Volatility and Return Forecasting: time series and options-based methods

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

by

Xingzhi Yao

Department of Economics

Lancaster University

December 2017

Acknowledgements

I wish to extend my sincere appreciation to many people for their help and support. First I would like to thank my supervisors Dr. Marwan Izzeldin and Professor David Peel for their encouragement, support and academic guidance during these three and a half years. Particular gratitude goes to Dr. Marwan Izzeldin who has offered me so many wonderful opportunities to participate in the top workshops and conferences, and to teach in the econometrics labs and tutorials. I must also thank Mr. Gerry Steele for very helpful editorial suggestions.

I would like to thank Ms. Caren Wareing and the rest of the administrative staff. A heartfelt thanks to my friends and fellow colleagues who made the Lancaster experience something special, in particular, Xuguang Li, Summer Guan, Caroline Khan, Likun Mao, Jinyu Li and Vasileios Pappas. Completing this thesis without the studentship from the Economic and Social Research Council (ESRC) would have been impossible.

Lastly and most importantly, I would like to thank my family for their ongoing love and understanding throughout this period. I am indebted to my best friend, Xiaoqiang Li, who was always there cheering me up and stood by me through the good times and bad. Special thanks to my husband, Zhenxiong Li, for his love, rare patience and company, without whom I would not have been able to balance my doctoral study with everything else.

Declaration of Authorship

I hereby declare that this thesis is my own work and has not been submitted for the award of a higher degree elsewhere. Part of the second chapter of this thesis has been accepted for publication in the Journal of Futures Markets (10.1002/fut.21881), with my main supervisor, Dr. Marwan Izzeldin, as a second author. In addition to the second chapter, this thesis contains no material previously published or written by any other person except where references have been made in the thesis.

Xingzhi Yao

October 2017

I confirm that 90% of the work, "Forecasting Using Alternative Measures of Model-Free Option-Implied Volatility", is conducted by Xingzhi Yao.

Marwan Izzeldin

October 2017

Abstract

This thesis attempts to model and forecast returns and realized volatility using two different methods: time series models that exploit the historical information set and options-based approach that provides a natural forecast of return variation from listed option prices. Both univariate and multivariate estimation of the time series models are considered in our analysis.

Chapter 1: This chapter introduces a modified fractionally co-integrated vector autoregressive model, M-FCVAR, that caters for systems with $I(0)$ and $I(d)$ variables under the presence of long memory in the co-integrating residuals. Model inference of the FCVAR and M-FCVAR are compared using Monte Carlo simulations and an empirical application. The M-FCVAR is found to yield better in-sample fit and more precise model estimates. Higher return predictability is observed over long horizons using the M-FCVAR in the empirical example. In addition, the shocks associated with the $I(0)$ variables could be permanent or transitory. We show that particular equation specifications are required to restrict these shocks when they produce only transitory effects on the $I(d)$ variables. The simulation results show that the inappropriate treatment of the shock to the $I(0)$ variable may negatively affect the precision in the estimation of the model parameters as well as the in-sample fit.

Chapter 2: This chapter evaluates the performance of various measures of model-free option-implied volatility in predicting returns and realized volatility. The critical role of the out-of-the money call options is highlighted through an investigation of the relevance of different components of the model-free implied

volatility. The Monte Carlo simulations show that: first, volatility forecasting performance of measures of implied volatility can be enhanced by employing an interpolation-extrapolation technique; second, for most measures considered, gains in their predictive power for future returns can be obtained by implementing an interpolation procedure. An empirical application using SPX options recorded from 2003 to 2013 further illustrates these claims.

Chapter 3: This chapter compares the performance of various least absolute shrinkage and selection operator (Lasso) based models in forecasting future log realized variance (RV) constructed from high-frequency returns. We conduct a comprehensive empirical study using the SPY and 10 individual stocks selected from 10 different sectors. In an in-sample analysis, we provide evidence for the invalidity of the lag structure implied by the heterogeneous autoregressive (HAR) model which has been heavily adopted in volatility forecast. In our out-of-sample study considering the full time period, the best forecasting performance is usually provided by the Lasso-based model and the idea of forecast combination tends to improve the forecasting accuracy of the Lasso-based model. Among all models of interest, the ordered Lasso AR using the forecast combination serves as the top performer most frequently in forecasting RV and its improvements over the HAR model are, in most cases, significant over monthly horizons. Moreover, we observe a strong impact of the financial crisis on the performance of the Lasso-based models. Nevertheless, the ordered Lasso AR with the forecast combination still retains its advantages in the post-crisis period, especially over long horizons. In line with the existing study, the superiority of the Lasso-based models is more evident in a larger forecasting window size. The conclusions outlined above are not affected by the variation in the sampling frequency upon which the RV series are based. However, as the sampling frequency increases, there tends to be more situations where the Lasso-based model achieves the top performance in the full sample analysis.

Contents

List of Tables

List of Figures

Introduction

Volatility of financial time series plays a central role in pricing derivatives, hedging and computing measures of risk. Volatility forecasting is therefore an important topic in finance and financial economics, which has held enormous attention of academics and market investors over the last few decades. The increased availability of high-frequency data has spurred great interest in the model-free measurement of variance based upon intraday returns, termed realized variance (RV). On the other hand, expected returns are considered crucial equity market indicators at the aggregate market level since they reflect the attitudes of investors towards risk and should carry predictive power for actual future returns theoretically. However, it still remains controversial in terms of whether equity returns are indeed predictable. The difficulty is that expected returns are not directly observable and thus one needs to estimate them by means of publicly available information. This thesis presents various methods to achieve better RV forecast and return predictions.

Two different approaches are employed to conduct the forecast of returns and RV. First, we consider time series models that exploit the historical information set to formulate return and volatility forecasts in chapter 1 and chapter 3, respectively. Second, in chapter 2, we concentrate on the market's expectation of future return variation from listed option prices, which is perceived as a market based volatility forecast and may possess information content in predicting future market returns.

Specifically, we adopt a model for the analysis of multivariate time series in Chapter 1 which accommodates both the long-run and short-run movements of

the variables. Dynamic dependencies in aggregate stock market returns, implied and realized volatilities can be well captured by this joint modelling framework. Univariate time series volatility models are used in Chapter 3 where we apply several model selection devices to select the relevant lags of the RV for the purpose of better volatility forecasts. In using the time series models, we account for the long-memory property of volatility, described by fractional integration and a slow hyperbolic decay in the autocorrelations. In chapter 1, in addition to the investigation of long memory of volatilities, we also evaluate their long-run relationship via both parametric and semiparametric testing methods. In chapter 3, the observed long-memory behaviour is approximated by aggregating across short-memory heterogenous autoregressive processes. In a departure from chapter 1 and 3, chapter 2 constructs volatility forecasts extracted from combinations of option prices which do not depend on any pricing formula. Since the option price incorporates all available information in an efficient market, these model-free volatility expectations are highly correlated with the future RV and can be seen as priced risk factors in the cross-section of stock returns.

In chapter 1, we modify the fractionally co-integrated vector autoregressive (FCVAR) model proposed by Johansen (2008) to allow for the coexistence of $I(0)$ and $I(d)$ variables under the presence of long memory in the co-integrating errors. The proposed model is termed the M-FCVAR. We investigate the model inference of the FCVAR and M-FCVAR in Monte Carlo simulations covering a wide range of fractional integration orders as well as in an empirical example. With a more appropriate treatment of the $I(0)$ variable in the system of fractionally co-integrated processes, the M-FCVAR is found to yield less biased model estimates and better in-sample fit in the simulation study. In addition to this, we pay particular attention to the properties of the shocks arising from the $I(0)$ variables in the (M-)FCVAR framework, which could either be permanent or transitory. The existing work does not seem to recognize that a particular model design is required to ensure that the shocks associated with the $I(0)$ variables exert only transitory

effects on the long-memory variables. Taking into account all the possibilities one may encounter in practice, we provide equation specifications to restrict the shocks from the $I(0)$ variables. The simulation evidence suggests that one may obtain biased model estimates and low in-sample fit if the properties of the shocks to the $I(0)$ variables are incorrectly accounted for. Our empirical application consists of the intraday data for the SPX and SPY indices and the daily data for the volatility index (VIX), where a joint modelling of the three series, i.e. two fractionally integrated variances and one $I(0)$ returns, is implemented in both the FCVAR and M-FCVAR. With more precise estimates of the model parameters, i.e. fractional integration order, degree of fractional co-integration and co-integrating vectors, returns are shown more predictable under the M-FCVAR over long horizons.

Catering for a mixture of $I(0)$ and $I(d)$ variables, the M-FCVAR can easily find many applications in finance and financial economics. Apart from the example using the RV, VIX and market returns introduced above, the M-FCVAR can be further employed to examine the stock market return predictability suggested by the fluctuations in the aggregate consumption-wealth ratio. This is motivated by the work of Lettau and Ludvigson (2001) who indicate that the aggregate consumption-wealth ratio can be expressed with regard to several fractionally co-integrated variables and that the transitory deviations from the common trend in these variables serve as a strong predictor of future returns.

In chapter 2, we examine the performance of various measures of model-free option-implied volatility in the forecast of future returns and RV. By decomposing model-free implied volatility into several components with the use of different segments of option strike range, we investigate the role of each component in the forecasting practice and highlight the importance of the out-of-the money call options. In addition, we conduct Monte Carlo simulations to ascertain the impact of discrete strike prices on the forecasting performance of implied volatility measures. Simulation results show that: first, volatility forecast improves with a wider range of strikes; second, the range of strikes produces negative effects

on the predictive power of implied volatilities for future returns; third, a finer partition of strikes leads to better return predictions. These findings warrant the use of an interpolation and extrapolation procedure as an attempt to enhance the forecasting power of implied volatilities for future RV while only an interpolation method is needed in return predictions. In the empirical application based on SPX options from 2003 to 2013, the aforementioned interpolation/extrapolation procedure is found to significantly enhance the performance of implied volatilities for forecasting future RV and lead to better return predictions for most measures in the post-crisis period. The effectiveness of such procedure is also verified in our simulation study.

In chapter 3, we evaluate the performance of least absolute shrinkage and selection operator (Lasso) based models in forecasting future RV. The empirical study adopts the RV series of the SPY and ten individual stocks. We first show that the heterogeneous autoregressive (HAR) model does not fully agree with the Lasso-type models in terms of the lag structure, which brings into question whether the HAR is appropriate for modelling and forecasting future volatility. Compared with the HAR and its extensions, the Lasso-based model usually performs best and the idea of forecast combination tends to improve the accuracy of the volatility forecast. Among various Lasso-based models, the ordered Lasso AR using the forecast combination serves as the top performer most frequently and its gains over the HAR model are generally significant over monthly horizons. The global financial crisis is found to exert non-trivial impact on the performance of the Lasso-based models. However, the ordered Lasso AR with the forecast combination still retains its superiority in the post-crisis period, especially over long forecasting horizons. We also provide evidence that the Lasso-based models tend to perform better in a larger window size. Furthermore, as the sampling frequency upon which the RV series are based increases, the advantages of the Lasso-based models are more evident using the full sample.

Chapter 1

A Modified Fractionally Co-integrated VAR for Modelling Systems with $I(d)$ and $I(0)$ Variables

1.1 Introduction

Many financial and economic variables are appropriately described by a fractionally integrated process, denoted $I(d)$ (e.g. see the discussion and many references in Nielsen (2010)). In particular, equity and index volatility are well characterized by an $I(d)$ process (Andersen and Bollerslev (1997) and Comte and Renault (1998)). Implied volatility obtained from option prices displays many of the stylized facts of equity and index volatility and has been found to be a relevant predictor of the corresponding asset volatility. Implied volatility, the VIX index in particular, has featured in a number of volatility forecasting exercises using both short- and long-memory specifications (Bandi and Perron (2006) and Busch, Christensen, and Nielsen (2011a)). Another use of the implied volatility is to explore the long-run co-movements between the VIX and the realized volatility of S&P 500, where the difference between the implied-realized variation measures is termed the 'variance risk premium'. This idea has been adopted by Bollerslev et al. (2013) (BOST hereafter) who are pioneers in predicting stock market returns using a framework based on the fractionally co-integrated vector autoregressive (FCVAR) model of Johansen (2008) and Johansen and Nielsen (2012). BOST (2013) show that the gains of this approach arise from the joint modelling of the multivariate time series and the capture of the predictability inherent in the variance risk premium.

The FCVAR serves as a direct model of fractional co-integration and provides a central tool for the analysis of long-run equilibrium relationships among the $I(d)$ variables. Compared with conventional $I(1)/I(0)$ co-integration, fractional co-integration allows linear combinations of $I(d)$ processes to give $I(d-b)$ processes with $d \geq b > 0$ and with d and/or b as fractional numbers. The FCVAR has been applied in several studies. For example: Caporin, Ranaldo, and Santucci de Magistris (2013) demonstrate the superiority of the FCVAR framework in forecasting extreme stock prices by accommodating the fractional co-integration between high and low prices and the daily range obtained from the co-integrating

residuals; Rossi and Santucci de Magistris (2013) employ the FCVAR to analyze the long-run relationship between futures and spot range-based volatility measures; and Jones, Nielsen, and Popiel (2014) exploit the FCVAR to examine the relation between political support and macroeconomic conditions.

The work of BOST (2013) is uniquely distinctive in that it involves a mixture of $I(d)$ and $I(0)$ variables. In that presentation, the estimation of the FCVAR model is simplified by letting $d = b$; i.e. there is no memory in the co-integrating residuals. According to Definition 2 in Johansen (2008), the FCVAR allows for variation in the integration order of the variables within the system. Consequently, when $d = b$, the inclusion of the $I(0)$ variables is natural in the FCVAR, which is similar to the coexistence of the $I(1)$ and $I(0)$ variables in the traditional co-integrated VAR. However, the case of $d > b$ poses a challenge for the analysis of the FCVAR as the fractional differencing operator Δ^{d-b} is applied, not only to the real co-integrating vectors, but also to the $I(0)$ variables serving as pseudo co-integrating vectors. This results in the anti-persistence of the latter. In addition, assumptions need to be made with regard to the nature of the shocks emanating from the $I(0)$ variables when they enter the system of the FCVAR model. Theoretically, the impact of the shocks associated with the $I(0)$ variables can be either permanent or transitory on the $I(d)$ variables. However, we show that these shocks exert transitory effects in the FCVAR, only when particular equation specifications are adopted; otherwise, the shocks to the $I(0)$ variables would have nonzero long-run impact on the $I(d)$ variables. The same interaction between the $I(0)$ and $I(1)$ variables has been observed by Fisher, Huh, and Pagan (2016) (FHP hereafter) but in a VECM type of framework. FHP (2016) provide specifications for the traditional VECM that prevent the shocks associated with the $I(0)$ variables from having permanent effects on the $I(1)$ variables. However, their analysis is limited to situations where there are equal numbers of exogenous $I(1)$ variables and common factors.

This chapter proposes modifications to the FCVAR model of Johansen (2008) and Johansen and Nielsen (2012), which are more suitable for systems with $I(d)$

and $I(0)$ variables when there exists long memory in the co-integrating residuals; i.e. $d > b$. Specifically, the fractional differencing operator (Δ^{d-b}) is applied to the $I(d)$ variables within the system prior to the estimation of the FCVAR model. This procedure does not alter the representation theorem and the calculation of maximum likelihood estimators of the FCVAR. Without that adjustment, long memory is induced in the model-implied $I(0)$ variables, and this may further result in biased estimates. The chapter also provides the theoretical framework that outlines the changes required in the specifications to restrict shocks arising from the $I(0)$ variables, so that their effects on the $I(d)$ variables are only transitory. Complementary to FHP (2016), the chapter covers a variety of situations where the number of exogenous variables is fewer than or equal to the number of common factors or where there are only endogenous variables present.

To the best of our knowledge, this chapter is the first to consider a modified FCVAR, henceforth M-FCVAR, to allow for inference and prediction in the presence of $I(0)$ and $I(d)$ variables. In a simulation study, we show that, compared with the FCVAR, the M-FCVAR generally yields a better in-sample fit and less biased estimates of parameters d, b and co-integrating vectors in different sample sizes. In addition, the ignorance of the property of the shock arising from the $I(0)$ variable may damage the precision in the estimation of model parameters and lower the in-sample fit. The comparison between the FCVAR and M-FCVAR is also illustrated using an empirical application based on high-frequency data, in which case market returns are found more predictable over long horizons under the suggested M-FCVAR.

The rest of this chapter is organized as follows. Section 1.2 reviews the relevant literature. Section 1.3 presents methods adopted in this chapter together with the M-FCVAR model specifications and modifications. The Monte Carlo study is outlined in section 1.4. Section 1.5 describes the data and reports the empirical results. Section 1.6 concludes. Algorithms of the impulse response functions and the model-implied R^2 are given in section 1.6.

1.2 Literature Review

1.2.1 Background

Fractional co-integration, an extension of the co-integration to processes with fractional degrees of integration, has received substantial research attention recently. It has been applied in the topics of exchange rates, volatility of financial series, interest rates, electricity prices and political studies, see Gil-Alana and Hualde (2009) for an overview of the relevant studies. Despite various applications of the fractional co-integration, the main focus has been on the long-run relationship between implied-realized volatilities.

Implied volatility is universally considered the best market expectation of the future volatility over the remaining life of the relevant option. Not surprisingly, there has been enormous interest in examining the unbiasedness of the implied volatility forecast of subsequent realized volatility. The relation between the two volatility proxies can be evaluated via the regression

$$\sigma_t^{RV} = \alpha + \beta \sigma_t^{IV} + \varepsilon_t \tag{1.1}$$

where σ_t^{IV} denotes implied volatility at time t and σ_t^{RV} represents realized volatility from t till the option's expiration time. As noted by Christensen and Nielsen (2006) and Nielsen (2007), the unbiasedness hypothesis implies a β coefficient of unity. Traditional tests for this hypothesis using the OLS technique generally result in the conclusion that σ_t^{IV} provides biased forecast of σ_t^{RV} by obtaining the slope parameter β not equal to one, see Christensen and Prabhala (1998) and Poteshman (2000).

Realized and implied volatilities are found to display long-memory properties, see Comte and Renault (1998), Comte, Coutin, and Renault (2012), Ray and Tsay (2000), Andersen et al. (2001a) and Andersen et al. (2001b), among others. The fractional co-integration between the implied and realized volatilities is documented

in the work of Bandi and Perron (2006), Christensen and Nielsen (2006) and Nielsen (2007), among others. The presence of fractional co-integration suggests that both σ_t^{RV} and σ_t^{IV} are fractionally integrated and that ε_t in equation (1.1) is serially uncorrelated or displays short memory. Furthermore, the studies listed above provide evidence for the long-run unbiasedness, i.e. $\beta = 1$, using different frequency domain methods accounting for the fractional property of the volatilities. Specifically, fractional integration in the region of non-stationarity is found in the work of Bandi and Perron (2006) whereas the stationary region is indicated in Christensen and Nielsen (2006) and Nielsen (2007).

It is worth noting that the OLS fails to give consistent estimates of the relation in equation (1.1) in the case of stationary fractional co-integration. This is due to the fact that, in such situation, both the regressor and the error exhibit long memory and thus correlation between them may exist even over long horizons, see Robinson (1994) and Robinson and Marinucci (2003). In the case of non-stationary fractional co-integration, the OLS converges slower than the narrow-band least squares (NBLS) proposed by Robinson (1994), see Robinson and Marinucci (2001) and Robinson and Marinucci (2003). In addition, the NBLS leads to consistent estimates but non-standard limit distributions in the non-stationary range. To sum up, the earlier findings in terms of the biased relation between the two volatility proxies using the OLS are not reliable since the predictive regression in (1.1) is usually viewed as stationary fractional co-integration. Similar conclusions supporting the long-run unbiasedness hypothesis can be found in Kellard, Dunis, and Sarantis (2010) where the integration order of volatility has confidence intervals spanning the stationary/non-stationary boundary and Nielsen and Frederiksen (2011) where the presence of a volatility risk premium, i.e. $\sigma_t^{RV} - \sigma_t^{IV}$, correlated with implied volatility is accounted for to remove the bias in the NBLS estimator in regression (1.1).

The difference between the implied and realized variances, the so-called variance risk premium, has been interpreted as a measure of the representative agent's

risk aversion, see Bollerslev, Tauchen, and Zhou (2009) and Drechsler and Yaron (2010), among others. The variance risk premium is found to capture attitudes toward uncertainty about economic fundamentals and thus predict financial market risk premia and financial returns. For instance. Bollerslev, Tauchen, and Zhou (2009) demonstrate that the variance risk premium is able to capture a nontrivial fraction of variation in quarterly stock market returns and can result in even greater return predictability when combined with other conventional predictor variables. BOST (2013) document a non-trivial return predictability over interdaily and monthly horizons using the FCVAR model based on 5-minute intraday data. They also show that the observed strong predictive power for future market returns is explained by the joint modelling of returns and variances within the FCVAR as well as the predictability contained in the variance risk premium. Furthermore, Bollerslev et al. (2014) provide evidence that such pronounced return predictability suggested by the variance risk premium is not induced by the statistical finite sample biases.

Despite the importance of fractional co-integration from both theoretical and practical perspectives in economics, the testing and estimation of the fractional co-integrating relation have encountered many difficulties. Although the work of Engle, Lilien, and Robins (1987) provides the concept of common trends between fractionally integrated processes, subsequent studies are confined to situations where the variables are integrated of order one. Progress in the area of fractional co-integration is only achieved when Robinson and Marinucci (2003), Christensen and Nielsen (2006) and Nielsen and Frederiksen (2011), among others develop the regression-based semiparametric approach to examine whether two long-memory processes are fractionally co-integrated. Subsequently, Robinson and Yajima (2002) and Nielsen and Shimotsu (2007) introduce a testing procedure to investigate the presence of the co-fractional relation by estimating the co-integrating rank of the matrix of two, or more, fractionally differenced variables. Studies by Johansen (2008) and Johansen and Nielsen (2012) have made further improvements in the

study of fractional co-integration by developing a parametric multivariate FCVAR model which explicitly captures both the long-run and short-run relationships of the long-memory processes.

1.2.2 Fractional Integration and Fractional Co-integration

We first introduce the fractional integration processes from which the concept of fractional co-integration stems. Fractional integration describes a strong dependency between observations which exhibit high persistence that the standard ARMA framework is unable to capture. This process is neither an $I(1)$ unit root process nor an $I(0)$ process but rather an $I(d)$ process, where d is between zero and one, see Baillie (1996) and Robinson (2003) for more details. Assume a covariance stationary time series X_t with the spectral density $f(\lambda)$. The series X_t is a long-memory process integrated of order d ($d \neq 0$) if

$$f(\lambda) \sim G\lambda^{-2d}, \text{ as } \lambda \to 0_+ \tag{1.2}$$

where $G \in (0, \infty)$ is a finite and nonzero matrix with strictly positive diagonal elements. The autocovariance functions of X_t decay hyperbolically as shown by

$$Cov(X_t,\ X_{t-\tau}) \sim \tau^{2d-1}, \text{ as } \tau \to \infty \tag{1.3}$$

The parameter d determines the memory of the process. For $0 < d < 0.5$, the series is covariance stationary and contains long memory, implying that shocks will decay hyperbolically rather than geometrically. By contrast, for $0.5 < d < 1$, the series is no longer stationary, yet still mean reverting. For $-0.5 < d < 0$, the process is stationary but antipersistent, giving rise to the zero spectral density at the origin frequency instead of infinity.

Applications of long memory to financial and economic data have been extensively explored with the development of techniques for modelling the fractionally integrated process and measuring the memory parameter d. The most widely accepted methods

are the log-periodogram regression of Geweke and Porter-Hudak (1983) and the local-Whittle likelihood procedure of Kuensch (1987). Both are semiparametric and thus immune to model mis-specification problems. On the other hand, the spirit of parametric methods is first to build a long-memory model and then to jointly estimate the model. Popular models are the fractional Brownian motion proposed by Mandelbrot and Van Ness (1968), the fractional white noise and the autoregressive fractionally integrated moving average (ARFIMA) model developed by Granger (1980), Granger and Joyeux (1980) and Hosking (1981). The ARFIMA has been heavily employed to capture the long-memory property of the realized volatility, see Andersen et al. (2003), Choi, Yu, and Zivot (2010) and Degiannakis and Floros (2013), among others.

Fractional co-integration generalizes the standard co-integration with $I(1)$ series and $I(0)$ linear co-integrating relationships by allowing for more flexibility in the order of integration. Specifically, fractional co-integration can be defined by assuming two series, y_t and x_t, which are both integrated of order d_x, where d_x can be a fractional number rather than integer one as commonly assumed in the concept of conventional co-integration, and a linear combination, $u_t = y_t - \beta x_t$, is $I(d_u)$. When $0 \leqslant d_u < d_x$, y_t and x_t are fractionally co-integrated. In particular, the model with $d_x - d_u < 0.5$ is characterized as weak fractional co-integration by Hualde and Robinson (2010). Next, we review some recent studies in testing and estimating the fractional co-integration from different perspectives.

1.2.3 Testing and Estimation Methods

There is a growing literature devoted to the testing and estimation of fractional co-integration. A group of contributions is characterized as semiparametric. With long-run components of each series at origin frequencies, some studies adopt the regression-based approach to estimate the co-integrating vectors and integration orders of both regressors and residuals. With the focus on the space of co-integration rather than the co-integrating regression, other studies estimate the co-integrating

rank in the long-memory systems and require no knowledge about the co-integrating vectors or memory parameters. Parametric maximum likelihood techniques have also been used to provide the joint estimation of the multivariate fractionally integrated system. However, there seems no consensus on the optimal testing procedure for fractional co-integration. One may find difficulties in having consistent and conclusive outcome when using different methodologies. The following section provides an overview of popular methods of fractional co-integration analysis and outlines the problems one may encounter in practice.

Semiparametric Approach

A widely accepted procedure is to consider a semiparametric approach characterized by using a degenerating band of low frequencies for estimation. The semiparametric approach does not require the accurate specification and estimation of the whole sample, i.e. it achieves consistency without relying on a parametric model. Two different methods in the semiparametric fashion are introduced below, which only require information related to the behaviour of the spectral density around the origin.

Regression-Based Approach Regression-based methods generally extend the work of Engle and Granger (1987) to the case where the order of integration is not restrictive to integer one, see Marinucci and Robinson (2001) and Gil-Alana (2003). The key step is to obtain the integration orders of the underlying series and the regression residuals from the estimated co-integrating relationship and then to examine whether the persistence reduces or not. The complication of the regression-based approach is that, unlike the standard situation under OLS regression, the regressors and the errors may be both stationary and fractionally integrated and thus are likely to be correlated in the long term. The implication then is that the OLS estimator is no longer consistent (Robinson (1994), Robinson and Marinucci (2003) and Robinson (1997)). To circumvent this problem, Robinson

(1994) develops a semiparametric narrow-band least squares (NBLS) estimator in the frequency domain and implements OLS on a degenerating part of the periodogram around zero frequency, i.e., the so-called narrow-band. In that paper, Robinson shows that the NBLS estimator is consistent in the stationary case. Christensen and Nielsen (2006) show that its asymptotic distribution is normal when $d_x + d_u < 0.5$ and when the coherence between regressors and errors is zero at the origin frequency, i.e. in the long run. The results on the NBLS estimator for the regressors which are non-stationary long memory are provided by Robinson and Marinucci (2003); and Chen and Hurvich (2003a) add polynomial trends by using a tapered NBLS estimator based on differenced data. Concentrating on the periodogram around the origin, this semiparametric approach enjoys the advantage of being variant to the short- and medium-run dynamics.

Kellard, Dunis, and Sarantis (2010) improve the NBLS estimator by developing a new fractional co-integration test which is robust in both the stationary and non-stationary context. The new estimator is shown to be approximately normally distributed in finite sample, which holds across the stationary and non-stationary regions. Extending the stationary setting of Christensen and Nielsen (2006) under a condition of zero long-run coherence between the regressors and co-integrating errors, Nielsen and Frederiksen (2011) focus on weak fractional co-integration, including non-stationarity, in the absence of this condition, in which case a bias term arises in the NBLS estimator. They show that the bias can be estimated and thus corrected by a fully modified NBLS (FMNBLS) procedure with a careful choice of bandwidth parameters. The regression-based method is sometimes not straightforward to implement in empirical studies where integration orders are not within a particular region as discussed above or where the co-integrating errors (residuals) are not well defined. By contrast, the co-integrating rank test does not require estimating the co-integrating vector(s) and thus may serve as a good alternative.

Spectral Matrix Approach The co-integrating rank test examines the presence of the co-fractional relation from the perspective of the long-run covariance matrix. This approach only requires the spectral density matrix at the origin frequency, but it displays great sensitivity to the selection of bandwidth parameters. Relative to the regression-based approach, it does not estimate the co-integrating vectors and only produces a consistent estimate of the co-integrating rank. The estimate of the rank greater than one and less than the number of variables indicates the existence of the co-fractional relation. Hence, this approach is not appropriate when specific information about the co-fractional relation, e.g. strength of the relation, is required.

Robinson and Yajima (2002) are the pioneers in implementing the co-integrating rank estimation in the region of stationarity. Chen and Hurvich (2003b) investigate the rank of an averaged periodogram matrix of tapered and differenced observations and fix the number of frequencies used in the periodogram averages as the sample size increases, which applies to both stationary and non-stationary situations. Their assumption of strictly positive rank is relaxed in subsequent work by Chen and Hurvich (2006) who consider the null of no fractional co-integration, i.e. rank equal to zero. Nielsen and Shimotsu (2007) also attempt to accommodate (asymptotically) stationary and nonstationary fractionally integrated processes. They use the exact local Whittle analysis of Shimotsu and Phillips (2005), which generalizes the local Whittle estimator of Kuensch (1987) to allow for any value of the memory parameter d. The estimate of co-integrating rank is achieved by examining the rank of the spectral density matrix of the d^{th} differenced processes around the zero frequency.

