	3.55	(0.5680)	6
Trimmed-*CER*	0.0331	(0.0960)	8	3.38	(0.1050)	8
Best-*CER*	**0.0386**	(1.0000)	1	**3.76**	(1.0000)	1
TW-$Q_{0.05}$	0.0319	(0.3830)	11	3.31	(0.2770)	10
TW-$Q_{0.01}$	0.0291	(0.0860)	26	3.01	(0.0470)	21
Trimmed-$Q_{0.05}$	0.0311	(0.1210)	15	3.21	(0.1480)	13
Trimmed-$Q_{0.01}$	0.0302	(0.1030)	18	3.10	(0.0920)	18
Best-$Q_{0.05}$	0.0344	(0.7010)	6	3.58	(0.5320)	4
Best-$Q_{0.01}$	0.0239	(0.1360)	28	2.57	(0.1140)	28
TW-*Sharpe*	0.0351	(0.6940)	5	3.53	(0.4520)	7
Trimmed-*Sharpe*	0.0331	(0.1130)	8	3.38	(0.0750)	8
Best-*Sharpe*	0.0376	(0.7760)	2	3.69	(0.7350)	2

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hsu et al. (2010) StepSPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 1,000 and the smoothing parameter for the mean block length is 0.1. The confidence level used is α=10%. 3. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 4. EXP and SQRT denote the exponential and the square root transformation, respectively.

We continue the analysis with the evaluation of the 5- and 22-step ahead forecasts. The results are presented in Tables 4.5 and 4.6. Regarding the 5-step ahead forecasting horizon, the FIGARCH model that accounts for long memory and asymmetric effects is considered as the best performing model for both loss functions. Under the *Sharpe* ratio, six economic combinations are considered amongst the best performers, while the transformed least squares combinations are ranked last. For the 22-step ahead forecasting horizon, the *CER* evaluation criterion selects the Best-*CER* combination, followed by the Best-*Sharpe* and the TW-*CER* combinations. The results are statistically significant, according to StepSPA test, implying that the proposed combination methodologies can produce larger portfolio returns with lower risk. It is interesting to note that the OLS combinations are proved inadequate and lead an investor to lower gains according to the economic loss functions.

Table 4.5 5-step ahead Economic Forecasting Performance for GARCH combinations

	CER			Sharpe		
	Loss function	StepSPA	Rank	Loss function	StepSPA	Rank
GARCH	0.0410	(0.8480)	8	4.39	(0.6810)	14
EGARCH	0.0401	(0.8530)	16	3.65	(0.3000)	30
GJR	0.0400	(0.8540)	18	4.07	(0.3290)	27
APARCH	0.0420	(0.9730)	2	4.21	(0.6330)	24
FIGARCH	**0.0433**	(1.0000)	1	**4.77**	(1.0000)	1
HYGARCH	0.0389	(0.7630)	26	3.95	(0.0100)	29
Mean	0.0417	(0.9930)	4	4.32	(0.5550)	19
Geometric Mean	0.0414	(0.9950)	6	4.25	(0.2960)	23
Harmonic Mean	0.0412	(0.9960)	7	4.19	(0.3310)	25
Trimmed Mean	0.0417	(0.9960)	5	4.31	(0.5760)	20
Median	0.0405	(0.3460)	14	4.19	(0.1920)	26
AFTER	0.0408	(0.8820)	9	4.54	(0.8790)	9
DMSFE	0.0380	(0.2990)	29	4.28	(0.0030)	22
IMSFE	0.0400	(0.9020)	19	4.60	(0.3210)	7
Kernel	0.0363	(0.4100)	30	3.97	(0.1140)	28
OLS_{EXP}	0.0281	(0.5580)	31	2.60	(0.1070)	31
OLS	0.0390	(0.7610)	25	4.31	(0.1560)	21
OLS_{SQRT}	0.0136	(0.4460)	32	2.00	(0.0010)	32
TW	0.0392	(0.8230)	23	4.33	(0.2580)	18
Trimmed MSPE	0.0399	(0.9070)	20	4.47	(0.3530)	12
TW-*CER*	0.0400	(0.6180)	17	4.61	(0.5080)	6
Trimmed-*CER*	0.0405	(0.7970)	11	4.71	(0.9360)	2
Best-*CER*	0.0405	(0.7650)	13	4.51	(0.6570)	10
TW-$Q_{0.05}$	0.0398	(0.9360)	21	4.50	(0.3110)	11
TW-$Q_{0.01}$	0.0403	(0.9410)	15	4.59	(0.6520)	8
Trimmed-$Q_{0.05}$	0.0384	(0.5470)	28	4.35	(0.0200)	17
Trimmed-$Q_{0.01}$	0.0385	(0.3440)	27	4.46	(0.1410)	13
Best-$Q_{0.05}$	0.0390	(0.4420)	24	4.37	(0.4310)	16
Best-$Q_{0.01}$	0.0394	(0.6650)	22	4.38	(0.5160)	15
TW-*Sharpe*	0.0408	(0.9190)	10	4.70	(0.9950)	4
Trimmed-*Sharpe*	0.0405	(0.8190)	11	4.71	(0.9270)	2
Best-*Sharpe*	0.0419	(0.9020)	3	4.68	(0.9100)	5

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hsu et al. (2010) StepSPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 1,000 and the smoothing parameter for the mean block length is 0.1. The confidence level used is α=10%. 3. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \le w_{it} \le 1.5$ for the economic combinations. 4. EXP and SQRT denote the exponential and the square root transformation, respectively.

Table 4.6 22-step ahead Economic Forecasting Performance for GARCH combinations

	CER			Sharpe		
	Loss function	StepSPA	Rank	Loss function	StepSPA	Rank
GARCH	0.0257	(0.1650)	28	3.49	(0.2270)	27
EGARCH	0.0192	(0.2330)	29	2.41	(0.3280)	29
GJR	0.0266	(0.2330)	27	3.41	(0.3340)	28
APARCH	0.0282	(0.3240)	24	3.66	(0.4530)	22
FIGARCH	0.0331	(0.6600)	5	4.35	(0.7550)	7
HYGARCH	0.0304	(0.7110)	19	3.52	(0.9430)	25
Mean	0.0287	(0.4930)	22	3.66	(0.4440)	21
Geometric Mean	0.0280	(0.3190)	26	3.56	(0.4120)	24
Harmonic Mean	0.0319	(0.2710)	10	4.30	(0.4970)	9
Trimmed Mean	0.0290	(0.5000)	21	3.70	(0.6150)	20
Median	0.0285	(0.4020)	23	3.65	(0.5640)	23
AFTER	0.0311	(0.5310)	17	4.02	(0.6900)	18
DMSFE	0.0312	(0.2940)	16	4.14	(0.3890)	14
IMSFE	0.0319	(0.3030)	11	4.30	(0.5380)	10
Kernel	0.0148	(0.0630)	30	1.99	(0.0070)	30
OLS_{EXP}	-0.0099	(0.2400)	31	0.89	(0.6700)	31
OLS	0.0302	(0.4750)	20	3.85	(0.6580)	19
OLS_{SQRT}	-0.0132	(0.1580)	32	0.55	(0.5210)	32
TW	0.0316	(0.2990)	13	4.14	(0.4240)	15
Trimmed MSPE	0.0318	(0.3400)	12	4.22	(0.4720)	12
TW-*CER*	0.0342	(0.2240)	3	4.72	(0.3570)	3
Trimmed-*CER*	0.0324	(0.2230)	6	4.44	(0.4470)	5
Best-*CER*	**0.0383**	(1.0000)	1	5.29	(0.5940)	2
TW-$Q_{0.05}$	0.0306	(0.2440)	18	4.03	(0.3890)	17
TW-$Q_{0.01}$	0.0320	(0.3770)	8	4.28	(0.6080)	11
Trimmed-$Q_{0.05}$	0.0313	(0.2920)	15	4.15	(0.4180)	13
Trimmed-$Q_{0.01}$	0.0315	(0.1040)	14	4.31	(0.2160)	8
Best-$Q_{0.05}$	0.0281	(0.3180)	25	3.52	(0.5130)	26
Best-$Q_{0.01}$	0.0320	(0.5560)	9	4.14	(0.7470)	16
TW-*Sharpe*	0.0335	(0.3150)	4	4.70	(0.3850)	4
Trimmed-*Sharpe*	0.0324	(0.2610)	6	4.44	(0.4050)	5
Best-*Sharpe*	0.0364	(0.6300)	2	**5.36**	(1.0000)	1

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hsu et al. (2010) StepSPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 1,000 and the smoothing parameter for the mean block length is 0.1. The confidence level used is α=10%. 3. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \le w_{it} \le 1.5$ for the economic combinations. 4. EXP and SQRT denote the exponential and the square root transformation, respectively.

Risk Management Evaluation

This section provides the results of volatility forecasting performance using a VaR-based loss function, the Smoothed-*Q*. The parameter of smooth function is set equal to 25, while the VaR is calculated for two confidence levels, 5% and 1%. For the 1-step ahead forecasts, the results are presented in Table 4.7. Several models seem to adequate predict the VaR for 5% confidence level, while for the 1% confidence level, there is no model that calculates the VaR adequately. The results indicate that the Best-$Q_{0.05}$ outperform the other models, in terms of the Smoothed-*Q* loss function under a 5% confidence level. The large Hansen's p-value makes it the preferred method of calculation, despite three single models, i.e. the EGARCH, GJR and APARCH models are amongst the best performers. Taking into consideration the 1% confidence level, the results clearly indicate that the OLS combination generates the smaller losses. The results are noteworthy because contrary to the previous section that accounts for the economic gains derived from combination forecasts, an OLS combination performs better in a risk management framework. Furthermore, the simple combination techniques seem to perform quite better compared to economic combinations and the single models.

Table 4.7 1-step ahead VaR Forecasting Performance for GARCH combinations

	Smoothed-$Q_{0.05}$				Smoothed-$Q_{0.01}$			
	Percentage of violations (5%)	Loss function	SPA	Rank	Percentage of violations (1%)	Loss function	SPA	Rank
GARCH	4.85%	0.2231	(0.0035)	32	1.43%	0.0605	(0.3798)	28
EGARCH	4.38%	0.2132	(0.8294)	2	1.67%	0.0595	(0.5919)	21
GJR	4.61%	0.2162	(0.4450)	6	1.67%	0.0597	(0.5484)	24
APARCH	**4.30%**	0.2152	(0.5818)	3	1.51%	0.0597	(0.5453)	23
FIGARCH	5.33%	0.2223	(0.0243)	30	1.75%	0.0599	(0.4740)	25
HYGARCH	5.33%	0.2211	(0.0503)	27	1.51%	0.0597	(0.5190)	22
Mean	4.77%	0.2168	(0.3112)	12	1.51%	0.0584	(0.9165)	5
Geometric Mean	4.93%	0.2173	(0.2308)	15	1.51%	0.0584	(0.9104)	6
Harmonic Mean	5.01%	0.2177	(0.1731)	18	1.51%	0.0585	(0.8920)	7
Trimmed Mean	4.61%	0.2171	(0.2579)	13	1.51%	0.0585	(0.8862)	8
Median	4.69%	0.2174	(0.2277)	16	1.51%	0.0588	(0.8062)	12
AFTER	5.41%	0.2205	(0.0396)	24	1.43%	0.0593	(0.5929)	19
DMSFE	4.93%	0.2167	(0.3265)	10	1.51%	0.0583	(0.9422)	4
IMSFE	4.77%	0.2166	(0.3519)	9	1.43%	0.0586	(0.8660)	9
Kernel	5.17%	0.2218	(0.0363)	28	2.07%	0.0635	(0.0553)	31
OLS_{EXP}	6.21%	0.2163	(0.4575)	7	2.31%	0.0605	(0.3508)	29
OLS	5.25%	0.2176	(0.2186)	17	1.59%	**0.0578**	(1.0000)	1
OLS_{SQRT}	7.56%	0.2207	(0.0792)	25	2.94%	0.0652	(0.0255)	32
TW	5.01%	0.2161	(0.4279)	5	1.59%	0.0581	(0.9736)	2
Trimmed MSPE	5.01%	0.2164	(0.3491)	8	1.43%	0.0582	(0.9385)	3
TW-*CER*	4.93%	0.2186	(0.1098)	22	1.59%	0.0587	(0.7661)	10
Trimmed-*CER*	4.77%	0.2179	(0.1821)	19	1.43%	0.0589	(0.7334)	14
Best-*CER*	5.01%	0.2210	(0.0307)	26	**1.35%**	0.0594	(0.5735)	20
TW-$Q_{0.05}$	4.69%	0.2155	(0.4945)	4	1.51%	0.0591	(0.7043)	18
TW-$Q_{0.01}$	4.77%	0.2183	(0.1398)	21	1.51%	0.0590	(0.7493)	16
Trimmed-$Q_{0.05}$	4.69%	0.2168	(0.3477)	11	1.43%	0.0589	(0.7733)	13
Trimmed-$Q_{0.01}$	4.93%	0.2172	(0.2779)	14	1.43%	0.0591	(0.7078)	17
Best-$Q_{0.05}$	4.77%	**0.2130**	(1.0000)	1	1.43%	0.0602	(0.4345)	27
Best-$Q_{0.01}$	5.01%	0.2228	(0.0128)	31	1.43%	0.0617	(0.1900)	30
TW-*Sharpe*	4.93%	0.2188	(0.1031)	23	1.59%	0.0588	(0.7252)	11
Trimmed-*Sharpe*	4.77%	0.2179	(0.1784)	19	1.43%	0.0589	(0.7330)	14
Best-*Sharpe*	5.17%	0.2220	(0.0131)	29	**1.35%**	0.0600	(0.4844)	26

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. The Smoothed-Q loss function is calculated for VaR level $\alpha=0.01$ and α=0.05. We set the smoothness parameter $\delta=25$. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

The results in Tables 4.8 and 4.9, considering the 5-step and 22-step ahead forecasts, show differences in the most preferred model. In both confidence levels, the single EGARCH model clearly dominates the rest combination and single models. Its superior forecasting performance is indicated by the large SPA value and can be attributed to the fact that the EGARCH model allows for testing asymmetries. Consequently, we conclude that for the GARCH combinations, the economic combinations based on *CER* and *Sharpe* ratio seem to provide economic gains for the 1-step and 22-step ahead forecast horizons, while the combinations based on risk management loss functions do not improve in a significant amount the forecasting performance in both statistical and economic terms.

Table 4.8 5-step ahead VaR Forecasting Performance for GARCH combinations

	Smoothed-$Q_{0.05}$				Smoothed-$Q_{0.01}$			
	Percentage of violations (5%)	Loss function	SPA	Rank	Percentage of violations (1%)	Loss function	SPA	Rank
GARCH	4.71%	0.2288	(0.0000)	32	1.68%	0.0658	(0.0434)	31
EGARCH	4.39%	**0.2130**	(1.0000)	1	1.36%	**0.0602**	(1.0000)	1
GJR	4.47%	0.2167	(0.3217)	8	1.44%	0.0617	(0.5628)	8
APARCH	4.39%	0.2165	(0.3453)	7	1.28%	0.0620	(0.5172)	19
FIGARCH	4.71%	0.2257	(0.0012)	30	1.28%	0.0637	(0.1727)	26
HYGARCH	4.87%	0.2245	(0.0020)	28	1.36%	0.0637	(0.1438)	27
Mean	4.47%	0.2195	(0.0797)	20	1.28%	0.0618	(0.5476)	11
Geometric Mean	4.55%	0.2193	(0.0827)	18	1.44%	0.0618	(0.5486)	12
Harmonic Mean	4.55%	0.2192	(0.0901)	15	1.52%	0.0619	(0.5068)	18
Trimmed Mean	4.47%	0.2192	(0.0892)	16	1.60%	0.0619	(0.5174)	17
Median	4.47%	0.2194	(0.0848)	19	1.60%	0.0623	(0.3926)	21
AFTER	4.55%	0.2154	(0.4628)	4	1.36%	0.0618	(0.5870)	9
DMSFE	4.39%	0.2193	(0.0952)	17	1.52%	0.0619	(0.5431)	13
IMSFE	**4.31%**	0.2197	(0.0847)	21	1.20%	0.0620	(0.4735)	20
Kernel	5.83%	0.2227	(0.0183)	25	2.00%	0.0645	(0.1634)	30
OLS_{EXP}	6.86%	0.2154	(0.4598)	3	2.31%	0.0608	(0.6436)	2
OLS	4.63%	0.2145	(0.4918)	2	1.36%	0.0611	(0.7841)	3
OLS_{SQRT}	8.30%	0.2236	(0.0262)	27	3.35%	0.0702	(0.0071)	32
TW	4.63%	0.2183	(0.1538)	11	1.44%	0.0613	(0.7358)	4
Trimmed MSPE	4.55%	0.2189	(0.1189)	14	1.44%	0.0619	(0.5502)	14
TW-*CER*	4.39%	0.2226	(0.0138)	24	**1.20%**	0.0629	(0.2583)	24
Trimmed-*CER*	**4.31%**	0.2216	(0.0287)	22	**1.20%**	0.0627	(0.2840)	22
Best-*CER*	4.71%	0.2256	(0.0018)	29	1.28%	0.0639	(0.2100)	28
TW-$Q_{0.05}$	4.39%	0.2175	(0.2289)	9	1.36%	0.0617	(0.6124)	6
TW-$Q_{0.01}$	4.39%	0.2178	(0.1956)	10	1.36%	0.0619	(0.5076)	16
Trimmed-$Q_{0.05}$	**4.31%**	0.2186	(0.1482)	12	**1.20%**	0.0618	(0.5663)	10
Trimmed-$Q_{0.01}$	**4.31%**	0.2188	(0.1377)	13	**1.20%**	0.0619	(0.5259)	15
Best-$Q_{0.05}$	4.63%	0.2157	(0.4294)	6	1.36%	0.0615	(0.6647)	5
Best-$Q_{0.01}$	4.63%	0.2154	(0.4667)	5	1.36%	0.0617	(0.6181)	7
TW-*Sharpe*	4.39%	0.2228	(0.0112)	26	**1.20%**	0.0630	(0.2463)	25
Trimmed-*Sharpe*	**4.31%**	0.2216	(0.0278)	22	**1.20%**	0.0627	(0.2868)	22
Best-*Sharpe*	4.63%	0.2263	(0.0013)	31	1.28%	0.0641	(0.1791)	29

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. The Smoothed-Q loss function is calculated for VaR level $\alpha=0.01$ and α=0.05. We set the smoothness parameter $\delta=25$. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Table 4.9 22-step ahead VaR Forecasting Performance for GARCH combinations

	Smoothed-$Q_{0.05}$				Smoothed-$Q_{0.01}$			
	Percentage of violations (5%)	Loss function	SPA	Rank	Percentage of violations (1%)	Loss function	SPA	Rank
GARCH	4.94%	0.2414	(0.0021)	32	1.62%	0.0665	(0.0496)	27
EGARCH	5.34%	**0.2247**	(1.0000)	1	1.29%	**0.0602**	(1.0000)	1
GJR	4.94%	0.2282	(0.4087)	6	1.21%	0.0616	(0.4953)	3
APARCH	4.85%	0.2272	(0.5179)	3	**1.13%**	0.0609	(0.6878)	2
FIGARCH	5.26%	0.2403	(0.0022)	31	1.70%	0.0653	(0.1418)	24
HYGARCH	5.50%	0.2361	(0.0242)	27	1.38%	0.0644	(0.2226)	21
Mean	5.10%	0.2316	(0.1600)	20	1.21%	0.0620	(0.3079)	5
Geometric Mean	5.10%	0.2316	(0.1510)	19	1.29%	0.0621	(0.2259)	8
Harmonic Mean	4.77%	0.2315	(0.1629)	18	1.46%	0.0641	(0.1502)	20
Trimmed Mean	5.10%	0.2315	(0.1659)	16	1.29%	0.0620	(0.2726)	7
Median	5.10%	0.2312	(0.1847)	14	1.29%	0.0620	(0.2750)	6
AFTER	5.58%	0.2281	(0.3993)	5	1.62%	0.0627	(0.1945)	11
DMSFE	5.02%	0.2302	(0.2434)	11	1.46%	0.0638	(0.1764)	18
IMSFE	4.77%	0.2315	(0.1676)	17	1.46%	0.0641	(0.1469)	19
Kernel	6.39%	0.2333	(0.0953)	23	2.51%	0.0715	(0.0053)	31
OLS_{EXP}	7.28%	0.2339	(0.1021)	24	2.91%	0.0699	(0.0171)	30
OLS	5.50%	0.2278	(0.4476)	4	1.62%	0.0623	(0.2626)	9
OLS_{SQRT}	9.06%	0.2390	(0.0255)	29	3.88%	0.0802	(0.0005)	32
TW	5.02%	0.2292	(0.3158)	9	1.46%	0.0625	(0.2023)	10
Trimmed MSPE	4.94%	0.2303	(0.2360)	13	1.46%	0.0634	(0.2168)	13
TW-*CER*	4.69%	0.2348	(0.0564)	25	1.54%	0.0657	(0.0946)	25
Trimmed-*CER*	4.69%	0.2332	(0.1027)	21	1.38%	0.0650	(0.1483)	22
Best-*CER*	**4.53%**	0.2379	(0.0168)	28	1.46%	0.0674	(0.0265)	28
TW-$Q_{0.05}$	4.85%	0.2291	(0.3241)	8	1.38%	0.0629	(0.1116)	12
TW-$Q_{0.01}$	4.94%	0.2298	(0.2764)	10	1.62%	0.0635	(0.2107)	16
Trimmed-$Q_{0.05}$	4.94%	0.2303	(0.2264)	12	1.46%	0.0635	(0.2212)	14
Trimmed-$Q_{0.01}$	4.85%	0.2313	(0.1760)	15	1.46%	0.0637	(0.1861)	17
Best-$Q_{0.05}$	5.34%	0.2272	(0.5607)	2	1.46%	0.0618	(0.4024)	4
Best-$Q_{0.01}$	5.58%	0.2284	(0.3716)	7	1.54%	0.0635	(0.1251)	15
TW-*Sharpe*	4.77%	0.2354	(0.0444)	26	1.46%	0.0659	(0.0828)	26
Trimmed-*Sharpe*	4.69%	0.2332	(0.1019)	21	1.38%	0.0650	(0.1442)	22
Best-*Sharpe*	**4.53%**	0.2403	(0.0058)	30	1.46%	0.0681	(0.0115)	29

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. The Smoothed-Q loss function is calculated for VaR level $\alpha=0.01$ and α=0.05. We set the smoothness parameter $\delta=25$. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

4.7.2 Out-of-Sample Forecasting Performance for HAR combinations

Statistical Evaluation

To test the forecasting accuracy of the HAR combinations, we use the symmetric MAE and MSE loss functions, as well as the asymmetric HRLF and QLIKE loss functions. Table 4.10 reports the SPA and MCS tests results of the HAR models and the combinations based on HAR models for 1-step-ahead forecasts. From both MAE and MSE loss functions, we can conclude that the best performing model is the square root transformed least squares regression model. It is interesting to note that for the MAE loss function only the best performing model passes the SPA test, while it is the only model included in the optimal set. Turning to the economic combinations, although seven of them are ranked to the first 10 models, they cannot beat the OLS_{SQRT} model. Under the MSE loss function, the null hypothesis that none of the models can beat the benchmark cannot be rejected, while four combinations and two HAR models are included to the optimal set.

For the asymmetric HR loss function, the HAR-RSV model is considered as the best performing model followed by the OLS_{SQRT} and the trimmed mean combination. It is worth noting that only for the Best-*Sharpe* combination the null hypothesis is rejected. In this case, three models are included to the optimal set. Second, we consider the QLIKE loss function, which indicates the superior performance of the Trimmed Mean combination that passes both the SPA and the MCS tests.

Table 4.10 1-step ahead Statistical Forecasting Performance for HAR combinations

	MAE			MSE			HRLF *(b=-1)*			QLIKE		
	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank
HAR-RV	2.29	(0.0000)	30	26.55*	(0.6041)	4	1.00	(0.5536)	15	0.23	(0.6471)	23
LHAR-RV	2.29	(0.0000)	29	27.29	(0.5442)	12	1.01	(0.4672)	18	0.22	(0.6992)	20
HAR-RV-J	2.16	(0.0000)	12	29.24	(0.1166)	24	1.05	(0.2585)	24	0.22	(0.7176)	21
HAR-CJ	2.12	(0.0000)	2	29.64	(0.4477)	25	1.10	(0.1155)	25	0.24	(0.5012)	24
HAR-RSV	2.24	(0.0000)	27	24.33*	(0.6669)	2	**0.92***	(1.0000)	1	0.22	(0.8163)	13
Mean	2.17	(0.0000)	16	26.67	(0.3578)	6	0.95	(0.8391)	3	0.21	(0.9802)	2
Geometric Mean	2.17	(0.0000)	18	27.28	(0.2209)	11	0.97	(0.7192)	5	0.21	(0.9477)	5
Harmonic Mean	2.18	(0.0000)	19	27.89	(0.1403)	19	1.00	(0.6040)	17	0.22	(0.8852)	10
Trimmed Mean	2.17	(0.0000)	16	26.67*	(0.3571)	6	0.95*	(0.8393)	3	**0.21***	(0.9761)	1
Median	2.26	(0.0000)	28	26.92	(0.5800)	8	1.00	(0.5748)	16	0.22	(0.7510)	19
AFTER	2.19	(0.0000)	24	26.60*	(0.4771)	5	1.02	(0.4421)	19	0.23	(0.6252)	22
DMSFE	2.17	(0.0000)	14	27.33	(0.2044)	13	0.99	(0.6601)	9	0.21	(0.9609)	7
IMSFE	2.17	(0.0000)	13	27.39	(0.1906)	16	0.99	(0.5897)	12	0.21	(0.9588)	8
Kernel	2.58	(0.0000)	31	54.74	(0.0333)	31	1.64	(0.0130)	31	0.33	(0.3519)	30
OLS_{EXP}	2.19	(0.0000)	23	30.16	(0.4391)	28	1.20	(0.0876)	30	0.41	(0.0963)	31
OLS	2.18	(0.0000)	20	26.46*	(0.4643)	3	0.98	(0.6446)	8	0.22	(0.8816)	14
OLS_{SQRT}	**1.76***	(1.0000)	1	**23.78***	(1.0000)	1	0.93*	(0.8334)	2	0.26	(0.4893)	29
TW	2.16	(0.0000)	11	27.21	(0.2290)	10	0.98	(0.7055)	7	0.21	(0.9763)	3
Trimmed MSPE	2.17	(0.0000)	15	27.02	(0.2634)	9	0.98	(0.7528)	6	0.21	(0.9768)	4
TW-*CER*	2.15	(0.0000)	9	28.34	(0.1255)	23	1.03	(0.3325)	23	0.22	(0.7474)	18
Trimmed-*CER*	2.15	(0.0000)	6	27.37	(0.1880)	14	0.99	(0.6067)	10	0.22	(0.9268)	9
Best-*CER*	2.22	(0.0000)	25	30.34	(0.4200)	30	1.14	(0.1018)	29	0.24	(0.4655)	28
TW-$Q_{0.05}$	2.15	(0.0000)	10	28.23	(0.1251)	20	1.03	(0.3751)	21	0.22	(0.7818)	15
TW-$Q_{0.01}$	2.14	(0.0000)	4	28.29	(0.1297)	21	1.02	(0.3708)	20	0.22	(0.7736)	16
Trimmed-$Q_{0.05}$	2.14	(0.0000)	3	27.50	(0.1821)	17	0.99	(0.5794)	13	0.21	(0.9548)	6
Trimmed-$Q_{0.01}$	2.15	(0.0000)	5	27.52	(0.1776)	18	0.99	(0.5543)	14	0.22	(0.8854)	12
Best-$Q_{0.05}$	2.18	(0.0000)	21	30.14	(0.4363)	26	1.13	(0.1130)	26	0.24	(0.4968)	25
Best-$Q_{0.01}$	2.18	(0.0000)	22	30.15	(0.4357)	27	1.13	(0.1111)	27	0.24	(0.4929)	26
TW-*Sharpe*	2.15	(0.0000)	8	28.34	(0.1310)	22	1.03	(0.3438)	22	0.22	(0.7474)	17
Trimmed-*Sharpe*	2.15	(0.0000)	7	27.37	(0.1984)	15	0.99	(0.6085)	11	0.22	(0.9168)	11
Best-*Sharpe*	2.22	(0.0000)	26	30.34	(0.4272)	29	1.14	(0.0969)	28	0.24	(0.4718)	27

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \le w_{it} \le 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Table 4.11 shows the performance of single and combination models at the forecast horizon of 5 days. Under the MAE loss function, only the OLS_{SQRT} model passes the SPA test, while it is the only included model in the MCS. Using the MSE loss function, two single models, the HAR-CJ and the HAR-RV-J produce the smaller losses. In this case, only two models, including an economic combination do not pass the SPA test, while almost all the simple combinations and four economic combinations are included to the MCS. The superior forecasting performance is also indicated by Stock and Watson (2004). Taking into consideration, the asymmetric HR loss function, we note the superior predictive ability of the TW scheme, while the transformed least squares combinations are proved inadequate. In this case, only an economic combination, the Best-$Q_{00.1}$ does not pass the SPA test, as well as the Kernel regression, the OLS_{SQRT} and the HAR-RSV that are ran-ked amongst the worst performers. It is interesting to note that the OLS_{SQRT} model that was among the best performing models for the 1-step ahead forecasts, is ranked last in this case. For the QLIKE loss function, the TW combination is still the best performing model, while the OLS_{SQRT} is the worst performing combination. In this case all the economic combinations pass the SPA test, while nine of them are included to the optimal set.

Considering the 22-step ahead forecasting performance in Table 4.12, the results are quite different. For the MAE loss function, the OLS_{SQRT} model is still the best performing model followed by the Kernel regression, where both of them are included to the optimal set. Tsangari (2007) indicated the superior forecasting performance of more complex combination techniques such as the Kernel regression. Using the MSE loss function, the Kernel regression is considered as the best combination that passes both the SPA and MCS test. Although, more models are included to the MCS, only three models pass the SPA test, the Kernel regression, the Best-$Q_{0.05}$ and the single

HAR-CJ. For the asymmetric loss functions, the TW model is the best performing model, while more economic combinations pass both tests and are ranked amongst the best performers.

Table 4.11 5-step ahead Statistical Forecasting Performance for HAR combinations

	MAE			MSE			HRLF *(b=-1)*			QLIKE		
	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank
HAR-RV	2.76	(0.0000)	29	39.01	(0.1528)	28	1.65	(0.1448)	26	0.35	(0.1264)	26
LHAR-RV	2.76	(0.0000)	28	39.00	(0.1544)	27	1.64	(0.1551)	25	0.35	(0.1236)	27
HAR-RV-J	2.67	(0.0000)	6	37.01*	(0.6969)	2	1.58*	(0.7072)	3	0.34*	(0.4596)	22
HAR-CJ	2.66	(0.0001)	3	**36.58***	(1.0000)	1	1.59*	(0.5906)	9	0.34	(0.2479)	24
HAR-RSV	2.78	(0.0000)	31	39.56	(0.0971)	29	1.66	(0.0816)	29	0.35	(0.0968)	28
Mean	2.69	(0.0002)	10	37.49	(0.4844)	12	1.59	(0.5025)	13	0.34	(0.4731)	16
Geometric Mean	2.67	(0.0001)	8	37.29*	(0.5663)	6	1.59*	(0.5726)	11	0.34	(0.4665)	18
Harmonic Mean	2.66	(0.0004)	5	37.14*	(0.6487)	5	1.58*	(0.6245)	7	0.34	(0.4327)	19
Trimmed Mean	2.69	(0.0001)	10	37.49*	(0.4913)	12	1.59*	(0.5089)	13	0.34*	(0.4772)	16
Median	2.75	(0.0001)	27	38.82	(0.1799)	26	1.64	(0.1863)	24	0.34	(0.1491)	25
AFTER	2.70	(0.0000)	14	37.46*	(0.6046)	11	1.58*	(0.6946)	6	0.33*	(0.9356)	5
DMSFE	2.71	(0.0001)	16	37.57	(0.4463)	14	1.59*	(0.5599)	8	0.33*	(0.8680)	8
IMSFE	2.71	(0.0000)	20	37.70	(0.3980)	19	1.59	(0.4692)	15	0.33*	(0.8287)	9
Kernel	2.62	(0.0297)	2	39.84	(0.1244)	31	1.71	(0.0318)	30	0.37	(0.0323)	30
OLS_{EXP}	2.74	(0.0000)	24	38.55	(0.2417)	25	1.65	(0.1521)	27	0.35	(0.1833)	29
OLS	2.67	(0.0001)	7	37.12*	(0.6971)	4	1.57*	(0.8718)	2	0.33*	(0.9959)	2
OLS_{SQRT}	**2.44***	(1.0000)	1	38.23	(0.2811)	21	1.72	(0.0062)	31	0.40	(0.0051)	31
TW	2.66	(0.0003)	4	37.11*	(0.6744)	3	**1.57***	(1.0000)	1	**0.33***	(1.0000)	1
Trimmed MSPE	2.70	(0.0002)	13	37.59	(0.4440)	15	1.59*	(0.5336)	12	0.33*	(0.7163)	10
TW-*CER*	2.71	(0.0001)	22	37.65	(0.4370)	17	1.59	(0.4091)	20	0.33*	(0.7359)	12
Trimmed-*CER*	2.75	(0.0000)	25	38.49	(0.2126)	23	1.62	(0.2742)	22	0.34	(0.4115)	20
Best-*CER*	2.71	(0.0000)	19	37.32*	(0.6836)	8	1.59	(0.5526)	18	0.33*	(0.6640)	15
TW-$Q_{0.05}$	2.70	(0.0001)	15	37.66	(0.4108)	18	1.59*	(0.5464)	10	0.33*	(0.9870)	3
TW-$Q_{0.01}$	2.71	(0.0002)	17	37.78	(0.3684)	20	1.59	(0.4720)	16	0.33*	(0.9225)	7
Trimmed-$Q_{0.05}$	2.74	(0.0001)	23	38.45	(0.2220)	22	1.61	(0.3495)	21	0.33*	(0.6441)	13
Trimmed-$Q_{0.01}$	2.69	(0.0000)	12	37.39*	(0.5150)	9	1.58*	(0.7170)	5	0.33*	(0.9566)	6
Best-$Q_{0.05}$	2.68	(0.0001)	9	37.40*	(0.6030)	10	1.58*	(0.6647)	4	0.33*	(0.8807)	4
Best-$Q_{0.01}$	2.77	(0.0001)	30	39.68	(0.0837)	30	1.66	(0.0970)	28	0.34	(0.3327)	23
TW-*Sharpe*	2.71	(0.0002)	21	37.65	(0.4377)	16	1.59	(0.4039)	19	0.33*	(0.7314)	11
Trimmed-*Sharpe*	2.75	(0.0000)	25	38.49	(0.2145)	23	1.62	(0.2607)	22	0.34	(0.4164)	20
Best-*Sharpe*	2.71	(0.0000)	18	37.31*	(0.6841)	7	1.59*	(0.5514)	17	0.33*	(0.6693)	14

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Table 4.12 22-step ahead Statistical Forecasting Performance for HAR combinations

	MAE			MSE			HRLF *(b=-1)*			QLIKE		
	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank
HAR-RV	3.16	(0.0000)	27	45.30	(0.0364)	27	2.06	(0.4128)	28	0.43	(0.6054)	26
LHAR-RV	3.13	(0.0000)	17	44.63*	(0.0754)	12	2.03*	(0.7000)	6	0.42*	(0.7139)	17
HAR-RV-J	3.11	(0.0000)	11	44.65	(0.0743)	13	2.03*	(0.7445)	5	0.42*	(0.7589)	9
HAR-CJ	3.04	(0.0000)	3	43.61*	(0.2748)	2	2.06	(0.3798)	26	0.66	(0.1335)	31
HAR-RSV	3.17	(0.0000)	31	45.81	(0.0200)	30	2.08	(0.2962)	30	0.43	(0.5699)	27
Mean	3.11	(0.0000)	8	44.51	(0.0805)	9	2.03	(0.6919)	11	0.42	(0.6783)	21
Geometric Mean	3.10	(0.0000)	6	44.43*	(0.0880)	7	2.03*	(0.7344)	9	0.43	(0.6792)	23
Harmonic Mean	3.09	(0.0000)	4	44.35*	(0.0936)	6	2.03	(0.6805)	13	0.46	(0.6140)	29
Trimmed Mean	3.11	(0.0000)	8	44.51*	(0.0799)	9	2.03*	(0.6916)	11	0.42	(0.6852)	21
Median	3.11	(0.0000)	10	44.79	(0.0647)	18	2.04	(0.5608)	17	0.42	(0.6869)	20
AFTER	3.12	(0.0000)	15	44.67	(0.0654)	14	2.04	(0.6501)	15	0.42	(0.6672)	19
DMSFE	3.13	(0.0000)	16	44.78	(0.0665)	17	2.03*	(0.7445)	10	0.42*	(0.7816)	7
IMSFE	3.13	(0.0000)	19	44.85	(0.0613)	19	2.04	(0.6824)	14	0.42	(0.7510)	10
Kernel	2.83*	(0.0557)	2	**42.47***	(1.0000)	1	2.04*	(0.4752)	19	0.45	(0.3866)	28
OLS_{EXP}	3.17	(0.0000)	29	45.87	(0.0181)	31	2.06	(0.5312)	25	0.42*	(0.7760)	4
OLS	3.11	(0.0000)	7	44.34*	(0.0890)	4	2.03*	(0.8304)	4	0.42	(0.6674)	18
OLS_{SQRT}	**2.75***	(1.0000)	1	44.47*	(0.0661)	8	2.22	(0.0220)	31	0.54	(0.3725)	30
TW	3.09	(0.0000)	5	44.35*	(0.0979)	5	**2.01***	(1.0000)	1	**0.42***	(1.0000)	1
Trimmed MSPE	3.11	(0.0000)	12	44.68	(0.0746)	15	2.02*	(0.8612)	2	0.42*	(0.8984)	2
TW-*CER*	3.15	(0.0000)	22	45.09	(0.0494)	22	2.04	(0.5291)	21	0.42	(0.6919)	16
Trimmed-*CER*	3.15	(0.0000)	23	45.26	(0.0409)	24	2.05	(0.4765)	22	0.42	(0.7328)	11
Best-*CER*	3.17	(0.0000)	30	45.56	(0.0304)	29	2.07	(0.3795)	29	0.43	(0.5493)	25
TW-$Q_{0.05}$	3.12	(0.0000)	14	44.52*	(0.0889)	11	2.03*	(0.8465)	3	0.42*	(0.8158)	5
TW-$Q_{0.01}$	3.14	(0.0000)	20	45.00	(0.0505)	20	2.04	(0.6397)	16	0.42*	(0.7751)	6
Trimmed-$Q_{0.05}$	3.13	(0.0000)	18	44.71	(0.0716)	16	2.03*	(0.7616)	8	0.42*	(0.7639)	8
Trimmed-$Q_{0.01}$	3.15	(0.0000)	23	45.26	(0.0416)	24	2.05	(0.4815)	22	0.42	(0.7363)	11
Best-$Q_{0.05}$	3.11	(0.0000)	13	44.28*	(0.1078)	3	2.03*	(0.7416)	7	0.42*	(0.6983)	14
Best-$Q_{0.01}$	3.16	(0.0000)	26	45.20	(0.0434)	23	2.04	(0.5691)	18	0.42*	(0.8675)	3
TW-*Sharpe*	3.14	(0.0000)	21	45.07	(0.0471)	21	2.04	(0.5406)	20	0.42	(0.6915)	15
Trimmed-*Sharpe*	3.15	(0.0000)	23	45.26	(0.0421)	24	2.05	(0.4836)	22	0.42	(0.7404)	11
Best-*Sharpe*	3.17	(0.0000)	28	45.44	(0.0344)	28	2.06	(0.2837)	27	0.43	(0.5564)	24

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \le w_{it} \le 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Economic Evaluation

Our analysis continues with the examination of the economic gains generated from the HAR models and the combinations based on them, To test whether there is a significantly larger difference in terms of *CER* and *Sharpe* ratio, the StepSPA test proposed by Hsu et al. (2010) is used. Table 4.13 reports our findings for both *CER* and *Sharpe* ratio for the one-step ahead forecast horizon. Differently from the previous section we find no advantage for the use of economic combinations in terms of HAR models. Interestingly, the regression combination approaches are found to produce higher economic gains for a mean-variance investor. We find that the OLS_{EXP} model has the larger *CER* and *Sharpe* ratios compared to the rest models. Therefore, the OLS_{EXP} model can increase the economic value in the oil futures market. Furthermore, the OLS_{EXP} model is the only model that passes the StepSPA test indicating superior forecasting performance compared. More importantly, we find that in 1-step ahead forecasting horizon, the economic combinations are proved inadequate except for the Trimmed-$Q_{0.05}$.