Parametric Approach

A fully parametric approach is more efficient in using the entire sample instead of focusing on the origin frequency of the periodogram only. However, it shows non-trivial inconsistency if the parametric model is mis-specified. In the traditional

$I(1)/I(0)$ co-integration, the standard tool to handle the relationship among the multivariate time series is the vector error correction model (VECM) as introduced by Engle and Granger (1987). The representation is given by

$$\Delta X_t = \alpha\beta' X_{t-1} + \Sigma_{i=1}^{k}\Gamma_i \Delta X_{t-i} + \varepsilon_t \tag{1.4}$$

where X_t is $p-$dimensional $I(1)$ series and ε_t is $p-$dimensional independent and identically distributed (i.i.d.) with mean zero and covariance matrix Ω.

Multivariate score tests (or Lagrange multiplier tests) for fractional integration have been developed by Johansen (1995) and Nielsen (2005), as a prerequisite for further detailed investigation of fractional co-integration. Substantial efforts have since been made to improve and optimize the parametric estimation of the fractional co-integration. The most famous model is the Fractionally Co-integrated Vector Autoregressive (FCVAR) model (or the so-called Fractional Vector Error Correction model (FVECM) in some studies) proposed by Johansen (2008) and further analyzed by Johansen and Nielsen (2012).

Some important studies related to the development of the parametric framework for fractional co-integration include: Breitung and Hassler (2002) suggest a test for the rank of fractional co-integration in the FCVAR while assuming the integration order is known and that the errors are i.i.d. Gaussian; Avarucci and Velasco (2009) introduce the Wald test to determine the co-integration rank in a system of nonstationary fractionally integrated variables within the FCVAR-type framework; Łasak (2010) considers a profile likelihood method to estimate the parameters of the Granger (1986) model and to test for the null hypothesis of no co-integration. This estimation method has been extended by Johansen and Nielsen (2012) to the FCVAR model; Franchi (2010) investigates a richer co-fractional structure by extending the representation theory of the FCVAR model in Johansen (2008); and Łasak and Velasco (2015) introduce a novel two-step procedure for co-integrating rank estimation, which allows different co-integration relations to display different persistence. This can be seen as an extension of the study by Łasak (2010) and

Johansen and Nielsen (2012). Essentially the FCVAR is derived by replacing the lag and difference operators in the VECM model with their fractional counterparts. It has been adopted by BOST(2013), Jones, Nielsen, and Popiel (2014), Dolatabadi, Nielsen, and Xu (2015) and Dolatabadi, Nielsen, and Xu (2016), among others. It exhibits the advantage of allowing for multivariate analysis, flexible selection of parameters and good performance in forecasting.

A mixture of variables with different integration orders in the analysis of fractional co-integration is frequently encountered in practice. The study by FHP (2016) considers a VECM model with both $I(1)$ and $I(0)$ variables but little work has been done in the FCVAR framework with the coexistence of $I(d)$ and $I(0)$ variables. FHP(2016) classify the shocks arising from the $I(0)$ variables into the permanent and transitory, termed the P0 and T0 shocks. They show that, in a system containing $I(0)$ and $I(1)$ variables which are co-integrated, the co-integration may no longer exist when the T0 shocks become P0 shocks. A device is suggested by FHP(2016) to calculate the permanent component of the $I(1)$ variables when the shocks associated with the $I(0)$ variables have either transitory or permanent effects on the $I(1)$ variables. Furthermore, in order to restrict the shocks from the $I(0)$ variables to exert only transitory impact on the other variables, they show that the true error correction terms and the $I(0)$ variables must appear as differences in the equations of the VECM where the response variables are permanent components. However, their work is confined to the situation where there are equal numbers of exogenous $I(1)$ variables and common factors, i.e. permanent components.

1.3 Methodology

The following section presents the methods adopted in this chapter. We employ the FCVAR model to accommodate the system of $I(d)$ variables. When moving to the system of $I(d)$ and $I(0)$ variables, we consider a modified FCVAR, M-FCVAR, to cater for the introduction of the $I(0)$ variables and allow for long memory in

the co-integrating error. Properties of the shocks arising from the $I(0)$ variables are also accounted for in various situations. As an attempt to overcome the identification problem of the (M-)FCVAR, we obtain the fractional integration estimate using the exact local Whittle estimator of Shimotsu and Phillips (2005). The presence of fractional co-integration is examined using the exact local Whittle rank test by Nielsen and Shimotsu (2007) coupled with a modified Wald test for the equality of the orders of fractional integration.

1.3.1 Fractional Integration Estimation

Fractional co-integration originates from several variables exhibiting long-memory properties. Hence, we first introduce the procedure of estimating the order of fractional integration, which serves as the basis for the subsequent analysis of co-fractional relations. A fractionally integrated process is defined as $I(d)$ if its dth difference is integrated of order zero, where d can be any real number. In spite of a number of approaches proposed to estimate the long-memory parameter d, the semiparametric procedure has long been widely explored and applied since it requires no assumptions about the short-run dynamics and thus stays robust to mis-specification problems. This chapter adopts the exact local Whittle estimator of Shimotsu and Phillips (2005), which extends the work of local Whittle analysis by Kuensch (1987) and Robinson (1995) to allow for any value of the fractional differencing parameter, d.

To define the frequency-domain local Whittle estimator, we assume that a process X_t has the spectral density, $f(\lambda)$, defined in equation (1.2). Let the fractionally integrated X_t be generated by the model

$$\Delta^d X_t = (1-L)^d X_t = \mu_t \mathbf{1}\{t \geqslant 1\},\ t = 0, \pm 1, ... \tag{1.5}$$

where $\mathbf{1}\{.\}$ represents the indicator function, L is the lag operator and μ_t is assumed to be a covariance stationary process whose spectral density, $f_\mu(\lambda)$, is bounded and bounded away from zero at the origin frequency $\lambda = 0$, i.e. $f_\mu(\lambda) \sim G$

for $\lambda \sim 0$ (Robinson (1995)). An alternative representation of X_t based on $\mu_1, ..., \mu_n$ can be derived by inverting and expanding the equation (1.5),

$$\begin{aligned} X_t &= \Delta^{-d}\mu_t \mathbf{1}\{t \geqslant 1\} \\ &= (1-L)^{-d}\mu_t \mathbf{1}\{t \geqslant 1\} \end{aligned} \tag{1.6}$$

The discrete Fourier transform (DFT), $\omega_x(\lambda_j)$, and the periodogram, $I_x(\lambda_j)$ of X_t, t =1,$\cdots$,T at the fundamental frequencies can be written as

$$\begin{aligned} \omega_x(\lambda_j) &= (2\pi T)^{-1/2}\Sigma_{t=1}^{T} X_t e^{it\lambda_j},\ \lambda_j = \frac{2\pi j}{T},\ j = 1, ..., m < \frac{T}{2} \\ I_x(\lambda_j) &= |\omega_x(\lambda_j)|^2 \end{aligned} \tag{1.7}$$

One advantage of semiparametric estimation over the parametric approach is that it employs frequencies near the origin only and treats the periodogram away from the zero nonparametrically. The conventional local Whittle (LW) estimator by Kuensch (1987) and Robinson (1995) relies on the Gaussian objective function,

$$Q_m(G, d) = \frac{1}{m}\Sigma_{j=1}^{m}[\log(G\lambda_j^{-2d}) + \frac{\lambda_j^{2d}}{G} I_x(\lambda_j)] \tag{1.8}$$

where $\lambda_j = \frac{2\pi j}{T}$, $j = 1, ..., m < \frac{T}{2}$. The LW estimate is thus derived by minimizing the function $Q_m(G, d)$. Robinson (1995) proves that the asymptotic standard errors of the LW estimator are $\sqrt{m}(\widehat{d}_{T,m} - d) \Rightarrow N(0, \frac{1}{4})$. In spite of its enhanced efficiency over the GPH estimator (Geweke and Porter-Hudak (1983)) within the stationary region, both GPH and LW estimators display nonstandard behaviour when $d > \frac{3}{4}$ (Kim and Phillips (2006)).

Shimotsu and Phillips (2005) provide a procedure which can be applied in the stationary and nonstationary regions and it estimates (G, d) by minimizing the objective function

$$Q_m(G, d) = \frac{1}{m}\Sigma_{j=1}^{m}[\log(G\lambda_j^{-2d}) + \frac{1}{G} I_{\Delta^d x}(\lambda_j)] \tag{1.9}$$

where $I_{\Delta^d x}(\lambda_j)$ is the periodogram of $\Delta^d X_t$. Concentrating $Q_m(G, d)$ with respect to G, we obtain the exact local Whittle (ELW) estimator given by

$$\widetilde{d} = \underset{d\in[\Delta_1,\Delta_2]}{\arg\min} R(d) \tag{1.10}$$

where Δ_1 and Δ_2 are the lower and upper bounds of the admissible values of d and

$$\begin{aligned} R(d) &= \log \widehat{G}(d) - 2d\frac{1}{m}\Sigma_{j=1}^{m} \log \lambda_j \\ \widehat{G}(d) &= \frac{1}{m}\Sigma_{j=1}^{m} I_{\Delta^d x}(\lambda_j) \end{aligned} \tag{1.11}$$

This ELW estimator has been proved to be consistent and asymptotically normally distributed when the underlying value of $d \in (\Delta_1, \Delta_2)$ and $\Delta_2 - \Delta_1 \leq \frac{9}{2}$ with the mild assumptions on bandwidth m and stationary μ_t.

The desirable properties of the ELW estimator by Shimotsu and Phillips (2005) are based on the assumption that X_t is generated by the process in equation (1.5) and that the mean/initial value of the process is known. When the series is accompanied by a linear time trend or an unknown mean/initial condition, a more appropriate choice of estimation is the two-step ELW estimator by Shimotsu (2010). However, Shimotsu (2010) indicates that the ELW estimator by Shimotsu and Phillips (2005) remains consistent for $d \in (-\frac{1}{2}, 1)$ and is asymptotically normal for $d \in (-\frac{1}{2}, \frac{3}{4})$ if the unknown mean is replaced by the sample average. In our case, the memory estimates of the variance series considered in the empirical study are within the standard region of ELW estimation, and thus we carry out the empirical analysis using the ELW estimator of Shimotsu and Phillips (2005).

1.3.2 Fractional Co-integration Estimation

Exact Local Whittle Rank Test

Before moving to the modelling and estimation of the fractional co-integration by the FCVAR, we obtain the estimate of the co-integration rank based on an exact local Whittle approach. Specifically, we determine the co-integrating rank of the spectral density matrix of the dth differenced process around the origin frequency. This procedure is first proposed by Robinson and Yajima (2002) and later extended by Nielsen and Shimotsu (2007), who account for both stationary and non-stationary situations.

Robinson and Yajima (2002) stress that the test for homogeneity of orders of fractional integration could deliver misleading conclusions if the co-integration is not accounted for. Hence, we start the analysis by first estimating the co-integrating rank. Assume there is a p-vector fractional process X_t where each element is fractionally integrated of order $d_1, ..., d_p$, respectively. The work of Nielsen and Shimotsu (2007) builds on the assumption of equal integration orders and thus $d_1, ..., d_p$ are represented by d_*, where $\widetilde{d}_* = \frac{1}{p}\sum_{a=1}^{p} \widehat{d}_a$ with each $\widehat{d}_a$ given by equation 1.10. The consistent estimator of the spectral density at the origin is given by

$$\widehat{G}(d_*) = \frac{1}{m_1}\sum_{j=1}^{m_1} \mathrm{Re}[I_{\Delta(L;d_*,...,d_*)x}(\lambda_j)] \tag{1.12}$$

where $\lambda_j = \frac{2\pi j}{T}$ and $I_{\Delta(L;d_*,...,d_*)x}(\lambda_j)$ is the periodogram of $(\Delta^{d_*}X_{1t,...,}\Delta^{d_*}X_{pt})'$.

Here, $\widehat{G}(d_*)$ uses a new bandwidth parameter m_1 instead of m present in equation (1.8). Let $\widehat{\delta}_a$ be the ath eigenvalues of $\widehat{G}(d_*)$ and the co-integration rank r can be determined by following the procedure of Robinson and Yajima (2002)

$$\begin{aligned} \widehat{r} &= \arg\min_{u=0,...,p-1} L(u) \\ L(u) &= v(T)(p-u) - \sum_{a=1}^{p-u} \widehat{\delta}_a \end{aligned} \tag{1.13}$$

where $v(T)$ should be positive and meet the assumption as follows

$$v(T) + \frac{1}{m_1^{1/2} v(T)} \to 0 \tag{1.14}$$

Nielsen and Shimotsu (2007) show that a higher rank estimate is more likely to be selected when a larger $v(T)$ is applied. In order to obtain a more conservative estimate of r, we choose to employ a small $v(T) = m_1^{-0.4}$.

Once the presence of the co-fractional relation has been investigated by the rank estimation, we can examine the equality of the orders of fractional integration by testing the null $H_0 : d_a = d_*$, $a = 1, ...p$. The test statistic is given by

$$\widehat{T}_0 = m(S\widehat{d})'(S\frac{1}{4}\widehat{D}^{-1}(\widehat{G} \odot \widehat{G})\widehat{D}^{-1}S' + h(T)^2 I_{p-1})^{-1}(S\widehat{d}) \tag{1.15}$$

where $\odot$ represents the Hadamard product, $S = [I_{p-1}, -\iota]'$, ι is the $(p-1)$-vector of ones, $h(T) = 1/\log(T)$ is of more frequent application, and $\widehat{D} = diag(\widehat{G}_{11}, \cdots \widehat{G}_{pp})$. The memory estimates $\widehat{d}$ of variables in the vector are derived by the univariate exact local Whittle estimator by Shimotsu and Phillips (2005), with m Fourier frequencies being employed. The selection of parameters, (m, m_1, $v(T)$), will be specified in our empirical example. If variables are not fractionally co-integrated, $\widehat{T}_0 \to_d \chi^2(p-1)$, while $\widehat{T}_0 \to_p 0$ if they are co-integrated.

The FCVAR model

To further examine the long-run and short-run dynamics among the fractionally integrated $I(d)$ variables, we adopt the framework by Johansen (2008) and Johansen and Nielsen (2012). Consider a vector $X_t \in I(d)$ containing p elements, the FCVAR model is in the form of

$$\Delta^d X_t = \alpha\beta'\Delta^{d-b} L_b X_t + \sum_{c=1}^{k} \Gamma_c \Delta^d L_b^c X_t + \varepsilon_t \tag{1.16}$$

where ε_t is $p-$dimensional $i.i.d.(0,\Omega)$. Let $L_b = 1-\Delta^b$ be the fractional lag operator and Δ^d be the fractional difference operator with $\Delta^d = (1-L)^d$

$$\begin{aligned}(1-L)^d &= \sum_{i=0}^{\infty}\theta_i(d)L^i, \text{ with } \theta_i(d) = (-1)^i\binom{d}{i} = \frac{\Gamma(-d+i)}{\Gamma(-d)\Gamma(i+1)} \quad (1.17)\\ &= 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 + \cdots\end{aligned}$$

where $\Gamma(.)$ is the gamma function. The error correction term is denoted by $\beta'\Delta^{d-b}X_t$, where β is a $(p\times r)$ matrix consisting of r co-integrating vectors and r is the so-called co-integration rank. The linear combination $\beta'X_t$ is integrated of order $(d-b)$ with $d\geq b>0$. This suggests that the co-integrating combination reduces the integration order of X_t by b, where b measures the degree of fractional co-integration. The matrix α is of order $(p\times r)$ and contains the parameters representing the speed of adjustment towards long-run equilibrium. The short-run dynamics are measured by the lag coefficients $(\Gamma_1,\ldots,\Gamma_k)$.

The FCVAR model is estimated by means of a profile likelihood technique (see Johansen and Nielsen (2012)). The maximum likelihood estimators (MLE) and maximized likelihood are calculated by minimizing the profile likelihood $\ell(\psi,r)$ as a function of $\psi=(d,\ b)$. Once d and b are determined, all the other parameters, $\widehat{\alpha}$, $\widehat{\beta}$, and $\widehat{\Gamma}_c$ for $c=1,\cdots,k$ can be concentrated out by regression and reduced rank regression. Recall the FCVAR in equation (1.16) and define $Z_{0,t}=\Delta^d X_t$, $Z_{1,t}=(\Delta^{d-b}-\Delta^d)X_t$ and $Z_{k,t}=\left\{\Delta^d L_b^c X_t\right\}_{c=1}^k$. For fixed $\psi=(d,\ b)$, the MLE is found by reduced rank regression of $Z_{0,t}$ on $Z_{1,t}$ corrected for $Z_{k,t}$. More specifically, we need to obtain the residuals of the respective regressions of $Z_{0,t}$ and $Z_{1,t}$ on $Z_{k,t}$, denoted as $R_{0,t}$ and $R_{1,t}$, to construct the profile likelihood function $\ell(\psi,r)$. In the conventional situation where all the variables are fractional of order d, the regression of $Z_{0,t}$ on $Z_{k,t}$ is balanced in the sense that both the regressand and regressor are $I(0)$, while the regression of $Z_{1,t}$ on $Z_{k,t}$ is not balanced with the $I(b)$ regressand and $I(0)$ regressor, i.e. a reduction of b in the integration order from LHS to RHS. The choice of $\widehat{\beta}$ is the choice of linear combinations of $Z_{1,t}$ which

have the largest squared partial correlations with $Z_{0,t}$ after correcting for fractional lags.

The likelihood ratio (LR) test can be used to determine the co-integration rank r. Letting $\Pi = \alpha\beta'$, we have the LR test statistic of the null hypothesis $H_r : rank(\Pi) = r$ against the alternative $H_p : rank(\Pi) = p$. The profile likelihood function is maximized both under the null and alternative hypothesis and then the LR test statistic is such that

$$LR_T(q) = 2\log(\ell(\widehat{\psi}_p, p)/\ell(\widehat{\psi}_r, r)) \tag{1.18}$$

where $\ell(\widehat{\psi}_p, p) = \max_\psi \ell(\psi, p)$; $\ell(\widehat{\psi}_r, r) = \max_\psi \ell(\psi, r)$ and $q = p - r$. Johansen and Nielsen (2012) show that $LR_T(q)$ depends heavily on the parameter b in that

$$\begin{cases} LR_T(q) \rightarrow \chi^2(q^2), \ 0 < b < 1/2 \text{ (weak fractional co-integration)} \\ LR_T(q) \sim \text{ non-standard}, \ b > 1/2 \text{ (strong fractional co-integration)} \end{cases} \tag{1.19}$$

Due to the non-standard asymptotic distribution of the test statistic in the case of strong fractional co-integration, we follow the program developed by MacKinnon and Nielsen (2014) to obtain the asymptotic P values for the LR co-integrating rank tests. In addition, the selection of lag value k is of critical importance in the specification of the FCVAR model. We determine the order of lag by following the BIC information criteria while ensuring that the short-run coefficients Γ_k are significantly different from zero and that the residuals are stationary and serially uncorrelated.

A crucial problem of the FCVAR model is the lack of identification on the likelihood function as suggested by Carlini and Santucci de Magistris (2017), i.e., there may exist equivalent sub-models associated with different sets of parameters. When the lag order, k, is unknown and potentially over-specified, Carlini and Santucci de Magistris (2017) present a strong relationship between the lag length and the indeterminacy of the FCVAR. They also provide a necessary and sufficient

condition for identification of the FCVAR model corresponding to any lag structure. Such condition, $\mathcal{F}(\mathrm{d})$, is defined by

$$|\alpha_{\perp}'\Gamma\beta_{\perp}| \neq 0 \tag{1.20}$$

where $\alpha_{\perp}'\alpha = 0$, and $\Gamma = I - \sum_{c=1}^{k}\Gamma_c$. When the cointegrating rank is unknown, they further show that the FCVAR with full rank and k lags is equivalent to that with rank 0 and $k+1$ lags, in which case the $\mathcal{F}(\mathrm{d})$ condition delivers no information in terms of the model identification. Whether the rank is known or not, the identification issue for any lag greater than one can be solved by imposing a lower-bound restriction on d where the lower bound is based upon a semiparametric estimate, e.g. the estimator by Shimotsu and Phillips (2005), of the integration order, termed as the $\widetilde{d}$. The lower bound $\delta_{\min}$ is given by

$$\delta_{\min} = \widetilde{d} - c \times \widetilde{d} \tag{1.21}$$

where $c = 0.15$ is recommended in the work of Carlini and Santucci de Magistris (2017).

The M-FCVAR model

This section introduces the M-FCVAR model and provides the specifications which ensure that the shocks associated with the $I(0)$ variables do not exert long-run effects on the $I(d)$ variables.

The FCVAR in equation (1.16) does not require that all components of X_t exhibit the same order of integration, which accords with the situation where there can be a mixture of $I(1)$ and $I(0)$ variables in the traditional VECM. For instance, following Example 3 in Johansen (2008), we construct a system containing two

$I(0.4)$ variables, X_{1t} and X_{2t}, and one $I(0)$ variable, X_{3t}, given by

$$\begin{aligned} X_{1t} &= \Delta_+^{-0.4}\varepsilon_{1t} - \Delta_+^{-0.2}\varepsilon_{2t} + \varepsilon_{3t} \\ X_{2t} &= \Delta_+^{-0.4}\varepsilon_{1t} + \Delta_+^{-0.2}\varepsilon_{2t} + \varepsilon_{3t} \\ X_{3t} &= \varepsilon_{1t} + \varepsilon_{2t} + \varepsilon_{3t} \end{aligned} \tag{1.22}$$

where $X_t = (X_{1t},\, X_{2t},\, X_{3t})'$, ε_t is i.i.d. $(0,\, I_3)$ and $\Delta_+^{-d}\varepsilon_t = \sum_{i=0}^{t-1}(-1)^i\binom{-d}{i}\varepsilon_{t-i}$. The long-run transfer function for $\Delta^{0.4}X_t$, i.e. the matrix of responses of the variables to the shocks, is

$$C(1) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and the spectrum is $C(1)'C(1) \neq 0$. Hence, $\Delta^{0.4}X_t \in \mathcal{F}(0)$ and $X_t \in \mathcal{F}(0.4)$ according to Definitions 1 and 2 in Johansen (2008), which suggests that the representation theorem and the properties of MLE of the FCVAR remain unchanged when the $I(0)$ variables are introduced into the system of fractional variables.

The following section gives an outline of the problem that may arise when the FCVAR in equation (1.16) is used to accommodate a system containing $I(d)$ and $I(0)$ variables. As a standard method employed in the literature of treating an $I(0)$ variable in the VECM, we adopt the idea of 'pseudo' co-integrating relation. Specifically, we involve the extra co-integration vector being unit vector with unity in the position corresponding to the $I(0)$ variable and zeros elsewhere. Without loss of generality, we assume that there are n $I(d)$ variables and q $I(0)$ variables in X_t which contains p elements, giving $p = n + q$. Among the $I(d)$ variables, there are r co-integrating relations and thus $l = n - r$ permanent components. Here, we refer to the r co-integrating relations as 'true' co-integrating relations, as opposed to the q 'pseudo' co-integrating relations that arise from the $I(0)$ variables, which are treated as 'fractionally co-integrated with themselves'. We let $X_t = (x_{1t},\, x_{2t},\, x_{3t})'$ where $(x_{1t},\, x_{2t})'$ is the $n \times 1$ vector of $I(d)$ variables which will be classified

as endogenous or weakly exogenous variables in the subsequent analysis and x_{3t} is the $q \times 1$ vector of $I(0)$ variables. We then construct

$$\alpha = \begin{pmatrix} \alpha_1^* & \delta_1^* \\ \alpha_2^* & \delta_2^* \\ \alpha_3^* & \delta_3^* \end{pmatrix}_{p\times(r+q)} \qquad \beta' = \begin{pmatrix} \beta_1' & \beta_2' & 0 \\ 0 & 0 & I_q \end{pmatrix}_{(r+q)\times p} \tag{1.23}$$

The first row of block matrices in β' are the coefficients in the 'true' co-fractional relations among $I(d)$ variables, while those in the second row correspond to the 'pseudo' co-fractional relations. The FCVAR in equation (1.16) is no longer appropriate for modelling a system containing a mixture of $I(d)$ and $I(0)$ variables when $d > b$, in which case the term $\beta'\Delta^{d-b}X_t$ contains anti-persistent error correction terms, i.e. terms which are integrated of a negative order, due to the presence of $I(0)$ variables in X_t. The mis-specification problem can also be seen by considering the representation theorem of the FCVAR (1.16). Given α and β as defined in equation (1.23) and $\Gamma = I - \sum_{c=1}^{k} \Gamma_c$, the matrix $C = \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp'$ contains only zeros in the last q rows corresponding to the q $I(0)$ variables in X_t. Following the work of Johansen and Nielsen (2012), the FCVAR in equation (1.16) has the solution

$$X_t = C\Delta_+^{-d}\varepsilon_t + \Delta_+^{-(d-b)}Y_t^+ + \mu_t \tag{1.24}$$

for $d \geq 1/2$ where the operator Δ_+^{-d} is used to define a nonstationary process and Y_t is fractional of order zero. The solution of the FCVAR model for the last q equations, i.e. $I(0)$ x_{3t}, then reduces to

$$x_{3t}^{FCVAR} = e3'\Delta_+^{-(d-b)}Y_t^+ + e3'\mu_t \tag{1.25}$$

where $e3' = (0_{q\times n}, I_{q\times q})$. It is clear that the x_{3t}^{FCVAR} is integrated of order $(d-b)$, which erroneously exhibits long memory if $d > b$ due to the mis-specifications. This problem remains in the case of $d < 1/2$ where the solution of the FCVAR becomes $X_t = C\Delta^{-d}\varepsilon_t + \Delta^{-(d-b)}Y_t$.

The M-FCVAR is developed to address the issue described above. We apply the fractional differencing operator Δ^{d-b} to each of the long-memory variables in X_t and construct a new system X_t^*. Adapting the notation of the FCVAR, we obtain the M-FCVAR as follows

$$\Delta^b X_t^* = \alpha\beta' L_b X_t^* + \sum_{c=1}^{k} \Gamma_c \Delta^b L_b^c X_t^* + \varepsilon_t \tag{1.26}$$

Here, the M-FCVAR differs from the FCVAR only in the way that the fractional $I(d)$ variables have been transformed to $I(b)$ variables. Hence, the representation theorem of the M-FCVAR in equation (1.26) is the same as that of the FCVAR except that d is always equal to b. We can then show that the M-FCVAR model-implied $x_{3t}^{M-FCVAR}$ remains $I(0)$ according to equation (1.25) since $d = b$. Both the FCVAR and M-FCVAR are estimated by means of a profile likelihood technique (see Johansen and Nielsen (2012)). For each fixed combination of $\psi = (d,\ b)$ in the estimation of the M-FCVAR, we first construct X_t^* by applying $\Delta^{(d-b)}$ to the fractionally integrated variables in X_t. We then define $Z_{0,t} = \Delta^b X_t^*$, $Z_{1,t} = (1 - \Delta^b)X_t^*$ and $Z_{k,t} = \left\{\Delta^b L_b^c X_t^*\right\}_{c=1}^{k}$. The MLE is found by reduced rank regression of $Z_{0,t}$ on $Z_{1,t}$ corrected for $Z_{k,t}$.

For the case of the coexistence of $I(d)$ and $I(0)$ variables in the system of the (M-)FCVAR, there will be shocks coming from the $I(0)$ variables. The effects of those shocks can either be transitory or permanent on the $I(d)$ variables. In the rest of this section, we show that the shocks associated with the $I(0)$ variables produce zero long-run impact on the fractional variables only when a particular model design is considered. We follow the approach of FHP (2016) to control for the long-run effects of the $I(0)$ variables and extend their work by allowing for variation in the number of weakly exogenous $I(d)$ variables within the (M-)FCVAR framework.

For simplicity, we consider the M-FCVAR with only one lag ($k = 1$) as follows

$$\Delta^b X_t^* = \alpha\beta' L_b X_t^* + \Gamma_1 \Delta^b L_b X_t^* + \varepsilon_t \tag{1.27}$$

We start with the case where only n $I(d)$ variables, x_{1t} and x_{2t}, are present in the M-FCVAR. Applying the operator Δ^{d-b} to x_{1t} and x_{2t}, we obtain $X_t^* = (x_{1t}^*, x_{2t}^*)'$ which is fractionally integrated of order b. We define $\alpha'_{true} = \begin{pmatrix} \alpha_1^* & \alpha_2^* \end{pmatrix}$ and its orthogonal complement $\alpha'_{true\perp} = \begin{pmatrix} \theta^* & \lambda^* \end{pmatrix}$ as a $l \times n$ matrix, which gives $\alpha'_{true\perp}\alpha_{true} = \theta^*\alpha_1^* + \lambda^*\alpha_2^* = \mathbf{0}$. There are a number of different ways of defining the common stochastic trend. Here, we follow Gonzalo and Granger (1995) in estimating the permanent and transitory (PT) components of X_t^*. As in equation (1.27), X_t^* can be explained with regard to a smaller number, $l = n - r$, of $I(b)$ variables, defined as common factors $f_t = \alpha'_{true\perp}(x_{1t}^*, x_{2t}^*)'$ and r $I(0)$ variables $z_t = \beta'(x_{1t}^*, x_{2t}^*)'$ conditional on that $(\alpha'_{true\perp}, \beta)'$ is nonsingular. The Gonzalo and Granger (1995) definition requires the $I(b)$ common factors f_t to be linear combinations of the observable variables and also requires that z_t does not Granger-cause f_t in the long run. The PT approach exhibits two main advantages such as: (1) f_t is unique and can be easily obtained from the (M-)FCVAR; (2) hypothesis testing on the common trends is straightforward and follows a chi-square distribution. The Gonzalo and Granger (1995) decomposition has been widely applied in various studies such as Baillie et al. (2002), Banerjee, Marcellino, and Osbat (2004), Blanco, Brennan, and Marsh (2005), Bollerslev et al. (2013) and Dolatabadi, Nielsen, and Xu (2015), among others.

We then add x_{3t}, the $q \times 1$ $I(0)$ variables, into the system of the M-FCVAR and adopt α and β specified in equation (1.23). The 'true' error correction term, ξ_t, is given by

$$\xi_t = \beta_1' x_{1t}^* + \beta_2' x_{2t}^* \tag{1.28}$$

and the 'pseudo' error correction term is x_{3t}.

Proposition 1 *In the equations of the M-FCVAR (1.26) where the $I(b)$ dependent variables constitute the common component, the 'pseudo' error correction term, x_{3t} in levels form, is not present, which ensures that shocks associated with x_{3t} have only zero long-run effects on the fractionally integrated variables.*

Proposition 1 suggests that: if all the fractional variables, x_{1t}^* and x_{2t}^*, are endogenous, there is a requirement for $\delta_1^* = \mathbf{0}$ and $\delta_2^* = \mathbf{0}$; if any fractional variables are weakly exogenous, loadings on the 'true' error correction term equate to zero for equations where the response variables are exogenous; and loadings on the 'pseudo' error correction term equate to zero for equations where the dependent variables constitute the permanent components as defined in Gonzalo and Granger (1995). We take into account each of the three possibilities: (a) all the fractional variables are endogenous; (b) there are as many weakly exogenous fractional variables as common trends; (c) the number of weakly exogenous fractional variables are fewer than the number of common trends. Case (b) is analogous to the example given by FHP (2016). The remainder of this section discusses how the $I(0)$ variables should be accounted for in the M-FCVAR model if the shocks to the $I(0)$ variables produce only transitory impact on the fractional variables. However, it is worth mentioning that the Proposition 1 also applies to the FCVAR model introduced earlier since the FCVAR differs from the M-FCVAR only in terms of the integration orders of the long-memory variables within the system.

Proof. [Case a] To ensure that the shocks arising from the $I(0)$ variables exert transitory effects only on the $I(b)$ variables, the common permanent components of a system with a mixture of $I(b)$ and $I(0)$ variables should remain the same as those in a system containing $I(b)$ variables only given by

$$f_t = \alpha'_{true\perp} \begin{pmatrix} x_{1t}^* \\ x_{2t}^* \end{pmatrix} = \theta^* x_{1t}^* + \lambda^* x_{2t}^* \tag{1.29}$$

We then extract permanent components within the model (1.27) by multiplication with the matrix $\begin{pmatrix} \theta^* & \lambda^* & \underbrace{0 \cdots 0}_{q} \end{pmatrix}$. Next, we substitute $x_{2t}^* = (\beta_2')^{-1}(\xi_t - \beta_1' x_{1t}^*)$

and obtain

$$\begin{aligned}\Delta^b f_t &= (\theta^*\delta_1^* + \lambda^*\delta_2^*)L_b x_{3t} \\ &+ \left(\theta^*\Gamma_{11} + \lambda^*\Gamma_{21} - (\theta^*\Gamma_{12} + \lambda^*\Gamma_{22})(\beta_2')^{-1}\beta_1'\right)\Delta^b L_b x_{1t}^* \\ &+ (\theta^*\Gamma_{12} + \lambda^*\Gamma_{22})(\beta_2')^{-1}\Delta^b L_b \xi_t + (\theta^*\Gamma_{13} + \lambda^*\Gamma_{23})\Delta^b L_b x_{3t} + \varepsilon_t^f\end{aligned} \tag{1.30}$$

Let $w_t = (\Delta^b f_t, \xi_t, x_{3t})'$, so that equation (1.30) can be expressed in the general form

$$w_t = B_1 L_b w_t + B_2 L_b^2 w_t + \varepsilon_t \tag{1.31}$$

We then substitute equation $f_t = \theta^* x_{1t}^* + \lambda^* x_{2t}^*$ into (1.31) and expand it to

$$\begin{aligned}\Delta^b f_t &= (B_{12}^1 + B_{12}^2)L_b\xi_t + (B_{13}^1 + B_{13}^2)L_b x_{3t} \\ &+ \left(B_{11}^1\theta^* - B_{11}^2\lambda^*(\beta_2')^{-1}\beta_1' - B_{11}^1\beta_1' + B_{11}^2\theta^*\right)\Delta^b L_b x_{1t}^* \\ &+ \left(B_{11}^1\lambda^*(\beta_2')^{-1} - B_{12}^2 + B_{12}^2\lambda^*(\beta_2')^{-1}\right)\Delta^b L_b\xi_t - B_{13}^2 L_b\Delta^b x_{3t} \\ &- B_{11}^2\theta^*\Delta^{2b}L_b x_{1t}^* - B_{11}^2\lambda^*(\beta_2')^{-1}\Delta^{2b}L_b\xi_t - B_{11}^2\lambda^*(\beta_2')^{-1}\beta_1'\Delta^{2b}L_b x_{1t}^* + \varepsilon_{1t}\end{aligned} \tag{1.32}$$

with each term in analogous positions to those in equation (1.30). Comparing equation (1.32) with (1.30), we have

$$\begin{aligned}0 &= B_{12}^1 + B_{12}^2 \\ \theta^*\delta_1^* + \lambda^*\delta_2^* &= B_{13}^1 + B_{13}^2\end{aligned} \tag{1.33}$$

We define the characteristic polynomial of equation (1.31)

$$B(z) = I_p - B_1(1-(1-z)^b) - B_2(1-(1-z)^b)^2 \tag{1.34}$$

and thus $B(1) = I_p - B_1 - B_2$.

The moving average representation of w_t is given by

$$w_t = B(L)^{-1}\varepsilon_t = C(L)\varepsilon_t \tag{1.35}$$

where $C(L) = I_p + C_1 L + C_2 L^2 + C_3 L^3 + \cdots$, so that

$$C(1)B(1) = I \tag{1.36}$$

Shocks associated with the $I(0)$ variables, x_{3t} and ξ_t, are transitory, suggesting that the response of the first vector, f_t, to shocks arising from x_{3t} and ξ_t at infinity should be equal to zero. Thus we should have $C_{12}(1) = \mathbf{0}$ and $C_{13}(1) = \mathbf{0}$. Linking these conditions to equation (1.36), we obtain $C_{11}(1)B_{12}(1) = \mathbf{0}_{12}$ and $C_{11}(1)B_{13}(1) = \mathbf{0}_{13}$. The element $C_{11}(1)$ represents the response of x_{1t}^* to the permanent shock on the long-term run, which is non-zero by definition. Consequently, we have $B_{12}(1) = \mathbf{0}_{12}$ and $B_{13}(1) = \mathbf{0}_{13}$, which leads to

$$\begin{aligned} B_{12}^1 + B_{12}^2 &= \mathbf{0} \\ B_{13}^1 + B_{13}^2 &= \mathbf{0} \end{aligned} \tag{1.37}$$

Substituting equation (1.37) into (1.33), we obtain that $\delta_1^* = \mathbf{0}$ and $\delta_2^* = \mathbf{0}$ since θ^* and λ^* are not null matrices. ■

If there are as many weakly exogenous long-memory variables as permanent components (common trends), e.g. x_{1t}^* is the $l \times 1$ $I(b)$ exogenous variables, it requires that $\alpha_1^* = \mathbf{0}$ and $\delta_1^* = \mathbf{0}$ for shocks associated with the $I(0)$ x_{3t} to have transitory effects only. This suggests that, in the equations where the response variables are exogenous, both 'true' and 'pseudo' error correction terms, ξ_t and x_{3t}, do not appear.

Proof. [Case b] We now have x_{1t}^* as the $l \times 1$ vector of $I(b)$ variables whose shocks have permanent effects, x_{2t}^* as the $r \times 1$ vector of $I(b)$ variables whose shocks have transitory effects and x_{3t} as the $q \times 1$ vector of $I(0)$ variables whose shocks are also transitory. Accordingly, α_1^* is an $l \times r$ null matrix while α_2^* is an $r \times r$ matrix of full rank. Thus the orthogonal complement of α_{true} becomes $\alpha'_{true\perp} = \begin{pmatrix} I_{l\times l} & 0_{l\times r} \end{pmatrix}$.

It then follows that equation (1.30) can be written as

$$\begin{aligned} \Delta^b f_t &= \Delta^b x^*_{1t} \qquad (1.38) \\ &= \alpha^*_1 L_b \xi_t + \delta^*_1 L_b x_{3t} + (\Gamma_{11} - \Gamma_{12}(\beta'_2)^{-1}\beta'_1)\Delta^b L_b x^*_{1t} \\ &\quad + \Gamma_{12}(\beta'_2)^{-1}\Delta^b L_b \xi_t + \Gamma_{13}\Delta^b L_b x_{3t} + \varepsilon^f_t \end{aligned}$$

Comparing equation (1.32) with (1.38), we obtain

$$\begin{aligned} \alpha^*_1 &= B^1_{12} + B^2_{12} \qquad (1.39) \\ \delta^*_1 &= B^1_{13} + B^2_{13} \end{aligned}$$

Substituting equation (1.37) into (1.39), we conclude that $\alpha^*_1 = \mathbf{0}$ and $\delta^*_1 = \mathbf{0}$. ■

Finally, another possible situation is where there are $0 < j < l$ weakly exogenous long-memory variables with respect to the co-integrating relationships. We partition X^*_t and α as

$$X^*_t = \begin{pmatrix} x^*_{1(j)t} \\ x^*_{1(l-j)t} \\ x^*_{2t} \\ x_{3t} \end{pmatrix} \quad \alpha = \begin{pmatrix} \alpha^*_{1(j)} & \delta^*_{1(j)} \\ \alpha^*_{1(l-j)} & \delta^*_{1(l-j)} \\ \alpha^*_2 & \delta^*_2 \\ \alpha^*_3 & \delta^*_3 \end{pmatrix}_{p\times(r+q)}$$

where $x^*_{1(j)t}$ is the $(j \times 1)$ vector of exogenous long-memory variables and the rest of variables are all endogenous. $\alpha^*_{1(j)}$ ($\alpha^*_{1(l-j)}$) and $\delta^*_{1(j)}$ ($\delta^*_{1(l-j)}$) are, respectively, the first j (last $(l-j)$) rows of α^*_1 and δ^*_1. It requires that $\alpha^*_{1(j)} = \mathbf{0}$, $\delta^*_{1(j)} = \mathbf{0}$, $\delta^*_{1(l-j)} = \mathbf{0}$ and $\delta^*_2 = \mathbf{0}$ so that shocks associated with the $I(0)$ x_{3t} have zero long-run effects only on the other variables.

Proof. [Case c] Given that there are $j < l$ weakly exogenous long-memory variables, $\alpha^*_{1(j)}$ is the $(j \times r)$ null matrix. Therefore, the orthogonal complement

of α_{true} can be written as

$$\alpha'_{true\perp} = \begin{pmatrix} I_{j\times j} & 0_{j\times(l-j)} & 0_{j\times r} \\ 0_{(l-j)\times j} & \theta^*_{(l-j)\times(l-j)} & \lambda^*_{(l-j)\times r} \end{pmatrix}$$

where θ^* and λ^* are not null matrices. We then obtain the common permanent components of the system as

$$\Delta^b f_t = \Delta^b \begin{pmatrix} x^*_{1(j)t} & \left(\theta^* x^*_{1(l-j)t} + \lambda^* x^*_{2t}\right) \end{pmatrix}' \tag{1.40}$$

which can be taken as a combination of cases (a) and (b). We consider the components $x^*_{1(j)t}$ and $\left(\theta^* x^*_{1(l-j)t} + \lambda^* x^*_{2t}\right)$ separately and require that $\alpha^*_{1(j)} = \mathbf{0}$, $\delta^*_{1(j)} = \mathbf{0}$, $\delta^*_{1(l-j)} = \mathbf{0}$ and $\delta^*_2 = \mathbf{0}$ by following the procedure described in cases (a) and (b). ■

1.4 Simulation Study

To further illustrate the superiority of the M-FCVAR model when applied in a system with a mixture of $I(d)$ and $I(0)$ variables, a simulation study is conducted to compare the FCVAR and M-FCVAR in terms of the model fit and parameter estimation.

We consider a simple fractional and co-fractional process $X_t = (X_{1t}, X_{2t}, X_{3t})'$ similar to Example 3 of Johansen (2008):

$$\begin{aligned} X_{1t} &= \Delta_+^{-d}\varepsilon_{1t} + \Delta_+^{-d}\varepsilon_{2t} + \Delta_+^{-d}\varepsilon_{3t} - \Delta_+^{-(d-b)}\varepsilon_{2t} + \varepsilon_{3t} \\ X_{2t} &= \Delta_+^{-d}\varepsilon_{1t} + \Delta_+^{-d}\varepsilon_{2t} + \Delta_+^{-d}\varepsilon_{3t} + \Delta_+^{-(d-b)}\varepsilon_{2t} + \varepsilon_{3t} \\ X_{3t} &= \varepsilon_{1t} + \varepsilon_{2t} + \varepsilon_{3t} \end{aligned} \tag{1.41}$$

where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})'$ is i.i.d. $(0, I_3)$. In equation (1.41), $\Delta^d X_t \in \mathcal{F}(0)$, which implies that $X_t \in \mathcal{F}(d)$. In addition, the shock arising from the $I(0)$ variable X_{3t} produces non-zero long-run effects on both X_{1t} and X_{2t} and the two fractional

variables X_{1t} and X_{2t} are fractionally co-integrated of order $CI(d, b)$. In estimation of the FCVAR and M-FCVAR models, we employ the natural normalization of the β matrix as

$$\beta' = \begin{pmatrix} 1 & \beta_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where $\beta_1 = -1$ in our case. Here, the relation between X_{3t} and $(X_{1t} + \beta_1 X_{2t})$ is analogous to the relation between returns and the variance risk premium introduced earlier. The Monte Carlo simulation is based on 10000 replications, with the sample size T =(2500, 1000, 500). It should be noted that the M-FCVAR is equivalent to the FCVAR only when $d = b$ so that, in practice, it is necessary to determine the presence of long memory, i.e. $d - b$, in the co-integrating residuals of the fractional variables before introducing the $I(0)$ variable into the system. In the simulation, we vary d in the range 0.4 to 0.8 as commonly seen in the empirical studies and consider cases with the gap between d and b equal to 0.1 and 0.2. Since the shock to X_{3t} has permanent effect on the fractional variables in equation (1.41), we do not impose zero restrictions on the α matrix in the estimation of the FCVAR and M-FCVAR. In each case, we set rank as two, i.e. one 'true' co-integrating relation and one 'pseudo' co-integrating relation, and let lag equal to zero. The MSE of model estimates and the BIC are reported in Table 1(a).

In Table 1(a), the estimates of the model parameters, $\widehat{d}$ and $\widehat{\beta}_1$, are shown to become more precise as the sample size increases. Under the same fractional integration order d, the precision in the estimates of model parameters as well as the in-sample fit improve as b increases. The advantage of the M-FCVAR is more evident in Table 1(b), where the gains are measured by the reduction in the values of MSE and BIC of the M-FCVAR relative to those of the FCVAR. We find that the M-FCVAR achieves a better in-sample fit, i.e. lower BIC, and produce less biased estimates in almost all cases considered. In addition, such superiority generally improves with the gap between d and b. This example illustrates the advantages of the M-FCVAR in a mixture of $I(d)$ and $I(0)$ variables under the

presence of long memory in the co-integrating error. Next, we provide a different example to justify the usefulness of Proposition 1, i.e. the need to restrict the impact of the shock to the $I(0)$ variable.

To let the shock arising from the $I(0)$ variable has only transitory effect on the other variables, we design the experiment as follows

$$\begin{aligned} X_{1t} &= \Delta_+^{-d}\varepsilon_{1t} - \Delta_+^{-(d-b)}\varepsilon_{2t} \\ X_{2t} &= \Delta_+^{-d}\varepsilon_{1t} + \Delta_+^{-(d-b)}\varepsilon_{2t} \\ X_{3t} &= \varepsilon_{2t} + \varepsilon_{3t} \end{aligned} \tag{1.42}$$

where ε_t is i.i.d. $(0, I_3)$. This mimics case (a) in section 1.3.2 where the common permanent component is made up of X_{1t} and X_{2t} and the shock to X_{3t} has zero long-run impact on X_{1t} and X_{2t}. According to Proposition 1, zero restrictions $\delta_1^* = \delta_2^* = 0$ are required to prevent the shock associated with X_{3t} from having permanent effect on X_{1t} and X_{2t}. In this example, we only consider the M-FCVAR due to its superiority over the FCVAR outlined above. To ascertain the influence of the properties of the shock to the $I(0)$ variable upon the estimation results, we implement the M-FCVAR with and without the restrictions $\delta_1^* = \delta_2^* = 0$, in which cases the impact on the $I(d)$ variables of the $I(0)$ variable is transitory and permanent, respectively. The MSE and BIC of the M-FCVAR in these two scenarios are reported in Table 2(a) and the percentage gains of the M-FCVAR with zero restrictions are provided in Table 2(b). In almost all cases, higher precision in the parameter estimation and better in-sample fit are observed for the M-FCVAR where the zero restrictions are imposed on the α matrix to restrict the impact of the shock coming from the $I(0)$ variable.

1.5 Empirical Study

1.5.1 Data and Volatility Metrics

Our empirical analysis is based on high-frequency data−for the period September 22, 2003 to December 31, 2013−of the aggregate S&P 500 composite index[1] and the SPDR S&P 500 EFT TRUST (SPY) index, and the daily data of the CBOE VIX volatility index obtained from Tick Data Inc..

The annualized VIX volatility index is transformed into a monthly squared version

$$VIX_t^2 = \log\left(\frac{30}{365}(VIX_t^{CBOE})^2\right) \tag{1.43}$$

where the time subscript t refers to the daily observation. To obtain the 5-minute realized variance of the S&P 500 and SPY, we calculate the intraday returns within each 5-minute interval

$$r_{t,j} = 100\left(\frac{\log(p_{t,j}^c) + \log(p_{t,j+1}^o)}{2} - \frac{\log(p_{t,j-1}^c) + \log(p_{t,j}^o)}{2}\right) \tag{1.44}$$

where $p_{t,j}^c$ ($p_{t,j}^o$) is the closing (opening) price of the jth intraday interval. The daily return is thus given by $r_t = \sum_{j=1}^M r_{t,j}$, where M is the number of intervals per trading day. Here we rely on a simple realized variance estimator proposed by Barndorff-Nielsen and Shephard (2002), which equals the sum of intraday squared returns, $rv_t = \sum_{j=1}^M r_{t,j}^2$. In order to measure the return variation during the overnight period, we add the squared overnight return, derived as the squared close-to-open logarithmic price change, to the realized variance, rv_t, obtained over the trading day. As indicated by Andersen and Bondarenko (2007), the impact of the lack of detailed information on the price movement overnight is trivial due to the relatively low volatility during non-trading periods. Given that the VIX index reflects market expectations of the one-month cumulative variation of the corresponding aggregate market index, we compute the one-month forward horizon

[1] We extend the data employed in BOST (2013) by two years.

of the variance measure in order to match the VIX. Both VIX_t^2 and RV_t are in logarithmic form.

$$RV_t = \log\left(\sum_{i=1}^{22} rv_{t+i}\right) \tag{1.45}$$

With the construction of RV_t, loss of one month's observations at the end of the period reduces the sample size to 2566 observations.

Standard summary statistics for the return and variance series are reported in Table 1.3. In line with the existing literature, daily returns are serially uncorrelated with a mean value approximately equal to zero. However, the variance series show substantially slow decaying rates in their autocorrelations, which is indicative of strong long-memory properties. In addition, the mean value of VIX_t^2 is greater than that of RV_t^{SP500} and RV_t^{SPY}, suggesting a negative variance risk premium.

1.5.2 Estimation Results

In the subsequent analysis, we first employ a semiparametric approach to estimate the fractional integration order of each of variances and examine the presence of the fractional co-integration between the implied-realized variances for the case of S&P 500 and SPY, respectively. As suggested by Carlini and Santucci de Magistris (2017), these procedures help to solve the identification problem of the FCVAR model. We then accommodate the long-run and short-run dynamics of $(RV_t,\ VIX_t^2)'$ using the FCVAR model. A direct comparison between the FCVAR and M-FCVAR is provided once returns are introduced to the system $(RV_t,\ VIX_t^2)'$. Finally, we evaluate the return predictability implied by the FCVAR and M-FCVAR models. The model estimates for the system $(RV_t,\ VIX_t^2,\ r_t)'$ cannot be easily used for the out-of-sample forecasting because the RV_t is not known at time t. Hence, we also estimate the models with $(RV_{t-22},\ VIX_t^2,\ r_t)'$ and the relation between RV_{t-22} and VIX_t^2 is investigated using the same method as we implement for $(RV_t,\ VIX_t^2)'$. The time span for the case of $(RV_{t-22},\ VIX_t^2,\ r_t)'$ is from August 21, 2003 to November 27, 2013, so that the total observations remain 2566.

We begin by determining the fractional integration d of the RV_t, RV_{t-22} and VIX_t^2 employing the exact local Whittle estimator of Shimotsu and Phillips (2005). The results are outlined in Table 1.4 where the selection of the parameter m is determined by the log-log periodograms of the variance series. To capture the long-run dependency, we set the truncation parameter[2] to $j_{\max} = 35$, from which the periodogram of each of the variances drops steeply, as indicated in Figure 1.1.

Table 1.4 suggests that all series considered, RV_t, RV_{t-22} and VIX_t^2, exhibit the long-memory property, where the magnitude of the memory of VIX_t^2 is greater at $\widehat{d}_{VIX_t^2} = 0.765$. This is consistent with Bandi and Perron (2006) and Kellard, Dunis, and Sarantis (2010) who show that the implied and realized volatility relation may lie in a non-stationary region[3] where $\widehat{d} \geq 0.5$. As a necessary condition for the presence of fractional co-integration, we examine the equality of integration orders of $(RV_t, VIX_t^2)'$ and $(RV_{t-22}, VIX_t^2)'$ for both cases of S&P 500 and SPY. A large value of the test statistic $\widehat{T}_0$ is evidence against the null hypothesis of the equality of $\widehat{d}$. Specifically, in comparing the $\widehat{T}_0$ with the 95% critical value of the $\chi^2(1)$ distribution (3.841), we accept the null and take the average $\widehat{d}$ as $\frac{1}{2}(\widehat{d}_{RV_t} + \widehat{d}_{VIX_t^2})$ or $\frac{1}{2}(\widehat{d}_{RV_{t-22}} + \widehat{d}_{VIX_t^2})$. To solve the identification problem of the FCVAR framework, we impose a lower-bound restriction on d according to equation (1.21), i.e. $\widehat{d} - 0.15 \times \widehat{d}$. We also set the average $\widehat{d}$ as a starting value in the numerical optimization of the likelihood function of both the FCVAR and M-FCVAR. Results of the fractional co-integration determination of Nielsen and Shimotsu (2007) are presented in columns 5-7 of Table 1.4. Given that $L(1) < L(0)$ for the $(RV_t, VIX_t^2)'$ and $(RV_{t-22}, VIX_t^2)'$ of both S&P 500 and SPY, i.e. $\widehat{r} = 1$, we confirm the presence of the co-fractional relation in all cases analyzed from the perspective of the frequency domain.

Before proceeding to the systems with $I(d)$ and $I(0)$ variables, we first consider

[2] We choose $m = [T^{0.45}] = j_{\max} = 35$ in the univariate exact local Whittle likelihood procedure of Shimotsu and Phillips (2005).

[3] It should be noted that lower d estimates of the realized and implied volatility, e.g. below 0.5, are found in Christensen and Nielsen (2006) and BOST (2013) using different volatility constructions.

a simple situation containing $(RV_t,\ VIX_t^2)'$ or $(RV_{t-22},\ VIX_t^2)'$ while using the FCVAR with deterministic terms, as suggested by Johansen and Nielsen (2015)

$$\Delta^d(X_t-\mu)=\alpha\beta'\Delta^{d-b}L_b(X_t-\mu)+\sum_{c=1}^{k}\Gamma_c\Delta^d L_b^c(X_t-\mu)+\varepsilon_t \tag{1.46}$$

where the level parameter μ serves to reduce the effects of pre-sample observations. Lag k is selected by following a procedure detailed in Dolatabadi, Nielsen, and Xu (2016). Estimation results of S&P 500 are outlined in Table 1.5 where Panel (A) and Panel (B) correspond to $(RV_t^{SP500}, VIX_t^2)'$ and $(RV_{t-22}^{SP500}, VIX_t^2)'$, respectively. These serve as a benchmark for comparisons when the $I(0)$ returns are not present. In Panel (A) using the rank test of Johansen and Nielsen (2012), we first confirm the presence of fractional co-integration between the RV_t^{SP500} and VIX_t^2 by having the rank equal to one, which is consistent with the conclusion obtained from Table 1.4 where a semiparametric method is considered. We then find $\widehat{b}<\widehat{d}$, which is indicative of long memory in the co-integrating residuals. The adjustment parameters in matrix $\widehat{\alpha}$ are significantly different from zero with the expected signs and $\beta'\neq\begin{pmatrix}1 & -1\end{pmatrix}$. To further evaluate the endogeneity and exogeneity properties of the RV_t^{SP500} and VIX_t^2, and the long-run relation between the two, we implement Likelihood Ratio (LR) tests of the following hypotheses

$$\begin{aligned}
H_\beta &: \text{ long-run unbiasedness in implied-realized variances } \beta'=\begin{pmatrix}1 & -1\end{pmatrix}\\
H_\alpha^1 &: \text{ realized variance is weakly exogenous } (\alpha_1^*=0)\\
H_\alpha^2 &: \text{ implied variance is weakly exogenous } (\alpha_2^*=0)
\end{aligned}$$

and list the outcomes in Table 1.5. As H_β, H_α^1 and H_α^2 are all rejected at the 5% significance level, this suggests that the RV_t^{SP500} and VIX_t^2 are both endogenous and that the VIX_t^2 is a biased forecast[4] of RV_t^{SP500}. Moving to Panel (B), we

[4]The latter finding contradicts those of Bandi and Perron (2006) and BOST (2013). This may arise from two aspects. First, the series of interest are not the same. Where Bandi and Perron (2006) concentrate on the monthly nonoverlapping observations of the VXO and the RV_t of S&P 100 based on the daily returns, BOST (2013) use variance series recorded every 5 minutes. We use similar data as in BOST (2013) but the RV_t is recorded daily. Second, the model specification

observe similar results such as the presence of fractional co-integration and long memory in the co-integrating errors. However, we fail to reject H_β and H_α^2 at the 5% significance level. First, this supports the hypothesis for the unbiased relation between RV_{t-22}^{SP500} and VIX_t^2. Second, it implies that the RV_{t-22}^{SP500} is endogenous and VIX_t^2 is weakly exogenous.

Next, we include the $I(0)$ returns into the system of $(RV_t^{SP500}, VIX_t^2)'$ and in Table 1.6 we present the difference in model estimation between the FCVAR and M-FCVAR for this situation. Note that we now have one 'true' co-integrating relation between the RV_t^{SP500} and VIX_t^2 and one 'pseudo' co-integrating relation arising from the $I(0)$ returns. In addition to the hypothesis of the long-run unbiasedness between the RV_t^{SP500} and VIX_t^2, H_β, we are also interested in the nature of the shock arising from returns, which is examined by the following test.

H_δ: shock to returns produces only a transitory effect on the variance series

$$\delta_1^* = \delta_2^* = 0$$

If the impact of the shock to returns on the variances is transitory, we need to impose zero restrictions on δ_1^* and δ_2^* since both the RV_t^{SP500} and VIX_t^2 are endogenous. According to Proposition 1, the $I(0)$ variable r_t does not appear in levels form in the equations of the FCVAR or M-FCVAR where the RV_t^{SP500} (VIX_t^2) is the dependent variable to ensure the shock arising from returns only delivers a zero long-run effect. In our empirical study, we account for the properties of the shock to returns in both the FCVAR and M-FCVAR because the impact of the shock to returns on the variances is crucial for the precision in the estimation of model parameters as shown in the simulation study in section 1.4.

With the results of the likelihood ratio tests in Table 1.6, H_β is rejected at the

adopted is different. Where Bandi and Perron (2006) apply a semiparametric narrow band least squares estimator which does not parameterize the short-run dynamics, BOST (2013) estimate the FCVAR with the fixed value of d and a restricted constant. However, we allow the parameter d to be jointly estimated with the other parameters and choose to introduce the level parameter μ as in equation (1.46).

5% significance level under the M-FCVAR but not under the FCVAR. The latter finding is different from that obtained in the FCVAR estimation for $(RV_t^{SP500}$, $VIX_t^2,)'$ in Table 1.5. In addition, H_δ is rejected under both models, suggesting that the shock associated with returns has a nonzero long-run effect on the implied and realized variances. For this case, the parameters δ_1^* and δ_2^* capture the long-run dynamic "leverage effect" and the rejection of H_δ accords with Corsi and Renò (2012) and Bollerslev, Sizova, and Tauchen (2012). Unlike the FCVAR in Panel (A), the M-FCVAR in Panel (B) achieves a better in-sample fit by having lower BIC. Moreover, the estimates ($\widehat{d}$, $\widehat{b}$ and $\widehat{\beta}$) under the M-FCVAR are much closer to those listed in Table 1.5 where returns are not added into the system. This indicates that the M-FCVAR is less sensitive than the FCVAR model to the introduction of the $I(0)$ returns.