Table 4.13 1-step ahead Economic Forecasting Performance for HAR combinations

	CER			Sharpe		
	Loss function	StepSPA	Rank	Loss function	StepSPA	Rank
HAR-RV	0.0427	(0.0000)	30	4.09	(0.0000)	30
LHAR-RV	0.1107	(0.0010)	4	9.40	(0.0000)	2
HAR-RV-J	0.0777	(0.0000)	11	6.15	(0.0000)	18
HAR-CJ	0.1237	(0.0070)	2	6.80	(0.0000)	8
HAR-RSV	0.0649	(0.0000)	21	5.77	(0.0000)	21
Mean	0.0770	(0.0000)	13	6.51	(0.0000)	13
Geometric Mean	0.0796	(0.0000)	9	6.58	(0.0000)	11
Harmonic Mean	0.0860	(0.0000)	7	6.78	(0.0000)	9
Trimmed Mean	0.0770	(0.0000)	13	6.51	(0.0000)	13
Median	0.0666	(0.0000)	18	5.85	(0.0000)	20
AFTER	0.0596	(0.0000)	27	5.42	(0.0000)	27
DMSFE	0.0797	(0.0000)	8	7.28	(0.0000)	5
IMSFE	0.0772	(0.0000)	12	7.07	(0.0000)	7
Kernel	0.0260	(0.0000)	31	2.61	(0.0000)	31
OLS_{EXP}	**0.2064**	(1.0000)	1	**14.21**	(1.0000)	1
OLS	0.0655	(0.0000)	20	5.96	(0.0000)	19
OLS_{SQRT}	0.1139	(0.0000)	3	6.62	(0.0000)	10
TW	0.0861	(0.0000)	6	7.81	(0.0000)	4
Trimmed MSPE	0.0783	(0.0000)	10	7.16	(0.0000)	6
TW-*CER*	0.0667	(0.0000)	17	6.17	(0.0000)	16
Trimmed-*CER*	0.0602	(0.0000)	24	5.64	(0.0000)	22
Best-*CER*	0.0566	(0.0000)	28	5.32	(0.0000)	28
TW-$Q_{0.05}$	0.0715	(0.0000)	15	6.52	(0.0000)	12
TW-$Q_{0.01}$	0.0690	(0.0000)	16	6.33	(0.0000)	15
Trimmed-$Q_{0.05}$	0.0873	(0.0000)	5	7.88	(0.0000)	3
Trimmed-$Q_{0.01}$	0.0601	(0.0000)	25	5.63	(0.0000)	23
Best-$Q_{0.05}$	0.0604	(0.0000)	22	5.48	(0.0000)	25
Best-$Q_{0.01}$	0.0604	(0.0000)	22	5.48	(0.0000)	25
TW-*Sharpe*	0.0666	(0.0000)	19	6.17	(0.0000)	17
Trimmed-*Sharpe*	0.0601	(0.0000)	25	5.63	(0.0000)	23
Best-*Sharpe*	0.0563	(0.0000)	29	5.31	(0.0000)	29

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hsu et al. (2010) StepSPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 1,000 and the smoothing parameter for the mean block length is 0.1. The confidence level used is α=10%. 3. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 4. EXP and SQRT denote the exponential and the square root transformation, respectively. 5. The CER ratio is annualized, while the SR is multiplied by 100.

Tables 4.14 and 4.15 present the results for the 5- and 22-step ahead forecasts, respectively. Considering the 5-step ahead forecasts, in Table 4.14, a non-parametric technique, the Kernel regression approach has the larger economic value for both ratios. It is worth noting that the economic combinations based on either *CER* or *Sharpe* ratio lead to increased economic gains compared to single models and other combinations. Ma et al. (2018c) indicated the superior forecasting performance of combination methods in economic terms, but they combined models based solely on statistical measures. In contrast to the previous sections, the transformed least squares combinations are ranked last and indicate inadequate performance. Under the StepSPA test, we note that only the LHAR-RV model does not pass the test.

Regarding the 22-step ahead forecasting horizon, we can see that a single model, the HAR-RSV model is found to absolutely outperform the other single and combination models. The Best-*CER* and the Best-*Sharpe* combinations are ranked second and third respectively. The forecasting evaluation criteria reject the null hypothesis for only the HAR-RV-J model, while the rest pass the test. Three models, the OLS_{SQRT}, the Harmonic Mean and the HAR-CJ are ranked last as they produce economic losses for an investor. In contrast to previous findings the OLS_{SQRT} model is proved inadequate for the 22-step ahead forecasting horizon. We note that although an OLS model is the best performing model that generates higher economic gains for 1-step ahead forecasts, it is ranked amongst the worst performers in longer forecast horizons.

Table 4.14 5-step ahead Economic Forecasting Performance for HAR combinations

	CER			Sharpe		
	Loss function	StepSPA	Rank	Loss function	StepSPA	Rank
HAR-RV	0.0320	(0.6130)	25	3.30	(0.6840)	24
LHAR-RV	0.0286	(0.0460)	29	2.99	(0.0350)	29
HAR-RV-J	0.0314	(0.5840)	28	3.21	(0.6210)	28
HAR-CJ	0.0366	(0.8090)	18	3.62	(0.8410)	18
HAR-RSV	0.0319	(0.5610)	26	3.28	(0.5630)	26
Mean	0.0324	(0.7200)	21	3.32	(0.7180)	21
Geometric Mean	0.0324	(0.7780)	23	3.31	(0.8210)	23
Harmonic Mean	0.0321	(0.5620)	24	3.28	(0.7390)	25
Trimmed Mean	0.0324	(0.7120)	21	3.32	(0.7180)	21
Median	0.0318	(0.6510)	27	3.27	(0.7300)	27
AFTER	0.0385	(0.9620)	8	3.93	(0.9950)	8
DMSFE	0.0387	(0.9860)	4	3.98	(0.9980)	6
IMSFE	0.0386	(0.9790)	5	3.98	(0.9890)	7
Kernel	**0.0406**	(1.0000)	1	**4.27**	(1.0000)	1
OLS_{EXP}	0.0230	(0.2420)	30	2.49	(0.2770)	30
OLS	0.0379	(0.8870)	10	3.88	(0.9370)	14
OLS_{SQRT}	0.0078	(0.2270)	31	1.64	(0.5210)	31
TW	0.0379	(0.8790)	11	3.88	(0.9500)	13
Trimmed MSPE	0.0379	(0.8420)	9	3.92	(0.8410)	11
TW-*CER*	0.0386	(0.9740)	6	4.00	(0.9870)	4
Trimmed-*CER*	0.0377	(0.2120)	14	3.92	(0.2880)	9
Best-*CER*	0.0388	(0.8750)	2	4.03	(0.9100)	2
TW-$Q_{0.05}$	0.0375	(0.8010)	16	3.84	(0.8350)	17
TW-$Q_{0.01}$	0.0378	(0.8740)	13	3.86	(0.9360)	15
Trimmed-$Q_{0.05}$	0.0372	(0.1910)	17	3.84	(0.2390)	16
Trimmed-$Q_{0.01}$	0.0378	(0.5150)	12	3.89	(0.5610)	12
Best-$Q_{0.05}$	0.0343	(0.6550)	20	3.45	(0.7290)	20
Best-$Q_{0.01}$	0.0351	(0.7500)	19	3.50	(0.8270)	19
TW-*Sharpe*	0.0386	(0.9720)	7	3.99	(0.9920)	5
Trimmed-*Sharpe*	0.0377	(0.2390)	14	3.92	(0.2770)	9
Best-*Sharpe*	0.0388	(0.8890)	3	4.03	(0.9020)	3

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hsu et al. (2010) StepSPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 1,000 and the smoothing parameter for the mean block length is 0.1. The confidence level used is α=10%. 3. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 4. EXP and SQRT denote the exponential and the square root transformation, respectively.

Table 4.15 22-step ahead Economic Forecasting Performance for HAR combinations

	CER			Sharpe		
	Loss function	StepSPA	Rank	Loss function	StepSPA	Rank
HAR-RV	0.0260	(0.1900)	6	3.06	(0.1390)	19
LHAR-RV	0.0258	(0.2750)	10	2.98	(0.3630)	21
HAR-RV-J	0.0230	(0.0870)	26	2.70	(0.0510)	26
HAR-CJ	-2.3542	(0.1820)	31	-2.58	(0.3470)	31
HAR-RSV	**0.0312**	(1.0000)	1	**3.52**	(1.0000)	1
Mean	0.0240	(0.1210)	23	2.79	(0.1260)	24
Geometric Mean	0.0150	(0.1970)	28	1.95	(0.2270)	28
Harmonic Mean	-0.4220	(0.2770)	30	-2.21	(0.4220)	30
Trimmed Mean	0.0240	(0.1260)	23	2.79	(0.1030)	24
Median	0.0252	(0.0750)	18	2.92	(0.1000)	23
AFTER	0.0251	(0.2080)	20	3.12	(0.2820)	17
DMSFE	0.0258	(0.9640)	11	3.19	(0.8690)	9
IMSFE	0.0258	(0.9660)	9	3.19	(0.8570)	8
Kernel	0.0251	(0.7130)	19	3.16	(0.4070)	15
OLS_{EXP}	0.0184	(0.0530)	27	2.39	(0.1710)	27
OLS	0.0261	(0.8800)	4	3.24	(0.7580)	4
OLS_{SQRT}	-0.0242	(0.0410)	29	-0.14	(0.2180)	29
TW	0.0261	(0.9100)	5	3.21	(0.7590)	6
Trimmed MSPE	0.0257	(0.9040)	12	3.17	(0.7750)	10
TW-*CER*	0.0260	(0.9660)	7	3.21	(0.8860)	5
Trimmed-*CER*	0.0256	(0.6680)	13	3.16	(0.7560)	12
Best-*CER*	0.0265	(0.8290)	2	3.27	(0.6820)	2
TW-$Q_{0.05}$	0.0250	(0.4760)	21	3.11	(0.6500)	18
TW-$Q_{0.01}$	0.0254	(0.7490)	17	3.15	(0.8220)	16
Trimmed-$Q_{0.05}$	0.0255	(0.7530)	16	3.17	(0.6240)	11
Trimmed-$Q_{0.01}$	0.0256	(0.6710)	13	3.16	(0.7900)	12
Best-$Q_{0.05}$	0.0238	(0.4040)	25	2.96	(0.6200)	22
Best-$Q_{0.01}$	0.0245	(0.6970)	22	3.04	(0.6910)	20
TW-*Sharpe*	0.0259	(0.6540)	8	3.21	(0.6970)	7
Trimmed-*Sharpe*	0.0256	(0.6880)	13	3.16	(0.7830)	12
Best-*Sharpe*	0.0264	(0.2850)	3	3.26	(0.3470)	3

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hsu et al. (2010) StepSPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 1,000 and the smoothing parameter for the mean block length is 0.1. The confidence level used is α=10%. 3. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 4. EXP and SQRT denote the exponential and the square root transformation, respectively.

Risk Management Evaluation

Considering a risk management perspective, we evaluate the results from HAR models and combinations based on HAR models, using the Smoothed-*Q* loss function. In table 4.16, the errors related to Smoothed-*Q* are quite different considering the two confidence levels for 1-step ahead forecasting horizon. Considering the percentage of violations, the OLS_{EXP} combination clearly dominates the rest as it produces the lower number of violations. However, under the Smoothed-*Q* loss function, the OLS_{EXP} is the dominant model for 5%, while we cannot reject the null hypothesis of superior predictive ability according to SPA. The combinations based on Smoothed-*Q* loss function that take into consideration a 5% VaR exhibit an adequate performance based on the annotated ranking, but they do not pass the SPA test. Considering a 1% VaR, although the combinations based on Smoothed-*Q* are amongst the best performers, the Trimmed-$Q_{0.05}$ is suggested as the best performing model. In this case all the single models except for the HAR-RV-J and most combination schemes perform poorly even if they indicate superior forecasting ability under the SPA test.

Table 4.16 1-step ahead VaR Forecasting Performance for HAR combinations

	Smoothed-$Q_{0.05}$				Smoothed-$Q_{0.01}$			
	Percentage of violations (5%)	Loss Function	SPA	Rank	Percentage of violations (1%)	Loss function	SPA	Rank
HAR-RV	5.25%	0.2112	(0.0005)	29	1.59%	0.0551	(0.0661)	29
LHAR-RV	4.30%	0.2010	(0.0569)	2	1.11%	0.0524	(0.3892)	25
HAR-RV-J	5.17%	0.2027	(0.0040)	9	1.11%	0.0509	(0.9201)	8
HAR-CJ	5.57%	0.2032	(0.0118)	16	1.51%	0.0515	(0.6405)	15
HAR-RSV	5.17%	0.2090	(0.0005)	28	1.51%	0.0540	(0.1190)	28
Mean	5.09%	0.2038	(0.0075)	17	1.27%	0.0516	(0.7000)	18
Geometric Mean	5.25%	0.2042	(0.0048)	21	1.35%	0.0518	(0.5844)	22
Harmonic Mean	5.25%	0.2046	(0.0041)	22	1.35%	0.0520	(0.4888)	24
Trimmed Mean	5.09%	0.2038	(0.0066)	17	1.27%	0.0516	(0.6947)	18
Median	5.17%	0.2073	(0.0043)	27	1.43%	0.0524	(0.3483)	26
AFTER	5.17%	0.2072	(0.0011)	26	1.19%	0.0530	(0.2528)	27
DMSFE	4.77%	0.2029	(0.0060)	12	0.95%	0.0509	(0.8979)	7
IMSFE	4.77%	0.2029	(0.0063)	15	0.95%	0.0510	(0.8795)	11
Kernel	6.36%	0.2206	(0.0001)	31	2.70%	0.0665	(0.0010)	31
OLS_{EXP}	**3.98%**	**0.1920**	(1.0000)	1	**0.72%**	0.0514	(0.5859)	14
OLS	5.17%	0.2053	(0.0004)	25	1.03%	0.0518	(0.5986)	23
OLS_{SQRT}	7.48%	0.2172	(0.0001)	30	2.86%	0.0639	(0.0095)	30
TW	4.69%	0.2024	(0.0068)	6	0.88%	0.0507	(0.9195)	3
Trimmed MSPE	4.85%	0.2038	(0.0032)	19	1.11%	0.0515	(0.6800)	16
TW-*CER*	4.93%	0.2028	(0.0038)	11	0.95%	0.0509	(0.9561)	5
Trimmed-*CER*	4.85%	0.2048	(0.0016)	23	1.11%	0.0517	(0.6079)	20
Best-*CER*	5.17%	0.2029	(0.0027)	13	1.03%	0.0512	(0.8281)	12
TW-$Q_{0.05}$	4.85%	0.2021	(0.0068)	4	0.95%	0.0506	(0.9909)	2
TW-$Q_{0.01}$	4.85%	0.2025	(0.0049)	8	0.95%	0.0508	(0.9621)	4
Trimmed-$Q_{0.05}$	4.77%	0.2020	(0.0068)	3	0.95%	**0.0505**	(1.0000)	1
Trimmed-$Q_{0.01}$	4.85%	0.2039	(0.0023)	20	1.03%	0.0516	(0.6400)	17
Best-$Q_{0.05}$	5.17%	0.2024	(0.0042)	7	0.95%	0.0509	(0.8870)	9
Best-$Q_{0.01}$	5.17%	0.2021	(0.0040)	5	0.95%	0.0509	(0.8758)	10
TW-*Sharpe*	4.93%	0.2028	(0.0039)	10	0.95%	0.0509	(0.9533)	6
Trimmed-*Sharpe*	4.85%	0.2048	(0.0012)	24	1.11%	0.0518	(0.6006)	21
Best-*Sharpe*	5.17%	0.2029	(0.0038)	14	1.03%	0.0512	(0.8279)	13

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. The Smoothed-Q loss function is calculated for VaR level $\alpha=0.01$ and α=0.05. We set the smoothness parameter $\delta=25$. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \le w_{it} \le 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Tables 4.17 and 4.18 present the results for the 5- and 22-step ahead forecast horizon including the empirical percentage of violations during the out-of sample period for 5% and 1% VaR confidence levels. For the 5-step ahead forecasts, we note that the percentage of violations for all models is always higher than 5% and 1% suggesting that all models underforecast VaR and none of the models is adequately reliable as an internal VaR model. For the 5% confidence level, the problem is less severe for TW-$Q_{0.05}$ and TW-$Q_{0.01}$ where the empirical percentage is close to 5%. Based on the Smoothed-Q loss function, the TW model followed by the AFTER, the DMSFE, the IMSFE and the Trimmed-$Q_{0.01}$ model produce the smallest errors, while the single modes exhibit poor forecasting performance and they are ranked last. It is worth to note that for the 1% confidence level, the AFTER combination produce smaller errors, while the transformed least squares models are ranked last. The superior forecasting performance of the AFTER model is indicated by the large Hansen's p-value. The results are noteworthy because the transformed OLS schemes (i.e. the OLS_{SQRT} and the OLS_{EXP}) perform worse in terms of the SPA test. To this end, we argue that the combinations based on the ranking of the single models perform better as they capture better the structural breaks compared with the rest combinations.

The results for the 22-step ahead forecasting horizon, are presented in Table 4.18. we note, again, that all the models underforecast VaR and none of the models is considered as an adequate VaR model. For both confidence levels, the OLS_{EXP} combination indicates superior forecasting performance. It is noteworthy that although the economic combinations indicate good forecasting performance according to SPA test, only a limited number of them are ranked among the best models, while the single models and the OLS_{SQRT} model are ranked last.

Table 4.17 5-step ahead VaR Forecasting Performance for HAR combinations

	Smoothed-$Q_{0.05}$				Smoothed-$Q_{0.01}$			
	Percentage of violations (5%)	Loss function	SPA	Rank	Percentage of violations (1%)	Loss function	SPA	Rank
HAR-RV	5.67%	0.2249	(0.3582)	27	2.08%	0.0661	(0.0960)	24
LHAR-RV	5.83%	0.2250	(0.3526)	28	2.00%	0.0661	(0.0996)	25
HAR-RV-J	5.91%	0.2243	(0.4219)	21	2.15%	0.0656	(0.1371)	23
HAR-CJ	6.23%	0.2258	(0.2622)	29	2.15%	0.0652	(0.2884)	21
HAR-RSV	5.67%	0.2247	(0.3761)	25	2.08%	0.0662	(0.0982)	27
Mean	5.75%	0.2237	(0.5514)	18	**1.92%**	0.0645	(0.3370)	14
Geometric Mean	5.83%	0.2242	(0.4440)	20	2.00%	0.0649	(0.2295)	20
Harmonic Mean	5.83%	0.2246	(0.3477)	24	2.08%	0.0652	(0.1894)	22
Trimmed Mean	5.75%	0.2237	(0.5434)	18	**1.92%**	0.0645	(0.3312)	14
Median	5.67%	0.2249	(0.3551)	26	2.00%	0.0662	(0.0926)	28
AFTER	5.35%	0.2218	(0.8898)	2	2.00%	**0.0619**	(1.0000)	1
DMSFE	5.27%	0.2218	(0.9489)	3	2.00%	0.0629	(0.7564)	4
IMSFE	5.27%	0.2220	(0.9314)	4	**1.92%**	0.0630	(0.7036)	6
Kernel	6.30%	0.2272	(0.1215)	30	2.23%	0.0667	(0.0963)	30
OLS_{EXP}	5.75%	0.2245	(0.4181)	23	2.23%	0.0661	(0.1595)	26
OLS	5.43%	0.2222	(0.9088)	6	2.08%	0.0627	(0.8191)	3
OLS_{SQRT}	8.70%	0.2302	(0.0360)	31	3.43%	0.0735	(0.0011)	31
TW	5.35%	**0.2212**	(1.0000)	1	2.08%	0.0625	(0.8779)	2
Trimmed MSPE	5.35%	0.2222	(0.8105)	7	2.08%	0.0636	(0.5899)	8
TW-*CER*	5.27%	0.2224	(0.8442)	9	2.00%	0.0637	(0.5093)	10
Trimmed-*CER*	5.35%	0.2229	(0.7115)	12	**1.92%**	0.0646	(0.2869)	17
Best-*CER*	5.27%	0.2229	(0.6840)	14	2.15%	0.0644	(0.3392)	12
TW-$Q_{0.05}$	**5.11%**	0.2226	(0.7944)	11	2.00%	0.0637	(0.5066)	9
TW-$Q_{0.01}$	**5.11%**	0.2224	(0.8775)	8	2.00%	0.0634	(0.6253)	7
Trimmed-$Q_{0.05}$	5.27%	0.2231	(0.6780)	16	**1.92%**	0.0646	(0.2763)	19
Trimmed-$Q_{0.01}$	5.27%	0.2221	(0.9124)	5	**1.92%**	0.0630	(0.7112)	5
Best-$Q_{0.05}$	5.43%	0.2233	(0.5848)	17	2.23%	0.0646	(0.3101)	16
Best-$Q_{0.01}$	5.43%	0.2244	(0.4389)	22	2.47%	0.0666	(0.0707)	29
TW-*Sharpe*	5.27%	0.2224	(0.8312)	10	2.00%	0.0637	(0.5083)	11
Trimmed-*Sharpe*	5.35%	0.2229	(0.7050)	12	**1.92%**	0.0646	(0.2904)	17
Best-*Sharpe*	5.27%	0.2230	(0.6660)	15	2.15%	0.0644	(0.3432)	13

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. The Smoothed-Q loss function is calculated for VaR level $\alpha=0.01$ and α=0.05. We set the smoothness parameter $\delta=25$. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Table 4.18 22-step ahead VaR Forecasting Performance for HAR combinations