As for the case of $(RV_{t-22}^{SP500}, VIX_t^2, r_t^{SP500})'$, similar procedures are undertaken and results are provided in Table 1.7. At the 5% significance level, we fail to reject H_β under the FCVAR but reject it under the M-FCVAR. The latter finding seems inconsistent with that from the case of $(RV_{t-22}^{SP500}, VIX_t^2)'$ in Table 1.5. However, the M-FCVAR dominates the FCVAR in Table 1.7 by providing $\widehat{\beta}$ much closer to that obtained from the bivariate case of $(RV_{t-22}^{SP500}, VIX_t^2)$ where returns are not involved. There is only trivial difference in the P value of the H_β test between the FCVAR (0.054) in Panel (B) of Table 1.5 and the M-FCVAR (0.046) in Table 1.7. Since the VIX_t^2 is weakly exogenous and RV_{t-22}^{SP500} is endogenous as suggested by the results in Table 1.5, we examine whether the shock arising from the $I(0)$ returns has a permanent impact on the variances by conducting the following test

$H_{\alpha\delta}$: shock to returns produces only a transitory effect on the variance series

$$\alpha_2^* = \delta_2^* = 0$$

Under both the FCVAR and M-FCVAR models, we fail to reject $H_{\alpha\delta}$, indicating that the shock to returns exerts only a transitory effect on the variance series. The restrictions $\alpha_2^* = \delta_2^* = 0$ are thus imposed on both models considered in

order to restrict the impact of the shock associated with returns. Again, we find that the M-FCVAR achieves a better in-sample fit and that ($\widehat{d}$, $\widehat{b}$ and $\widehat{\beta}$) given by the M-FCVAR stays closer to that from Panel (B) of Table 1.5 which serves as a benchmark model.

In the analysis above with the focus on the case of S&P 500, the superiority of the M-FCVAR is observed in terms of the in-sample fit and insensitivity to the introduction of $I(0)$ returns. Next, we consider the case of SPY as a robustness check. As before, we start with the system containing two variances only in the estimation of the FCVAR. Panel (A) of Table 1.8 shows that the VIX_t^2 is an biased forecast of RV_t^{SPY} and that both VIX_t^2 and RV_t^{SPY} are endogenous in the system of the FCVAR. In Panel (B), although biasedness still holds for (RV_{t-22}^{SPY}, $VIX_t^2)'$, RV_{t-22}^{SPY} is found endogenous whereas VIX_t^2 is weakly exogenous, similar to the case of S&P 500. We then account for the situation where returns are introduced to (RV_t^{SPY}, $VIX_t^2)'$ and present the results of the FCVAR in Panel (A) and those of the M-FCVAR in Panel (B) of Table 1.9. Under both the FCVAR and M-FCVAR models, we show that H_β and H_δ are rejected at the 5% significance level, indicative of the biased relation between RV_t^{SPY} and VIX_t^2 as well as the nonzero long-run effect of the shock to returns on the variance series. In line with the case of S&P 500, the M-FCVAR results in better in-sample fit compared with the FCVAR and the estimates ($\widehat{d}$, $\widehat{b}$ and $\widehat{\beta}$) of the M-FCVAR are much closer to those obtained from the case of (RV_t^{SPY}, $VIX_t^2)'$ in Table 1.8. Another comparison between the FCVAR and M-FCVAR is made using the case of (RV_{t-22}^{SPY}, VIX_t^2, $r_t^{SPY})'$ in Table 1.10, where the shock coming from returns is found to produce only a transitory effect on the variances. Conclusions with regard to the superiority of the M-FCVAR in a mixture of $I(d)$ and $I(0)$ variables remain intact for the system (RV_{t-22}^{SPY}, VIX_t^2, $r_t^{SPY})'$.

1.5.3 Permanent and Transitory Shocks

To illustrate the dynamic dependencies implied by the FCVAR and M-FCVAR models, we consider the Impulse Response Functions (IRF) related to three different shocks within the system. For simplicity, we concentrate on the case of S&P 500 only. To separate the permanent and transitory components, we multiply the ε_t in equation (1.46) with the matrix $G = \begin{pmatrix} \alpha'_\perp & \beta' \end{pmatrix}'$ following the work of Gonzalo and Granger (1995), known as the PT decomposition discussed in section 1.3.2. We then further orthogonalize the shocks using the idea of Gonzalo and Ng (2001), see more details in Appendix. The implications of the three shocks are now more apparent. We demonstrate the IRF for variance series in the system of $(RV_t^{SP500}, VIX_t^2, r_t^{SP500})'$ in Figure 1.2. Under both the FCVAR and M-FCVAR models, the first shock is permanent and associated with RV_t^{SP500}, VIX_t^2 and r_t^{SP500} since we impose no zero restrictions on the α matrix in Table 1.6. The second shock is transitory and arises from the 'true' error correction term. The third shock is also transitory and originates from the 'pseudo' error correction term (r_t^{SP500}) only. In Figure 1.2, under both the FCVAR and M-FCVAR, the effect of the permanent shock depicted in the first column persists over long periods while the two transitory shocks in the second and third columns, respectively, decay at a faster rate. The third column shows that the shock associated with r_t initially has positive (negative) impact on VIX_t^2 (RV_t) and the effects on the variances almost dissipate after three months. IRF with respect to the original shocks with no PT decomposition are plotted in Figure 1.3. It is more evident that the shock to returns, presented in the third column, delivers nonzero long-run effect on the variances within the system of $(RV_t^{SP500}, VIX_t^2, r_t^{SP500})'$.

In Tables 1.5 and 1.7, we conclude that the RV_{t-22}^{SP500} is endogenous and VIX_t^2 is weakly exogenous and that the shock to r_t^{SP500} has only a transitory effect on the common component within the system. A more clear picture in terms of the impact of the shocks can be found in Figure 1.4 where the IRF for the two variances with respect to the permanent and transitory shocks, using the PT

decomposition, are provided. The first shock is permanent and associated with the VIX_t^2 only since the VIX_t^2, as a weakly exogenous variable, serves as the common long-memory component within the system. The transitory shocks in the second and third column have the same meaning as those in Figure 1.2 for the case of $(RV_t^{SP500}, VIX_t^2, r_t^{SP500})'$. In Figure 1.5 where we consider the IRF with respect to the original shocks without using the PT decomposition, the shocks to RV_{t-22}^{SP500} and r_t^{SP500} exert only transitory effects whereas the effect of the shock to VIX_t^2 persists for long time periods.

1.5.4 Return Predictability

To outline the superiority of the M-FCVAR over the FCVAR for predicting future returns, we calculate the predictive $R^2(h)$, where h denotes the time horizon, using the impulse response functions (IRF) as in the work of BOST (2013), see more details in Appendix. Values of the R^2 represent the fraction of the variance of the model-implied returns which are predictable. Two benchmark models[5], the long-memory adjusted VAR and AR, are also taken into consideration. We present the $R^2(h)$ for systems of $(RV_t, VIX_t^2, r_t)'$ and $(RV_{t-22}, VIX_t^2, r_t)'$ of S&P 500 and SPY in Figures 1.6 and 1.7, respectively, where the horizon h ranges from 1 day to 100 days.

The upper panel of Figure 1.6 gives the values of the $R^2(h)$ for the system of $(RV_t^{SP500}, VIX_t^2, r_t^{SP500})'$ and the lower panel corresponds to the system of $(RV_{t-22}^{SP500}, VIX_t^2, r_t^{SP500})'$. As compared with the FCVAR and M-FCVAR models, VAR and AR result in almost no return predictability over all the horizons. The forecasting superiority of the FCVAR framework, relative to alternative models, is also documented in BOST (2013). The upper panel also shows that the M-FCVAR is superior to the FCVAR in the degree of return predictability once $h > 4$. The clearest advantages of the M-FCVAR is observed around the one-month horizon, after which its superiority decreases gradually with the horizon h. The hump

[5]We include $(\Delta^d RV_t, \Delta^d VIX_t^2, r_t)'$ for VAR estimation and let $r_t = \beta r_{t-1} + e_t$ for AR estimation.

shape in the R^2 of the FCVAR and M-FCVAR as a function of the horizon h is consistent with Bollerslev, Tauchen, and Zhou (2009) and Bollerslev et al. (2014). As for the case of $(RV_{t-22}^{SP500}, VIX_t^2, r_t^{SP500})'$, the superiority of the M-FCVAR is evident only when $h > 12$. Moving to the case of SPY in Figure 1.7, the M-FCVAR dominates the FCVAR in predicting returns when $h > 5$ for the system $(RV_t^{SPY}, VIX_t^2, r_t^{SPY})'$ while such advantage is only observed when $h > 32$ for the system $(RV_{t-22}^{SPY}, VIX_t^2, r_t^{SPY})'$.

1.6 Conclusion

We modify the fractionally co-integrated VAR (FCVAR) of Johansen (2008) for modelling systems with $I(0)$ and $I(d)$ variables, where there exists long memory in the co-integrating residuals. The problem occurring particularly in the use of the FCVAR with $I(0)$ and $I(d)$ variables is associated with the anti-persistent error correction term when $d > b$, which brings fractional property to the model-implied $I(0)$ variables. To better accommodate the systems with $I(0)$ and $I(d)$ variables when $d > b$, we propose a modified FCVAR, i.e., the M-FCVAR model, where the fractional differencing operator Δ^{d-b} is applied to the fractional $I(d)$ variables. In addition, we examine the nature of shocks arising from the $I(0)$ variables in the co-fractional system. We show that the shocks associated with the $I(0)$ variables have transitory effects on the fractionally integrated variables only when particular equation specifications are considered and that the long-run impact of such shocks is nonzero if no parameter restrictions are imposed on the FCVAR (M-FCVAR) model. Our Monte Carlo study shows that the M-FCVAR delivers less biased model estimates and better in-sample fit. Apart from this, inappropriate treatment of the shock to the $I(0)$ variable is found to lower the precision in the estimation of model parameters as well as the in-sample fit. A comparison between the FCVAR and M-FCVAR is also undertaken in an empirical application where market returns are found more predictable using the M-FCVAR over long horizons.

1.7 Appendix

1.7.1 Impulse Response Functions

We derive Impulse-Response Functions (IRF) below to gain a better understanding of the dependencies implied by the M-FCVAR. We follow the steps outlined in BOST (2013) where the FCVAR is adopted. First, we re-write equation (1.26) by expanding the fractional differencing operator Δ^d defined in equation (1.17) as follows

$$\sum_{i=0}^{\infty}\theta_i(b)L^iX_t^* = \alpha\beta'(1-\sum_{i=0}^{\infty}\theta_i(b)L^i)X_t^* + \sum_{c=1}^{k}\Gamma_c(1-\sum_{i=0}^{\infty}\theta_i(b)L^i)^c\sum_{i=0}^{\infty}\theta_i(b)L^iX_t^* + \varepsilon_t \tag{1.47}$$

For the case of lag $k=1$, we can demonstrate X_t^* in its infinite lag form by

$$\begin{aligned} X_t^* &= -\sum_{i=1}^{\infty}\theta_i(b)L^iX_t^* + \alpha\beta'(1-1-\sum_{i=1}^{\infty}\theta_i(b)L^i)X_t^* \\ &\quad + \sum_{c=1}^{k}\Gamma_c(1-1-\sum_{i=1}^{\infty}\theta_i(b)L^i)^c\sum_{i=0}^{\infty}\theta_i(b)L^iX_t^* + \varepsilon_t \\ &= -\sum_{i=1}^{\infty}\theta_i(b)L^iX_t^* - \alpha\beta'\sum_{i=1}^{\infty}\theta_i(b)L^iX_t^* \\ &\quad -\Gamma_1\sum_{i=1}^{\infty}(\sum_{l=0}^{i-1}\theta_{i-l}(b)\theta_l(b))L^iX_t^* + \varepsilon_t \\ &= \sum_{i=1}^{\infty}(-I\theta_i(b) - \alpha\beta'\theta_i(b) - \Gamma_1\sum_{l=0}^{i-1}\theta_{i-l}(b)\theta_l(b))L^iX_t^* + \varepsilon_t \end{aligned} \tag{1.48}$$

Next, we further expand equation (1.48) by allowing for multiple lags, $k \geq 1$, and provide the detailed algorithm below

Lag i	1	2	3	$\cdots$
$\lvert L_b \rvert$	$\theta_1(b)$	$\theta_2(b)$	$\theta_3(b)$	$\cdots$
$\lvert \Delta^b L_b \rvert = \mathbf{K}_{1,i}$	$\theta_1(b)$	$\theta_1^2(b) + \theta_2(b)$	$2\theta_1(b)\theta_2(b) + \theta_3(b)$	$\cdots$
$\lvert \Delta^b L_b^2 \rvert = \mathbf{K}_{2,i}$	0	$\theta_1(b)\mathbf{K}_{1,1}$	$\theta_2(b)\mathbf{K}_{1,1} + \theta_1(b)\mathbf{K}_{1,2}$	$\cdots$
$\lvert \Delta^b L_b^3 \rvert = \mathbf{K}_{3,i}$	0	0	$\theta_1(b)\mathbf{K}_{2,2} + \theta_2(b)\mathbf{K}_{2,1}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\lvert \Delta^b L_b^c \rvert = \mathbf{K}_{c,i}$		$\sum_{l=1}^{i-1} \theta_l(b)\mathbf{K}_{c-1,i-l}$		
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

(1.49)

where each element represents the coefficient at the corresponding lag i. Now, X_t^* can be written as

$$\begin{aligned} X_t^* &= \sum_{i=1}^{\infty}(-I\theta_i(b) - \alpha\beta'\theta_i(b) - \Gamma_1\mathbf{K}_{1,i} + \Gamma_2\mathbf{K}_{2,i} - \Gamma_3\mathbf{K}_{3,i} + \cdots)L^i X_t^* + \varepsilon_t \\ &= \sum_{i=1}^{\infty}(-I\theta_i(b) - \alpha\beta'\theta_i(b) + \sum_{c=1}^{k}(-1)^c\Gamma_c\mathbf{K}_{c,i})L^i X_t^* + \varepsilon_t \\ &= \sum_{i=1}^{\infty}\Xi_i L^i X_t^* + \varepsilon_t \end{aligned} \tag{1.50}$$

Let $e3' \equiv (0,0,1)$ so that $x_{3t} = e3'X_t^*$. We express the M-FCVAR model-implied infinite moving average representation for x_{3t} below

$$x_{3t} = e3' \sum_{j=0}^{\infty} \Phi_j \varepsilon_{t-j} \tag{1.51}$$

where the impulse responses, Φ_j can be constructed by the coefficients Ξ_i such as

$$\begin{aligned} \Phi_0 &= I \\ \Phi_1 &= \Xi_1 \\ \Phi_2 &= \Xi_2 + \Xi_1^2 \\ &\vdots \\ \Phi_j &= \sum_{i=0}^{j-1} \Xi_{j-i}\Phi_i \end{aligned} \tag{1.52}$$

for $\Xi_i = -I\theta_i(b) - \alpha\beta'\theta_i(b) + \sum_{c=1}^{k}(-1)^c\Gamma_c\mathbf{K}_{c,i}$.

We can then consider the IRF associated with the shocks to the permanent and transitory components within the FCVAR or M-FCVAR model. Exploiting the impulse response matrices that we derive in equation (1.52), we obtain the following representation

$$X_t^* = \Phi(L)\varepsilon_t = \begin{pmatrix} \Phi_{11}(L) & \Phi_{12}(L) & \Phi_{13}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) & \Phi_{23}(L) \\ \Phi_{31}(L) & \Phi_{32}(L) & \Phi_{33}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \\ \varepsilon_{t3} \end{pmatrix} \tag{1.53}$$

which suggests that the number of shocks corresponds to the number of variables in the system. Employing $\alpha'_\perp$, which is the orthogonal complement of α, we construct

$$G = \begin{pmatrix} \alpha'_\perp \\ \beta' \end{pmatrix}$$

Then the $l \times 1$ ($l = n - r$) vector $u_t^P = \alpha'_\perp\varepsilon_t$ and the $(r+q) \times 1$ vector $u_t^T = \beta'\varepsilon_t$ are the permanent and transitory shocks, respectively; see Gonzalo and Granger (1995). However, the shocks generated by the system might be mutually correlated and thus we follow Gonzalo and Ng (2001) to obtain 'orthogonalized' permanent and transitory shocks. Let H denote the Cholesky decomposition of $cov(G\varepsilon_t)$ and thus $\widetilde{\eta}_t = H^{-1}(G\varepsilon_t)$, which is equivalent to imposing $p \times (p-1)/2$ zero restrictions

on the off-diagonals of $cov(\widetilde{\eta}_t)$. We can now express X_t^* as follows

$$X_t^* = \Phi(L)G^{-1}HH^{-1}G\varepsilon_t = \widetilde{D}(L)\widetilde{\eta}_t \tag{1.54}$$

1.7.2 Predictive R-square

To quantify the predictability of the $I(0)$ x_{3t}, e.g. market returns, we write the M-FCVAR model (1.26) in moving average form. We then further decompose x_{3t} into an expected and unexpected component and represent the model-implied R^2 by the fraction of the variance of x_{3t} which is predictable.

To decompose x_{3t} into the expected and unexpected parts, first recall equation (1.51) and write it in the form of continuously compounded x_{3t} over h horizons. For convenience, we replace x_{3t} with x_t in the subsequent equations with no confusion hereinafter.

$$x_t^h = \sum_{j=0}^{h-1} x_{t+j} = e3' \sum_{j=0}^{h-1} \sum_{i=0}^{\infty} \Phi_i \varepsilon_{t+j-i} \tag{1.55}$$

On the basis of equation (1.55), we derive the decomposition (see Campbell (1991) for more details)

$$x_t^h = \underbrace{e3' \sum_{j=0}^{h-1} \sum_{i=j+1}^{\infty} \Phi_i \varepsilon_{t+j-i}}_{\text{expected}} + \underbrace{e3' \sum_{j=0}^{h-1} \sum_{i=0}^{j} \Phi_i \varepsilon_{t+j-i}}_{\text{unexpected}} \tag{1.56}$$

The predictive R^2 for the x_t over h horizons implied by the model is given by

$$R_h^2 = \frac{\sum_{k=1}^{\infty} e3'(\sum_{j=0}^{h-1} \Phi_{k+j})\Omega(\sum_{j=0}^{h-1} \Phi_{k+j})'e3}{\sum_{k=-(h-1)}^{\infty} e3'(\sum_{j=\max(0,-k)}^{h-1} \Phi_{k+j})\Omega(\sum_{j=\max(0,-k)}^{h-1} \Phi_{k+j})'e3} \tag{1.57}$$

representing the fraction of the variance of x_t which is predictable.

Table 1(a): Monte Carlo Simulation Results. This table reports the MSE of the model estimates $\widehat{\lambda} = (\widehat{d}, \widehat{b}, \widehat{\beta}_1)$ and the BIC under both the FCVAR and M-FCVAR models. The Monte Carlo experiment is based on 10000 replications, with T=(2500, 1000, 500). No restrictions are imposed on the α matrix since the shock coming from the $I(0)$ variable produces a permanent effect on the fractional variables in the DGP. In the simulations, β_1=-1. The values of d and b are provided in the table.

	FCVAR										**M-FCVAR**									
d	0.4	0.4	0.5	0.5	0.6	0.6	0.7	0.7	0.8	0.8	0.4	0.4	0.5	0.5	0.6	0.6	0.7	0.7	0.8	0.8
b	0.2	0.3	0.3	0.4	0.4	0.5	0.5	0.6	0.6	0.7	0.2	0.3	0.3	0.4	0.4	0.5	0.5	0.6	0.6	0.7
T=2500																				
MSE_$\widehat{d}$ (1.0E-3)	0.358	0.230	0.245	0.224	0.240	0.217	0.228	0.205	0.217	0.197	0.248	0.183	0.176	0.171	0.161	0.164	0.147	0.153	0.133	0.142
MSE_$\widehat{b}$ (1.0E-3)	16.009	4.350	14.554	4.049	13.438	3.803	12.274	3.662	11.261	3.592	0.382	0.378	0.366	0.327	0.371	0.309	0.375	0.295	0.377	0.292
MSE_$\widehat{\beta}_1$ (1.0E-2)	2.544	0.324	0.728	0.125	0.284	0.047	0.097	0.016	0.031	0.005	0.654	0.173	0.156	0.053	0.052	0.018	0.018	0.006	0.007	0.002
BIC	24730	24715	24735	24721	24747	24723	24759	24725	24774	24730	24722	24713	24722	24717	24727	24716	24729	24715	24732	24717
T=1000																				
MSE_$\widehat{d}$ (1.0E-3)	1.319	0.631	0.689	0.590	0.610	0.569	0.584	0.522	0.563	0.518	1.032	0.594	0.624	0.561	0.581	0.544	0.572	0.504	0.578	0.499
MSE_$\widehat{b}$ (1.0E-3)	18.328	5.691	16.332	5.079	14.812	4.717	13.770	4.414	12.863	4.249	6.684	3.913	5.133	2.958	4.263	2.306	3.780	1.870	3.399	1.646
MSE_$\widehat{\beta}_1$ (1.0E-2)	11.596	0.699	1.207	0.222	0.425	0.092	0.172	0.037	0.063	0.013	4.543	0.550	0.532	0.177	0.188	0.064	0.070	0.024	0.025	0.009
BIC	9857	9852	9859	9851	9862	9852	9868	9854	9873	9854	9853	9851	9853	9850	9854	9849	9856	9850	9856	9849
T=500																				
MSE_$\widehat{d}$ (1.0E-3)	1.979	1.270	1.362	1.209	1.201	1.117	1.226	1.081	1.193	1.082	1.680	1.193	1.206	1.113	1.042	1.036	1.059	0.988	1.063	1.005
MSE_$\widehat{b}$ (1.0E-3)	8.161	5.218	8.577	5.001	8.621	4.809	8.621	4.683	8.520	4.525	5.484	4.697	5.078	3.997	4.764	3.425	4.498	3.026	4.141	2.699
MSE_$\widehat{\beta}_1$ (1.0E-2)	17.351	1.038	1.425	0.401	0.603	0.169	0.263	0.073	0.113	0.030	6.693	1.170	1.162	0.398	0.388	0.148	0.152	0.061	0.062	0.024
BIC	4996	4992	4998	4993	4999	4994	5001	4995	5004	4995	4994	4991	4994	4992	4994	4992	4994	4993	4995	4993

Table 1(b): Percentage Gains of the MSE and BIC. The gains of the M-FCVAR are computed as ΔMSE_$\widehat{\lambda}$ =[MSE($\widehat{\lambda}_{\text{FCVAR}}$)-MSE($\widehat{\lambda}_{\text{M-FCVAR}}$)]/MSE($\widehat{\lambda}_{\text{FCVAR}}$) and ΔBIC =[BIC(FCVAR)-BIC(M-FCVAR)]/BIC(FCVAR).

		Gains (%)									
	d	0.4	0.4	0.5	0.5	0.6	0.6	0.7	0.7	0.8	0.8
	b	0.2	0.3	0.3	0.4	0.4	0.5	0.5	0.6	0.6	0.7
T=2500	ΔMSE_$\widehat{d}$	30.806	20.646	28.018	23.465	33.223	24.318	35.748	24.974	38.594	28.259
	ΔMSE_$\widehat{b}$	97.614	91.303	97.486	91.922	97.235	91.882	96.943	91.953	96.649	91.882
	ΔMSE_$\widehat{\beta}_1$	74.289	46.519	78.544	57.554	81.769	61.550	81.889	60.385	78.790	56.484
	ΔBIC	0.033	0.010	0.054	0.018	0.083	0.026	0.122	0.039	0.169	0.052
T=1000	ΔMSE_$\widehat{d}$	21.755	5.766	9.512	4.863	4.665	4.367	2.032	3.514	-2.684	3.653
	ΔMSE_$\widehat{b}$	63.532	31.243	68.572	41.770	71.218	51.119	72.550	57.629	73.576	61.261
	ΔMSE_$\widehat{\beta}_1$	60.822	21.316	55.964	20.111	55.852	30.822	59.140	35.361	60.087	34.734
	ΔBIC	0.037	0.013	0.058	0.018	0.085	0.027	0.122	0.038	0.165	0.050
T=500	ΔMSE_$\widehat{d}$	15.118	6.044	11.469	7.870	13.200	7.203	13.602	8.598	10.896	7.172
	ΔMSE_$\widehat{b}$	32.802	9.982	40.792	20.079	44.742	28.774	47.827	35.386	51.390	40.355
	ΔMSE_$\widehat{\beta}_1$	61.425	-12.731	18.449	0.729	35.713	12.398	42.114	16.948	44.861	19.982
	ΔBIC	0.050	0.013	0.069	0.018	0.096	0.025	0.126	0.034	0.167	0.044

Table 2(a): Monte Carlo Simulation Results. This table reports the MSE of the model estimates $\widehat{\lambda} = (\widehat{d}, \widehat{b}, \widehat{\beta}_1)$ and the BIC under the M-FCVAR model with and without zero restrictions on the α matrix. The Monte Carlo experiment is based on 10000 replications, with T=(2500, 1000, 500). The shock coming from the $I(0)$ variable produces only a transitory effect on the fractional variables in the DGP. In the simulations, β_1=-1. The values of d and b are provided in the table.

	M-FCVAR 1 (no zero restrictions)										**M-FCVAR 2 (zero restrictions)**									
d	0.4	0.4	0.5	0.5	0.6	0.6	0.7	0.7	0.8	0.8	0.4	0.4	0.5	0.5	0.6	0.6	0.7	0.7	0.8	0.8
b	0.2	0.3	0.3	0.4	0.4	0.5	0.5	0.6	0.6	0.7	0.2	0.3	0.3	0.4	0.4	0.5	0.5	0.6	0.6	0.7
T=2500																				
MSE_$\widehat{d}$ (1.0E-3)	0.401	0.250	0.253	0.243	0.244	0.255	0.247	0.246	0.250	0.227	0.208	0.204	0.208	0.204	0.202	0.205	0.210	0.207	0.206	0.186
MSE_$\widehat{b}$ (1.0E-3)	4.424	1.933	2.079	1.402	1.374	1.063	1.006	0.823	0.890	0.765	1.923	0.998	1.159	0.766	0.741	0.571	0.574	0.488	0.524	0.461
MSE_$\widehat{\beta}_1$ (1.0E-2)	2.330	0.289	0.299	0.112	0.106	0.042	0.041	0.016	0.015	0.005	0.910	0.289	0.298	0.112	0.107	0.042	0.041	0.016	0.015	0.005
BIC	24731	24731	24829	24710	24808	24827	24726	24726	24710	24737	24718	24717	24816	24697	24795	24814	24713	24713	24697	24724
T=1000																				
MSE_$\widehat{d}$ (1.0E-3)	1.471	0.660	0.688	0.641	0.631	0.618	0.646	0.652	0.623	0.643	1.351	0.647	0.681	0.637	0.624	0.615	0.644	0.648	0.621	0.640
MSE_$\widehat{b}$ (1.0E-3)	11.271	5.632	5.696	3.683	3.855	2.880	3.295	2.339	2.290	2.041	10.457	5.361	5.570	3.635	3.773	2.869	3.258	2.325	2.284	2.028
MSE_$\widehat{\beta}_1$ (1.0E-2)	28.075	0.794	2.627	0.317	0.323	0.133	0.135	0.053	0.055	0.021	10.781	0.783	1.812	0.317	0.322	0.132	0.134	0.053	0.055	0.021
BIC	9861	9960	9861	9849	9948	9849	9863	9849	9850	9850	9849	9948	9849	9837	9936	9837	9851	9837	9838	9838
T=500																				
MSE_$\widehat{d}$ (1.0E-3)	3.147	1.496	1.546	1.255	1.327	1.294	1.327	1.302	1.319	1.281	2.952	1.380	1.373	1.223	1.293	1.285	1.310	1.289	1.306	1.273
MSE_$\widehat{b}$ (1.0E-3)	19.059	12.056	12.046	7.754	7.971	6.242	6.028	5.042	5.146	4.486	18.407	11.788	11.138	7.577	7.745	6.024	5.812	4.990	5.096	4.438
MSE_$\widehat{\beta}_1$ (1.0E-2)	1266.757	6.806	92.716	0.795	0.816	0.318	0.307	0.139	0.134	0.059	49.170	4.370	3.165	0.785	0.804	0.311	0.307	0.139	0.132	0.059
BIC	4995	4894	4895	4998	4896	4897	4994	4893	4899	4996	4985	4883	4885	4988	4886	4886	4983	4882	4888	4986

Table 2(b): Percentage Gains of the MSE and BIC. The gains of the M-FCVAR with zero restrictions on the α matrix relative to that without zero restrictions are computed as
ΔMSE_$\widehat{\lambda}$ =[MSE($\widehat{\lambda}_{\text{M-FCVAR 1}}$)-MSE($\widehat{\lambda}_{\text{M-FCVAR 2}}$)]/MSE($\widehat{\lambda}_{\text{M-FCVAR 1}}$) and ΔBIC =[BIC(M-FCVAR 1)-BIC(M-FCVAR 2)]/BIC(M-FCVAR 1).

		Gains (%)									
	d	0.4	0.4	0.5	0.5	0.6	0.6	0.7	0.7	0.8	0.8
	b	0.2	0.3	0.3	0.4	0.4	0.5	0.5	0.6	0.6	0.7
T=2500	ΔMSE_$\widehat{d}$	48.116	18.354	17.639	16.170	17.544	19.279	14.975	15.882	17.558	17.872
	ΔMSE_$\widehat{b}$	56.546	48.341	44.226	45.406	46.123	46.288	42.967	40.706	41.089	39.798
	ΔMSE_$\widehat{\beta}_1$	60.942	0.014	0.361	-0.291	-0.223	0.401	0.018	0.353	0.229	0.857
	ΔBIC	0.053	0.054	0.054	0.054	0.054	0.054	0.055	0.055	0.054	0.055
T=1000	ΔMSE_$\widehat{d}$	8.185	1.880	1.024	0.711	1.097	0.514	0.284	0.539	0.328	0.477
	ΔMSE_$\widehat{b}$	7.225	4.814	2.219	1.316	2.134	0.386	1.131	0.570	0.258	0.640
	ΔMSE_$\widehat{\beta}_1$	61.601	1.277	31.041	0.185	0.328	0.134	0.322	0.039	0.185	0.015
	ΔBIC	0.120	0.118	0.118	0.120	0.119	0.120	0.119	0.120	0.120	0.120
T=500	ΔMSE_$\widehat{d}$	6.196	7.789	11.178	2.567	2.540	0.720	1.286	1.041	0.947	0.618
	ΔMSE_$\widehat{b}$	3.420	2.222	7.539	2.274	2.830	3.505	3.584	1.027	0.959	1.056
	ΔMSE_$\widehat{\beta}_1$	96.118	35.796	96.586	1.167	1.507	2.121	0.060	0.303	0.758	0.098
	ΔBIC	0.205	0.209	0.213	0.206	0.212	0.211	0.209	0.213	0.215	0.208

Table 1.3: This table reports standard summary statistics for the returns, monthly realized variance based on 5-minute intraday returns of S&P 500 and SPY as well as daily risk-neutral variance. In order to measure the return variation during the overnight period, we add the squared overnight return, computed as the squared close-to-open logarithmic price change, to the realized variance obtained over the trading day. RV_t is recorded daily but contains monthly (future) variance. The sample under analysis covers the period of Sep 22, 2003 to Nov 27, 2013, with a total of 2566 observations.