	Smoothed-$Q_{0.05}$				Smoothed-$Q_{0.01}$			
	Percentage of violations (5%)	Loss function	SPA	Rank	Percentage of violations (1%)	Loss function	SPA	Rank
HAR-RV	6.07%	0.2387	(0.1069)	25	2.18%	0.0686	(0.4131)	21
LHAR-RV	5.99%	0.2375	(0.2344)	20	2.35%	0.0678	(0.6154)	18
HAR-RV-J	6.07%	0.2387	(0.1284)	27	2.35%	0.0696	(0.1961)	27
HAR-CJ	6.63%	0.2416	(0.0424)	30	2.67%	0.0725	(0.0530)	30
HAR-RSV	6.07%	0.2389	(0.0942)	28	2.27%	0.0692	(0.2761)	26
Mean	5.99%	0.2383	(0.1352)	21	2.43%	0.0687	(0.3707)	22
Geometric Mean	6.07%	0.2386	(0.1149)	24	2.51%	0.0688	(0.3312)	24
Harmonic Mean	6.15%	0.2395	(0.0730)	29	2.59%	0.0699	(0.1559)	28
Trimmed Mean	5.99%	0.2383	(0.1356)	21	2.43%	0.0687	(0.3665)	22
Median	6.07%	0.2384	(0.1268)	23	2.35%	0.0689	(0.3274)	25
AFTER	5.74%	0.2350	(0.5361)	9	2.18%	0.0672	(0.8115)	8
DMSFE	5.66%	0.2348	(0.5219)	5	2.27%	0.0670	(0.8575)	4
IMSFE	5.66%	0.2348	(0.5151)	6	2.27%	0.0670	(0.8592)	3
Kernel	6.31%	0.2387	(0.1586)	26	3.07%	0.0714	(0.0996)	29
OLS_{EXP}	5.66%	**0.2329**	(1.0000)	1	**2.10%**	**0.0663**	(1.0000)	1
OLS	5.74%	0.2352	(0.5002)	13	2.27%	0.0672	(0.7738)	9
OLS_{SQRT}	9.47%	0.2451	(0.0091)	31	4.61%	0.0833	(0.0001)	31
TW	**5.50%**	0.2343	(0.6578)	2	2.35%	0.0672	(0.8225)	7
Trimmed MSPE	5.66%	0.2346	(0.5752)	3	2.27%	0.0671	(0.8319)	5
TW-*CER*	5.74%	0.2352	(0.4930)	14	2.27%	0.0673	(0.8081)	11
Trimmed-*CER*	5.83%	0.2351	(0.5122)	10	2.27%	0.0673	(0.7931)	13
Best-*CER*	5.74%	0.2362	(0.3285)	19	2.27%	0.0677	(0.6472)	17
TW-$Q_{0.05}$	5.58%	0.2349	(0.5047)	7	2.18%	0.0671	(0.8557)	6
TW-$Q_{0.01}$	5.66%	0.2350	(0.5367)	8	2.18%	0.0673	(0.8073)	12
Trimmed-$Q_{0.05}$	5.58%	0.2347	(0.5528)	4	2.18%	0.0669	(0.9003)	2
Trimmed-$Q_{0.01}$	5.83%	0.2351	(0.5164)	10	2.27%	0.0673	(0.7895)	13
Best-$Q_{0.05}$	**5.50%**	0.2359	(0.3893)	17	2.27%	0.0685	(0.4480)	20
Best-$Q_{0.01}$	5.66%	0.2356	(0.4039)	16	**2.10%**	0.0679	(0.5926)	19
TW-*Sharpe*	5.74%	0.2352	(0.4973)	15	2.27%	0.0673	(0.8067)	10
Trimmed-*Sharpe*	5.83%	0.2351	(0.5303)	10	2.27%	0.0673	(0.7891)	13
Best-*Sharpe*	5.74%	0.2361	(0.3377)	18	2.27%	0.0677	(0.6586)	16

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. The Smoothed-Q loss function is calculated for VaR level $\alpha=0.01$ and α=0.05. We set the smoothness parameter $\delta=25$. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

4.7.3 Out-of-Sample Forecasting Performance for combinations based on both HAR-type and GARCH-type models

Statistical Evaluation

We consider 1-, 5- and 22-step ahead volatility forecasts based on HAR-type, GARCH-type and combinations based on both classes of models. We assess the forecasting accuracy of different single and combination models based on standard statistical loss functions, such as the MAE and MSE, and asymmetric loss functions, such as the HRLF and the QLIKE. Table 4.19 shows the performances of single and combination models for 1-step ahead forecasts. For the MAE and MSE loss functions, the OLS_{SQRT} combination clearly dominates the rest. Especially for the case of MAE, the OLS_{SQRT} combination is the only forecasting scheme that passes the SPA test and is included to the MCS. On the contrary, under the MSE loss function, almost all the combinations pass the SPA test and are included to the MCS. From the GARCH-type models only the FIGARCH model is included to the MCS, although from the HAR-type all are incorporated to it, indicating the better performance of models based on high-frequency data. Considering the HRLF and the QLIKE measures, there is a single model that clearly dominates the rest; this is the HAR-RSV model which produces the smallest losses. This result is opposite to the findings of Becker and Clements (2008) that concluded that combination forecasts can beat the single models based on high-frequency data. For the HRLF, some economic combinations based on Smoothed-*Q* loss function and five other combinations pass the SPA test, while the null hypothesis is not rejected for all the HAR-type models. The results for the QLIKE measure, assure the inadequate performance of the GARCH-type models.

Table 4.19 1-step ahead Statistical Forecasting Performance for HAR-GARCH combinations

	MAE			MSE			HRLF (*b=-1*)			QLIKE		
	Loss Function	SPA	Rank	Loss Function	SPA	Rank	Loss Function	SPA	Rank	Loss Function	SPA	Rank
HAR-RV	2.29	(0.0000)	13	26.55*	(0.5073)	4	1.00*	(0.5362)	4	0.23*	(0.6987)	5
LHAR-RV	2.29	(0.0000)	12	27.29*	(0.3534)	6	1.01*	(0.4412)	5	0.22*	(0.7638)	3
HAR-RV-J	2.16	(0.0000)	4	29.24*	(0.1507)	22	1.05*	(0.2069)	9	0.22*	(0.8420)	4
HAR-CJ	2.12	(0.0000)	3	29.64*	(0.1339)	23	1.10*	(0.1109)	20	0.24	(0.5481)	9
HAR-RSV	2.24	(0.0000)	9	24.33*	(0.8049)	2	**0.92***	(1.0000)	1	**0.22***	(1.0000)	1
GARCH	2.89	(0.0000)	35	33.34	(0.2830)	35	1.39	(0.0002)	35	0.31	(0.0458)	35
EGARCH	2.70	(0.0000)	30	32.05	(0.3434)	33	1.31	(0.0052)	33	0.29	(0.1516)	31
GJR	2.71	(0.0000)	31	31.77	(0.3519)	31	1.29	(0.0120)	30	0.29	(0.1811)	29
APARCH	2.75	(0.0000)	33	31.77	(0.3487)	30	1.31	(0.0051)	32	0.29	(0.1376)	32
FIGARCH	2.74	(0.0000)	32	30.93*	(0.3786)	29	1.28	(0.0429)	29	0.29	(0.1538)	30
HYGARCH	2.79	(0.0000)	34	31.83	(0.3512)	32	1.30	(0.0232)	31	0.28	(0.1829)	28
Mean	2.43	(0.0000)	22	28.16*	(0.1634)	7	1.08*	(0.1050)	11	0.25	(0.4272)	20
Geometric Mean	2.39	(0.0000)	16	28.39*	(0.1368)	12	1.08*	(0.0951)	12	0.24	(0.4461)	14
Harmonic Mean	2.36	(0.0000)	15	28.66*	(0.1266)	18	1.08*	(0.0904)	14	0.24	(0.4669)	13
Trimmed Mean	2.41	(0.0000)	18	28.24*	(0.1476)	8	1.08*	(0.0918)	13	0.25	(0.4324)	21
Median	2.44	(0.0000)	24	28.68*	(0.1092)	19	1.11*	(0.0494)	21	0.25	(0.4026)	22
AFTER	2.19	(0.0000)	8	26.62*	(0.5278)	5	1.03*	(0.3679)	6	0.23*	(0.6918)	6
DMSFE	2.41	(0.0000)	17	28.65*	(0.1153)	17	1.10*	(0.0627)	19	0.24	(0.4408)	15
IMSFE	2.42	(0.0000)	21	28.49*	(0.1221)	16	1.09*	(0.0707)	16	0.24	(0.4374)	16
Kernel	2.60	(0.0000)	27	56.04	(0.0336)	37	1.67	(0.0111)	37	0.33	(0.0470)	37
OLS_{EXP}	2.06	(0.0001)	2	30.21*	(0.1277)	27	1.17*	(0.0802)	26	0.32	(0.1175)	36
OLS	2.19	(0.0000)	7	26.48*	(0.5509)	3	0.99*	(0.5903)	2	0.22*	(0.9875)	2
OLS_{SQRT}	**1.78***	(1.0000)	1	**24.32***	(1.0000)	1	0.99*	(0.5357)	3	0.28	(0.1634)	27
TW	2.35	(0.0000)	14	28.29*	(0.1527)	9	1.07*	(0.1021)	10	0.24	(0.5012)	10
Trimmed MSPE	2.42	(0.0000)	19	28.42*	(0.1345)	13	1.09*	(0.0713)	15	0.24	(0.4410)	17
TW-*CER*	2.64	(0.0000)	29	30.24	(0.3981)	28	1.21	(0.0036)	28	0.27	(0.3004)	25
Trimmed-*CER*	2.47	(0.0000)	25	28.81*	(0.1070)	20	1.12	(0.0355)	22	0.25	(0.3925)	23
Best-*CER*	2.93	(0.0000)	37	33.85	(0.2735)	36	1.40	(0.0005)	36	0.30	(0.0645)	34
TW-$Q_{0.05}$	2.26	(0.0000)	11	28.36*	(0.1610)	10	1.05*	(0.1824)	7	0.23	(0.6215)	7
TW-$Q_{0.01}$	2.26	(0.0000)	10	28.39*	(0.1634)	11	1.05*	(0.1831)	8	0.23	(0.6055)	8
Trimmed-$Q_{0.05}$	2.42	(0.0000)	20	28.45*	(0.1276)	15	1.09*	(0.0694)	17	0.24	(0.4371)	18
Trimmed-$Q_{0.01}$	2.43	(0.0000)	23	28.45*	(0.1278)	14	1.09*	(0.0686)	18	0.25	(0.4406)	19
Best-$Q_{0.05}$	2.18	(0.0000)	5	30.14*	(0.1259)	25	1.13*	(0.1242)	24	0.24*	(0.5292)	11
Best-$Q_{0.01}$	2.18	(0.0000)	6	30.15*	(0.1308)	26	1.13*	(0.1250)	25	0.24	(0.5402)	12
TW-*Sharpe*	2.64	(0.0000)	28	29.96	(0.4081)	24	1.20	(0.0055)	27	0.27	(0.2999)	26
Trimmed-*Sharpe*	2.47	(0.0000)	26	28.82*	(0.1004)	21	1.12	(0.0368)	23	0.25	(0.3932)	24
Best-*Sharpe*	2.91	(0.0000)	36	32.10*	(0.3347)	34	1.35	(0.0125)	34	0.30	(0.0798)	33

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \le w_{it} \le 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

The analysis continues with the 5- and 22-step ahead forecasts in Tables 4.20 and 4.21. For the 5-step ahead forecasts, the best performing model according to the two asymmetric loss functions and the MSE, there is a dominant model. This is an economic combination, the Best-$Q_{0.05}$ followed by several other economic combinations based on Smoothed-Q loss function. In the meanwhile, for most economic combinations, the null hypothesis that none of the rest models is better than the benchmark is not rejected, while some of them are included to the MCS. These results reassure previous findings from Ma et al. (2018d) who suggested that the combination of low-frequency and high-frequency data yields significantly better forecasting performance.

Regarding the 22-step ahead forecasts, we find that the best performing model varies across different loss functions (Table 4.21). For example, for the asymmetric loss functions, the single EGARCH model indicates superior performance compared to other single and combination models. It is interesting to note that none of the HAR-type models pass the SPA under the HRLF. However, the standard MSE loss function, indicates the superior forecasting performance of an economic combination, the Best-$Q_{0.05}$. Furthermore, the Best-$Q_{0.05}$, the Best-$Q_{0.01}$, the three least squares schemes, the AFTER combination and the single EGARCH model are the only models included to the MCS.

Table 4.20 5-step ahead Statistical Forecasting Performance for HAR-GARCH combinations

	MAE			MSE			HRLF *(b=-1)*			QLIKE		
	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank
HAR-RV	2.76	(0.0000)	17	39.01	(0.0734)	33	1.65	(0.0113)	29	0.35	(0.2116)	25
LHAR-RV	2.76	(0.0000)	16	39.00	(0.0770)	32	1.64	(0.0126)	28	0.35	(0.1997)	26
HAR-RV-J	2.67	(0.0000)	9	37.01	(0.2472)	27	1.58	(0.0850)	23	0.34*	(0.4384)	18
HAR-CJ	2.66	(0.0000)	8	36.58	(0.3004)	24	1.59	(0.0618)	25	0.34	(0.2184)	22
HAR-RSV	2.78	(0.0000)	22	39.56	(0.0471)	34	1.66	(0.0070)	32	0.35	(0.1773)	27
GARCH	3.21	(0.0000)	35	39.00	(0.1160)	31	1.76	(0.0002)	36	0.39	(0.0017)	37
EGARCH	2.81	(0.0000)	25	34.19*	(0.8338)	4	1.48*	(0.6474)	5	0.33*	(0.5955)	7
GJR	2.87	(0.0000)	29	34.95*	(0.5502)	9	1.53*	(0.3502)	13	0.33*	(0.5058)	13
APARCH	2.92	(0.0000)	30	34.68*	(0.5960)	8	1.53*	(0.3545)	14	0.34	(0.3859)	20
FIGARCH	3.03	(0.0000)	33	37.57	(0.1988)	28	1.67	(0.0038)	33	0.37	(0.0209)	33
HYGARCH	3.07	(0.0000)	34	38.24	(0.1307)	30	1.69	(0.0017)	34	0.36	(0.0666)	30
Mean	2.78	(0.0000)	23	35.52	(0.4297)	16	1.53	(0.3238)	12	0.33	(0.4168)	12
Geometric Mean	2.74	(0.0000)	14	35.40*	(0.4306)	13	1.52*	(0.3119)	10	0.33*	(0.5135)	10
Harmonic Mean	2.71	(0.0000)	12	35.34*	(0.4705)	11	1.51*	(0.3363)	9	0.33*	(0.6012)	9
Trimmed Mean	2.77	(0.0000)	18	35.62	(0.4140)	19	1.53	(0.2742)	16	0.33	(0.3992)	14
Median	2.78	(0.0000)	20	35.51	(0.4325)	14	1.54	(0.2034)	19	0.34	(0.3254)	21
AFTER	2.61	(0.0000)	5	34.13*	(0.8305)	3	1.48*	(0.5748)	4	0.33*	(0.6692)	5
DMSFE	2.76	(0.0000)	15	35.34	(0.4707)	12	1.52	(0.3489)	11	0.33*	(0.4934)	11
IMSFE	2.81	(0.0000)	26	35.75	(0.3893)	21	1.55	(0.1829)	20	0.34	(0.3527)	19
Kernel	2.61	(0.0073)	4	41.69	(0.0678)	37	1.66	(0.0405)	31	0.35	(0.1141)	28
OLS_{EXP}	2.56	(0.0041)	2	37.99	(0.1393)	29	1.66	(0.0225)	30	0.37	(0.0391)	32
OLS	2.63	(0.0000)	6	34.22*	(0.7984)	5	1.47*	(0.6947)	3	0.32*	(0.8393)	2
OLS_{SQRT}	**2.35***	(1.0000)	1	35.51*	(0.4749)	15	1.58	(0.0954)	24	0.38	(0.0229)	35
TW	2.68	(0.0000)	11	35.22*	(0.5202)	10	1.50*	(0.3854)	8	0.33*	(0.5883)	6
Trimmed MSPE	2.77	(0.0000)	19	35.55	(0.4226)	17	1.53	(0.2769)	15	0.34	(0.3852)	15
TW-*CER*	2.98	(0.0000)	31	36.71	(0.2854)	25	1.62	(0.0252)	26	0.36	(0.0767)	29
Trimmed-*CER*	2.85	(0.0000)	27	35.98	(0.3472)	22	1.56	(0.1164)	21	0.34	(0.2381)	23
Best-*CER*	3.22	(0.0000)	36	39.63	(0.0821)	35	1.75	(0.0002)	35	0.38	(0.0098)	34
TW-$Q_{0.05}$	2.68	(0.0000)	10	34.54*	(0.6403)	6	1.48*	(0.4467)	6	0.33*	(0.7816)	3
TW-$Q_{0.01}$	2.73	(0.0000)	13	34.63*	(0.5941)	7	1.50*	(0.5593)	7	0.33*	(0.5687)	8
Trimmed-$Q_{0.05}$	2.78	(0.0000)	24	35.67	(0.3953)	20	1.54	(0.2509)	18	0.34	(0.3823)	16
Trimmed-$Q_{0.01}$	2.78	(0.0000)	21	35.55	(0.4237)	18	1.53	(0.2563)	17	0.34	(0.3570)	17
Best-$Q_{0.05}$	2.58	(0.0003)	3	**33.75***	(1.0000)	1	**1.45***	(1.0000)	1	**0.32***	(1.0000)	1
Best-$Q_{0.01}$	2.65	(0.0000)	7	33.96*	(0.8979)	2	1.47*	(0.7184)	2	0.33	(0.7112)	4
TW-*Sharpe*	2.98	(0.0000)	32	36.72	(0.2861)	26	1.62	(0.0193)	27	0.36*	(0.0659)	31
Trimmed-*Sharpe*	2.85	(0.0000)	27	35.98	(0.3485)	22	1.56	(0.1161)	21	0.34	(0.2464)	23
Best-*Sharpe*	3.24	(0.0000)	37	39.71	(0.0752)	36	1.77	(0.0004)	37	0.39	(0.0035)	36

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Table 4.21 22-step ahead Statistical Forecasting Performance for HAR-GARCH combinations

	MAE			MSE			HRLF (*b=-1*)			QLIKE		
	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank
HAR-RV	3.16	(0.0000)	15	45.30	(0.0007)	29	2.06	(0.0023)	29	0.43	(0.5415)	26
LHAR-RV	3.13	(0.0000)	14	44.63	(0.0028)	27	2.03	(0.0070)	27	0.42	(0.5664)	24
HAR-RV-J	3.11	(0.0000)	11	44.65	(0.0014)	28	2.03	(0.0036)	26	0.42	(0.5849)	20
HAR-CJ	3.04	(0.0000)	8	43.61	(0.0025)	26	2.06	(0.0016)	28	0.66	(0.1068)	37
HAR-RSV	3.17	(0.0000)	17	45.81	(0.0006)	30	2.08	(0.0012)	30	0.43	(0.5365)	27
GARCH	3.88	(0.0000)	35	51.14	(0.0000)	35	2.34	(0.0000)	35	0.48	(0.3748)	34
EGARCH	3.06	(0.0000)	9	38.85*	(0.7778)	5	**1.75***	(1.0000)	1	**0.39***	(1.0000)	1
GJR	3.28	(0.0000)	26	40.74	(0.1801)	9	1.86	(0.2633)	8	0.40*	(0.6948)	5
APARCH	3.34	(0.0000)	30	40.74	(0.2391)	8	1.87	(0.2497)	10	0.41	(0.6337)	10
FIGARCH	3.56	(0.0000)	31	46.55	(0.0000)	33	2.17	(0.0000)	33	0.46	(0.4238)	32
HYGARCH	3.63	(0.0000)	34	47.61	(0.0000)	34	2.18	(0.0000)	34	0.44	(0.4852)	28
Mean	3.23	(0.0000)	24	42.58	(0.0062)	19	1.94	(0.0210)	20	0.41	(0.6101)	15
Geometric Mean	3.18	(0.0000)	19	42.24	(0.0138)	17	1.92	(0.0371)	14	0.41	(0.6256)	11
Harmonic Mean	3.13	(0.0000)	13	41.97	(0.0180)	13	1.92	(0.0430)	13	0.42	(0.6335)	25
Trimmed Mean	3.21	(0.0000)	20	42.52	(0.0100)	18	1.94	(0.0252)	19	0.41	(0.6122)	12
Median	3.18	(0.0000)	18	42.23	(0.0120)	16	1.93	(0.0233)	16	0.42	(0.6031)	17
AFTER	2.98	(0.0000)	6	38.80*	(0.8657)	3	1.80*	(0.6291)	4	0.41	(0.7195)	8
DMSFE	3.17	(0.0000)	16	42.12	(0.0155)	15	1.91	(0.0399)	12	0.41	(0.6787)	9
IMSFE	3.28	(0.0000)	25	43.23	(0.0026)	22	1.97	(0.0041)	22	0.42	(0.5685)	19
Kernel	2.73	(0.0001)	3	41.99	(0.0981)	14	1.85*	(0.2362)	7	0.40*	(0.6302)	6
OLS_{EXP}	2.65	(0.0081)	2	40.72*	(0.1738)	7	1.93	(0.0753)	15	0.44	(0.4357)	29
OLS	2.98	(0.0000)	5	38.81*	(0.8697)	4	1.79*	(0.7064)	3	0.40*	(0.7763)	4
OLS_{SQRT}	**2.51***	(1.0000)	1	39.43*	(0.5890)	6	1.93	(0.1138)	17	0.51	(0.4054)	36
TW	3.08	(0.0000)	10	40.75	(0.1068)	10	1.84*	(0.2990)	6	0.40*	(0.7924)	2
Trimmed MSPE	3.22	(0.0000)	21	42.64	(0.0055)	20	1.94	(0.0168)	18	0.41	(0.6348)	13
TW-*CER*	3.56	(0.0000)	32	46.50	(0.0000)	32	2.12	(0.0000)	31	0.45	(0.4515)	30
Trimmed-*CER*	3.32	(0.0000)	28	43.55	(0.0015)	23	1.98	(0.0031)	23	0.42	(0.5626)	21
Best-*CER*	3.96	(0.0000)	36	54.28	(0.0000)	36	2.38	(0.0000)	36	0.48	(0.3695)	33
TW-$Q_{0.05}$	3.12	(0.0000)	12	40.78	(0.1088)	11	1.86	(0.1557)	9	0.41	(0.7314)	7
TW-$Q_{0.01}$	3.23	(0.0000)	22	41.34	(0.0592)	12	1.90	(0.0535)	11	0.42	(0.5914)	18
Trimmed-$Q_{0.05}$	3.23	(0.0000)	23	42.78	(0.0055)	21	1.95	(0.0113)	21	0.42	(0.6127)	16
Trimmed-$Q_{0.01}$	3.32	(0.0000)	27	43.56	(0.0020)	24	1.98	(0.0036)	24	0.42	(0.5569)	22
Best-$Q_{0.05}$	2.97	(0.0000)	4	**38.63***	(1.0000)	1	1.78*	(0.7931)	2	0.40*	(0.7380)	3
Best-$Q_{0.01}$	3.02	(0.0000)	7	38.71*	(0.9198)	2	1.80*	(0.6102)	5	0.41	(0.6483)	14
TW-*Sharpe*	3.56	(0.0000)	33	46.48	(0.0000)	31	2.12	(0.0000)	32	0.45	(0.4431)	31
Trimmed-*Sharpe*	3.32	(0.0000)	29	43.56	(0.0017)	25	1.99	(0.0029)	25	0.42	(0.5632)	23
Best-*Sharpe*	4.04	(0.0000)	37	54.57	(0.0000)	37	2.43	(0.0000)	37	0.50	(0.3736)	35

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Economic Evaluation

To compare the models' economic performance, we also evaluate the economic value of the models in the context of portfolio management. Table 4.22 reports the portfolio performance of various forecasting models applied to the crude oil futures price volatility. Examining the economic gains derived from the single models and the combinations based on HAR-type and GARCH-type models, we find that the best performing model varies across different forecasting horizons. The OLS_{EXP} combination is the dominant approach, that leads to both higher economic gains and indicates superior forecasting performance based on StepSPA test. From the economic combinations only some based on Smoothed-*Q* indicate adequate forecasting performance according to the ranking of the models, while almost all HAR models are ranked quite higher than the GARCH models.

Table 4.22 1-step ahead Economic Forecasting Performance for HAR-GARCH combinations

	CER			Sharpe		
	Loss function	StepSPA	Rank	Loss function	StepSPA	Rank
HAR-RV	0.0427	(0.0000)	28	4.09	(0.0000)	28
LHAR-RV	0.1107	(0.0020)	3	9.40	(0.0000)	2
HAR-RV-J	0.0777	(0.0000)	5	6.15	(0.0000)	5
HAR-CJ	0.1237	(0.0120)	2	6.80	(0.0010)	3
HAR-RSV	0.0649	(0.0000)	7	5.77	(0.0000)	9
GARCH	0.0307	(0.0000)	33	3.15	(0.0000)	32
EGARCH	0.0220	(0.0000)	36	2.32	(0.0000)	37
GJR	0.0278	(0.0000)	34	2.85	(0.0000)	34
APARCH	0.0236	(0.0000)	35	2.53	(0.0000)	35
FIGARCH	0.0364	(0.0000)	31	3.57	(0.0000)	31
HYGARCH	0.0318	(0.0000)	32	3.06	(0.0000)	33
Mean	0.0522	(0.0000)	20	4.73	(0.0000)	19
Geometric Mean	0.0524	(0.0000)	19	4.67	(0.0000)	22
Harmonic Mean	0.0552	(0.0000)	17	4.77	(0.0000)	18
Trimmed Mean	0.0493	(0.0000)	21	4.50	(0.0000)	23
Median	0.0464	(0.0000)	25	4.28	(0.0000)	27
AFTER	0.0595	(0.0000)	13	5.41	(0.0000)	13
DMSFE	0.0477	(0.0000)	24	4.30	(0.0000)	26
IMSFE	0.0565	(0.0000)	14	5.36	(0.0000)	14
Kernel	0.0220	(0.0000)	37	2.32	(0.0000)	36
OLS_{EXP}	**0.2025**	(1.0000)	1	**11.73**	(1.0000)	1
OLS	0.0655	(0.0000)	6	5.96	(0.0000)	6
OLS_{SQRT}	0.1099	(0.0000)	4	6.27	(0.0000)	4
TW	0.0640	(0.0000)	8	5.84	(0.0000)	7
Trimmed MSPE	0.0544	(0.0000)	18	5.16	(0.0000)	17
TW-*CER*	0.0453	(0.0000)	26	4.41	(0.0000)	24
Trimmed-*CER*	0.0487	(0.0000)	22	4.70	(0.0000)	20
Best-*CER*	0.0386	(0.0000)	29	3.76	(0.0000)	29
TW-$Q_{0.05}$	0.0621	(0.0000)	9	5.81	(0.0000)	8
TW-$Q_{0.01}$	0.0599	(0.0000)	12	5.60	(0.0000)	10
Trimmed-$Q_{0.05}$	0.0554	(0.0000)	16	5.29	(0.0000)	15
Trimmed-$Q_{0.01}$	0.0559	(0.0000)	15	5.28	(0.0000)	16
Best-$Q_{0.05}$	0.0604	(0.0000)	10	5.48	(0.0000)	11
Best-$Q_{0.01}$	0.0604	(0.0000)	10	5.48	(0.0000)	11
TW-*Sharpe*	0.0451	(0.0000)	27	4.39	(0.0000)	25
Trimmed-*Sharpe*	0.0486	(0.0000)	23	4.69	(0.0000)	21
Best-*Sharpe*	0.0376	(0.0000)	30	3.69	(0.0000)	30

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hsu et al. (2010) StepSPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 1,000 and the smoothing parameter for the mean block length is 0.1. The confidence level used is α=10%. 3. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 4. EXP and SQRT denote the exponential and the square root transformation, respectively.

Tables 4.23 and 4.24 present the results for the 5- and 22-step forecasts. Regarding the 5-step ahead forecasting horizon, the single FIGARCH model that captures the long memory effect, is the best performing model for both *CER* and *Sharpe* ratio, while all the HAR models are proved inadequate and are ranked last. It is worth to note that the economic combinations based on either *CER* or *Sharpe* ratio, are ranked higher, while they found to be superior to the rest models under the StepSPA test.

The results for the 22-step ahead forecasts, are presented in Table 4.24. In this case, the Best-*CER* and the Best-*Sharpe* combinations are found to outperform the rest combinations and the single models on *CER* and *Sharpe* ratio, respectively. It is interesting to note that the rest *CER* and *Sharpe* combinations are ranked amongst the best performers while the transformed least squares regressions and the single models are the worst performers. As a result, based on the results from the previous and the current section, we conclude that the economic combinations seem to provide more accurate results for longer (i.e. the 22-step ahead) forecast horizons, when low-frequency and high-frequency data are combined in economic terms while the statistical combinations are proved adequate for short-term forecasting horizons.

Table 4.23 5-step ahead Economic Forecasting Performance for HAR-GARCH combinations

	CER			Sharpe		
	Loss function	StepSPA	Rank	Loss function	StepSPA	Rank
HAR-RV	0.0320	(0.5470)	32	3.30	(0.0300)	32
LHAR-RV	0.0286	(0.0880)	35	2.99	(0.0070)	35
HAR-RV-J	0.0314	(0.5450)	34	3.21	(0.0280)	34
HAR-CJ	0.0366	(0.8290)	28	3.62	(0.1180)	28
HAR-RSV	0.0319	(0.4780)	33	3.28	(0.0350)	33
GARCH	0.0410	(0.8780)	4	4.39	(0.7070)	6
EGARCH	0.0401	(0.8780)	7	3.65	(0.2500)	27
GJR	0.0400	(0.9290)	8	4.07	(0.3930)	15
APARCH	0.0420	(0.9740)	2	4.21	(0.6370)	11
FIGARCH	**0.0433**	(1.0000)	1	**4.77**	(1.0000)	1
HYGARCH	0.0389	(0.7910)	13	3.95	(0.0270)	19
Mean	0.0380	(0.8640)	20	3.92	(0.0400)	21
Geometric Mean	0.0375	(0.8470)	23	3.84	(0.0300)	24
Harmonic Mean	0.0372	(0.8710)	27	3.78	(0.0370)	26
Trimmed Mean	0.0375	(0.7690)	24	3.86	(0.0240)	23
Median	0.0373	(0.7970)	26	3.81	(0.0540)	25
AFTER	0.0383	(0.9440)	16	3.93	(0.1660)	20
DMSFE	0.0380	(0.7040)	21	4.10	(0.0000)	14
IMSFE	0.0395	(0.9140)	10	4.32	(0.3410)	9
Kernel	0.0327	(0.4330)	31	3.42	(0.0110)	31
OLS_{EXP}	-0.0023	(0.1900)	37	1.29	(0.0020)	36
OLS	0.0374	(0.8340)	25	3.90	(0.0310)	22
OLS_{SQRT}	-0.0019	(0.1660)	36	1.26	(0.0000)	37
TW	0.0383	(0.7850)	17	4.04	(0.2340)	17
Trimmed MSPE	0.0388	(0.9010)	14	4.20	(0.2090)	12
TW-*CER*	0.0397	(0.5710)	9	4.49	(0.3990)	5
Trimmed-*CER*	0.0393	(0.8320)	11	4.36	(0.4250)	7
Best-*CER*	0.0405	(0.7110)	5	4.51	(0.5550)	4
TW-$Q_{0.05}$	0.0380	(0.8910)	22	4.00	(0.0660)	18
TW-$Q_{0.01}$	0.0382	(0.8780)	19	4.06	(0.0780)	16
Trimmed-$Q_{0.05}$	0.0383	(0.5060)	18	4.15	(0.0030)	13
Trimmed-$Q_{0.01}$	0.0386	(0.5610)	15	4.23	(0.2500)	10
Best-$Q_{0.05}$	0.0343	(0.7150)	30	3.45	(0.1150)	30
Best-$Q_{0.01}$	0.0351	(0.7690)	29	3.50	(0.1170)	29
TW-*Sharpe*	0.0403	(0.8320)	6	4.56	(0.8710)	3
Trimmed-*Sharpe*	0.0393	(0.7910)	11	4.36	(0.4530)	7
Best-*Sharpe*	0.0419	(0.8970)	3	4.68	(0.8650)	2

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hsu et al. (2010) StepSPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 1,000 and the smoothing parameter for the mean block length is 0.1. The confidence level used is α=10%. 3. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 4. EXP and SQRT denote the exponential and the square root transformation, respectively.

Table 4.24 22-step ahead Economic Forecasting Performance for HAR-GARCH combinations

	CER			Sharpe		
	Loss function	SPA	Rank	Loss function	SPA	Rank
HAR-RV	0.0260	(0.2620)	25	3.06	(0.2120)	27
LHAR-RV	0.0258	(0.3490)	26	2.98	(0.4030)	29
HAR-RV-J	0.0230	(0.0690)	30	2.70	(0.0630)	31
HAR-CJ	-2.3542	(0.1950)	37	-2.58	(0.3750)	37
HAR-RSV	0.0312	(0.8370)	7	3.52	(0.9600)	18
GARCH	0.0257	(0.1610)	27	3.49	(0.2870)	21
EGARCH	0.0192	(0.3310)	32	2.41	(0.4140)	32
GJR	0.0266	(0.2420)	23	3.41	(0.4290)	24
APARCH	0.0282	(0.3420)	19	3.66	(0.6040)	16
FIGARCH	0.0331	(0.8270)	3	4.35	(0.8880)	5
HYGARCH	0.0304	(0.8340)	9	3.52	(0.9550)	19
Mean	0.0279	(0.6730)	21	3.43	(0.6880)	23
Geometric Mean	0.0247	(0.5040)	29	3.04	(0.3690)	28
Harmonic Mean	-0.1094	(0.1960)	36	-1.52	(0.4030)	36
Trimmed Mean	0.0282	(0.6190)	20	3.46	(0.7930)	22
Median	0.0227	(0.3170)	31	2.85	(0.2490)	30
AFTER	0.0262	(0.1830)	24	3.35	(0.2110)	25
DMSFE	0.0292	(0.3240)	16	3.72	(0.4660)	14
IMSFE	0.0299	(0.2920)	12	3.87	(0.4150)	11
Kernel	0.0124	(0.0650)	33	1.75	(0.0070)	33
OLS_{EXP}	-0.0113	(0.2510)	34	0.71	(0.7340)	34
OLS	0.0292	(0.3620)	17	3.69	(0.5680)	15
OLS_{SQRT}	-0.0145	(0.1750)	35	0.49	(0.5610)	35
TW	0.0289	(0.4000)	18	3.63	(0.4380)	17
Trimmed MSPE	0.0297	(0.3660)	14	3.80	(0.4940)	12
TW-*CER*	0.0330	(0.1820)	4	4.45	(0.3520)	3
Trimmed-*CER*	0.0301	(0.3210)	11	3.91	(0.4660)	9
Best-*CER*	**0.0383**	(1.0000)	1	5.29	(0.6840)	2
TW-$Q_{0.05}$	0.0276	(0.2300)	22	3.51	(0.2480)	20
TW-$Q_{0.01}$	0.0309	(0.3200)	8	4.07	(0.5670)	7
Trimmed-$Q_{0.05}$	0.0292	(0.3000)	15	3.74	(0.3840)	13
Trimmed-$Q_{0.01}$	0.0298	(0.2600)	13	3.87	(0.3840)	10
Best-$Q_{0.05}$	0.0250	(0.1880)	28	3.11	(0.2880)	26
Best-$Q_{0.01}$	0.0320	(0.5870)	6	4.14	(0.8120)	6
TW-*Sharpe*	0.0326	(0.2300)	5	4.45	(0.4850)	4
Trimmed-*Sharpe*	0.0301	(0.3170)	10	3.91	(0.5160)	8
Best-*Sharpe*	0.0364	(0.7880)	2	**5.36**	(1.0000)	1

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hsu et al. (2010) StepSPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 1,000 and the smoothing parameter for the mean block length is 0.1. The confidence level used is α=10%. 3. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 4. EXP and SQRT denote the exponential and the square root transformation, respectively.

Risk Management Evaluation

We continue our analysis through the evaluation of combination volatility forecasts based on both HAR and GARCH models under a risk management framework. The results for the 1-step ahead forecasts are presented to Table 4.25. We note that the percentage of violations is adequate for several combinations and single models under a 5% confidence level, while for the 1% confidence level most models under-forecast VaR. In particular, only the Best-$Q_{0.05}$ and Best-$Q_{0.01}$ schemes forecast adequately VaR. For the 5% confidence level, the OLS_{EXP} combination produces errors of lower magnitude. It is interesting to note that four combinations based on Smoothed-*Q* loss function are amongst the best performers, while the GARCH models are proved inadequate. The SPA test reveals that only the best performing models provide significantly higher forecasting accuracy. For the 1% confidence level, the single HAR-RV-J model dominates the rest, while four combinations based on Smoothed-*Q*, i.e. the TW-$Q_{0.05}$, the TW-$Q_{0.01}$, the Best-$Q_{0.05}$ and Best-$Q_{0.01}$, are among the best performing models.

The results for the 5- and 22-step ahead forecasts (Table 4.26 and Table 4.27) indicate that there is a single model that clearly dominates all the other models; this is the EGARCH model. It is interesting to note that all the HAR-type models are the worst performers, while the SPA test indicates their statistical insignificance. Furthermore, most combinations based on Smoothed-*Q* loss functions are among the best models, while the SPA test reveals their statistical significance.

Table 4.25 1-step ahead VaR Forecasting Performance for HAR-GARCH combinations

	Smoothed-$Q_{0.05}$				Smoothed-$Q_{0.01}$			
	Percentage of violations (5%)	Loss function	SPA	Rank	Percentage of violations (1%)	Loss function	SPA	Rank
HAR-RV	5.25%	0.2112	(0.0220)	23	1.59%	0.0551	(0.2280)	22
LHAR-RV	**4.30%**	0.2010	(0.3613)	2	1.11%	0.0524	(0.5876)	8
HAR-RV-J	5.17%	0.2027	(0.1922)	5	1.11%	**0.0509**	(1.0000)	1
HAR-CJ	5.57%	0.2032	(0.2153)	6	1.51%	0.0515	(0.8195)	4
HAR-RSV	5.17%	0.2090	(0.0366)	15	1.51%	0.0540	(0.3168)	12
GARCH	4.85%	0.2231	(0.0007)	37	1.43%	0.0605	(0.0440)	35
EGARCH	4.38%	0.2132	(0.0223)	26	1.67%	0.0595	(0.0361)	29
GJR	4.61%	0.2162	(0.0073)	30	1.67%	0.0597	(0.0328)	32
APARCH	**4.30%**	0.2152	(0.0121)	29	1.51%	0.0597	(0.0314)	31
FIGARCH	5.33%	0.2223	(0.0011)	36	1.75%	0.0599	(0.0557)	33
HYGARCH	5.33%	0.2211	(0.0020)	34	1.51%	0.0597	(0.0679)	30
Mean	5.01%	0.2098	(0.0483)	17	1.27%	0.0544	(0.2206)	16
Geometric Mean	5.25%	0.2102	(0.0399)	18	1.43%	0.0547	(0.1941)	17
Harmonic Mean	5.33%	0.2105	(0.0381)	19	1.43%	0.0550	(0.1745)	19
Trimmed Mean	5.09%	0.2106	(0.0333)	20	1.43%	0.0549	(0.1743)	18
Median	5.09%	0.2127	(0.0179)	25	1.43%	0.0559	(0.1929)	24
AFTER	5.17%	0.2073	(0.0546)	10	1.19%	0.0530	(0.5695)	9
DMSFE	5.17%	0.2116	(0.0282)	24	1.43%	0.0551	(0.1636)	23
IMSFE	4.61%	0.2086	(0.0611)	12	1.19%	0.0538	(0.3359)	11
Kernel	6.52%	0.2210	(0.0020)	33	2.86%	0.0642	(0.0088)	36
OLS_{EXP}	5.41%	**0.1973**	(1.0000)	1	1.51%	0.0574	(0.1710)	27
OLS	5.17%	0.2050	(0.0712)	8	1.03%	0.0516	(0.8819)	5
OLS_{SQRT}	7.72%	0.2166	(0.0150)	31	2.70%	0.0658	(0.0133)	37
TW	4.85%	0.2073	(0.0906)	11	1.19%	0.0531	(0.4602)	10
Trimmed MSPE	4.77%	0.2091	(0.0535)	16	1.19%	0.0541	(0.2843)	13
TW-*CER*	4.61%	0.2146	(0.0104)	27	1.27%	0.0563	(0.2015)	25
Trimmed-*CER*	4.61%	0.2106	(0.0343)	21	1.19%	0.0551	(0.1784)	20
Best-*CER*	5.01%	0.2210	(0.0022)	32	1.35%	0.0594	(0.0811)	28
TW-$Q_{0.05}$	4.77%	0.2046	(0.1676)	7	1.03%	0.0517	(0.8415)	6
TW-$Q_{0.01}$	4.85%	0.2052	(0.1438)	9	1.11%	0.0519	(0.7816)	7
Trimmed-$Q_{0.05}$	4.53%	0.2089	(0.0593)	13	1.11%	0.0542	(0.2626)	15
Trimmed-$Q_{0.01}$	4.69%	0.2089	(0.0513)	14	1.11%	0.0542	(0.2759)	14
Best-$Q_{0.05}$	5.17%	0.2024	(0.2213)	4	**0.95%**	0.0509	(0.9393)	2
Best-$Q_{0.01}$	5.17%	0.2021	(0.2356)	3	**0.95%**	0.0509	(0.9350)	3
TW-*Sharpe*	4.69%	0.2147	(0.0091)	28	1.27%	0.0564	(0.1883)	26
Trimmed-*Sharpe*	4.61%	0.2107	(0.0338)	22	1.19%	0.0551	(0.1790)	21
Best-*Sharpe*	5.17%	0.2220	(0.0017)	35	1.35%	0.0600	(0.0617)	34