	Mean	StDev.	Autocorrelations					
			1	2	3	5	22	100
Return								
r_t^{SP500}	0.005	0.889	-0.133	-0.047	0.027	-0.050	0.075	0.017
r_t^{SPY}	0.007	1.023	-0.095	-0.084	0.031	-0.028	0.048	0.030
Variance								
VIX_t^2	3.339	0.773	0.985	0.974	0.964	0.948	0.846	0.570
RV_t^{SP500}	2.094	0.942	0.998	0.994	0.989	0.976	0.789	0.425
RV_t^{SPY}	2.476	0.898	0.998	0.994	0.989	0.976	0.786	0.427

Table 1.4: This table reports the estimates of fractional integration and co-integrating rank, for the cases of $(VIX_t^2,\ RV_t^{SP500})$ and (VIX_t^2, RV_t^{SPY}) as well as their counterparts using RV_{t-22}. Columns 2-4 summarize the results of the univariate exact local Whittle likelihood procedure of Shimotsu and Phillips (2005) and columns 5-7 report those of the rank determination procedure by Nielsen and Shimotsu (2007). Ave.$\widehat{d}$ is the simple average of $\widehat{d}$ of each pair of $(VIX_t^2,\ RV_t)$ and $(VIX_t^2,\ RV_{t-22})$; $\widehat{T}_0$ denotes the statistic of the modified Wald test for the null hypothesis that $\widehat{d}_{VIX_t^2} = \widehat{d}_{RV_t}$ or $\widehat{d}_{VIX_t^2} = \widehat{d}_{RV_{t-22}}$, which is calculated with $h(T) = 1/log(T)$. The value of the function $L(u)$ is calculated by employing a small $v(T) = m_1^{-0.4}$; $r = 1$ when $L(1) < L(0)$, suggesting the presence of co-fractional relation. The sample size is $T = 2566$.

	Fractional Integration Estimate ($m = T^{0.45}$)			Co-integration Rank Test ($m = T^{0.45}$, $m_1 = T^{0.4}$)		
	$\widehat{d}$	Ave. $\widehat{d}$	$\widehat{T}_0$	$L(0)$	$L(1)$	$\widehat{r}$
VIX_t^2	0.765					
RV_t^{SP500}	0.661	0.713	1.355	-1.429	-1.557	1
RV_t^{SPY}	0.645	0.705	1.822	-1.429	-1.560	1
RV_{t-22}^{SP500}	0.663	0.714	1.947	-1.429	-1.658	1
RV_{t-22}^{SPY}	0.646	0.705	2.656	-1.429	-1.658	1

Table 1.5: Estimation Results in an Empirical Example of S&P 500: only variances are considered. This table summarizes the maximum likelihood estimates, results of hypothesis tests as well as rank test of the FCVAR, which serves as a benchmark for comparison with Tables 1.6 and 1.7. Standard errors are in parentheses below the model estimates. The sample size is $T = 2566$.

Panel (A) FCVAR: $X_t = (RV_t^{SP500}, VIX_t^2)'$

$\Delta^d(X_t - \mu) = \alpha\beta'\Delta^{d-b}L_b(X_t - \mu) + \sum_{c=1}^{k}\Gamma_c\Delta^d L_b^c(X_t - \mu) + \varepsilon_t$

Hypothesis Test				Co-integration Rank Test	
	H_β	H_α^1	H_α^2	Rank (r)=0	Rank (r)=1
LR statistic	5.294	10.936	140.295	148.464	1.601
P value	0.021	0.001	0.000	0.000	0.493

BIC	d	b	α'	β'	Γ	μ'
-11791	0.694 (0.023)	0.580 (0.035)	$\begin{bmatrix} -0.021 & 0.150 \\ (0.009) & (0.036) \end{bmatrix}$	$\begin{bmatrix} 1 & -0.892 \end{bmatrix}$	$\begin{bmatrix} 1.120 & -0.129 \\ (0.055) & (0.019) \\ -0.284 & 0.255 \\ (0.092) & (0.046) \end{bmatrix}$	$\begin{bmatrix} 2.282 & 3.519 \\ (0.301) & (0.313) \end{bmatrix}$

Panel (B) FCVAR: $X_t = (RV_{t-22}^{SP500}, VIX_t^2)'$

Hypothesis Test				Co-integration Rank Test	
	H_β	H_α^1	H_α^2	Rank (r)=0	Rank (r)=1
LR statistic	3.720	106.655	0.891	129.656	2.835
P value	0.054	0.000	0.345	0.000	0.092

BIC	d	b	α'	β'	Γ	μ'
-11946	0.739 (0.022)	0.438 (0.035)	$\begin{bmatrix} -0.146 & 0.027 \\ (0.041) & (0.029) \end{bmatrix}$	$\begin{bmatrix} 1 & -0.880 \end{bmatrix}$	$\begin{bmatrix} 1.289 & 0.033 \\ (0.092) & (0.032) \\ 0.150 & 0.232 \\ (0.069) & (0.067) \end{bmatrix}$	$\begin{bmatrix} 2.017 & 3.246 \\ (0.232) & (0.229) \end{bmatrix}$

Table 1.6: Estimation Results in an Empirical Example of S&P 500 using (RV_t^{SP500}, VIX_t^2, r_t^{SP500}): This table presents the difference in model estimation between the FCVAR and M-FCVAR when the $I(0)$ returns are added. Panel (A) provides the hypothesis test, rank test and the estimation results of the FCVAR and Panel (B) corresponds to those of the M-FCVAR. Standard errors are in parentheses below the model estimates. The sample size is $T = 2566$.

Panel (A) FCVAR: $X_t = (RV_t^{SP500}, VIX_t^2, r_t^{SP500})'$

$\Delta^d(X_t - \mu) = \alpha\beta'\Delta^{d-b}L_b(X_t - \mu) + \sum_{c=1}^{k}\Gamma_c\Delta^d L_b^c(X_t - \mu) + \varepsilon_t$

Hypothesis Test			Co-integration Rank Test		
	H_β	H_δ	Rank=0	Rank=1	Rank=2
LR statistic	3.732	9.430	620.837	139.077	0.564
P value	0.053	0.009	0.000	0.000	0.816

BIC	d	b	α'	β'	Γ	μ'
-6178	0.656 (0.020)	0.655 (0.027)	$\begin{bmatrix} -0.009 & 0.088 & -0.294 \\ (0.004) & (0.017) & (0.070) \\ 0.006 & 0.018 & -1.357 \\ (0.013) & (0.037) & (0.517) \end{bmatrix}$	$\begin{bmatrix} 1 & -0.917 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.047 & -0.094 & -0.004 \\ (0.034) & (0.014) & (0.002) \\ -0.142 & 0.302 & -0.007 \\ (0.061) & (0.044) & (0.007) \\ 0.231 & 0.027 & 0.123 \\ (0.332) & (0.220) & (0.066) \end{bmatrix}$	$\begin{bmatrix} 2.186 & 3.416 & 0.001 \\ (0.261) & (0.267) & (0.011) \end{bmatrix}$

Panel (B) M-FCVAR: $X_t^* = (\Delta^{d-b}RV_t^{SP500}, \Delta^{d-b}VIX_t^2, r_t^{SP500})'$

$\Delta^b(X_t^* - \mu) = \alpha\beta' L_b(X_t^* - \mu) + \sum_{c=1}^{k}\Gamma_c\Delta^b L_b^c(X_t^* - \mu) + \varepsilon_t$

Hypothesis Test			Co-integration Rank Test		
	H_β	H_δ	Rank=0	Rank=1	Rank=2
LR statistic	5.411	11.876	235.769	146.105	1.177
P value	0.020	0.003	0.000	0.000	0.588

BIC	d	b	α'	β'	Γ	μ'
-6187	0.704 (0.021)	0.560 (0.018)	$\begin{bmatrix} -0.025 & 0.171 & -0.633 \\ (0.011) & (0.040) & (0.181) \\ 0.009 & 0.024 & -1.414 \\ (0.023) & (0.060) & (0.422) \end{bmatrix}$	$\begin{bmatrix} 1 & -0.886 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.141 & -0.125 & -0.006 \\ (0.046) & (0.017) & (0.003) \\ -0.323 & 0.307 & -0.012 \\ (0.076) & (0.053) & (0.009) \\ 1.029 & -0.217 & 0.136 \\ (0.516) & (0.260) & (0.069) \end{bmatrix}$	$\begin{bmatrix} 2.296 & 3.535 & 0.001 \\ (0.291) & (0.302) & (0.033) \end{bmatrix}$

Table 1.7: Estimation Results in an Empirical Example of S&P 500 using (RV_{t-22}^{SP500}, VIX_t^2, r_t^{SP500}): variances and returns are considered. This table presents the difference in model estimation between the FCVAR and M-FCVAR when the $I(0)$ returns are added. Panel (A) provides the hypothesis test, rank test and the estimation results of the FCVAR and Panel (B) corresponds to those of the M-FCVAR. Standard errors are in parentheses below the model estimates. The sample size is $T = 2566$.

Panel (A) FCVAR: $X_t = (RV_{t-22}^{SP500}, VIX_t^2, r_t^{SP500})'$

$\Delta^d(X_t - \mu) = \alpha\beta'\Delta^{d-b}L_b(X_t - \mu) + \sum_{c=1}^{k}\Gamma_c\Delta^d L_b^c(X_t - \mu) + \varepsilon_t$

Hypothesis Test			Co-integration Rank Test		
	H_β	$H_{\alpha\delta}$	Rank=0	Rank=1	Rank=2
LR statistic	1.447	5.440	707.904	105.400	0.483
P value	0.229	0.066	0.000	0.000	0.810

BIC	d	b	α'	β'	Γ	μ'
-6319	0.669 (0.020)	0.581 (0.039)	$\begin{bmatrix} -0.048 & 0.000 & 0.087 \\ (0.012) & (0.000) & (0.064) \\ -0.005 & 0.000 & -1.581 \\ (0.047) & (0.000) & (1.776) \end{bmatrix}$	$\begin{bmatrix} 1 & -0.938 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.058 & 0.071 & 0.001 \\ (0.046) & (0.017) & (0.003) \\ 0.188 & 0.295 & 0.004 \\ (0.043) & (0.047) & (0.005) \\ -0.680 & 0.210 & 0.207 \\ (0.282) & (0.269) & (0.105) \end{bmatrix}$	$\begin{bmatrix} 1.990 & 3.221 & 0.004 \\ (0.225) & (0.226) & (0.018) \end{bmatrix}$

Panel (B) M-FCVAR: $X_t^* = (\Delta^{d-b}RV_{t-22}^{SP500}, \Delta^{d-b}VIX_t^2, r_t^{SP500})'$

$\Delta^b(X_t^* - \mu) = \alpha\beta' L_b(X_t^* - \mu) + \sum_{c=1}^{k}\Gamma_c\Delta^b L_b^c(X_t^* - \mu) + \varepsilon_t$

Hypothesis Test			Co-integration Rank Test		
	H_β	$H_{\alpha\delta}$	Rank=0	Rank=1	Rank=2
LR statistic	3.992	5.100	39.907	143.179	0.150
P value	0.046	0.078	0.000	0.000	0.698

BIC	d	b	α'	β'	Γ	μ'
-6348	0.741 (0.018)	0.413 (0.016)	$\begin{bmatrix} -0.183 & 0.000 & 0.139 \\ (0.034) & (0.000) & (0.255) \\ -0.005 & 0.000 & -1.418 \\ (0.010) & (0.000) & (0.225) \end{bmatrix}$	$\begin{bmatrix} 1 & -0.864 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.378 & -0.012 & -0.002 \\ (0.061) & (0.028) & (0.006) \\ 0.187 & 0.232 & 0.007 \\ (0.053) & (0.068) & (0.007) \\ -1.328 & 0.214 & 0.077 \\ (0.509) & (0.438) & (0.133) \end{bmatrix}$	$\begin{bmatrix} 2.047 & 3.276 & 0.003 \\ (0.225) & (0.223) & (0.123) \end{bmatrix}$

Table 1.8: Estimation Results in an Empirical Example of SPY: only variances are considered. This table summarizes the maximum likelihood estimates, results of hypothesis tests as well as rank test of the FCVAR, which serves as a benchmark for comparison with Tables 1.9 and 1.10. Standard errors are in parentheses below the model estimates. The sample size is $T = 2566$.

Panel (A) FCVAR: $X_t = (RV_t^{SPY}, VIX_t^2)'$

$\Delta^d(X_t - \mu) = \alpha\beta'\Delta^{d-b}L_b(X_t - \mu) + \sum_{c=1}^{k}\Gamma_c\Delta^d L_b^c(X_t - \mu) + \varepsilon_t$

Hypothesis Test				Co-integration Rank Test	
	H_β	H_α^1	H_α^2	Rank (r)=0	Rank (r)=1
LR statistic	9.704	11.404	156.725	165.702	1.430
P value	0.002	0.001	0.000	0.000	0.627

BIC	d	b	α'	β'	Γ	μ'
-12069	0.692 (0.023)	0.590 (0.033)	$\begin{bmatrix} -0.021 & 0.161 \\ (0.009) & (0.036) \end{bmatrix}$	$\begin{bmatrix} 1 & -0.870 \end{bmatrix}$	$\begin{bmatrix} 1.117 & -0.122 \\ (0.052) & (0.017) \\ -0.345 & 0.255 \\ (0.095) & (0.046) \end{bmatrix}$	$\begin{bmatrix} 2.610 & 3.465 \\ (0.273) & (0.292) \end{bmatrix}$

Panel (B) FCVAR: $X_t = (RV_{t-22}^{SPY}, VIX_t^2)'$

Hypothesis Test				Co-integration Rank Test	
	H_β	H_α^1	H_α^2	Rank (r)=0	Rank (r)=1
LR statistic	6.243	117.051	0.803	139.252	2.682
P value	0.012	0.000	0.370	0.000	0.101

BIC	d	b	α'	β'	Γ	μ'
-12188	0.729 (0.022)	0.456 (0.033)	$\begin{bmatrix} -0.135 & 0.024 \\ (0.036) & (0.027) \end{bmatrix}$	$\begin{bmatrix} 1 & -0.866 \end{bmatrix}$	$\begin{bmatrix} 1.259 & 0.033 \\ (0.080) & (0.027) \\ 0.174 & 0.240 \\ (0.067) & (0.063) \end{bmatrix}$	$\begin{bmatrix} 2.416 & 3.261 \\ (0.221) & (0.227) \end{bmatrix}$

Table 1.9: Estimation Results in an Empirical Example of SPY using (RV_t^{SPY}, VIX_t^2, r_t^{SPY}): variances and returns are considered. This table presents the difference in model estimation between the FCVAR and M-FCVAR when the $I(0)$ returns are added. Panel (A) provides the hypothesis test, rank test and the estimation results of the FCVAR and Panel (B) corresponds to those of the M-FCVAR. Standard errors are in parentheses below the model estimates. The sample size is $T = 2566$.

Panel (A) FCVAR: $X_t = (RV_t^{SPY}, VIX_t^2, r_t^{SPY})'$

$\Delta^d(X_t - \mu) = \alpha\beta'\Delta^{d-b}L_b(X_t - \mu) + \sum_{c=1}^{k}\Gamma_c\Delta^d L_b^c(X_t - \mu) + \varepsilon_t$

Hypothesis Test			Co-integration Rank Test		
	H_β	H_δ	Rank=0	Rank=1	Rank=2
LR statistic	7.831	6.764	693.835	158.434	0.716
P value	0.005	0.034	0.000	0.000	0.766

BIC	d	b	α'	β'	Γ	μ'
-5907	0.664 (0.020)	0.648 (0.027)	$\begin{bmatrix} -0.011 & 0.109 & -0.476 \\ (0.004) & (0.019) & (0.092) \\ 0.005 & 0.012 & -1.366 \\ (0.020) & (0.056) & (0.865) \end{bmatrix}$	$\begin{bmatrix} 1 & -0.894 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.058 & -0.094 & -0.003 \\ (0.037) & (0.014) & (0.002) \\ -0.229 & 0.282 & -0.006 \\ (0.068) & (0.046) & (0.006) \\ 0.716 & 0.453 & 0.200 \\ (0.430) & (0.262) & (0.068) \end{bmatrix}$	$\begin{bmatrix} 2.464 & 3.309 & 0.004 \\ (0.243) & (0.257) & (0.014) \end{bmatrix}$

Panel (B) M-FCVAR: $X_t^* = (\Delta^{d-b}RV_t^{SPY}, \Delta^{d-b}VIX_t^2, r_t^{SPY})'$

$\Delta^b(X_t^* - \mu) = \alpha\beta' L_b(X_t^* - \mu) + \sum_{c=1}^{k}\Gamma_c\Delta^b L_b^c(X_t^* - \mu) + \varepsilon_t$

Hypothesis Test			Co-integration Rank Test		
	H_β	H_δ	Rank=0	Rank=1	Rank=2
LR statistic	9.946	9.499	257.789	72.685	1.291
P value	0.000	0.000	0.000	0.000	0.562

BIC	d	b	α'	β'	Γ	μ'
-5915	0.703 (0.022)	0.568 (0.018)	$\begin{bmatrix} -0.025 & 0.186 & -0.876 \\ (0.010 & (0.041) & (0.226) \\ 0.007 & 0.016 & -1.383 \\ (0.030) & (0.081) & (0.667) \end{bmatrix}$	$\begin{bmatrix} 1 & -0.862 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.138 & -0.119 & -0.006 \\ (0.045) & (0.016) & (0.003) \\ -0.396 & 0.290 & -0.010 \\ (0.079) & (0.054) & (0.008) \\ 1.686 & 0.215 & 0.211 \\ (0.616) & (0.300) & (0.067) \end{bmatrix}$	$\begin{bmatrix} 2.566 & 3.419 & 0.004 \\ (0.263) & (0.283) & (0.034) \end{bmatrix}$

Table 1.10: Estimation Results in an Empirical Example of SPY using (RV_{t-22}^{SPY}, VIX_t^2, r_t^{SPY}): variances and returns are considered. This table presents the difference in model estimation between the FCVAR and M-FCVAR when the $I(0)$ returns are added. Panel (A) provides the hypothesis test, rank test and the estimation results of the FCVAR and Panel (B) corresponds to those of the M-FCVAR. Standard errors are in parentheses below the model estimates. The sample size is $T = 2566$.

Panel (A) FCVAR: $X_t = (RV_{t-22}^{SPY}, VIX_t^2, r_t^{SPY})'$

$\Delta^d(X_t-\mu) = \alpha\beta'\Delta^{d-b}L_b(X_t-\mu)+\sum_{c=1}^{k}\Gamma_c\Delta^d L_b^c(X_t-\mu)+\varepsilon_t$

Hypothesis Test			Co-integration Rank Test		
	H_β	$H_{\alpha\delta}$	Rank=0	Rank=1	Rank=2
LR statistic	3.916	2.365	748.377	119.361	0.761
P value	0.048	0.306	0.000	0.000	0.704

BIC	d	b	α'	β'	Γ	μ'
-6019	0.682 (0.018)	0.553 (0.033)	$\begin{bmatrix} -0.066 & 0.000 & 0.087 \\ (0.015) & (0.000) & (0.090) \\ -0.002 & 0.000 & -1.732 \\ (0.160) & (0.000) & (7.132) \end{bmatrix}$	$\begin{bmatrix} 1 & -0.911 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.098 & 0.060 & 0.000 \\ (0.048) & (0.018) & (0.003) \\ 0.212 & 0.270 & 0.000 \\ (0.041) & (0.049) & (0.004) \\ -0.534 & 0.579 & 0.362 \\ (0.363) & (0.339) & (0.115) \end{bmatrix}$	$\begin{bmatrix} 2.359 & 3.204 & 0.006 \\ (0.218) & (0.225) & (0.024) \end{bmatrix}$

Panel (B) M-FCVAR: $X_t^* = (\Delta^{d-b}RV_{t-22}^{SPY}, \Delta^{d-b}VIX_t^2, r_t^{SPY})'$

$\Delta^b(X_t^*-\mu) = \alpha\beta'L_b(X_t^*-\mu)+\sum_{c=1}^{k}\Gamma_c\Delta^b L_b^c(X_t^*-\mu)+\varepsilon_t$

Hypothesis Test			Co-integration Rank Test		
	H_β	$H_{\alpha\delta}$	Rank=0	Rank=1	Rank=2
LR statistic	5.825	2.512	289.766	18.663	1.796
P value	0.016	0.285	0.000	0.000	0.180

BIC	d	b	α'	β'	Γ	μ'
-6034	0.723 (0.017)	0.460 (0.015)	$\begin{bmatrix} -0.133 & 0.000 & 0.109 \\ (0.029) & (0.000) & (0.158) \\ -0.005 & 0.000 & -1.372 \\ (0.151) & (0.000) & (3.042) \end{bmatrix}$	$\begin{bmatrix} 1 & -0.870 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.258 & 0.015 & 0.001 \\ (0.043) & (0.022) & (0.004) \\ 0.215 & 0.223 & -0.000 \\ (0.050) & (0.061) & (0.005) \\ -1.104 & 0.641 & 0.189 \\ (0.500) & (0.441) & (0.092) \end{bmatrix}$	$\begin{bmatrix} 2.389 & 3.234 & 0.005 \\ (0.217) & (0.223) & (0.086) \end{bmatrix}$

Figure 1.1: Sample Periodograms for (VIX_t^2, RV_t^{SP500} and RV_t^{SPY}). Only the flat regions are picked and considered in the procedure of the exact local Whittle estimation. The sample size is $T = 2566$.

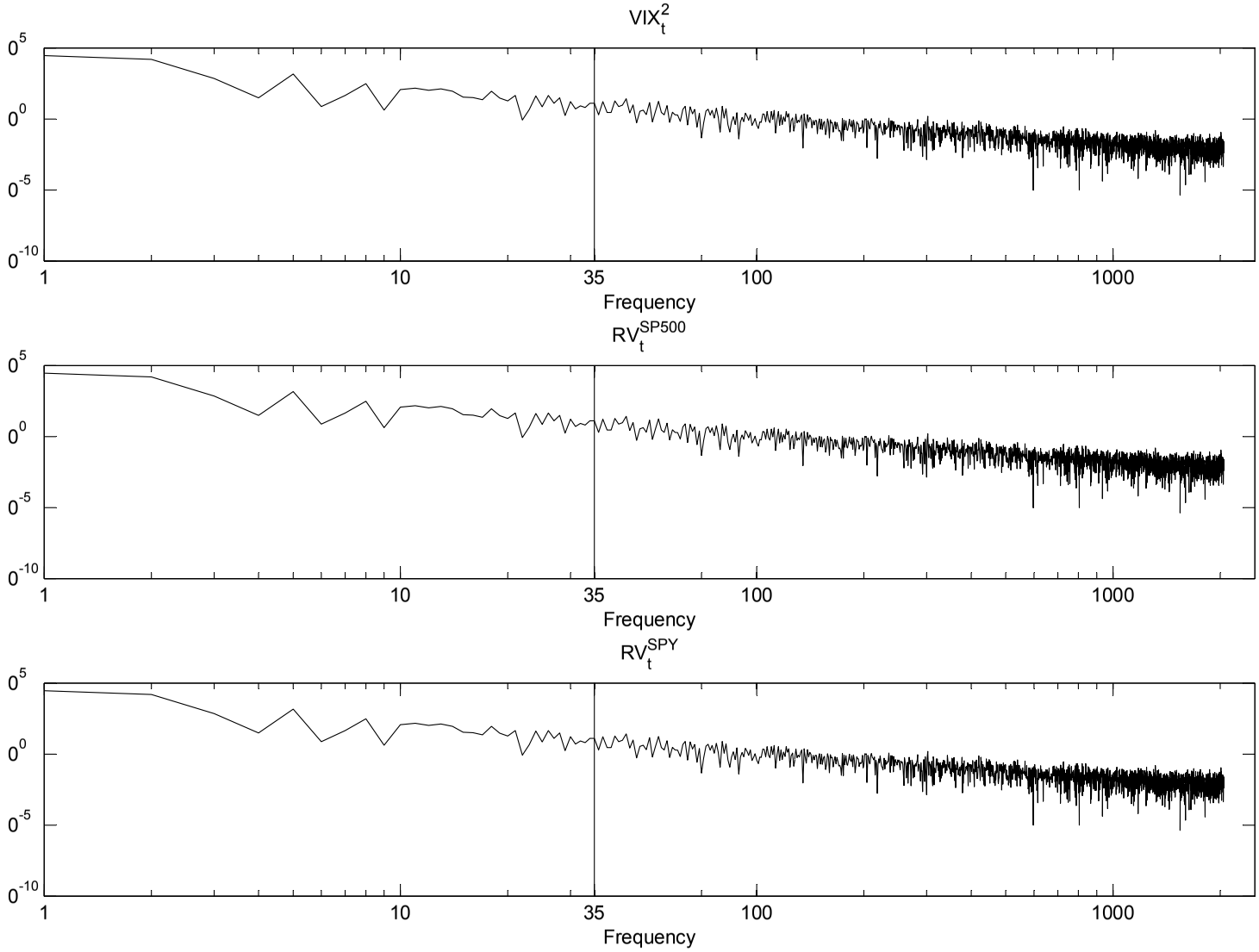


Figure 1.2: Impulse Response Functions for Long-memory Variables $\left(VIX_t^2,\ RV_t^{SP500}\right)$: shocks are applied with the permanent and transitory decomposition by Gonzalo and Granger (1995) and are further orthogonalized using the method in Gonzalo and Ng (2001). The figures plot the FCVAR and M-FCVAR model-implied impulse response functions (IRF) for RV_t^{SP500} (solid line) and VIX_t^2 (dashed line), associated with the common permanent component (first column) and two transitory components (second and third column). Model estimates are obtained from Table 1.6. The sample size is $T = 2566$.

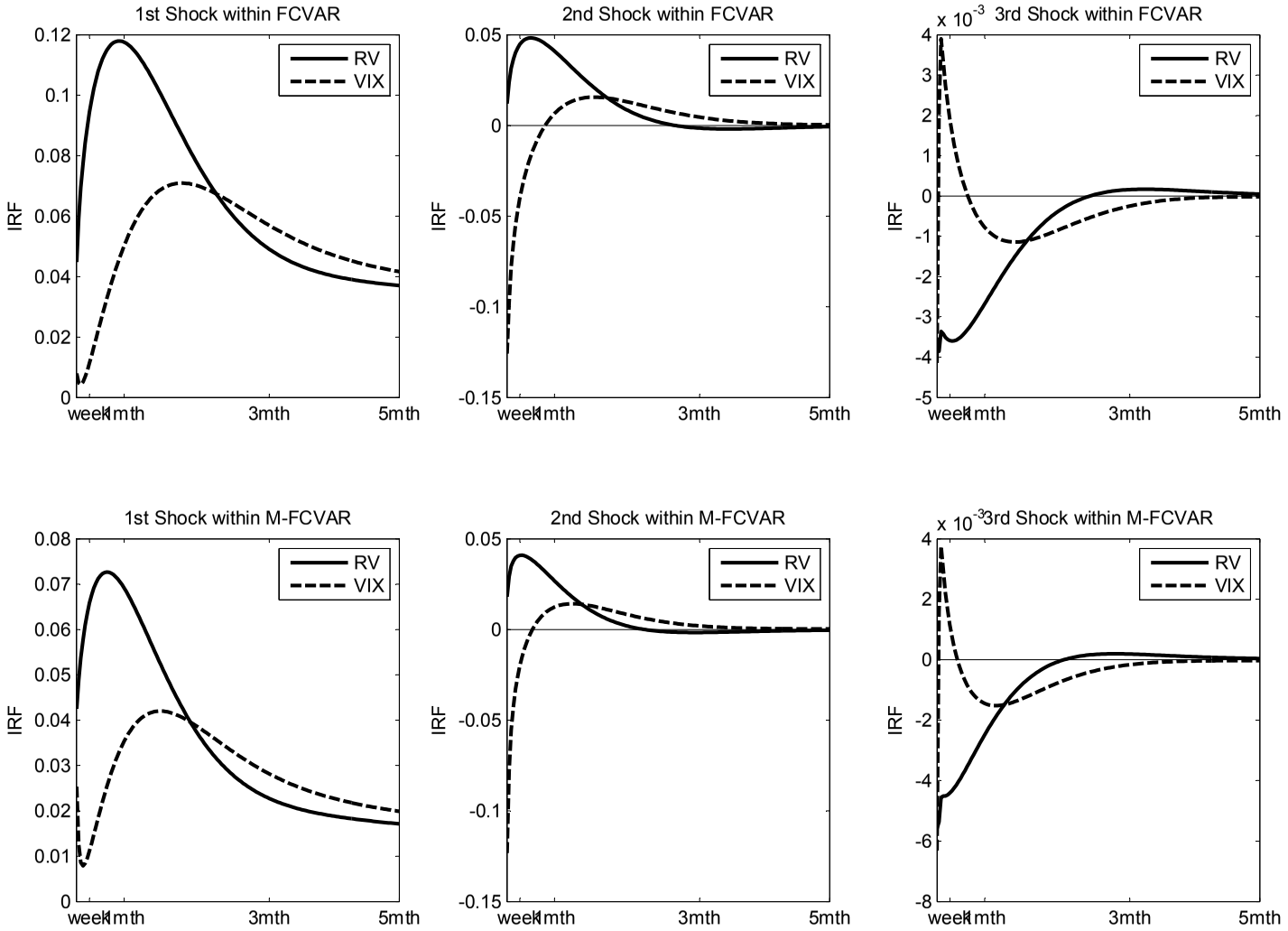


Figure 1.3: Impulse Response Functions for Long-memory Variables $\left(VIX_t^2,\ RV_t^{SP500}\right)$: shocks are taken from the model residuals directly. The figures plot the FCVAR and M-FCVAR model-implied impulse response functions (IRF) for RV_t^{SP500} (solid line) and VIX_t^2 (dashed line), associated with the RV_t^{SP500} (first column), VIX_t^2 (second column) and r_t^{SP500} (third column). Model estimates are obtained from Table 1.6. The sample size is $T = 2566$.