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. The Smoothed-Q loss function is calculated for VaR level $\alpha=0.01$ and α=0.05. We set the smoothness parameter $\delta=25$. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \le w_{it} \le 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Table 4.26 5-step ahead VaR Forecasting Performance for HAR-GARCH combinations

	Smoothed-$Q_{0.05}$				Smoothed-$Q_{0.01}$			
	Percentage of violations (5%)	Loss function	SPA	Rank	Percentage of violations (1%)	Loss function	SPA	Rank
HAR-RV	5.67%	0.2249	(0.0076)	29	2.08%	0.0661	(0.1018)	33
LHAR-RV	5.83%	0.2250	(0.0070)	30	2.00%	0.0661	(0.1027)	34
HAR-RV-J	5.91%	0.2243	(0.0069)	26	2.15%	0.0656	(0.1361)	31
HAR-CJ	6.23%	0.2258	(0.0103)	35	2.15%	0.0652	(0.2310)	30
HAR-RSV	5.67%	0.2247	(0.0122)	28	2.08%	0.0662	(0.0999)	35
GARCH	4.71%	0.2288	(0.0010)	37	1.68%	0.0658	(0.1031)	32
EGARCH	**4.39%**	**0.2130**	(1.0000)	1	1.36%	**0.0602**	(1.0000)	1
GJR	4.47%	0.2167	(0.4770)	5	1.44%	0.0617	(0.7257)	10
APARCH	4.39%	0.2165	(0.4947)	4	1.28%	0.0620	(0.6976)	12
FIGARCH	4.71%	0.2257	(0.0041)	34	1.28%	0.0637	(0.3998)	24
HYGARCH	4.87%	0.2245	(0.0165)	27	1.36%	0.0637	(0.3638)	26
Mean	5.03%	0.2196	(0.1612)	18	1.68%	0.0622	(0.6745)	15
Geometric Mean	5.11%	0.2198	(0.1329)	19	1.68%	0.0624	(0.6111)	17
Harmonic Mean	5.19%	0.2200	(0.1091)	21	1.76%	0.0626	(0.5654)	20
Trimmed Mean	5.11%	0.2199	(0.1317)	20	1.68%	0.0627	(0.5413)	21
Median	5.19%	0.2210	(0.0746)	22	1.60%	0.0633	(0.4210)	22
AFTER	5.11%	0.2177	(0.2995)	7	1.76%	0.0622	(0.6722)	16
DMSFE	4.47%	0.2184	(0.2675)	10	1.60%	0.0614	(0.8498)	5
IMSFE	4.47%	0.2190	(0.2051)	15	1.52%	0.0616	(0.7981)	8
Kernel	6.07%	0.2222	(0.0686)	25	2.08%	0.0646	(0.2807)	29
OLS_{EXP}	7.82%	0.2253	(0.0330)	31	2.63%	0.0675	(0.0669)	36
OLS	4.87%	0.2165	(0.2968)	3	1.60%	0.0610	(0.9485)	3
OLS_{SQRT}	8.22%	0.2255	(0.0197)	32	3.27%	0.0702	(0.0161)	37
TW	4.95%	0.2184	(0.2289)	11	1.92%	0.0609	(0.9255)	2
Trimmed MSPE	4.47%	0.2189	(0.2185)	14	1.68%	0.0615	(0.8236)	7
TW-*CER*	4.47%	0.2217	(0.0688)	23	1.36%	0.0625	(0.5848)	18
Trimmed-*CER*	4.47%	0.2195	(0.1884)	16	1.52%	0.0621	(0.6736)	13
Best-*CER*	4.71%	0.2256	(0.0112)	33	**1.28%**	0.0639	(0.3496)	27
TW-$Q_{0.05}$	5.03%	0.2182	(0.2639)	9	1.68%	0.0619	(0.7483)	11
TW-$Q_{0.01}$	4.87%	0.2176	(0.2808)	6	1.60%	0.0617	(0.7988)	9
Trimmed-$Q_{0.05}$	4.47%	0.2185	(0.2583)	12	1.60%	0.0614	(0.8400)	4
Trimmed-$Q_{0.01}$	4.47%	0.2186	(0.2504)	13	1.60%	0.0615	(0.8245)	6
Best-$Q_{0.05}$	5.19%	0.2179	(0.3508)	8	1.92%	0.0636	(0.3788)	23
Best-$Q_{0.01}$	4.95%	0.2160	(0.5514)	2	1.92%	0.0637	(0.3640)	25
TW-*Sharpe*	4.47%	0.2218	(0.0681)	24	1.36%	0.0626	(0.5730)	19
Trimmed-*Sharpe*	4.47%	0.2195	(0.1824)	16	1.52%	0.0621	(0.6818)	13
Best-*Sharpe*	4.63%	0.2263	(0.0086)	36	**1.28%**	0.0641	(0.3197)	28

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. The Smoothed-Q loss function is calculated for VaR level $\alpha=0.01$ and α=0.05. We set the smoothness parameter $\delta=25$. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Table 4.27 22-step ahead VaR Forecasting Performance for HAR-GARCH combinations

	Smoothed-$Q_{0.05}$				Smoothed-$Q_{0.01}$			
	Percentage of violations (5%)	Loss function	SPA	Rank	Percentage of violations (1%)	Loss function	SPA	Rank
HAR-RV	6.07%	0.2387	(0.0021)	31	2.18%	0.0686	(0.0122)	31
LHAR-RV	5.99%	0.2375	(0.0057)	28	2.35%	0.0678	(0.0211)	29
HAR-RV-J	6.07%	0.2387	(0.0054)	32	2.35%	0.0696	(0.0100)	33
HAR-CJ	6.63%	0.2416	(0.0025)	37	2.67%	0.0725	(0.0046)	36
HAR-RSV	6.07%	0.2389	(0.0021)	33	2.27%	0.0692	(0.0114)	32
GARCH	4.94%	0.2414	(0.0021)	36	1.62%	0.0665	(0.0957)	27
EGARCH	5.34%	**0.2247**	(1.0000)	1	1.29%	**0.0602**	(1.0000)	1
GJR	4.94%	0.2282	(0.4318)	3	1.21%	0.0616	(0.6624)	3
APARCH	4.85%	0.2272	(0.5468)	2	**1.13%**	0.0609	(0.7391)	2
FIGARCH	5.26%	0.2403	(0.0022)	35	1.70%	0.0653	(0.2262)	23
HYGARCH	5.50%	0.2361	(0.0323)	26	1.38%	0.0644	(0.2794)	18
Mean	5.26%	0.2336	(0.0540)	19	1.70%	0.0636	(0.2150)	6
Geometric Mean	5.34%	0.2337	(0.0394)	20	1.78%	0.0639	(0.1839)	14
Harmonic Mean	5.58%	0.2347	(0.0251)	24	2.02%	0.0650	(0.1608)	22
Trimmed Mean	5.34%	0.2337	(0.0477)	21	1.78%	0.0637	(0.2068)	8
Median	5.58%	0.2345	(0.0248)	23	1.86%	0.0638	(0.2428)	10
AFTER	5.91%	0.2285	(0.3800)	5	1.78%	0.0641	(0.2158)	16
DMSFE	5.18%	0.2308	(0.1918)	11	1.62%	0.0638	(0.1938)	11
IMSFE	5.10%	0.2327	(0.0940)	16	1.70%	0.0643	(0.2125)	17
Kernel	6.72%	0.2315	(0.2377)	12	2.67%	0.0698	(0.0237)	34
OLS_{EXP}	7.28%	0.2367	(0.0559)	27	3.48%	0.0713	(0.0117)	35
OLS	5.58%	0.2286	(0.3467)	6	1.70%	0.0625	(0.3727)	4
OLS_{SQRT}	8.98%	0.2387	(0.0270)	30	4.21%	0.0788	(0.0005)	37
TW	5.26%	0.2304	(0.2255)	8	1.70%	0.0636	(0.1395)	7
Trimmed MSPE	5.18%	0.2317	(0.1322)	13	1.62%	0.0639	(0.1789)	13
TW-*CER*	4.94%	0.2344	(0.0696)	22	1.54%	0.0656	(0.1212)	24
Trimmed-*CER*	5.10%	0.2326	(0.0984)	15	1.70%	0.0645	(0.1922)	19
Best-*CER*	**4.53%**	0.2379	(0.0239)	29	1.46%	0.0674	(0.0577)	28
TW-$Q_{0.05}$	5.34%	0.2305	(0.2032)	9	1.62%	0.0637	(0.1021)	9
TW-$Q_{0.01}$	5.10%	0.2305	(0.2230)	10	1.70%	0.0639	(0.1882)	12
Trimmed-$Q_{0.05}$	5.18%	0.2321	(0.1102)	14	1.70%	0.0640	(0.1827)	15
Trimmed-$Q_{0.01}$	5.10%	0.2327	(0.0985)	18	1.70%	0.0646	(0.1787)	21
Best-$Q_{0.05}$	5.66%	0.2298	(0.3435)	7	1.78%	0.0660	(0.1003)	26
Best-$Q_{0.01}$	5.58%	0.2284	(0.3954)	4	1.54%	0.0635	(0.3424)	5
TW-*Sharpe*	4.94%	0.2351	(0.0484)	25	1.54%	0.0657	(0.1069)	25
Trimmed-*Sharpe*	5.10%	0.2327	(0.0968)	17	1.70%	0.0645	(0.1963)	20
Best-*Sharpe*	**4.53%**	0.2403	(0.0068)	34	1.46%	0.0681	(0.0329)	30

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. The Smoothed-Q loss function is calculated for VaR level $\alpha=0.01$ and α=0.05. We set the smoothness parameter $\delta=25$. 4. We constrain the portfolio weight on the risky asset to lie between 0% and 150%, i.e. $0 \leq w_{it} \leq 1.5$ for the economic combinations. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

4.7.4 Out-of-Sample Forecasting Performance for best combinations across different loss functions

Several interesting conclusions are reached from Table 4.28 that summarizes the best performing models across all the forecasting horizons examined. Particularly, the best performing model for each different combination type is presented (i.e. the best model considering the GARCH combinations, the HAR combinations and the combinations based on both models). We find that the forecasting performance of each model/combination varies greatly with the change in horizons and the models used in each combination type.

For the 1-step ahead forecasting horizon, the transformed OLS combinations based on HAR models, i.e. the statistical combinations, indicate superior forecasting performance across almost all loss functions. This result is opposite to Ma et al. (2018d) who argue that HAR combinations do not lead to higher forecasting ability. Also, it is worth noting that according to the Smoothed-$Q_{0.01}$ loss function, an economic combination, the Trimmed-$Q_{0.05}$ is the best performer.

Considering the 5-step ahead forecasts, the best performer in statistical terms, is an economic combination derived from all models, the Best-$Q_{0.05}$ that produces the minimum losses. However, the economic evaluation suggests that two single models, the FIGARCH and the EGARCH model, lead to increased economic gains compared to combination techniques. The superior performance of the two single models is attributed to the fact that the FIGARCH model accounts for long memory and asymmetric effects, while the EGARCH model minimizes the VaR losses for the 5-step ahead forecasting period. As a result, although an economic combination point out adequate performance in statistical context, there is no evidence for using economic loss functions to optimally combine forecasts in economic terms for the 5-step ahead forecasts.

Taking into consideration the 22-step ahead forecasting horizon, GARCH combinations dominate the rest. Two economic combinations, the Best-*CER* and the Best-*Sharpe* lead to higher economic gains a mean-variance investor suggesting that economic combinations are more accurate in longer forecasting horizons as the different models combined incorporate different aspects of the market. Moreover, statistical combination techniques are superior according to symmetric statistical loss functions, while the asymmetric loss functions suggest the use of the asymmetric EGARCH model.

The main finding of our research is that it does worth combining volatility forecasts from either one model-type or more model-types, as they lead to increased forecasting accuracy and higher economic gains. The results suggest that low frequency models increase forecasting accuracy when they are combined through economic loss functions, especially in 1-step and 22-step ahead forecasting horizons. However, there is a single model that clearly dominates the rest, this is the EGARCH model that indicates superior forecasting accuracy under statistical and risk management evaluation. Although this result is not expected, a limited number of studies (Arouri et al., 2011; Wei et al., 2017) indicate that in some cases single models outperform combination techniques.

Table 4.28 Out-of-Sample performance of best combinations across all loss functions

	MAE	MSE	HRLF	QLIKE	CER	Sharpe	Smoothed-$Q_{0.05}$	Smoothed-$Q_{0.01}$
				h=1				
GARCH	OLS_{SQRT} (2.33)	OLS_{EXP} (29.93)	OLS_{SQRT} (1.24)	TW (0.28)	Best-*CER* (0.0386)	Best-*CER* (3.76)	Best-$Q_{0.05}$ (0.2130)	OLS (0.0578)
HAR	**OLS_{SQRT} (1.76)**	**OLS_{SQRT} (23.78)**	**HAR-RSV (0.92)**	**Trimmed Mean (0.21)**	**OLS_{EXP} (0.2064)**	**OLS_{EXP} (14.21)**	**OLS_{EXP} (0.1920)**	**Trimmed-$Q_{0.05}$ (0.0505)**
ALL	OLS_{SQRT} (1.78)	OLS_{SQRT} (24.32)	HAR-RSV (0.92)	HAR-RSV (0.22)	OLS_{EXP} (0.2025)	OLS_{EXP} (11.73)	OLS_{EXP} (0.1973)	HAR-RV-J (0.0509)
				h=5				
GARCH	OLS_{SQRT} (2.36)	Best-$Q_{0.05}$ (33.90)	EGARCH (1.48)	EGARCH (0.33)	**FIGARCH (0.0433)**	**FIGARCH (4.77)**	**EGARCH (0.2130)**	**EGARCH (0.0602)**
HAR	OLS_{SQRT} (2.44)	HAR-CJ (36.58)	TW (1.57)	TW (0.33)	Kernel (0.0406)	Kernel (4.27)	TW (0.2212)	AFTER (0.0619)
ALL	**OLS_{SQRT} (2.35)**	**Best-$Q_{0.05}$ (33.75)**	**Best-$Q_{0.05}$ (1.45)**	**Best-$Q_{0.05}$ (0.32)**	**FIGARCH (0.0433)**	**FIGARCH (4.77)**	**EGARCH (0.2130)**	**EGARCH (0.0602)**
				h=22				
GARCH	**OLS_{SQRT} (2.50)**	**Best-$Q_{0.05}$ (38.49)**	**EGARCH (1.75)**	**EGARCH (0.39)**	**Best-*CER* (0.0383)**	**Best-*Sharpe* (5.36)**	**EGARCH (0.2247)**	**EGARCH (0.0602)**
HAR	OLS_{SQRT} (2.75)	Kernel (42.47)	TW (2.01)	TW (0.42)	HAR-RSV (0.0312)	HAR-RSV (3.52)	OLS_{EXP} (0.2329)	OLS_{EXP} (0.0663)
ALL	OLS_{SQRT} (2.51)	Best-$Q_{0.05}$ (38.63)	**EGARCH (1.75)**	**EGARCH (0.39)**	**Best-*CER* (0.0383)**	**Best-*Sharpe* (5.36)**	**EGARCH (0.2247)**	**EGARCH (0.0602)**

Note. 1. The table presents the best performing model across all loss functions and combination types. 2. GARCH, HAR and ALL denote the combinations derived from GARCH, HAR and all models respectively. 3. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 4. Numbers in parentheses denote the value of the specific loss function.

4.8 Concluding Remarks

Forecasting the volatility of energy prices is one of the most asked question in energy economics. Given the uncertainty of energy markets the issue of selecting a single model in all cases of volatility forecasting is quite complicated. A vast literature proposes a large number of single models, based on either GARCH-type or HAR-type models, while a limited number of researches proposes combination forecasts as a good way to improve volatility forecasting accuracy.

In this chapter we investigated the benefits of combining forecasts using economic and risk management measures in the context of crude oil futures volatility, while their forecasting accuracy was compared with standard combination techniques generating 1-, 5- and 22-step ahead forecasts. The economic combination techniques are based on two portfolio loss functions, the *CER* and the *Sharpe* ratio, whilst a risk management loss function, the Smoothed-*Q* is also considered. Furthermore, we computed combinations based on HAR-type models (high-frequency data), combinations based on GARCH-type models (low-frequency data) and combinations based on both HAR and GARCH models. Lastly, we compared the forecasting ability of forecast combinations and single models in both statistical and economic terms.

From a statistical point, the volatility models are evaluated using both symmetric and asymmetric loss functions. The results suggest that the square root transformed OLS model outperforms the other single and combination techniques under most statistical loss functions. However, considering the multi-step ahead forecasts, the economic Best-$Q_{0.05}$ combination exhibits better forecasting performance for the 5-step ahead forecasts. Our results are even corroborated from the SPA and the MCS tests. It is worth noting that simple combinations indicate similar and sometimes worse performance to the single models suggesting the use of more complex combination techniques for volatility prediction. This finding is consistent with the literature, for example Claessen and Mittnik (2002), Fuertes et al. (2009), Li et al. (2013) and Yang et al. (2015). However, our results suggest that a single model, the EGARCH model, exhibits often better forecasting performance than the combination techniques and models based on the high-frequency dataset. The superior performance of the EGARCH model can be attributed to the fact that it takes into consideration the asymmetric response of volatility to positive and negative returns.

Regarding the combinations derived from economic and risk management loss functions, the examination of the economic loss functions results yields several useful conclusions. We summarize the results considering the combinations derived from the different model-types. Firstly, for the HAR combinations, there is no improvement to the economic gains under the economic combinations (except for one case), as more complex statistical techniques perform better lead to higher economic gains for all the examined horizons. Taking into consideration the GARCH combinations, the economic gains derived from the economic combinations, the Best-*CER* and the Best-*Sharpe*, are significantly superior under the *CER* and the *Sharpe* loss functions for the 1- and 22-step ahead forecasts. Surprisingly, for the 5-step ahead forecasting horizon, two single models, the FIGARCH and the EGARCH model outperform all the combination schemes. Finally, the economic combinations based on all models, are more accurate in statistical terms only for the 5-step ahead forecasting horizon. To this end, we conclude that although combining volatility forecasts leads to increased forecasting accuracy in both statistical and economic terms, economic combinations increase the economic gains for an investor only when low-frequency data are considered.

Chapter 5

Conclusions and Suggestions for Further Work

This thesis was primarily concerned with the development and application of new combination techniques in financial applications. The aim has been threefold: first to explore the existing methodologies to combine forecasts and underline the benefits derived from combination forecasts; second to use a large variety of combination techniques to optimally combine volatility forecasts and evaluate them through statistical and economic measures; and thirdly to develop an innovative methodology for optimally combine volatility forecasts based on economic loss functions. Although this thesis has dealt with a small number of the numerous issues derived from volatility forecasting, we have managed to establish some important results. This chapter draws together the evidence from the particular components of this study and provides some suggestion for future research.

Chapter 2 of this thesis reviews the existing literature on combination forecasting in financial econometrics. The review concludes that combination forecasts achieve improved forecasting accuracy compared with their single counterparts and influence significantly financial decisions. In addition, a number of stylized facts have been reported in the literature concerning the combination forecasts properties. The most important is that combination forecasts reduce uncertainty risk, while they are more robust to unknown instabilities. As a result, a large variety of alternative combination techniques ranging from regression approaches to time-varying weights based on the forecasting performance of the single models during a training period has been developed but there is no clear winner amongst different methodologies. The most exciting is that a simple average of forecasts is often reported as the best performing model for point forecasts.

In Chapter 3 of this thesis, we use an exhaustive variety of combination techniques to optimally combine volatility forecasts of S&P500 index. Firstly, models based on daily, high-frequency and implied volatility data are used. We use all data types to the combination techniques as it is expected to add to the forecasting performance of the combinations. Secondly, we use various simple and more complex combination techniques, widely used in the literature, to predict the volatility of S&P500 index. Thirdly, we explore whether combination forecasts outperform the single models based on both statistical and economic loss functions.

We find similar results for both statistical and economic evaluation. Combination forecasts improve volatility forecasting compared to single models. Using symmetric and asymmetric loss functions, the OLS-based schemes outperform the other single and combination models. Moreover, the tests for statistical significance assure their superior forecasting performance. Furthermore, we find that OLS-based schemes outperform again the rest models under the economic loss functions. To this end, the forecasts derived from regression model combinations exhibit superior forecasting performance in both statistical and economic terms.

Although this finding is opposite to the combination literature for point forecasts that suggests the simple combinations as the best performing model, it is related to the findings of volatility combination forecasting. For example Fuertes et al. (2009) and Li et al. (2013) suggest the use of regression approaches to optimally combine volatility forecasts in the context of stock market volatility. In these combinations, the optimal weights are updated daily and incorporate new different market microstructures that vary over time, while the regression-based forecasts account more properly for bias correction. Moreover, the combination of three sources of information, i.e. the combination of different information channels is more efficient than combining models based on the same dataset, as each volatility model reduces the "model uncertainty" and improves the combination forecast in both statistical and economic terms. On the contrary, the simple average does not take into consideration the forecasting error of each model and leads sometimes to increased errors as simple combinations do not account for the different properties of each volatility model. However, the results indicate that there is no clear winner across all loss functions, suggesting that different combination schemes are preferable based on the economic application to be used.

Chapter 4 of this thesis is primarily concerned with the development of a new methodology for combining volatility forecasts through economic and risk management loss functions. Firstly, combinations based on two economic loss functions, widely used in the portfolio evaluation procedure (i.e. the *CER* and *Sharpe* ratio), and a risk management loss function (i.e. the Smoothed-*Q*) are used to combine volatility forecasts for crude oil futures prices. Their forecasting ability is compared to standard combination techniques. Secondly, daily and intraday data are used in this application, resulting in three combination schemes: GARCH combinations, HAR combinations and combinations based on both GARCH and HAR models. Thirdly, all the combinations and the single models are evaluated through a statistical and economic framework.

Under a statistical perspective, we find that more complex techniques, such as the regression approaches, outperform the single and other combinations for all the three combinations considered. Only a combination based on the risk management loss function, the Best-$Q_{0.05}$ is found to exhibit a better performance in some cases. However, it is interesting to mention that some single models, such as the EGARCH model outperforms many combination schemes for longer forecast horizons.

The empirical results of this study indicate that the economic combinations are more profitable when GARCH combinations or combinations based on all models are used. However, the economic combinations do not perform stable across different forecast horizons. The reported evidence points out the use for economic combination techniques for 1-step and 22-step ahead forecasts. This result is reinforced by the statistical significance of the economic gains during the examined period. We argue that combining volatility forecasts of various GARCH-type or all models generate higher gains in terms of *CER* and *Sharpe* ratio for 22-step ahead forecasts. Interestingly, the combinations based on a risk management loss function are not found to forecast volatility adequately for risk management applications. Turning to HAR combinations, the economic combinations do not exhibt a clear advantage, as statistical combinations are more robust in economic terms even if they are ranked among the best performers. These results are opposite to Ma et al. (2018d) who argue that there is no advantage to combine HAR-type models as the simple HAR models outperform their combinations in statistical and economic terms. Finally, the combination of two information channels increases the economic gains for an investor for longer forecasting horizons. We conclude that although it does worth combine volatility forecasts through economic loss functions (especially in the case of low-frequency models) as there is a substantial improvement according to economic loss functions, the results are not stable across different model-types and forecasting horizons.

Several issues for further research arise from this thesis. Chapter 3 deals with the importance of economic evaluation in volatility forecasts combinations using a large variety of simple and more complex combination techniques. An alternative course of future research is to investigate whether there are economic gains from volatility forecasts under a multivariate modeling approach. For example, a series of studies including Ledoit et al. (2003), Bauwens et al. (2006), Laurent et al. (2012) amongst others compare the predictive ability of various multivariate models. However, there is no study that examines the economic gains of these models or their combinations.

The development of new combination techniques based on economic and risk management loss functions is examined in chapter 4 of the thesis. The results indicated a relative improvement in economic terms for GARCH combinations and combinations based on all models. An interesting topic for future research would be to study applications where the new combination techniques will be implemented to financial economic problems. For example, the proposed economic volatility combination forecasts could be used for hedging risk. Furthermore, these methodologies can be used for measuring volatility of several asset types, e.g. currencies, interest rates, commodities, option prices. One more extension is to compare the performance of the new methods with alternative combination schemes that take into consideration different economic loss functions. To this end, an interesting question that arises from the empirical evidence of this thesis is whether a different economic loss function or a different combination scheme could achieve higher economic gains.

References

Ahoniemi, K. and Lanne, M. (2013) “Overnight stock returns and realized volatility”, *International Journal of Forecasting*, **29**, pp. 592-604

Aiolfi, M. and Timmermann, A. (2006) “Persistence in forecasting performance and conditional combination strategies”, *Journal of Econometrics*, **135(1-2)**, pp. 31-53

Aksu, C. and Gunter, S.I. (1992) “An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts”, *International Journal of Forecasting*, **8**, pp. 27-43

Amendola, A. and Storti, G. (2016) “A comparison of different procedures for combining high-dimensional multivariate volatility forecasts”, *Topics in Theoretical and Applied Statistics*, pp. 263-273

Amendola, A. and Storti, G. (2008) “A GMM procedure for combining volatility forecasts”, *Computational Statistics & Data Analysis*, **52(6)**, pp. 3047-3060

Andersen, T.G., Bollerslev, T. and Diebold, F.X. (2007) “Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility”, *The Review of Economics and Statistics*, **89(4)**, pp. 701-720

Andersen, T.G. and Bollerslev, T. (1998) “Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts”, *International Economic Review*, **39(4)**, pp. 885-905

Arouri, M.H., Nguyen, D.K., Lévy, A. and Lahiani, A. (2012) "Forecasting the conditional volatility of oil spot and futures prices with structural breaks and long memory models", *Energy Economics*, **34**, pp. 283-293

Asai, M., McAleer, M. and Medeiros, M.C. (2012) "Asymmetry and long memory in volatility modeling", *Journal of Financial Econometrics*, **10(3)**, pp. 495-512

Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996) "Fractionally integrated generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, **74(1)**, pp. 3-30

Basel Committee on Banking and Supervision (2004) *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*. Bank for International Settlements, Basel, Switzerland

Basel Committee on Banking and Supervision (1996) *Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements*. Bank for International Settlements, Basel, Switzerland

Bates, J.M. and Granger, C.W.J. (1969) "The combination of forecasts", *Operations Research Quarterly*, **20**, pp. 451-468

Baumeister, C. and Kilian, I. (2015) "Forecasting the real price of oil in a changing world: A forecast combination approach", *Journal of Business and Economic Statistics*, **33**, pp. 338-351

Baumeister, C. and Kilian, I. (2012) "Real-Time forecasts of the real price of oil", *Journal of Business and Economic Statistics*, **30**, pp. 326-336

Baumeister, C., Kilian, L. and Lee, T.K. (2014) "Are there gains from pooling real-time oil price forecasts?", *Energy Economics*, **46**, pp. S33-S43

Bauwens, L., Laurent, S. and Rombouts, J.V.K. (2006) "Multivariate GARCH models: A survey", *Journal of Applied Econometrics*, **21(1)**, pp. 79-109

Becker, R. and Clements, A. (2008) "Are combination forecasts of S&P 500 volatility statistically superior?", *International Journal of Forecasting*, **24(1)**, pp. 122-133

Benavides, G. and Capistrán, C. (2012) "Forecasting exchange rate volatility: The superior performance of conditional combinations of time series and option implied forecasts", *Journal of Empirical Finance*, **19**, pp. 627-639

Black, F. and Scholes, M. (1973) "The pricing of options and corporate liabilities", *Journal of Political Economy*, **81(3)**, pp. 637-654

Bollerslev, T. (1986) "Generalized autoregressive conditional heteroscedasticity", *Journal of Econometrics*, **31**, pp. 307-327

Campbell, J.Y. and Thompson, S.B. (2008) "Predicting the equity premium out of sample: Can anything beat the historical average?", *Review of Financial Studies*, **21**, pp. 1509-1531

Capistrán, C. and Timmermann, A. (2009) "Forecast combination with entry and exit of experts", *Journal of Business and Economic Statistics*, **27(4)**, pp. 559-583

CBOE (2014) The CBOE Volatility Index-VIX, White Paper

CBOE (2003) The CBOE Volatility Index-VIX, White Paper.

Chan, L., Stock, J.H. and Watson, M.W. (1999) "A dynamic factor model framework for foacat combination", *Spanish Economic Review*, **1**, pp. 91-121

Charles, A. and Darné, O. (2017) "Forecasting crude-oil market volatility: Further evidence with jumps", *Energy Economics*, **67**, pp. 508-519

Cheng, G. and Yang, Y. (2015) "Forecast combination with outlier protection", *International Journal of Forecasting*, **31**, pp. 223-237

Chou, R.Y., Chou, H. and Liu, N. (2010) 'Range volatility models and their applications in finance', *in* Handbook of Quantitative Finance and Risk Management, Springer US, pp 1273-1281

Claessen, H. and Mittnik, S. (2002) "Forecasting stock market volatility and the informational efficiency of the DAX-index options market", *The European Journal of Finance*, **8(3)**, pp. 302-321

Clemen, R.T. (1989) "Combining forecasts: A review and annotated bibliography", *International Journal of Forecasting*, **5**, pp. 559-583

Clemen, R.T. (1986) "Linear constraints and the efficiency of combined forecasts", *Journal of Forecasting*, **5**, pp. 31-38

Corsi, F., Pirino, D. and Reno, R. (2010) "Threshold bipower variation and the impact of jumps on volatility foecasting", *Journal of Econometrics*, **159**, pp. 276-288

Corsi, F. (2009) "A simple approximate long memory model of realized volatility", *Journal of Financial Econometrics*, **7(2)**, pp. 174-196

Dark, J. (2015) “Futures hedging with Markov switching vector error correction FIEGARCH and FIAPARCH”, *Journal of Banking & Finance*, **61**, pp. 269-285

Davidson, J. (2004) “Moment and memory properties of linear conditional heteroskedasticity models, and a new model”, *Journal of Business & Economic Statistics*, **22(1)**, pp. 16-29

Davies, P. (2007) “What’s the value of an energy economist?”, Wellington, New Zealand: Speech presented at the International Association for Energy Economics

Degiannakis, S. and Filis, G. (2017) “Forecasting oil price realized volatility using information channels from other asset classes”, *Journal of International Money and Finance*, **76**, pp. 28-49

Diebold, F.X. (1988) “Serial correlation and the combination of forecasts”, *Journal of Business and Economic Statistics*, **6**, pp. 105-111

Diebold, F.X. and Pauly, P. (1987) “Structural change and the combination of forecasts”, *Journal of Forecasting*, **6**, pp. 21-40

Ding, Z., Granger, C.W.J. and Engle, R.F. (1993) “A long memory property on stock market returns and a new model”, *Journal of Empirical Finance*, **1**, pp. 83-106

Engle, R.F. (1982) “Autoregressive conditional heteroscedasticity with estimates of the variande of United Kingdom inflation”, *Econometrica*, **50(4)**, pp. 987-1007

Engle, R.F., Ghysels, E. and Sohn, B. (2013) “Stock market volatility and macroeconomic fundamentals”, *Review of Economics and Statistics*, **95(3)**, pp. 776-797

Engle, R.F., Kane, A. and Noh, J. (1997) “Index-option pricing with stochastic volatility and the value of the accurate variance forecasts”, *Review of Derivatives*, **1**, pp. 120-144

Fuertes, A.M., Izzeldin, M. and Kalotychou, E. (2009) “On forecasting daily stock volatility: The role of intraday information and market conditions”, *International Journal of Forecasting*, **25**, pp. 259-281

Genre, V., Kenny, G., Meyler, A. and Timmermann, A. (2013) “Combining expert forecasts: Can anything beat the simple average?”, *International journal of Forecasting*, **29(1)**, pp. 108-121

Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993) “On the relation between the expected value and the volatility of the nominal excess return on stocks”, *The Journal of Finance*, **48(5)**, pp. 1779-1801

Gong, X. and Lin, B. (2018) “The incremental information content of investor fear gauge for volatility forecasting in the crude oil futures market”, *Energy Economics*, **74**, pp. 370-386

González-Rivera, G., Lee, T.H. and Mishra, S. (2004) “Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood”, *International Journal of Forecasting*, **20**, pp. 629-645

Granger, C.W. and Ramanathan, R. (1984) “Improved method of combining forecasts”, *Journal of Forecasting*, **3(2)**, pp. 197-204

Granger, C.W.J. and Jeon, Y. (2004) “Thick modeling”, *Economic Modelling*, **21(2)**, pp. 323-343

Granger, C.W.J. and Pesaran, M.H. (2000) “Economic and statistical measures of forecast accuracy”, *Journal of Forecasting*, **19(7)**, pp. 537-560

Gurung, B., Singh, K.N., Paul, R.K., Panwar, S., Gurung, B. and Lepcha, L. (2017) “An alternative method for forecasting price volatility by combining models”, *Communications in Statistics-Simulation and Computation*, **46(6)**, pp. 4627-4636

Halbleib, R. and Pohlmeier, W. (2012) “Improving the Value at Risk forecasts: Theory and evidence from financial crisis”, *Journal of Economic Dynamics & Control*, **26**, 1212-1228

Hamilton, J.D. (2009) “Understanding crude oil prices”, *Energy Journal*, **30**, pp. 179-206

Hansen, P., Lunde, A. and Nason, J. (2011) “The Model Confidence Set”, *Econometrica*, **79(2)**, pp. 453-497

Hansen, B. (2008) “Least-squares forecast averaging”, *Journal of Econometrics*, **146(2)**, pp. 342-350

Hansen, P. (2005) “A test for Superior Predictive Ability”, *Journal of Business & Economic Statistics*, **23(4)**, pp. 365-380

Härdle, W. (1990) *Applied Nonparametric Regression*, Cambridge University Press

Harvey, D.I., Leybourne, S.J. and Newbol, P. (1998) “Tests for forecast encompassing”, *Journal of Business & Economic Statistics*, **16(2)**, pp. 254-259

Haugom, E., Langeland, H., Molnár, P. and Westgaard, S. (2014) “Forecasting volatility of the US oil market”, *Journal of Banking & Finance*, **47**, pp. 1-14

Hou, A. and Suardi, S. (2012) “A nonparametric GARCH model of crude oil price return volatility”, *Energy Economics*, **34**, pp. 618-626

Hsu, P.H., Hsu, Y.C. and Kuan, C.M. (2010) “Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias”, *Journal of Empirical Finance*, **17(3)**, pp. 471-484

Jiang, I., Hung, J. and Wang, C. (2013) “Volatility forecasters: Do volatility estimators and evaluation methods matter?”, *The Journal of Futures Market*, **34(11)**, pp. 1077-1094

Jose, V.R.R. and Winkler, R.L. (2008) “Simple robust averages of forecasts: Some empirical results”, *International Journal of Forecasting*, **24**, pp. 163-169

Kambouridis, D.,S., McMillan, D.G. and Tsakou, K. (2016) “Forecasting stock return volatility: A comparison of GARCH, Implied Volatility, and Realized Volatility models”, *Journal of Futures Markets*, **36(12)**, pp. 1127-1163

Kang, S.H., Kang, S.M. and Yoon, S.M. (2009) “Forecasting volatility of crude oil markets”, *Energy Economics*, **31**, pp. 119-125

Klein, T. and Walther, T. (2016) “Oil price volatility forecast with mixture memory GARCH”, *Energy Economics*, **58**, pp. 46-58

Koenker, R. and Basseti, G. (1978) “Regression quantiles”, *Econometrica*, **46(1)**, pp. 33-50

Kolassa, S. (2011) “Combining exponential smoothing forecasts using Akaine weights”, *International Journal of Forecasting*, **27(2)**, pp. 238-251

Laurent, S., Rombouts, J. and Violante, F. (2012) "On the forecasting accuracy of multivariate GARCH models", *Journal of Applied Econometrics*, **27(6)**, pp. 934-955

Ledoit, O., Santa-Clara, P. and Wolf, M. (2003) "Flexible multivariate GARCH modeling with an application to international stock markets", *Review of Economics and Statistics*, **85(3)**, pp. 735-747

Ledoit, O. and Wolf, M. (2003) "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection", *Journal of Empirical Finance*, **10(5)**, pp. 603-621

Li, Y., Huang, W.P. and Zhang, J. (2013) "Forecasting volatility in the Chinese stock market under model uncertainty", *Economic Modelling*, **35**, pp. 231-234

Lin, H., Wu, C. and Zhou, G. (2018) "Forecasting corporate bond returns with a large set of predictors: An iterated combination approach", *Management Science*, **64(9)**, pp. 4218-4238

Liu, C. and Maheu, J.M. (2009) "Forecasting realized volatility: A Bayesian model averaging approach", *Journal of Applied Econometrics*, **24**, pp. 709-733

Lupoletti, W.M. and Webb, R.H. (1986) "Defining and improving the accuracy of macroeconomic forecasts: Contributions from a VAR model", *The Journal of Business*, **59(2)**, pp. 263-285

Lux, T., Segnon, M. and Gupta, R. (2016) "Forecasting crude oil price volatility and value-at-risk: Evidence from historical and recent data", *Energy Economics*, **56**, pp. 117-133

Ma, F., Zhang, Y., Huang, D. and Lai, X. (2018a) “Forecasting oil futures price volatility: New evidence from realized range-based volatility”, *Energy Economics*, **75**, pp. 400-409

Ma, F., Wei, Y., Liu, L., Huang. D. (2018b) “Forecasting realized volatility of oil futures market: A new insight”, *Journal of Forecasting*, pp. 1-18

Ma, F., Liu, J., Wahab, M.I.M. and Zhang, Y. (2018c) “Forecasting the aggregate oil price volatility in a data-rich environment”, *Economic Modelling*, **72**, pp. 320-332

Ma, F., Li, Y., Liu, L. and Zhang, Y. (2018d) “Are low-frequency data really uninformative? A forecasting combination perspective”, *North American Journal of Economics and Finance*, **44**, pp. 92-108

Ma, F., Wahab, M.I.M., Huang, D. and Xu, W. (2017) “Forecasting the realized volatility of the oil futures market: A regime switching approach”, *Energy Economics*, **67**, pp. 136-145

Makridakis, S and Hibon, M. (2003), “The M3-Competition”, *International Journal of Forecasting*, **16**, pp. 433-436

Marsellino, M. (2004) “Forecast pooling for European macroeconomic variables”, *Oxford Bulletin of Economics and Statistics*, **66**, pp. 91-112

Martens, M. (2002) “Measuring and forecasting S&P 500 index-futures volatility using high-frequency data”, *Journal of Futures Markets*, **22(6)**, pp. 497-518

Martens, M. and Zein, J. (2004) “Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility”, *Journal of Futures Markets*, **24**, pp. 1005-1028

McAleer, M., Jiménez Martín, J.A. and Pérez-Amaral, T. (2011) “GFC-Robust risk management strategies under the Basel Accord”, *International Review of Economics & Finance*, **27**, 97-111

Mohammadi, H. and Su, L. (2010) “International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models”, *Energy Economics*, **32(5)**, pp. 1001-1008

Nelson, D.B. (1991) “Conditional heteroscedasticity in asset returns: A new approach”, *Econometrica*, **59**, pp. 347-370

Opschoor, A., Van Dijk, D. and Van derWel, M. (2017) “Combining density forecasts using focused scoring rules”, *Journal of Applied Econometrics*, **32**, 1298-1313

Palm, F.C. and Zellner, A. (1992) “To combine or not to combine? Issues of combining forecasts”, *Journal of Forecasting*, **11(8)**, pp. 687-701

Patton, A.J. (2011) “Volatility forecast comparison using imperfect volatility proxies”, *Journal of Econometrics*, **160**, pp. 246-256

Patton, A.J. and Sheppard, K. (2015) “Good volatility, bad volatility : Signed Jumps and the persistence of volatility”, *The Review of Economics and Statistics*, **97(3)**, pp. 683-697

Patton, A.J. and Sheppard, K. (2009) “Optimal combinations of realized volatility estimators”, International Journal of Forecasting, ***25(2)**, pp. 218-238*

Patton, A.J. and Timmermann, A. (2007) “Properties of optimal forecasts under asymmetric loss and nonlinearity”, *Journal of Econometrics*, **140(2)**, pp. 884-918

Pierdzioch, C., Döpke, J. and Hartmann, D. (2008) "Forecasting stock market volatility with macroeconomic variables in real time", *Journal of Economics and Business*, **60**, pp. 256-276

Politis, D.N. and Romano, J.P. (1994) "The stationary bootstrap", *Journal of the American Statistical Association*, **89**, pp. 1303-1313

Raftery, A.E., Kárný, M. and Ettler, P. (2010) "Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill", *Technometrics*, **52(1)**, pp. 52-66

Rapach, D.E., Strauss, J.K. and Zhou, G. (2010) "Out-of-sample equity premium prediction: Combination forecasts and links to real economy", *Review of Financial Studies*, **23(2)**, pp. 821-862

Romano, J.P. and Wolf, M. (2005) "Stepwise multiple testing as formalized data snooping", *Econometrica*, **73(4)**, pp. 1237-1282

Sadorsky, P. (2006) "Modelling and forecasting petroleum futures volatility", *Energy Economics*, **28**, pp. 467-488

Sánchez, I. (2008) "Adaptive combination of forecasts with application of wind energy", *International Journal of Forecasting*, **24(4)**, pp. 679-693

Sévi, B. (2014) "Forecasting the volatility of crude oil futures using intraday data", *Journal of Operational Research*, **235(1)**, pp. 643-659

Smith, J. and Wallis, K.F. (2009) "A simple explanation of the forecast combination puzzle", *Oxford Bulletin of Economics and Statistics*, **71(3)**, pp. 331-355

Stock, J.H. and Watson, M.W. (2004) "Combination forecasts of output growth in a seven-country data set", *Journal of Forecasting*, **23**, pp. 405-430