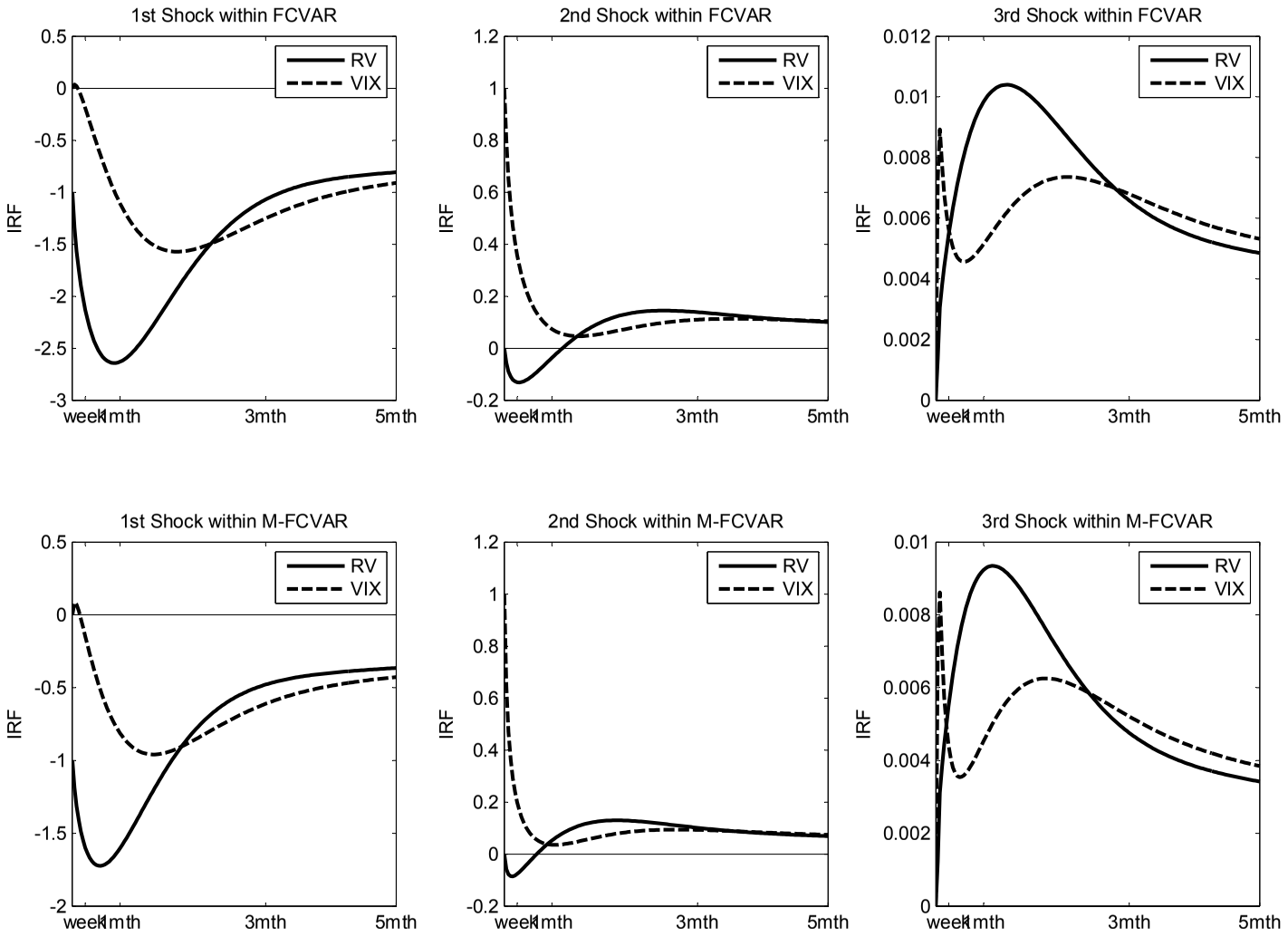


Figure 1.4: Impulse Response Functions for Long-memory Variables $\left(VIX_t^2,\ RV_{t-22}^{SP500}\right)$: shocks are applied with the permanent and transitory decomposition by Gonzalo and Granger (1995) and are further orthogonalized using the method in Gonzalo and Ng (2001). The figures plot the FCVAR and M-FCVAR model-implied impulse response functions (IRF) for RV_{t-22}^{SP500} (solid line) and VIX_t^2 (dashed line), associated with the common permanent component (first column) and two transitory components (second and third column). Model estimates are obtained from Table 1.7. The sample size is $T = 2566$.

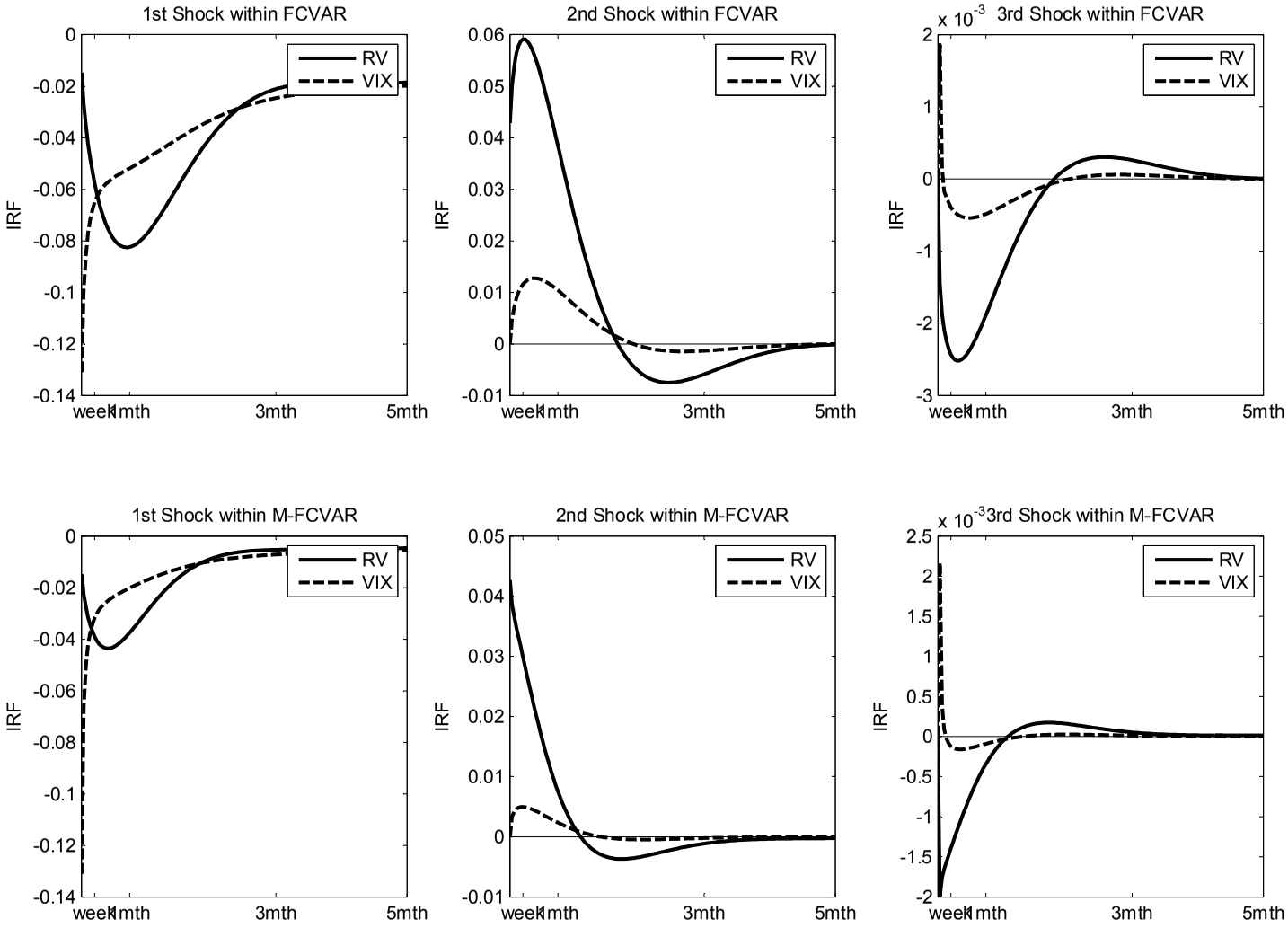


Figure 1.5: Impulse Response Functions for Long-memory Variables $\left(VIX_t^2,\ RV_{t-22}^{SP500}\right)$: shocks are taken from the model residuals directly. The figures plot the FCVAR and M-FCVAR model-implied impulse response functions (IRF) for RV_{t-22}^{SP500} (solid line) and VIX_t^2 (dashed line), associated with the RV_{t-22}^{SP500} (first column), VIX_t^2 (second column) and r_t^{SP500} (third column). Model estimates are obtained from Table 1.7. The sample size is $T = 2566$.

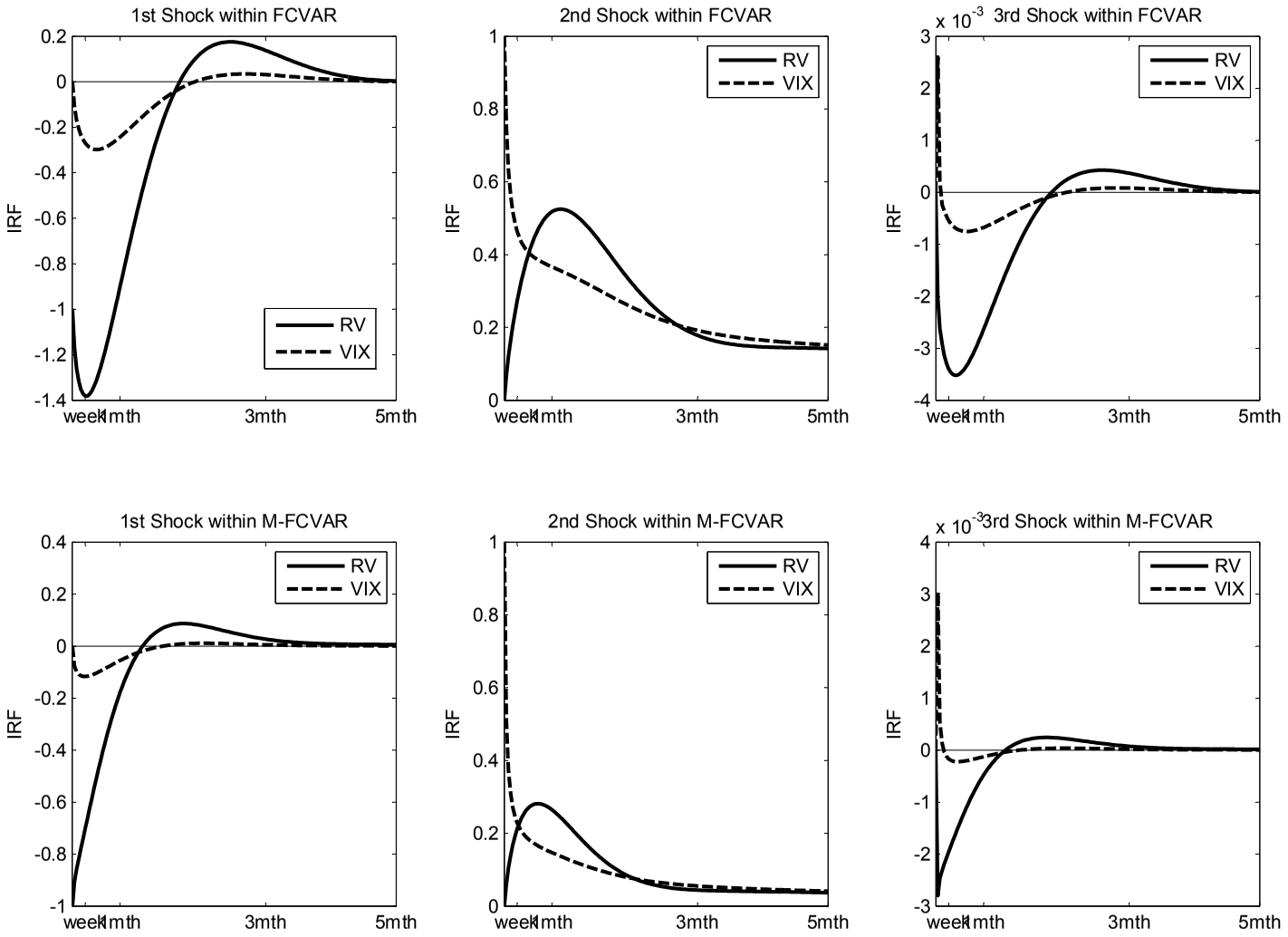


Figure 1.6: Return Predictability for S&P 500. Figures plot predictive R^2 implied by the M-FCVAR (solid line), FCVAR (dashed line), long-memory adjusted VAR (dash-dot line) and AR (dotted line) for returns over h horizons. The sample size is $T = 2566$.

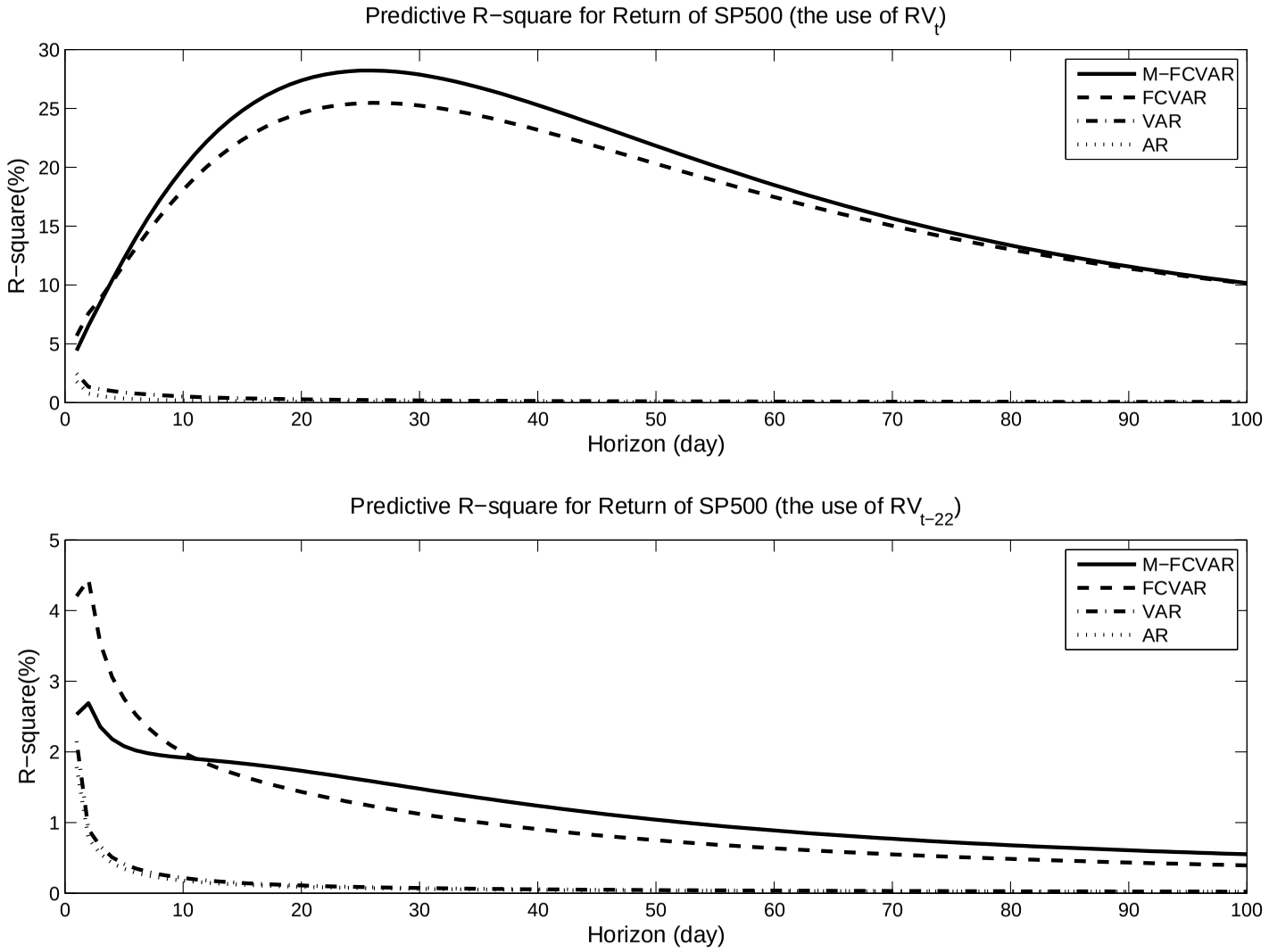
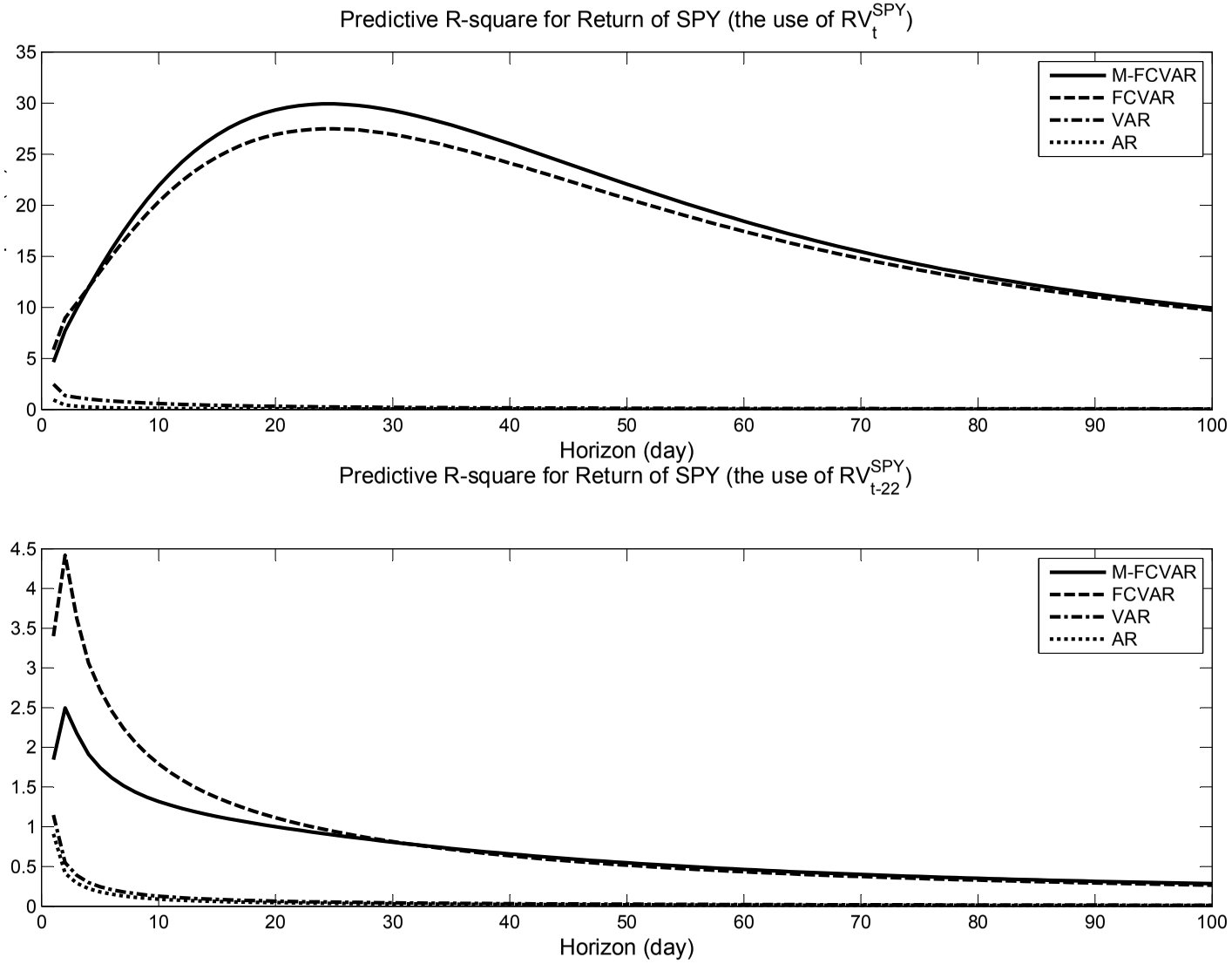


Figure 1.7: Return Predictability for SPY. Figures plot predictive R^2 implied by the M-FCVAR (solid line), FCVAR (dashed line), long-memory adjusted VAR (dash-dot line) and AR (dotted line) for returns over h horizons. The sample size is $T = 2566$.



Chapter 2

Forecasting Using Alternative Measures of Model-Free Option-Implied Volatility

2.1 Introduction

In an efficient market, the option price embodies all available useful information about future movements of the underlying asset. Hence, traders and hedge fund managers are primarily interested in option-implied volatility when making financial decisions. As a natural forecast of return variation over the remaining life of the relevant option, option-implied volatility has been frequently used in forecasting future volatility, see Poon and Granger (2003) for an extensive review of the studies on this topic. As opposed to the Black-Scholes (BS) implied volatility, model-free option-implied volatilities have gained substantial popularity because, relying upon no particular parametric model, they avoid potential mis-specification problems. See, for example, Britten-Jones and Neuberger (2000), Carr and Wu (2006) and Taylor, Yadav, and Zhang (2010).

One of the most widely used measures of model-free option-implied volatility is the VIX volatility index, disseminated by the Chicago Board of Options Exchange (CBOE). The VIX provides a measure of the expected value of the S&P 500 return variation under the risk-neutral measure and is designed to closely mimic the model-free implied volatility ($MFIV$). Derived by Britten-Jones and Neuberger (2000), the $MFIV$ is defined as an integral of cross-section of out-of-the money (OTM) European style put and call options over an infinite range of strikes for the given maturity. Jiang and Tian (2005) show that the $MFIV$ is a more efficient forecast for future realized volatility than the BS implied volatility and the historical realized volatility. However, Andersen and Bondarenko (2007) argue that the $MFIV$ and VIX are biased forecasts of future volatility since they contain non-trivial and time-varying risk premiums. As a more important part of their empirical study, Andersen and Bondarenko (2007) investigate the properties of the corridor implied volatility index (CX), which is obtained from the $MFIV$ by truncating the integration domain between two barriers. Being less sensitive to variation in the market variance risk premium, the CX with the narrowest

corridor width is found to dominate other implied volatility measures in the work of Andersen and Bondarenko (2007). Another advantage of the CX is that it is constructed only over intervals of the risk-neutral density (RND) where price quotes are directly observable. By contrast, the computation requirements for deriving the $MFIV$ are not satisfied by the existing data as options are traded only over a finite range of strikes. Andersen, Bondarenko, and Gonzalez-Perez (2015) further improve the construction of the CX by adopting the concept of an invariant coverage across time, which ensures that the CX is coherent in the time series dimension. As compared with the VIX, which is based upon strongly time-varying coverage of the tails of the RND, the CX uses a consistent range of strikes, which serves as a more accurate volatility indicator over time.

In addition to the use of implied volatilities in forecasting future volatility, prior studies also indicate that the VIX may carry some predictive power for future returns on stock market indices. For example, Giot (2005) finds that future returns are always positive (negative) for very high (low) levels of the VIX. This accords with the work of Guo and Whitelaw (2006) who provide evidence for the positive relationship between market returns and implied volatilities. The positive relationship between the VIX and future returns is also documented in Banerjee, Doran, and Peterson (2007) who suggest that both levels and innovations of the VIX are significantly related to future returns. That finding is indicative of a negative volatility risk premium, which is consistent with Ang et al. (2006) where stocks with high past sensitivities to the innovation in the VIX display on average future decreasing returns. The evidence that the VIX is a priced risk factor in the time series of returns helps to explain why the VIX may exhibit predictive power for future returns. Although a substantial empirical literature is devoted to the investigation of risk-return relations (see, e.g., the discussion in Rossi and Timmermann (2010), and the many references therein), most rely on the VIX as a directly observable proxy for risk. Other measures of model-free option-implied volatility are rarely considered.

In spite of the increasing popularity of the VIX index, measurement errors in its construction has been noted by Jiang and Tian (2005). The common problem inherent in the computation of the VIX as well as other measures of model-free implied volatility is that only a discrete set of strikes is actually traded in the market and that very low and high strikes are usually absent. To account for measurement errors induced by the limited number of strikes, Jiang and Tian (2005) apply the cubic spline method to interpolate between existing strikes and exploit a flat extrapolation scheme to infer option prices beyond the truncation point. Andersen and Bondarenko (2007) address the issue induced by the discrete set of strikes via the positive convolution approximation method proposed by Bondarenko (2003). Although interpolation and extrapolation techniques are widely accepted, it remains unclear how such techniques affect the performance of implied volatilities in predicting future returns and realized volatility. In addition, there appears to be no consensus on the roles played by the OTM call and put options in the forecast of future volatility and returns. Jackwerth (2000), Jones (2006) and Bates (2008) suggest that the OTM put options may be irrelevant to known risk factors affecting stock returns. Using a cubic spline interpolation and flat extrapolation methods, Dotsis and Vlastakis (2016) also find that the OTM put options, especially deep OTM puts, do not contain important information with respect to equity volatility risk. They also show that the OTM call options subsume all useful information embedded in the OTM puts for forecasting future realized volatility. However, Andersen, Fusari, and Todorov (2015) show that the left tail risk, driving a substantial part of the OTM put option dynamics, exhibits strong predictive power for future excess market returns over long horizons.

Against this background, this study examines the performance of model-free option-implied volatilities in predicting future returns and volatility and contribute to the existing literature in the following ways. First, this chapter is among the first to provide simulation evidence to justify the use of the interpolation/extrapolation procedure for better forecasting performance of implied volatilities. The usefulness

of this procedure is verified in both the simulation and empirical studies. The adoption of a stochastic volatility model with both jumps and volatility risk premium in the present study mimics more closely the observed data dynamics. This can be seen as an extension of the work of Zhang, Taylor, and Wang (2013) where a simple square-root model of Cox, Ingersoll, and Ross (1985) is employed to investigate the number of options upon the information content of the $MFIV$ in an in-sample analysis. Distinct from Zhang, Taylor, and Wang (2013), this chapter conducts comprehensive out-of-sample (OOS) volatility forecasts made by different implied volatility measures including the $MFIV$.

Second, to ascertain the relevance of the OTM call and put options, this chapter considers implied volatility measures constructed entirely from the cross-section of OTM put (call) options and measures which discard the deep OTM put (call) options. This is achieved by splitting the $MFIV$ into different components with the use of different intervals of the cross-section of OTM put and call option prices. Similar constructions of implied volatilities are conducted in Dotsis and Vlastakis (2016) who examine the price of volatility risk in the cross-section of stock returns. With a different focus from that of Dotsis and Vlastakis (2016), the present chapter compares the fraction of the time-series variation in future returns that are explained by various measures of implied volatility. Return predictability provided by implied volatilities is investigated in the pre- and post-crisis periods, respectively. The impact of the recent financial crisis is accounted for since the crisis represents an informative period during which uncertainty and risk aversion may have been more evident than the non-crisis period, see Hilal, Poon, and Tawn (2011) and Bates (2012).

A preview of the main findings of this study is as follows. Simulation results show that, with a wider range of strikes upon which model-free option-implied volatilities are based, the OOS volatility forecast becomes more accurate while returns tend to be less predictable. In addition, a finer partition of strikes usually leads to greater predictive power of implied volatilities for future returns. These

findings warrant the application of an interpolation and extrapolation scheme in the practice of volatility forecast and an interpolation method only in return predictions. In the empirical study using SPX options from 2003 to 2013, the aforementioned procedure, i.e. interpolation/extrapolation methods, significantly improves the performance of different measures of implied volatility considered in the OOS volatility forecast and gives rise to higher return predictability for most measures in the post-crisis period. With the use of this procedure, the SPX OTM call options substantially dominate the OTM put options with regard to their forecasting performance. The empirical findings outlined above are supported by the simulation evidence. However, when measures of implied volatility are derived from the listed options only, the superiority of the OTM put options over the OTM call options is noted in volatility forecast and post-crisis return predictions.

The rest of this chapter is organized as follows. Section 2.2 provides the construction of various model-free option-implied volatility measures and realized volatility considered in this study. Section 2.3 outlines the techniques adopted to address measurement errors in the construction of various implied volatilities. Section 2.4 presents the design and settings of the Monte Carlo study along with the results. Section 2.5 describes the data and section 2.6 reports the empirical results. Conclusion is provided in section 2.7.

2.2 Construction of Volatility Measures

This section provides an outline of the construction of various measures of volatility. Section 2.2.1 gives an introduction of the $MFIV$ and its components derived from OTM calls and OTM puts, respectively. The VIX index is then reviewed as a close approximation of the $MFIV$. Section 2.2.2 discusses the computation of model-free corridor implied volatilities where three different segments of the cross-section of OTM put and call option prices are adopted. Finally, in section 2.2.3, the high-frequency realized volatility is defined, which is used to obtain an accurate measure of the ex-post return variation of the underlying asset.