Stock, J.H. and Watson, M.W. (2003) "Forecasting output and inflation: The role of asset prices", *Journal of Economic Perspectives*, **41**, pp. 788-829

Swanson, N.R. and Zeng, T. (2001) "Choosing among competing econometric forecasts: Regression-based forecast combination using model selection", *Journal of Forecasting*, **20**, pp. 425-440

Tian, S. and Hamori, S. (2015) "Modeling interest rate volatility: A realized GARCH approach", *Journal of Banking & Finance*, **61**, pp. 158-171

Tibshirani, R. (1996) "Regression Shrinkage and Selection via the Lasso", *Journal of Royal Statistical Society. Series B (Methodological)*, **58(1)**, pp. 267-288

Timmermann, A. (2006) "Forecast Combinations", *Handbook of Economic Forecasting*, **1**, pp. 135-196

Tsangari, H. (2007) "An alternative methodology for combining different forecasting models", *Journal of Applied Statistics*, **34(4)**, pp. 403-421

Tsiotas, G. (2015) "A quasi-Bayesian model averaging approach for conditional quantile models", *Journal of Statistical Computation and Simulation*, **10**, 1963-1986

Varian, H.R. (1975) A Bayesian approach to real estate assessment, *in* Zellner, A. and Fienberg, S.E., ed.: *Studies in Bayesian Econometrics and Statistics in Honor of L.J. Savagerol.* pp. 195-208, North-Holland: Amsterdam

Wang, Y., Liu, L. and Wu, C. (2017) "Forecasting the real prices of crude oil using forecast combinations over time-varying parameter models", *Energy Economics*, **66**, pp. 337-348

Wang, Y., Wu, C. and Yang, L. (2016) "Forecasting crude oil market volatility: A Markov switching multifractal volatility approach", *International Journal of Forecasting*, **32**, pp. 1-9

Wei, Y., Liu, J., Lai, X. and Hu, Y. (2017) "Which determinant is the most informative in forecasting crude oil market volatility? Fundamental, speculation, or uncertainty?", *Energy Economics*, **68**, pp. 141-150

Wei, Y., Wang, Y. and Huang, D. (2010) "Forecasting crude oil market volatility: Further evidence using GARCH-class models", *Energy Economics*, **32**, pp. 1477-1484

Weiss, C.E., Raviv, E. and Roetzer, G. (2018) "Forecast combinations in R using the ForecastComb package", *The R Journal*, **10(2)**, pp. 262-281

Wen, F., Gong, X. and Cai, S. (2016) "Forecasting the volatility of crude oil futures using HAR-type models with structural breaks", *Energy Economics*, **59**, pp. 400-413

Yang, Y. (2004) "Combining forecasting procedures: some theoretical results", *Econometric Theory*, **20(1)**, pp. 176-222

Yang, K., Chen, L. and Tian, F. (2015) "Realized volatility forecast of stock index under structural breaks", *Journal of Forecasting*, **34**, pp. 57-82

Zellner, A. (1986) "Bayesian estimation and prediction using asymmetric loss functions", *Journal of American Statistical Association*, **81**, pp. 446-451

Zhang, Y., Wei, Y., Zhang, Y. and Jin, D. (2019) "Forecasting oil price volatility: Forecast combination versus shrinkage method", *Energy Economics*, **80**, pp. 423-433

Zhang, Y., Ma, F. and Shi, B. (2018) "Forecasting the prices of crude oil: An iterated combination approach", *Energy Economics*, **70**, pp. 472-483

Zhang, Y. and Ma, F. (2018) Out-of-sample prediction of the oil futures market volatility: A comparison of new and traditional combination approaches. Working Paper.

Zhou, H. and Yang, Y. (2004) "Combining time series models for forecasting", *International Journal of Forecasting*, **20**, pp. 69-84

Appendix

Table A.1: Statistical Evaluation under symmetric loss functions

	MAE			MSE		
	Loss function	SPA	Rank	Loss function	SPA	Rank
GARCH	0.6242	(0.0000)	28	2.2469*	(0.1930)	27
EGARCH	0.6925	(0.0000)	30	2.0737*	(0.6123)	20
FIGARCH	0.6500	(0.0000)	29	2.5466*	(0.0338)	30
MIDAS-RV	0.5976	(0.0000)	26	2.1044*	(0.5845)	22
MIDAS-CPI	0.5965	(0.0000)	25	2.0631*	(0.6909)	18
MIDAS-IP	0.5885	(0.0000)	23	2.0442*	(0.8093)	16
ARMA	0.5590	(0.0000)	17	2.1565*	(0.3947)	25
HAR	0.5429	(0.0000)	11	2.1490*	(0.4195)	24
ARFIMA	0.5415	(0.0000)	10	2.2846*	(0.1766)	28
VIX	0.5349	(0.0000)	7	2.0006*	(0.9526)	4
Mean	0.5697	(0.0000)	22	2.0205*	(0.9832)	6
Geometric Mean	0.5542	(0.0000)	13	2.0351*	(0.9390)	13
MSFE	0.5552	(0.0000)	14	2.0369*	(0.9409)	15
$OLS_{C\text{-}SQRT}$	0.4330*	(0.3335)	3	2.0710*	(0.6977)	19
$OLS_{NC\text{-}EXP}$	0.4283*	(0.5617)	2	2.2203*	(0.2991)	26
$NERLS_{C\text{-}SQRT}$	**0.4279***	(1.0000)	1	2.0007*	(0.9915)	5
$NRLS_{C}$	0.4978	(0.0000)	4	1.9909*	(0.9812)	3
Kernel	0.5157	(0.0000)	6	2.3261*	(0.1684)	29
IMSFE	0.5654	(0.0000)	18	2.0313*	(0.9374)	9
Nonlinear	0.5671	(0.0000)	20	2.0323*	(0.9291)	11
TW	0.5374	(0.0000)	8	**1.9831***	(1.0000)	1
Trimmed MSPE	0.5553	(0.0000)	16	2.0251*	(0.9698)	7
Harmonic Mean	0.5398	(0.0000)	9	2.0620*	(0.8346)	17
AFTER	0.5540	(0.0000)	12	2.1240*	(0.5320)	23
Shrinkage DMSFE	0.5553	(0.0000)	15	2.0365*	(0.9368)	14
Shrinkage $OLS_{NC\text{-}EXP}$	0.6191	(0.0000)	27	2.0307*	(0.8973)	8
Shrinkage $NRLS_{C}$	0.4984	(0.0000)	5	1.9906*	(0.9798)	2
Shrinkage IMSFE	0.5655	(0.0000)	19	2.0318*	(0.9350)	10
Shrinkage Nonlinear	0.5671	(0.0000)	21	2.0324*	(0.9289)	12
Shrinkage TW	0.5942	(0.0000)	24	2.1040*	(0.5863)	21

Note. 1. Values in bold denote that the corresponding model attains the best forecasting performance under the specific loss function. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. The Shrinkage factor is set equal to *d=0.50*. However, we included to our calculations the values 0.25 and 1 and all gave us similar results and ranking. 5. EXP and SQRT denote the exponential and the square root transformation, respectively.

Table A.2: Statistical Evaluation under asymmetric loss functions

	HRLF *(b=-1)*			QLIKE			LINEX (α=0.5)			LINEX (α=1)		
	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank	Loss function	SPA	Rank
GARCH	0.3410	(0.0771)	25	0.3598	(0.0001)	26	0.2572	(0.3875)	28	6.7130	(0.8289)	25
EGARCH	0.3398	(0.1912)	24	0.3755	(0.0000)	27	0.2170	(0.4913)	23	2.4278	(0.8362)	21
FIGARCH	0.3984	(0.0187)	30	0.4022	(0.0004)	30	0.4404	(0.2153)	29	33.5479	(0.4363)	29
MIDAS-RV	0.3064*	(0.4169)	20	0.3267	(0.0932)	19	0.2496	(0.4603)	27	7.4104	(0.7477)	26
MIDAS-CPI	0.3045*	(0.4423)	17	0.3306	(0.0770)	22	0.2486	(0.4108)	26	8.9119	(0.7082)	27
MIDAS-IP	0.3025*	(0.4534)	16	0.3264	(0.1026)	18	0.2068	(0.5283)	22	5.1104	(0.8582)	24
ARMA	0.3466	(0.0662)	27	0.3568	(0.0011)	25	0.1370	(0.7420)	12	1.2870	(0.8758)	12
HAR	0.3432	(0.0828)	26	0.3471	(0.0068)	24	0.1212	(0.7833)	7	0.7029	(0.8729)	5
ARFIMA	0.3795	(0.0464)	29	0.3896	(0.0029)	29	0.1566	(0.6743)	15	3.0749	(0.8647)	22
VIX	0.2960*	(0.7098)	5	0.3234	(0.0977)	15	0.0881	(0.8528)	4	0.3026	(0.9032)	3
Mean	0.3017*	(0.5481)	14	0.3269	(0.0720)	20	0.1427	(0.7312)	13	1.4628	(0.8983)	13
Geometric Mean	0.3025*	(0.5392)	15	0.3228	(0.0829)	14	0.1273	(0.7682)	9	1.0604	(0.9003)	8
MSFE	0.2979*	(0.5751)	9	0.3161	(0.1618)	7	0.1594	(0.6603)	20	2.2119	(0.8806)	20
$OLS_{C\text{-}SQRT}$	0.2974*	(0.3678)	7	0.3035	(0.1724)	4	0.0879*	(0.9012)	3	0.3355	(0.9287)	4
$OLS_{NC\text{-}EXP}$	0.3338*	(0.1399)	23	0.3248	(0.1196)	17	**0.0801***	(1.0000)	1	**0.2149***	(1.0000)	1
$NERLS_{C\text{-}SQRT}$	0.2828*	(0.7352)	3	**0.2810***	(1.0000)	1	0.0822*	(0.9588)	2	0.2742	(0.9258)	2
$NRLS_C$	0.2809*	(0.9386)	2	0.2855*	(0.5314)	2	0.1210	(0.8052)	6	1.2244	(0.9017)	10
Kernel	0.3649*	(0.0925)	28	0.3761	(0.0172)	28	0.6002	(0.1350)	30	244.9816	(0.1540)	30
IMSFE	0.2997*	(0.5874)	10	0.3206	(0.1271)	10	0.1581	(0.6652)	16	2.0124	(0.8841)	15
Nonlinear	0.3002*	(0.5773)	13	0.3214	(0.1198)	13	0.1595	(0.6503)	21	2.0347	(0.8851)	17
TW	0.2898*	(0.7679)	4	0.3059	(0.1371)	5	0.1302	(0.7705)	11	1.1236	(0.8840)	9
Trimmed MSPE	0.2971*	(0.5761)	6	0.3162	(0.1631)	8	0.1541	(0.6832)	14	1.9477	(0.8815)	14
Harmonic Mean	0.3063*	(0.4502)	19	0.3207	(0.0754)	11	0.1147	(0.8095)	5	0.7588	(0.8956)	7
AFTER Shrinkage	0.3047*	(0.5119)	18	0.3235	(0.0935)	16	0.2453	(0.4564)	25	12.6978	(0.6494)	28
DMSFE Shrinkage	0.2978*	(0.5723)	8	0.3161	(0.1591)	6	0.1593	(0.6610)	19	2.2014	(0.8850)	19
$OLS_{NC\text{-}EXP}$ Shrinkage	0.3128*	(0.3553)	22	0.3441	(0.0084)	23	0.1300	(0.7722)	10	0.7264	(0.8616)	6
$NRLS_C$ Shrinkage	**0.2809***	(1.0000)	1	0.2857*	(0.5226)	3	0.1212	(0.8032)	8	1.2271	(0.8963)	11
IMSFE Shrinkage	0.2998*	(0.5984)	11	0.3206	(0.1287)	9	0.1587	(0.6682)	17	2.0492	(0.8871)	18
Nonlinear Shrinkage	0.3002*	(0.5773)	12	0.3214	(0.1133)	12	0.1593	(0.6475)	18	2.0170	(0.8847)	16
Shrinkage TW	0.3091*	(0.4215)	21	0.3301	(0.0555)	21	0.2228	(0.4819)	24	5.0024	(0.8565)	23

Note. 1. Values in bold denote that the corresponding model has the lowest loss function under the specific criterion. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. The Shrinkage factor is set equal to *d=0.50*. However, we included to our calculations the values 0.25 and 1 and all gave us similar results and ranking. 5. EXP and SQRT denote the exponential and the square root transformation respectively.

Table A.3: Economic Evaluation using VaR-based loss functions

	Percentage of violations	Q-loss (α=0.01) Loss function	Rank	SPA	Smoothed-Q (α=0.01) Loss function	Rank	SPA	MRC Loss function	Rank	SPA
GARCH	1.79%	0.0325*	22	(0.1826)	0.0324*	22	(0.1872)	22.4777	28	(0.0000)
EGARCH	1.13%	0.0302*	5	(0.6675)	0.0301*	5	(0.6666)	23.5144	30	(0.0000)
FIGARCH	2.19%	0.0352	25	(0.0305)	0.0351	25	(0.0287)	23.3967	29	(0.0000)
MIDAS-RV	1.39%	0.0301*	3	(0.7512)	0.0300*	3	(0.7595)	21.2956	21	(0.0000)
MIDAS-CPI	1.26%	0.0301*	2	(0.7320)	0.0300*	2	(0.7457)	21.2388	19	(0.0000)
MIDAS-IP	1.33%	0.0302*	4	(0.7061)	0.0301*	4	(0.7165)	21.4918	23	(0.0000)
ARMA	1.72%	0.0328*	23	(0.1445)	0.0327*	23	(0.1484)	21.8463	24	(0.0000)
HAR	2.12%	0.0337*	24	(0.0776)	0.0336*	24	(0.0765)	22.1657	26	(0.0000)
ARFIMA	2.45%	0.0360	27	(0.0309)	0.0359	27	(0.0304)	22.2781	27	(0.0000)
VIX	1.59%	0.0313*	16	(0.2953)	0.0312*	17	(0.3103)	20.6856	7	(0.0000)
Mean	1.39%	0.0308*	13	(0.4358)	0.0307*	13	(0.4485)	21.3617	22	(0.0000)
Geometric Mean	1.46%	0.0313*	17	(0.2454)	0.0312*	16	(0.2685)	20.9900	16	(0.0000)
MSFE	1.59%	0.0309*	15	(0.3909)	0.0308*	15	(0.4190)	20.8104	11	(0.0000)
$OLS_{C\text{-}SQRT}$	3.45%	0.0367	29	(0.0047)	0.0365	29	(0.0072)	20.0429	2	(0.0016)
$OLS_{NC\text{-}EXP}$	3.78%	0.0379	30	(0.0047)	0.0377	30	(0.0066)	**19.7969***	1	(1.0000)
$NERLS_{C\text{-}SQRT}$	3.58%	0.0360	26	(0.0068)	0.0357	26	(0.0106)	20.3313	5	(0.0000)
$NRLS_{C}$	1.72%	0.0314*	19	(0.2264)	0.0313*	19	(0.2476)	20.3056	4	(0.0000)
Kernel	2.45%	0.0365*	28	(0.0343)	0.0363*	28	(0.0383)	20.7194	8	(0.0000)
IMSFE	1.52%	0.0307*	12	(0.5081)	0.0306*	12	(0.5334)	20.9786	14	(0.0000)
Nonlinear	1.52%	0.0307*	10	(0.5221)	0.0306*	10	(0.5361)	21.0096	17	(0.0000)
TW	1.66%	0.0304*	6	(0.6493)	0.0303*	6	(0.6758)	20.7349	10	(0.0000)
Trimmed MSPE	1.52%	0.0306*	8	(0.5059)	0.0305*	8	(0.5276)	20.7302	9	(0.0000)
Harmonic Mean	1.59%	0.0318*	21	(0.1388)	0.0317*	21	(0.1494)	20.8635	13	(0.0000)
AFTER	1.59%	0.0316*	20	(0.1534)	0.0316*	20	(0.1507)	20.5734	6	(0.0000)
Shrinkage DMSFE	1.59%	0.0309*	14	(0.3942)	0.0308*	14	(0.4152)	20.8121	12	(0.0000)
Shrinkage $OLS_{NC\text{-}EXP}$	1.33%	**0.0295***	1	(1.0000)	**0.0295***	1	(1.0000)	22.1492	25	(0.0000)
Shrinkage $NRLS_{C}$	1.66%	0.0314*	18	(0.2235)	0.0313*	18	(0.2341)	20.0999	3	(0.0005)
Shrinkage IMSFE	1.52%	0.0307*	11	(0.5155)	0.0306*	11	(0.5367)	20.9791	15	(0.0000)
Shrinkage Nonlinear	1.52%	0.0307*	9	(0.5204)	0.0306*	9	(0.5348)	21.0109	18	(0.0000)
Shrinkage TW	1.33%	0.0306*	7	(0.5531)	0.0305*	7	(0.5604)	21.2751	20	(0.0000)

Note. 1. Values in bold denote that the corresponding model has the lowest loss function under the specific criterion. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. The Shrinkage factor is set equal to $d=0.50$. However, we included to our calculations the values 0.25 and 1 and all gave us similar results and ranking. 5. The Q-loss function and the Smoothed-Q loss function are calculated for VaR level $\alpha=0.01$ and α=0.05. For the Smoothed-Q calculation we set the smoothness parameter $\delta=25$. 6. EXP and SQRT denote the exponential and the square root transformation respectively.

Table A.4: Economic Evaluation using utility-based and option-based loss functions

	Utility (gamma=3)			Option criterion		
	Loss function	Rank	SPA	Loss function	Rank	SPA
GARCH	-0.0469	29	(0.0000)	-5.3618	29	(0.0259)
EGARCH	-0.0466	30	(0.0000)	-5.2546*	28	(0.0530)
FIGARCH	-0.0469	28	(0.0000)	-7.2089	30	(0.0223)
MIDAS-RV	-0.0476	15	(0.0000)	-1.1160*	22	(0.2724)
MIDAS-CPI	-0.0476	22	(0.0000)	-0.3262*	18	(0.4130)
MIDAS-IP	-0.0476	16	(0.0000)	-1.2064*	23	(0.2646)
ARMA	-0.0474	25	(0.0000)	-0.1558*	17	(0.3762)
HAR	-0.0478	10	(0.0000)	0.5415*	15	(0.5037)
ARFIMA	-0.0481	8	(0.0000)	0.4382*	16	(0.4719)
VIX	-0.0475	23	(0.0000)	-1.6965*	25	(0.2139)
Mean	-0.0475	24	(0.0000)	-1.7075	26	(0.1317)
Geometric Mean	-0.0477	14	(0.0000)	-0.8866*	20	(0.2346)
MSFE	-0.0478	11	(0.0000)	3.1154*	6	(0.9850)
OLS_{C-SQRT}	**-0.0511***	1	(1.0000)	0.9602*	14	(0.5044)
OLS_{NC-EXP}	-0.0509	2	(0.0201)	-1.4282*	24	(0.2828)
$NERLS_{C-SQRT}$	-0.0508	3	(0.0046)	1.0319*	13	(0.5074)
$NRLS_{C}$	-0.0488	4	(0.0000)	3.6663*	3	(0.9913)
Kernel	-0.0482	6	(0.0000)	2.2968*	9	(0.7218)
IMSFE	-0.0476	17	(0.0000)	2.9738*	7	(0.8990)
Nonlinear	-0.0476	20	(0.0000)	1.3408*	11	(0.6491)
TW	-0.0481	7	(0.0000)	**3.8251***	1	(1.0000)
Trimmed MSPE	-0.0478	13	(0.0000)	3.3575*	4	(0.9781)
Harmonic Mean	-0.0479	9	(0.0000)	-0.3709*	19	(0.3215)
AFTER	-0.0476	19	(0.0000)	1.4242*	10	(0.6770)
Shrinkage DMSFE	-0.0478	12	(0.0000)	3.1909*	5	(0.9862)
Shrinkage OLS_{NC-EXP}	-0.0471	27	(0.0000)	-4.9923	27	(0.0574)
Shrinkage $NRLS_{C}$	-0.0488	5	(0.0000)	3.7227*	2	(0.9988)
Shrinkage IMSFE	-0.0476	18	(0.0000)	2.4370*	8	(0.8292)
Shrinkage Nonlinear	-0.0476	21	(0.0000)	1.0944*	12	(0.5972)
Shrinkage TW	-0.0474	26	(0.0000)	-0.9579*	21	(0.2789)

Note. 1. Values in bold denote that the corresponding model has the lowest loss function under the specific criterion. 2. Numbers in parentheses denote the p-value of the Hansen's SPA test. The null hypothesis is that none of the models is better than the benchmark. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. The confidence level used is α=10%. 3. We use * to denote that the model belongs to 10% MCS. The number of bootstrap replications to calculate the p-value is 10,000 and the block length is 2. 4. The Shrinkage factor is set equal to d=0.50. However, we included to our calculations the values 0.25 and 1 and all gave us similar results and ranking. 5. The risk aversion parameter for the Utility function is set equal to 3. However, we included values ranging from 1 to 5 leading to similar results. 6. EXP and SQRT denote the exponential and the square root transformation respectively.