2.2.1 Model-Free Implied Volatility and VIX

The concept of the $MFIV$ is derived by Britten-Jones and Neuberger (2000). Its computation for a given maturity involves market prices for a continuum of European-style options with strikes from zero to infinity, which takes the form

$$MFIV = \sqrt{\frac{2}{\tau}e^{r\tau}\left[\int_0^F \frac{P(\tau,K)}{K^2}dK + \int_F^\infty \frac{C(\tau,K)}{K^2}dK\right]} \tag{2.1}$$

where r is the annualized risk-free interest rate as measured by the corresponding U.S. Treasury bill rate, τ is time-to-maturity measured in annual units, F is the forward price for transaction at maturity τ, $P(\tau,K)$ and $C(\tau,K)$ are the mid-quotes for European put and call options with strike price K and maturity τ. By construction, only OTM options (call if $K > F$ and put otherwise) are taken into account. For a different purpose, Demeterfi et al. (1999) develop the idea of fair value of future volatility, represented by V_{ddkz}, upon which the VIX is based. V_{ddkz} is also independent of any option pricing model and can be extracted from option prices directly such as

$$V_{ddkz} = \sqrt{\frac{2}{\tau}\left\{r\tau - \left[\frac{S_0}{S_*}e^{r\tau} - 1\right] - \ln(S_*/S_0) + e^{r\tau}\int_0^{S_*}\frac{P(\tau,K)}{K^2}dK + e^{r\tau}\int_{S_*}^\infty \frac{C(\tau,K)}{K^2}dK\right\}} \tag{2.2}$$

where S_0 is the current asset price and S_* denotes the stock price close to the forward price. Jiang and Tian (2007) demonstrate that the V_{ddkz} is conceptually equivalent to the $MFIV$.

Motivated by Dotsis and Vlastakis (2016), the $MFIV$ can be further divided into two components; i.e., that from the OTM call options (VC) and that from the OTM put options (VP), which are given by

$$VC = \sqrt{\frac{2}{\tau}e^{r\tau}\int_F^\infty \frac{C(\tau,K)}{K^2}dK} \tag{2.3}$$

and

$$VP = \sqrt{\frac{2}{\tau}e^{r\tau}\int_0^F \frac{P(\tau,K)}{K^2}dK} \tag{2.4}$$

where $MFIV^2 = VC^2 + VP^2$.

The VIX index is based on the idea of fair value of future volatility developed by Demeterfi et al. (1999). The general formula for computation of the VIX index is given by

$$VIX = \sqrt{\frac{2}{\tau}\sum_i \frac{\Delta K_i}{K_i^2}e^{r\tau}Q(\tau,K_i) - \frac{1}{\tau}(\frac{F^*}{K_0}-1)^2} \tag{2.5}$$

where $\tau = 30/365$ is the option maturity, K_i is the strike price of the ith OTM option in the calculation, K_0 is the first strike price below the forward index level $F^*(K_0 \leq F^*)$, $Q(\tau,\ K_i)$ is the midpoint of the latest available bid and ask prices of the OTM option at strike K_i, and ΔK_i stands for the strike price interval as $\Delta K_i = (K_{i+1} - K_{i-1})/2$. The forward price, F^*, is calculated from at-the-money options according to put-call parity, $F^* = K_* + e^{r\tau}[C(K_*,\tau) - P(K_*,\tau)]$ and K_* is determined as the strike price for which the difference between the call and put prices is minimal. It is worth noting that, at the boundaries of strike prices, ΔK_i is adjusted as the difference between the two highest (or lowest) strike prices. In addition, at the strike price K_0, the option price $Q_i(\tau,K_i)$ is modified to be the average of call and put prices. The CBOE computes the VIX from an interpolation of two volatility indices with respect to two different maturities: τ_t^l and τ_t^u. The VIX index is finally obtained by taking a weighted average of these two VIX measures based on τ_t^l and τ_t^u

$$VIX = 100 \times \sqrt{\left[w_1(VIX_t^2(\tau_t^l)\tau_t^l) + w_2(VIX_t^2(\tau_t^u)\tau_t^u)\right] \times \frac{365}{30}} \tag{2.6}$$

where $w_1 = \frac{\tau_t^u - \tau}{\tau_t^u - \tau_t^l}$ and $w_2 = \frac{\tau - \tau_t^l}{\tau_t^u - \tau_t^l}$ so that $w_1 + w_2 = 1$.

Recall equation (2.2), we obtain the approximation below by applying Taylor's expansion of log function and discarding terms greater than the second order of

moments[1].

$$
\begin{aligned}
r\tau - \left(\frac{S_0}{S_*}e^{r\tau} - 1\right) - \ln(S_*/S_0) &= r\tau - \left(\frac{F}{K_0} - 1\right) - \ln(K_0/S_0) \\
&= \ln(F/K_0) - \left(\frac{F}{K_0} - 1\right) \\
&\approx -\frac{1}{2}\left(\frac{F}{K_0} - 1\right)^2
\end{aligned} \tag{2.7}
$$

where the forward price (F) at time τ is equal to $S_0 e^{r\tau}$ and S_* is chosen as the first strike price (K_0) below the forward price. Replacing the term $r\tau - \left(\frac{S_0}{S_*}e^{r\tau} - 1\right) - \ln(S_*/S_0)$ with $-\frac{1}{2}\left(\frac{F}{K_0} - 1\right)^2$, the discrete framework of equation (2.2) is identical to the general formula for the VIX calculation in equation (2.5).

The VIX in equation (2.5) may give rise to different approximation errors such as truncation, discretization, expansion and interpolation errors, see details in Jiang and Tian (2005) and Jiang and Tian (2007). Here, we consider truncation and discretization errors only since the others are widely regarded as negligible and are unlikely to have any material impact on the forecasting performance of implied volatilities.

The truncation error is due to the fact that very low or very high strike prices are not available in practice. Let K_L and K_U denote the lowest and highest strikes for a certain maturity, respectively. The infinite range of strike prices in equation (2.2) is approximated by the CBOE with a finite range $[K_L,\ K_U]$ as follows

$$
\int_0^{K_0} \frac{P(\tau,K)}{K^2}dK + \int_{K_0}^{\infty} \frac{C(\tau,K)}{K^2}dK \approx \int_{K_L}^{K_0} \frac{P(\tau,K)}{K^2}dK + \int_{K_0}^{K_U} \frac{C(\tau,K)}{K^2}dK \tag{2.8}
$$

The magnitude of the truncation error is measured as

$$
\delta_{trunc} = -\frac{2}{\tau}e^{r\tau}\left[\int_0^{K_L} \frac{P(\tau,K)}{K^2}dK + \int_{K_U}^{\infty} \frac{C(\tau,K)}{K^2}dK\right] \tag{2.9}
$$

[1]The construction of the VIX index is based on the assumption that the stochastic process of asset returns follow the Geometric Brownian Motion (GBM) and Ito's Lemma (IL), where both the GBM and IL assume returns to be continuous and symmetrically distributed and therefore the moments higher than the variance neither exist nor generate effects on the return process.

and such type of error gives rise to a downward bias in the computed volatility. It is noted by Jiang and Tian (2007) that the δ_{trunc} may vary significantly over time as the interval $[K_L,\ K_U]$ is not fixed. The reasons are twofold. First, the CBOE usually introduces new strikes when the underlying index is outside the range of listed strikes. Second, the CBOE utilizes filters to exclude the potentially problematic options in the VIX computation procedure.

The other type of error of interest is discretization error induced by the numerical integration conducted using a coarse grid of available strike prices. The CBOE computes the integrals in equation (2.2) as follows

$$\int_{K_L}^{K_0} \frac{P(\tau,K)}{K^2}dK + \int_{K_0}^{K_U} \frac{C(\tau,K)}{K^2}dK \approx \sum_i \frac{\Delta K_i}{K_i^2} Q(\tau,K_i) \tag{2.10}$$

The magnitude of the discretization error is

$$\delta_{disc} = \frac{2}{\tau}e^{r\tau}\left\{\sum_i \frac{\Delta K_i}{K_i^2}Q(\tau,K_i) - \left[\int_{K_L}^{K_0} \frac{P(\tau,K)}{K^2}dK + \int_{K_0}^{K_U} \frac{C(\tau,K)}{K^2}dK\right]\right\} \tag{2.11}$$

Jiang and Tian (2007) provide evidence that the δ_{disc} may lead to an overestimation of the underlying volatility. In practice, the model-free option-implied volatility measures are all subject to these measurement errors to some degree, due to the limited availability of market prices for a continuum of European-style options with strikes from zero to infinity.

2.2.2 Corridor Implied Volatility

The corridor implied volatility index (CX) is initially analyzed in the empirical work of Andersen and Bondarenko (2007). Unlike the $MFIV$, which requires the availability of options with strikes from zero to infinity, the CX only captures volatility over a certain segment of the underlying RND. For a fixed coverage $[B_L,$

$B_H]$, $0 \leq B_L \leq B_H \leq \infty$, the CX is computed as

$$CX = \sqrt{\frac{2e^{r\tau}}{\tau} \int_{B_L}^{B_H} \frac{M(K)}{K^2} dK} \tag{2.12}$$

where the time to maturity $\tau = 30$ days and $M(K)$ stands for the minimum of the put and call prices at current time such as

$$M(K) = \min(P(\tau, K),\ C(\tau, K))$$

In order to ensure an invariant portion of the strike range considered in the CX across time, Andersen, Bondarenko, and Gonzalez-Perez (2015) propose the ratio $R(K)$ to determine the integration barriers of the CX in equation (2.12) using directly observable prices of call and put options only,

$$R(K) = \frac{P(\tau, K)}{P(\tau, K) + C(\tau, K)} \tag{2.13}$$

For given lower and upper percentiles $p,\ q \in (0, 1)$, $B_L = K_p = R^{-1}(p)$ and $B_H = K_{1-q} = R^{-1}(1 - q)$. In the subsequent simulation and empirical studies, three measures of the CX computed from equation (2.12) are used where $[B_L,$ $B_H]$ takes the values $[R^{-1}(0.25),\ R^{-1}(0.75)]$, $[R^{-1}(0),\ R^{-1}(0.75)]$ and $[R^{-1}(0.25),$ $R^{-1}(1)]$. These implied volatilities are respectively represented by $CXNT$, $CXLT$ and $CXRT$. The definitions of implied volatilities considered in this chapter are listed in Table 2.1. All the measures are computed from options across two nearest maturities (less than 30 days and greater than 30 days) and the 30-day implied volatilities is computed by interpolating between the two separate maturities.

2.2.3 Realized Volatility

In addition to implied volatilities, this study employs monthly realized volatility and historical volatility. We employ a simple realized variance estimator proposed by Barndorff-Nielsen and Shephard (2002), which is equal to the sum of intraday

squared returns

$$rv_t = \sum_{j=1}^{M} r_{t,j}^2 \tag{2.14}$$

where $r_{t,j}$ stands for intraday returns within each 5-minute interval. The realized variance is then calculated over a period of one month in order to match the maturities of the corresponding implied volatilities

$$RV_t = \frac{1}{22} \sum_{i=1}^{22} rv_{t+i} \tag{2.15}$$

The measure RV_t is recorded daily but contains monthly (future) variance. The substantial serial correlation induced by the construction of RV_t in equation (2.15) will be accounted for in the subsequent analysis. Furthermore, the realized variance on the latest trading day, rv_{t-1}, is used as a proxy for historical variance, which may contain useful information for future return variation.

2.3 Error Adjustment Mechanisms

As introduced in section 2.2, the $MFIV$ is computed as an integral of option prices over an infinite range of strikes; and all the measures of implied volatility that are considered require numerical integration using the trapezoidal rule. However, only a limited number of strikes are actually traded in the market, which may result in inaccuracies in the computation of the option-implied volatilities, so further affecting their performance in predicting future volatility and returns. Specifically, very low and high strikes are usually not available in practice, which leads to the so-called truncation errors; and the set of discrete strikes can be rather sparse, which gives rise to the discretization errors. To account for the measurement errors discussed above, the use of an interpolation and extrapolation scheme is essential. The following section provides an introduction of the interpolation and extrapolation techniques adopted in the present chapter.

We first rely on the interpolation procedure to correct the discretization error

by employing a sufficiently fine partition of strikes. Exploiting the BS model to smooth and interpolate option prices has now become a common practice. Shimko (1993) pioneers in filling in option prices for a denser set of strikes by converting option prices into implied volatilities using the BS equation, interpolating and smoothing the curve with a simple quadratic function, and then translating the implied volatilities into option prices using the BS model again. It is worth noting that this procedure is not based on the assumption that the BS model is the underlying model of option prices and it simply performs as a computational device to guarantee a one-to-one mapping between option prices and implied volatilities. The main issue here is how to ensure the smoothness of the interpolated implied volatilities across a wide range of strikes, which satisfy the no-arbitrage conditions. Some popular methods heavily adopted in practice are such as the natural cubic spline (Bates (2000), Jiang and Tian (2005), Jiang and Tian (2007) and Neuberger (2012)), the clamped cubic spline (Malz (2014)) and the smoothing spline (Bliss and Panigirtzoglou (2002) and Figlewski (2008)).

Following the work of Jiang and Tian (2005), we choose the natural cubic spline in this chapter among the available interpolation tools. The natural cubic spline is applied to the known implied volatilities directly rather than option prices due to the fact that the relationship between option prices and strikes appears nonlinear. In the lower and upper price bound $[K_L, K_U]$, we construct a differentiable function $f(K)$ given by

$$f(K_i) = \sigma(K_i, \tau) \tag{2.16}$$

for $i = 1, 2, \cdots, T$, where implied volatilities $\sigma(K_i, \tau)$ are obtained from the BS formula for a given maturity τ. All known strikes from K_2 to K_{N-1} are taken as knot points in fitting the cubic splines. As specified by Jiang and Tian (2007), a natural cubic spline with a different cubic function is employed in each interval between any two consecutive strikes and the first and second derivatives of the cubic functions in any two adjacent intervals are set the same at the common knot points. As a result, the cubic spline fitting procedure provides a smooth implied

volatility function $f(K)$ between K_L and K_U with the continuity of the first and second derivatives at every strike within the interval. The second derivatives at the endpoints are equal to zero. Below, we detail the procedure of computing the BS implied volatility $\sigma(K_i, \tau)$.

The prices at time zero of European call and put options on a non-dividend paying stock are

$$C = S_0 N(d_1) - K e^{-r\tau} N(d_2) \tag{2.17}$$

and

$$P = K e^{-r\tau} N(-d_2) - S_0 N(-d_1) \tag{2.18}$$

where

$$\begin{aligned} d_1 &= \frac{\ln(S_0/K) + (r + \sigma^2/2)\tau}{\sigma/\sqrt{\tau}} \\ d_2 &= \frac{\ln(S_0/K) + (r - \sigma^2/2)\tau}{\sigma/\sqrt{\tau}} = d_1 - \sigma\sqrt{\tau} \end{aligned}$$

Here, we use the same notation as in the previous section and S_0 represents the current asset price. The function $N(x)$ represents the probability that a variable with a standard normal distribution, $\phi(0, 1)$, will be less than x. The volatility σ is the so-called BS implied volatility which can be derived once the value of a option (C or P), K, S_0, r and τ are available.

The interpolation of the option prices within the boundary of actual strikes is relatively straightforward. The major challenge is how to extrapolate the option prices towards the tails of the RND with precision. The flat-line extrapolation, i.e. implied volatilities beyond the truncation points are equal to those observed for the highest and lowest strikes, is adopted in several empirical studies, see Bliss and Panigirtzoglou (2002), Bliss and Panigirtzoglou (2004), Jiang and Tian (2005) and Dotsis and Vlastakis (2016). The flat-line extrapolation produces a lognormal shape of the tails of the RND, which contradicts with the extensive empirical evidence of fat tails in return distributions. Hence, the flat-line extrapolation is

considered inappropriate if the aim is to capture tail events accurately. Jiang and Tian (2007) further point out the disadvantages of this approach such as: (1) it results in the underestimation of implied volatilities outside the available strikes due to the neglect of famous volatility smile or skew; (2) it fails to meet no-arbitrage conditions. To overcome these drawbacks, Jiang and Tian (2007) impose a smooth pasting condition at the minimum (K_L) and maximum (K_U) strikes and adjust the slope of the extrapolated component to match the corresponding slope of the existing interior component at K_L and K_U. The implied volatility extrapolated this way achieves a linear structure.

A different strategy to extrapolate prices based on the existing options is to first estimate the RND and then extract prices beyond the truncation points from the estimated RND. In line with Andersen and Bondarenko (2007), this chapter estimates the RND using a nonparametric approach, the so-called positive convolution approximation (PCA) proposed by Bondarenko (2003). The PCA method for estimating the RND offers several benefits: (1) it guarantees no-arbitrage density estimates; (2) it avoids overfitting while allowing for small samples; (3) it involves simple computation algorithm only; (4) it is insensitive to the data generating process. The main idea of the PCA is to construct a set of admissible densities containing functions which can be expressed as a convolution of a fixed positive kernel and another density. The optimal density is that obtained from the admissible densities which generates the best fit to the listed option prices. The sub-section below briefly describes the RND estimation using the PCA approach.

The relationship between the RND and call/put options can be expressed as

$$h_0(S_\tau) = \frac{1}{e^{\int_0^\tau r_s ds}} \frac{\partial^2 C(\tau, K)}{\partial K^2}\bigg|_{K=S_\tau} = \frac{1}{e^{\int_0^\tau r_s ds}} \frac{\partial^2 P(\tau, K)}{\partial K^2}\bigg|_{K=S_\tau} \tag{2.19}$$

where S_τ represents the value of an underlying asset on trading date τ and r_s is the risk-free rate. For simplicity, it is assumed that the asset pays no dividends and $r_s = 0$ and thus $e^{\int_0^\tau r_s ds} = 1$. In the PCA approach, the first step is to construct the approximating set W_m representing all admissible or candidate densities, from

which the optimal density is selected. Let L^d denote the set of all probability densities, i.e., nonnegative functions that integrate to one. For a basis density $\phi(K) \in L^d$, a new density $\phi_m(K) := \frac{1}{m}\phi(\frac{K}{m})$ can be obtained by smoothing $\phi(K)$ with the bandwidth parameter m. Once $\phi_m(K)$ is fixed, the approximating set $W_m = W_{\phi_m}$ is given by

$$W_m := \left\{ g \in L^d \middle| \, g = \phi_m * \mu, \text{ for } \mu \in L^d \right\} \tag{2.20}$$

which contains functions g, expressed as a convolution of ϕ_m with positive functions μ. In the work of Bondarenko (2003), the basis density $\phi(K)$ is assumed as the standard normal distribution since different choices for $\phi(K)$ result in similar estimators. Although the space L^d accommodates very general shapes of densities, W_m is made up of only smooth and well-behaved densities where the bandwidth m determines the smoothness of densities in the set. If functions h (the true RND) and g are both integrable, the following equation holds.

$$h * g := \int_{-\infty}^{\infty} h(K - y)g(y)dy \tag{2.21}$$

Next, we search for the optimal density in the set W_m containing all candidate densities. An estimator of the RND is the function of $\widehat{h}(K) \in W_m$ which provides the best fit to a certain cross-section of put options $\{P_i\}$ with strikes $K_1 < \cdots K_T$, i.e. it achieves the objective function as follows

$$\underset{\widehat{h} \in W_m}{\text{Minimize}} \sum_{i=1}^{T} \left(P_i - D^{-2}\widehat{h}\,(K_i) \right)^2 \tag{2.22}$$

where $D^{-2}\widehat{h}\,(K_i)$ represents the second integral of $\widehat{h}\,(K_i)$. To measure how closely the density h can be approximated by another density in the set, Bondarenko (2003) defines the distance between h and W_m as

$$\rho(h, W_m) := \inf_{g \in W_m} \|h - g\| = \min_{g \in W_m} \|h - g\| \tag{2.23}$$

where $\|.\| = \|.\|^2$ is the L_2-norm. He further shows that the distance $\rho(h, W_m)$ exhibits basic properties, e.g. an approximation with a small m is always no worse than an approximation with a larger m. We first discretize the admissible set W_m by

$$W_m^{\Delta z} := \left\{ g \in L^d \middle| \, g(K) = \sum_{j=-\infty}^{\infty} a_j \phi_m(K - z_j), \; a_j \geq 0, \; \sum_{j=-\infty}^{\infty} a_j = 1 \right\} \tag{2.24}$$

where $j = 0, \pm 1, \cdots$, and $z_j = j\Delta z$ with Δz being the equally-spaced grid step. Bondarenko (2003) indicates that a sufficiently small Δz can lead to a arbitrarily close distance between the $W_m^{\Delta z}$ and W_m. Finally, the finite-dimensional approximation of the problem in equation (2.22) can be solved numerically by

$$\underset{\widehat{h} \in W_m^{\Delta z}([v,w])}{\text{Minimize}} \sum_{i=1}^{T} \left(P_i - D^{-2} \widehat{h}\,(K_i) \right)^2 \tag{2.25}$$

where $[v, w]$ is a large but finite interval on which the underlying density h is approximated. Once the estimated RND is obtained, option prices can be inferred for a continuum of strikes through the relationship in equation (2.19).

2.4 Monte Carlo Simulation

This section presents a Monte Carlo simulation study where different numbers of option prices are considered as the strike range and increment vary. The aim of this experiment is (i) to ascertain the impact of discrete strike prices on the performance of various implied volatility measures in forecasting future volatility and returns and (ii) to provide guidance for the use of interpolation and extrapolation technique in forecasting.

2.4.1 Simulation Design

The simulation exercise conducted in the present chapter is motivated by Zhang, Taylor, and Wang (2013) who examine the effect of the number of strikes on

the information content of the $MFIV$ using a simple model without jumps and volatility risk premium. In a departure from Zhang, Taylor, and Wang (2013), this study concentrates on the OOS volatility forecasting performance of implied volatilities and the predictive power of implied volatilities for future returns. A jump-diffusion model adopted by Duan and Yeh (2010) is used to simulate the asset price and the latent stochastic volatility by

$$\begin{aligned} d\ln S_t &= [r - q + \delta_s V_t - \frac{V_t}{2}]dt + \sqrt{V_t}dW_t + J_t dN_t - \lambda\mu_J dt \\ dV_t &= \kappa(\theta - V_t)dt + \upsilon V_t^{\gamma} dB_t \end{aligned} \tag{2.26}$$

where W_t and B_t are correlated Wiener processes, having correlation coefficient equal to ρ; N_t denotes a Poisson process with intensity λ, which is independent of W_t and B_t; J_t is an independent normal random variable with mean μ_J and standard deviation σ_J. The price, S_t, and volatility, V_t, processes are dependent through the correlated diffusive terms–W_t and B_t. The other parameters, r, q and δ_s are the risk-free rate, the dividend yield and the asset risk premium, respectively[2].

Option valuation is implemented using the corresponding model under the risk-neutral probability measure given by

$$\begin{aligned} d\ln S_t &= [r - q - \frac{V_t}{2} + \lambda^*(\mu_J^* + 1 - e^{\mu_J^* + \sigma_J^2})]dt + \sqrt{V_t}dW_t^* \\ &\quad + J_t^* dN_t^* - \lambda^*\mu_J^* dt \\ dV_t &= (\kappa\theta - \kappa^* V_t)dt + \upsilon V_t^{\gamma} dB_t^* \end{aligned} \tag{2.27}$$

where $\kappa^* = \kappa + \delta_V$ and $B_t^* = B_t + \delta_V/\upsilon \int_0^t V_s^{1-\gamma} ds$ with δ_V being the volatility risk premium. Again, W_t^* and B_t^* are the Wiener processes correlated with the coefficient ρ; N_t^* is a Poisson process with intensity λ^* independent of W_t^* and B_t^*; the independent normal random variable J_t^* has a new mean μ_J^* but an unchanged

[2] The mean of $J_t dN_t - \lambda\mu_J dt$ is zero due to the introduction of the term $\lambda\mu_J dt$, which serves to center the Poisson innovation.

standard deviation σ_J. The theoretical variance expectation under the model considered above, represented by VIX^2_{Theo}, can be computed as

$$VIX^2_t(\tau^l_t, \tau^u_t, n^l_t, n^u_t) \simeq VIX^2_{Theo} = \frac{\kappa\theta}{\kappa^*}\left(1 - \frac{1 - e^{-\kappa^*\tau}}{\kappa^*\tau}\right) + 2\phi^* + \frac{1 - e^{-\kappa^*\tau}}{\kappa^*\tau}V_t \tag{2.28}$$

see Duan and Yeh (2010) for more details. In this simulation experiment, we adopt the jump risk premium measured by $\delta_J = \phi^* - \phi$, where $\phi^* = \lambda^*\left(e^{\mu^*_J + \sigma^2_J/2} - \mu^*_J - 1\right)$ and $\phi = \lambda\left(e^{\mu_J + \sigma^2_J/2} - \mu_J - 1\right)$. The term δ_J reflects the compensation in the expected return for the jump risk, resulting from the change from the physical probability measure P to the risk-neutral pricing measure Q.

The empirical martingales simulation (EMS) method developed by Duan and Simonato (1998) is used to compute option prices, given that there is no closed-form option pricing formula for equation (2.27). The EMS method, which imposes upon the simulated sample a martingale property, exhibits two benefits such as: (1) the price simulated by the EMS satisfies rational option-pricing bounds; (2) it yields substantial reduction in Monte Carlo errors. Let $t_0 = 0$ be the current time, we generate the EMS prices at a sequence of future time points, t_1, t_2,$\cdots$, t_m using the following system

$$S^*_i(t_j,\ n) = S_0\frac{Z_i(t_j,\ n)}{Z_0(t_j,\ n)} \tag{2.29}$$

where

$$Z_i(t_j,\ n) = S^*_i(t_{j-1},\ n)\frac{\widehat{S}_i(t_j)}{\widehat{S}_i(t_{j-1})} \tag{2.30}$$

$$Z_0(t_j,\ n) = \frac{1}{n}e^{-rt_j}\sum_{i=1}^{n} Z_i(t_j,\ n) \tag{2.31}$$

$\widehat{S}_i(t)$ is the ith simulated asset price at time t prior to the EMS adjustment, and $\widehat{S}_i(t_0)$ and $S^*_i(t_0,\ n)$ are set equal to S_0. At one time, the EMS implements a simulation of n sample points, where $n = 1000$ in our case. After the EMS correction, the simulation proceeds to the next time point, and repeats the whole

process again. The estimate of the discounted EMS asset price is given by

$$S_0^*(t,\ n) = \frac{1}{n} e^{-rt} \sum_{i=1}^{n} S_i^*(t,\ n) \tag{2.32}$$

and therefore the estimate of the call option price is

$$C_0^*(t,\ n) = \frac{1}{n} e^{-rt} \sum_{i=1}^{n} \max\left[S_i^*(t,\ n) - K,\ 0\right] \tag{2.33}$$

The put option prices can be obtained by the put-call parity, i.e. $C + Ke^{-rt} = P + S_0$.

This study assumes one year has 252 trading days and that one day consists of 6.5 hours of open trading, as is the case on the NYSE and NASDAQ. A sparse sampling at a frequency of once every 5 minutes is used in this simulation study and therefore one day can be divided up into 78 intraday intervals, i.e., $\frac{6.5\times 3600}{300} = 78$. A daily series is extracted by sampling once every 78 data points. The asset price and the latent stochastic volatility are simulated according to the Euler discretized version[3] of equation (2.26). The simulation is simplified by assuming no dividends and a zero interest rate. The initial stock price (S_0) and latent stochastic volatility (V_0) are set respectively as 1000 and 0.08 (or 0.02). Here, we consider two situations where the magnitude of volatility is relatively low ($V_0 = 0.02$) and high ($V_0 = 0.08$) in order to evaluate the forecasting performance under different market conditions. The sample size of daily series is 2000. The parameter values are similar to those adopted by Duan and Yeh (2010).

[3]The asset price and volatility path will be discretized into constant-increment time steps of $\Delta t = \frac{1}{78\times 252}$. The discretization for the price and volatility processes through Euler scheme is given by

$$\begin{aligned} S_{i+1} &= S_i \exp[(r - q + \delta_s V_i - 0.5V_i)\Delta t + \sqrt{V_i \Delta t} W_t + J_t N_t - \lambda \mu_J \Delta t] \\ V_{i+1} &= V_i + (\kappa\theta - \kappa V_i)\Delta t + \rho \upsilon V_i^{\gamma} \sqrt{\Delta t} W_t + \sqrt{1-\rho^2}\upsilon V_i^{\gamma} \sqrt{\Delta t} B_t \end{aligned}$$

where J_t is i.i.d. $N(\mu_J,\ \sigma_J)$, N_t is $Poiss(\lambda\Delta t)$, W_t and B_t are two Brownian-motion processes, and ρ represents the instantaneous correlation between the return process and the volatility process. As introduced in the main text, $S_0 = 1000$ and $V_0 = 0.08$ (or $V_0 = 0.02$).

κ	θ	λ	$\mu_J(\%)$	$\sigma_J(\%)$	υ	ρ	γ	δ_s	κ^*	$\phi^*(\%)$	δ_V	$\delta_J(\%)$
2.500	0.080	55.000	0.300	0.500	1.400	-0.800	0.900	0.420	-13.000	0.035	-15.500	-0.059

Option prices are computed corresponding to two nearby maturities, 23 and 37 days. This experiment considers two fixed strike price increments (ΔK=5 and ΔK=1) and attempts with different moneyness ranges ([0.8, 1.2], [0.7, 1.3] and [0.6, 1.4]).

2.4.2 Simulation Results

Table 2.2 and 2.3 report the summary statistics of various volatility measures with $V_0 = 0.08$ and $V_0 = 0.02$, respectively. In both scenarios, it is evident that the mean of the implied volatility estimates increases with the moneyness range. This accords with the work of Jiang and Tian (2007), where the truncation errors usually result in an underestimation of the true volatility. The mean of the VIX, $MFIV$, $CXNT$, $CXLT$ and $CXRT$ decreases as the strike increment becomes smaller, which is consistent with the finding of overestimation of the underlying volatility induced by discretization errors in Jiang and Tian (2007). For most measures considered, the mean squared error[4] (MSE) tends to decrease with the strike range. Table 2.2 and 2.3 also show that measures of implied volatility become more volatile with the range of strikes while they, except the VC, tend to appear less volatile as the partition of strikes is smaller.

To evaluate the OOS volatility forecasting performance of various option-implied volatilities, a univariate Mincer-Zarnowitz regression is adopted as follows

$$y_{t+1} = \alpha_1 + \beta_1 x_t + \mu_{t+1,t} \tag{2.34}$$

where y_{t+1} represents the realized volatility containing the information of month $t+1$ and where x_t indicates each volatility estimate among all candidate estimates. To obtain OOS forecasts of the realized volatility measure, this study employs a rolling

[4]This is defined as the time-series average of the squared differences between the certain volatility estimate and the theoretical VIX index, VIX_{Theo}.

window of 1000 observations for the one-step-ahead forecasts. The daily realized volatility, i.e. the square root of the realized variance in equation (2.15), contains substantial induced serial correlation, which seriously affects the standard errors of the coefficient estimates. To overcome this problem, the Bartlett/Newey-West heteroskedasticity consistent covariance matrix estimator with 44 lags is used, see Andersen and Bondarenko (2007). Regressions are examined for both volatility and logarithms of volatility. The forecasts are evaluated by the MSE, which is robust to the presence of noise in the volatility proxy, see Patton (2011). The OOS R^2 of the Mincer-Zarnowitz regression is also taken into account, which corrects for bias by reflecting the variance but not the bias-squared component of the MSE.

Forecasting results for the case of $V_0 = 0.08$ are reported in Table 2.4. Clearly, the VIX_{Theo} dominates all the other candidate measures in terms of the volatility forecasting performance[5]. Forecasting performance increases with the strike range for all the measures, except that of the $CXNT$ and $CXLT$. It is not surprising that the $CXNT$ performs the same for different moneyness ranges since the options within the barriers $B_L = K_{0.25}$ and $B_H = K_{0.75}$ are not affected by the variation in the strike range. The worse performance of the $CXLT$ with a wider range of strikes may be attributed to the poor forecasting power of the deep OTM put options for future volatility. In addition, Table 2.4 shows that the strike increment ΔK tends to have a negative impact on the volatility forecasting power of the VIX, $MFIV$, $CXNT$, $CXLT$ and $CXRT$ but exerts a positive impact on that of the VC and VP. Overall, the effect of the strike range on the forecasting performance is considerable and that of the strike increment is negligible. The use of different loss functions, i.e. MSE and OOS R^2, gives the identical conclusion in terms of the role of the strike range and increment in the forecasting practice as well as the ranking of forecast performance among implied volatility candidates. These

[5] In several situations, the rv_{t-1} outperforms option-implied volatility estimates, which seems to contradict the findings of Jiang and Tian (2005) and Andersen and Bondarenko (2007). The explanation is that this experiment considers the case of very high volatility, i.e. $V_0 = 0.08$. When the initial latent stochastic volatility is set lower, the performance of daily lagged RV falls as compared with the other implied volatility measures.

findings motivate the application of an extrapolation procedure to extend the tails of the RND in an attempt to improve the volatility forecast accuracy. On the other hand, an interpolation method is considered necessary since the number of listed options may be rather small in practice. The lack of observed options may lead to inaccuracies in the estimation of the RND using the PCA method and thus result in failure in inferring the options beyond the truncation points. Moreover, the critical role of the OTM call options is noted in Table 2.4 where the VC serves as the top forecaster and the $CXRT$ substantially outperforms the $CXLT$. The deep OTM put options are found to weaken the forecasting power of the implied volatilities for future volatility due to the poorer performance of the $CXLT$ relative to the $CXNT$.

The next step is to apply the natural cubic spline to interpolate between available strikes and to implement the PCA method in order to obtain the option values beyond the range of listed strikes. The corresponding measures computed by options with the use of such procedure are prefixed by CP-. To examine the performance of the CP-measures in the forecasting practice for future volatility, this study focuses on the case of $\Delta K = 5$ and moneyness range=[0.8, 1.2] only. Specifically, a step of one unit of the index is used to numerically compute the integral in the interpolation procedure and four standard deviations from forward prices are adopted as an integration range[6]. The interval of strikes that are needed to extrapolate is $([F_0 - 4SD, K_{\min}]$ and $[K_{\max}, F_0 + 4SD])$ where $K_{\min}(K_{\max})$ represents the minimum (maximum) listed strike price in the market. Table 2.5 reports the volatility forecast performance, measured by both the MSE and OOS R^2, of various implied volatility measures and their corresponding CP-measures. The values in parentheses below the MSE are the mean difference of squared forecasting errors between the original implied volatility and its corresponding CP-measure. Numbers in bold indicate statistically significant differences at 5% by the Diebold-Mariano test. Columns 1-4 show that the CP-$MFIV$, CP-$CXRT$,

[6]The choice of the truncation point is motivated by the finding of Jiang and Tian (2005) who show that the truncation errors are virtually zero beyond $3.5SD$ from F_0.

CP-VC and CP-VP achieve significant gains in the forecasting performance for future volatility and that the ranking of forecasting power of the CP-measures remains unchanged from that of the original measures. Columns 5-8 present values of the OOS R^2 where the percentage changes of the R^2 are represented by the numbers in parentheses and where the gains of the CP-measures are indicated in bold. With the single exception of $CXNT$, the use of the interpolation and extrapolation method brings higher OOS R^2 for all the measures considered.

Our conclusions in terms of the impact of the number of strikes on the OOS volatility forecast remain intact when moving to the situation of low volatility with $V_0 = 0.02$. Specifically, in Table 2.6, we find that the accuracy of volatility forecast improves with a wider range of strikes for all measures considered except the $CXNT$. Comparing the values of the loss functions, the OOS R^2 in particular, the strike increment, ΔK, produces trivial and negative effects on the forecasting power of the VIX, $MFIV$, $CXNT$, $CXLT$ and $CXRT$ and positive effects on that of the VC and VP. Different from the case of high volatility with $V_0 = 0.08$, the top performance in the OOS volatility forecast is achieved by the $CXNT$ and $CXRT$ in the low volatility situation. In both cases of high and low volatility, the superiority of the OTM call options is evident given that the $CXRT$ (VC) substantially outperforms the $CXLT$ (VP). As can be observed from Table 2.7, once we apply the interpolation and extrapolation procedure, the MSE of the CP-measures are, in most cases, lower than those of their counterparts derived from the original options although some of these improvements in forecasts are insignificant. In addition, the interpolation and extrapolation procedure results in higher OOS R^2 for 9 out of 12 measures considered.

Another important application of the implied volatility is to predict future market returns. As in the work of Banerjee, Doran, and Peterson (2007), the 30- and 60-day future returns are regressed on daily levels[7] of the implied variance

[7] As a robustness check, the analysis of return predictions is also conducted by regressing future returns on the innovations of the implied variances, motivated by the work of Banerjee, Doran, and Peterson (2007). Conclusions remain unchanged.

estimates as follows

$$\frac{1}{h}\sum_{j=1}^{h} r_{t+j} = \alpha_2 + \beta_2 v_t + u_{t+h,t} \tag{2.35}$$

where v_t indicates each of the measures of implied variance levels. To account for residual correlation caused by overlapping returns, this study considers the Newey-West standard errors. The adjusted R^2 is employed to indicate the degree of return predictability; the values are reported in Table 2.8 where a high volatility scenario ($V_0 = 0.08$) is considered. First, results indicate that the return predictions by implied volatility measures deteriorate with the strike range. Second, with a finer partition of strikes, return predictive power generally improves, with the one exception of VC. From this evidence, only the interpolation method, which provides a smaller partition of strikes, is needed to achieve better return predictions by measures of implied volatility. Consistent with the work of Andersen, Fusari, and Todorov (2015), the deep OTM put options dominate the deep OTM call options in predicting future returns. This is indicated by the higher R^2s given by the $CXLT$ relative to those by the $CXRT$. In addition, the VC displays the strongest predictive power for future returns in most cases while the VP serves as the top performer only in the case of $\Delta K = 1$ when short horizon is considered. This suggests that OTM call options exhibit superior predictive power overall to that of the OTM put options for future returns. This is despite the superiority of the deep OTM puts over the deep OTM calls in this exercise.

The cubic spline is then applied to achieve a finer partition of strikes in the case of return predictions. Measures of implied volatility based upon the options using the interpolation method are prefixed by C-. To examine the effect of the interpolation procedure on return predictions, this study takes the case of $\Delta K = 5$ and moneyness range=[0.8, 1.2] as an example and reports the results of the return predictability in Table 2.9. Gains in the predictive power for future returns are only observed for C-$CXNT$, C-$CXLT$, C-$CXRT$ over 30-day and 60-day horizons, and for C-VP over 30-day horizon.

As a robustness check, we also evaluate the return predictability in a low

volatility situation with $V_0 = 0.02$. Consistent with the high volatility case, results in Table 2.10 indicate that the predictive power of various measures drops as the strike range increases whereas the return predictions tend to improve with a smaller partition of strikes, with the exceptions of VC over 60-day horizon and VP over 30-day horizon. In the low volatility scenario, the advantages of the OTM put options are noted in predicting future returns, suggested by the superior performance of the VP ($CXLT$) over VC ($CXRT$). Two points are worth noting when comparing the forecasting ability of the OTM calls and puts: first, in both low and high volatility scenarios, there are equal numbers of the OTM call and put options used in the return predictions. However, when the volatility level is low, a proportion of the deep OTM options are priced at zero on many trading days and the number of deep OTM calls with the zero price is substantially greater than that of the deep OTM puts; second, the deep OTM call and put options both carry some predictive power for future returns whether the initial volatility is set to 0.02 or 0.08. This is suggested by the higher R^2 delivered by the $CXLT$ and $CXRT$ relative to that by the $CXNT$; see Tables 2.8 and 2.10. Consequently, the phenomenon of options with zero price present in the low volatility setting may seriously weaken the power of the OTM calls in forecasting future returns. However, such phenomenon does not affect the conclusion for the superiority of the OTM calls in volatility forecasts, where the deep OTM puts are shown to lower the forecasting accuracy; see Tables 2.4 and 2.6. Options with zero price are rarely observed when the volatility is high, in which case the roles of the OTM calls and puts in the forecasting practice are more comparable. To sum up, the overall simulation evidence is in support of the superiority of the OTM calls in forecasting future volatility and returns, which is in agreement with our empirical findings to be discussed later.

As for the effectiveness of the interpolation, gains in return predictability are observed in 6 out of 12 cases in Table 2.11. However, given the positive impact of the strike increment on return predictions in Tables 2.8 and 2.10, the interpolation

procedure is expected to lead to more evident gains in the predictive power of various implied volatilities for future returns in the empirical case, where the partition of strikes is often much more sparse, i.e. greater than 5. Findings in section 2.6 confirm this hypothesis.

2.5 Data

The data sample spans from January 02, 2003–December 31, 2013, encompassing 2769 trading days. Data are taken from several sources. Closing bid and ask SPX option prices and dividend yield are obtained from Optionmetrics via the WRDS system. High-frequency data at 5-minute intervals for the SPX[8] are collected from the Tick Data Inc.. Daily one-month and three-month Treasury-bill yields[9], taken as the risk-free rates, are extracted from the Federal Reserve Bulletin. In addition, the average of bid and ask is taken as the best available measure of the option price to alleviate the bid-ask bounce problem. For the two nearby maturities, there is an average of 34 out of 97 (63 out of 97) OTM call (put) option quotes per day. Two commonly used data filters are applied. First, options with less than seven days remaining to maturity are excluded. These options may be subject to problems of liquidity and market microstructure. Second, options violating the boundary conditions, i.e. with BS implied volatilities below zero or above 100%, are excluded from the sample. Only OTM options are included since in-the-money options are less liquid and thus may induce bias into the computation of implied volatilities.

The CBOE calculates the VIX index using option prices updated every five minutes. However, the Optionmetrics database includes the last daily bid-ask quote only, which might not correspond to the data published by CBOE for the final end-of-day computation. Hence, as a more direct benchmark, this chapter

[8] In order to measure the return variation during the overnight period, the squared overnight returns, computed as the squared close-to-open logarithmic price change, are added to the realized variance obtained over the trading day.

[9] Following the work of Jiang and Tian (2007), the risk-free rate is linearly interpolated between these two yields. However, when the maturity is shorter (longer) than one (three) month, the one-month (three-month) yield is adopted.

derives a replicated VIX index, RX, using the exact CBOE procedure every day. Thereby, it follows the work of Andersen, Bondarenko, and Gonzalez-Perez (2015). The RX provides an equivalent of the VIX using the SPX option prices from the Optionmetrics data set. It is well known that the CBOE adopts a particular rule to exclude OTM options: once two puts (calls) with consecutive strikes are found to have zero bid option prices, no puts (calls) with lower (higher) strikes are taken into account. The model-free implied volatility index with a broader strike range, denoted by $MFIV$, can be obtained by discarding any options with a zero bid price and employing all OTM options with a positive bid quote, i.e. ignoring the cutoff rule by the CBOE. Hence, the $MFIV$ provides an upper bound for RX. In addition, the same notations are adopted for the other candidate measures as those in the simulation study[10].

For the 2769 trading days under consideration, the implied volatility measures are not available at some points due to a variety of reasons, including: (1) the requirement for the two nearby maturities is not satisfied; (2) the lack of OTM options; (3) boundary conditions are violated, which reduces the sample size to 2330. The construction of the RV_t leads to the loss of one month at the end. Finally, the sample data under analysis contains 2307 observations, for the period from January 02, 2003 to November 27, 2013.

2.6 Empirical Results

This section starts by reporting the basic statistical properties of different volatility measures. It then investigates their performance as predictors of the future realized volatility and market returns of the underlying S&P 500 index.

Table 2.12 reports the summary statistics[11] of the monthly volatility measures

[10] Throughout the empirical work, this paper makes use of the robust forward F as in the work of Andersen, Bondarenko, and Gonzalez-Perez (2015) rather than the "implied" forward F^* determined by the CBOE according to put-call parity. However, the F^* is still employed in computing the RX in order to approximate the VIX.

[11] In the empirical study, the $MFIV$ is computed in the same way as the CBOE VIX in equation (2.5) but it ignores the cutoff rule by the CBOE. The VC and VP are computed as equations (2.3) and (2.4). This explains why $MFIV^2 \neq VC^2 + VP^2$ in Table 2.12. The reason

which are annualized and recorded daily. First, the unconditional mean of most implied volatility measures clearly exceeds the mean of the RV. This is consistent with the extant literature establishing the presence of a negative volatility risk premium. Note also that the RV has the highest skewness and kurtosis statistics. This erratic nature is attributed to the unpredictable innovation term of the RV as noted in the work of Andersen and Bondarenko (2007). Second, the $CXLT$ (VP) is found to be more volatile and higher in magnitude than the $CXRT$ (VC) because deep OTM puts generally have the highest implied volatility, i.e. volatility smile. A similar phenomenon is observed in the case of the CP-measures. Such evidence is also given in Figure 2.1 which depicts the time-variation of various implied volatility candidates. In particular, the RX overlaps the $MFIV$ closely and thus high similarity is expected in their forecasting power for future realized volatility and returns. Finally, all volatility measures exhibit substantial persistence with extremely slow decay in their autocorrelations.

The correlation between various measures of implied volatility and realized volatility is provided in Table 2.13. Compared with the measures extracted from the listed options only, the corresponding CP-measures display higher correlation with the RV. This is indicative of superior forecasting power for future volatility. Contrast to the work of Zhang, Taylor, and Wang (2013) and Dotsis and Vlastakis (2016) who examine the information content of implied volatilities in in-sample regressions, this study concentrates on the OOS volatility forecasts. The results of the RV forecasts are presented in Table 2.14 where the forecasting performance is measured by the MSE and OOS R^2. Gains achieved by the CP-measures are generally more evident than those in the simulation study. In almost all cases, gains in MSE are significant at 5% level. The $CXNT$ dominates other measures that are based on the existing options. The CP-$CXNT$ ranks best among all CP-measures. As shown in the upper panel of Table 2.14, $CXLT$ (VP) outperforms the $CXRT$

for the use of the CBOE computation procedure, instead of the traditional $MFIV$ calculation method, is to make a direct comparison between the $MFIV$ and RX in terms of the forecasting performance. In doing so, we may ascertain the impact of the cutoff rule of the CBOE on the forecasting power of the VIX.

(VC) in the forecasting of future volatility. This can be attributed to the fact that only a very small number of OTM calls (34 out of 97 per day on average) are available in this empirical study. However, in the lower panel, where more options are involved with the use of interpolation and extrapolation scheme, the OTM call options are superior to the OTM puts, indicated by the better forecasting performance of the CP-$CXRT$ (CP-VC) than that of the CP-$CXLT$ (CP-VP). The evidence for the advantage of the OTM calls is in line with the simulation result discussed in section 2.4.2. Moreover, conclusions drawn from Table 2.14 remain intact when different loss functions for OOS forecasts are considered.

Finally, the return predictability is evaluated by various implied volatilities using equation (2.35) where the excess returns are considered as opposed to raw returns. To better understand the predictive power of implied volatilities for future returns in different market conditions, this study further splits the data sample into pre-crisis and post-crisis periods. The beginning of the financial crisis is set at September 01, 2007. As discussed in the simulation study, only interpolation is needed in the exercise of return predictions. Values of the adjusted R^2 implied by different return regressions are reported in Table 2.15. In the pre-crisis period, the interpolation improves the return predictive power for 4 out of 12 measures. In the post-crisis period, this result holds for 7 out of 12 measures. Moreover, the C-VC dominates all the other implied volatilities in terms of the performance for predicting future returns in the post-crisis period. The $CXRT$ performs the best in such forecasting practice in the pre-crisis period. Hence, the results suggest a few good substitutes for the VIX index as predictors for future returns. In the upper panel of Table 2.15, where measures are derived from the observed option prices only, the OTM call options exhibit greater predictive power for future returns than the OTM put options in the pre-crisis period while the OTM put options play a more dominant role in the post-crisis period. In the lower panel, where the cubic spline is used to interpolate between available strikes, OTM call options outperform OTM put option in predicting future returns in both pre- and post-crisis periods.

2.7 Conclusion

This chapter examines the forecasting power of various model-free option-implied volatilities for future returns and realized volatility using Monte Carlo simulations and an empirical study. By decomposing the model-free implied volatility into different components using various segments of the out-of-the money (OTM) put and call options, this study ascertains the role of each of the components. The chapter provides a simulation study on the impact of the strike range and increment on the predictive power of the implied volatilities. Results show that: first, the forecast accuracy for future volatility improves with the range of strikes; second, the strike range exerts a negative impact on the predictive power of the implied volatilities for future returns; third, a smaller partition of strikes tends to result in greater performance of implied volatilities in predicting returns. These findings warrant the application of an interpolation and extrapolation scheme in order to enhance the forecasting power of implied volatilities for future volatility while only an interpolation method is needed in the case of return predictions.

In both simulation and empirical studies, the superiority of the aforementioned method, i.e. interpolation/extrapolation techniques, is observed for most measures considered in forecasts of future returns and volatility. More interestingly, once this technique is implemented in the empirical case to overcome the problem of the lack of strikes, the OTM SPX call options clearly exhibit higher forecasting power than the OTM SPX put options. This accords with the evidence from the simulation experiment. On the other hand, the advantages of the OTM SPX put options are noted when implied volatilities are derived from the listed options only.

Table 2.1: This table lists the definitions of various measures of model-free implied volatility adopted in this study.

Measures	Definition
VIX_{Theo}	The theoretical VIX index under the stochastic volatility-jump model, see more details in Duan and Yeh (2010).
VIX	The VIX index published by the CBOE, see equation (2.5) in main text.
RX	The replicated VIX index, an equivalent of the VIX using the SPX options from the Optionmetrics data set.
$MFIV$	The model-free implied volatility computed from the full cross-section of OTM put and call options, see equation (2.1) in main text. In the empirical study, the MFIV is computed using equation (2.5) while ignoring the cutoff rule by the CBOE.
$CXNT$	The corridor implied volatility computed from the cross-section of OTM put and call options that truncate the left and right tails of the underlying risk-neutral density by 25%.
$CXLT$	The corridor implied volatility computed from the cross-section of OTM put and call options that truncate the right tail of the underlying risk-neutral density by 25%.
$CXRT$	The corridor implied volatility computed from the cross-section of OTM put and call options that truncate the left tail of the underlying risk-neutral density by 25%.
VC	The model-free implied volatility computed from the full cross-section of OTM call options, see equation (2.3) in main text.
VP	The model-free implied volatility computed from the full cross-section of OTM put options, see equation (2.4) in main text.
CP-	Measures constructed by options with the use of an interpolation and extrapolation scheme.
C-	Measures constructed by options with the use of an interpolation method.

Table 2.2: Simulation study with $V_0 = 0.08$: summary Statistics. This table reports the mean, standard deviation, lower quartile (25%), median (50%), and upper quartile (75%) of daily annualized volatility estimates over 2000 days. All the values are percentages. The mean squared estimation error, MSE, is the average of the squared differences between the volatility estimates and the theoretical VIX index, VIX_{Theo}. The strike price increment is denoted by ΔK and N_K refers to the number of available options on each estimation day.

				Mean	StdDev	25%	50%	75%	MSE
	VIX_{Theo}			39.1297	14.3567	29.4363	35.3602	43.8213	
	RV			23.6841	9.4726	17.2824	21.5069	26.7618	
$\Delta K = 5$		N_K	Moneyness Range						
	VIX	86	[0.8,1.2]	36.2688	11.3866	28.1721	33.8260	41.4539	29.1182
		128	[0.7,1.3]	37.7167	12.9644	28.6248	34.6614	43.2231	16.8826
		167	[0.6,1.4]	38.2435	13.8219	28.6925	34.8117	43.7395	14.2855
	$MFIV$	86	[0.8,1.2]	36.9711	11.3161	28.8961	34.5055	42.2275	25.6308
		128	[0.7,1.3]	38.6101	13.0396	29.4648	35.5447	44.2569	14.7022
		167	[0.6,1.4]	39.2324	13.9959	29.5486	35.7811	44.8681	13.1674
	$CXNT$	86	[0.8,1.2]	27.9266	10.2982	20.9701	25.4019	31.7878	150.4968
		128	[0.7,1.3]	27.9266	10.2982	20.9701	25.4019	31.7878	150.4968
		167	[0.6,1.4]	27.9266	10.2982	20.9701	25.4019	31.7878	150.4968
	$CXLT$	86	[0.8,1.2]	34.2586	10.9680	26.5616	31.8153	39.0299	47.3269
		128	[0.7,1.3]	35.6924	12.3049	27.0270	32.7474	40.9577	30.1992
		167	[0.6,1.4]	36.2805	13.1204	27.1798	32.9743	41.6377	24.6749
	$CXRT$	86	[0.8,1.2]	31.2108	10.6210	23.8673	28.6452	35.5908	84.5573
		128	[0.7,1.3]	31.5717	11.1677	23.8923	28.7841	36.0049	74.6089
		167	[0.6,1.4]	31.6768	11.4019	23.8933	28.8077	36.0561	71.2892
	VC	86	[0.8,1.2]	22.9491	7.2992	17.7665	21.2982	26.0928	316.1882
		128	[0.7,1.3]	23.4417	8.0621	17.8222	21.4521	26.5423	289.5784
		167	[0.6,1.4]	23.5835	8.3822	17.8222	21.4761	26.6604	280.8594
	VP	86	[0.8,1.2]	28.4255	8.5915	22.2635	26.6314	32.5155	161.1815
		128	[0.7,1.3]	30.1422	10.2282	22.8413	27.8142	34.7325	113.4245
		167	[0.6,1.4]	30.8299	11.1983	23.0182	28.0057	35.4366	95.2843
$\Delta K = 1$									
	VIX	422	[0.8,1.2]	36.1206	11.2236	28.1317	33.7444	41.3216	31.0087
		629	[0.7,1.3]	37.6563	12.8712	28.6170	34.6308	43.1692	17.3348
		1239	[0.6,1.4]	38.2166	13.7685	28.6902	34.8077	43.7299	14.3939
	$MFIV$	422	[0.8,1.2]	36.8572	11.1951	28.8665	34.4344	42.1000	26.8896
		629	[0.7,1.3]	38.5635	12.9668	29.4599	35.5384	44.2100	14.9503
		1239	[0.6,1.4]	39.2117	13.9525	29.5504	35.7789	44.8443	13.1920
	$CXNT$	422	[0.8,1.2]	27.6792	10.2485	20.7987	25.1750	31.5773	156.4750
		629	[0.7,1.3]	27.6792	10.2485	20.7987	25.1750	31.5773	156.4750
		1239	[0.6,1.4]	27.6792	10.2485	20.7987	25.1750	31.5773	156.4750
	$CXLT$	422	[0.8,1.2]	34.0638	10.8394	26.4723	31.6732	38.8431	50.1150
		629	[0.7,1.3]	35.5644	12.2193	26.9314	32.6610	40.8070	31.4616
		1239	[0.6,1.4]	36.1791	13.0634	27.0938	32.8640	41.5633	25.4219
	$CXRT$	422	[0.8,1.2]	31.0706	10.5707	23.7608	28.5266	35.5119	87.1688
		629	[0.7,1.3]	31.4414	11.1297	23.8061	28.6810	35.9053	76.8183
		1239	[0.6,1.4]	31.5491	11.3691	23.8061	28.6948	35.9247	73.3751
	VC	422	[0.8,1.2]	23.2036	7.3257	18.0626	21.5130	26.4031	307.5802
		629	[0.7,1.3]	23.7030	8.0965	18.0836	21.6831	26.9452	280.9600
		1239	[0.6,1.4]	23.8465	8.4200	18.0880	21.6981	27.0022	272.2445
	VP	422	[0.8,1.2]	28.5067	8.5052	22.4153	26.7591	32.5390	160.2375
		629	[0.7,1.3]	30.2902	10.1843	23.0286	28.0615	34.8485	110.9583
		1239	[0.6,1.4]	31.0050	11.1826	23.1910	28.3104	35.7387	92.3282

Table 2.3: Simulation study with $V_0 = 0.02$: summary Statistics. This table reports the mean, standard deviation, lower quartile (25%), median (50%), and upper quartile (75%) of daily annualized volatility estimates over 2000 days. All the values are percentages. The mean squared estimation error, MSE, is the average of the squared differences between the volatility estimates and the theoretical VIX index, VIX_{Theo}. The strike price increment is denoted by ΔK and N_K refers to the number of available options on each estimation day.

				Mean	StdDev	25%	50%	75%	MSE
	VIX_{Theo}			19.3489	7.9989	14.0826	17.1069	21.6287	
	RV			12.2629	5.1875	8.9082	11.1537	13.6244	
$\Delta K = 5$		N_K	Moneyness Range						
	VIX	86	[0.8,1.2]	19.4806	7.5630	14.3654	17.4954	21.8103	0.6033
		128	[0.7,1.3]	19.6859	7.9713	14.3876	17.5196	21.9666	0.4720
		167	[0.6,1.4]	19.7411	8.1086	14.3876	17.5284	21.9828	0.5314
	$MFIV$	86	[0.8,1.2]	19.9707	7.6562	14.7781	17.9503	22.3529	0.9571
		128	[0.7,1.3]	20.2228	8.1261	14.8095	18.0293	22.5510	1.1804
		167	[0.6,1.4]	20.3088	8.2839	14.8144	18.0675	22.5954	1.4096
	$CXNT$	86	[0.8,1.2]	14.5759	5.8987	10.7158	12.9623	16.2147	27.4309
		128	[0.7,1.3]	14.5759	5.8987	10.7158	12.9623	16.2147	27.4309
		167	[0.6,1.4]	14.5759	5.8987	10.7158	12.9623	16.2147	27.4309
	$CXLT$	86	[0.8,1.2]	18.4638	7.0875	13.6823	16.5993	20.6721	2.0536
		128	[0.7,1.3]	18.7123	7.4954	13.7150	16.6840	20.8756	1.0884
		167	[0.6,1.4]	18.8203	7.6364	13.8050	16.7368	20.9154	0.8671
	$CXRT$	86	[0.8,1.2]	16.4443	6.5692	12.1027	14.6720	18.3152	10.7190
		128	[0.7,1.3]	16.4791	6.6775	12.1027	14.6720	18.3155	10.2120
		167	[0.6,1.4]	16.4860	6.7022	12.1027	14.6720	18.3155	10.1080
	VC	86	[0.8,1.2]	12.0575	4.8589	8.8248	10.7171	13.4904	63.2844
		128	[0.7,1.3]	12.1048	5.0048	8.8248	10.7171	13.4937	61.6739
		167	[0.6,1.4]	12.1140	5.0380	8.8248	10.7171	13.4942	61.3412
	VP	86	[0.8,1.2]	15.3480	5.8677	11.3534	13.8409	17.2816	21.1217
		128	[0.7,1.3]	15.6404	6.3585	11.4077	13.9543	17.5008	16.9690
		167	[0.6,1.4]	15.7561	6.5303	11.4673	14.0229	17.5645	15.6137
$\Delta K = 1$									
	VIX	422	[0.8,1.2]	19.4446	7.5112	14.3553	17.4847	21.7785	0.6552
		629	[0.7,1.3]	19.6706	7.9527	14.3806	17.5132	21.9565	0.4644
		1239	[0.6,1.4]	19.7315	8.1031	14.3806	17.5227	21.9754	0.5222
	$MFIV$	422	[0.8,1.2]	19.9465	7.6143	14.7732	17.9444	22.3339	0.9677
		629	[0.7,1.3]	20.2167	8.1086	14.8104	18.0277	22.5428	1.1655
		1239	[0.6,1.4]	20.3067	8.2779	14.8143	18.0696	22.5958	1.4007
	$CXNT$	422	[0.8,1.2]	14.3088	5.8672	10.4671	12.7445	15.9307	30.1713
		629	[0.7,1.3]	14.3088	5.8672	10.4671	12.7445	15.9307	30.1713
		1239	[0.6,1.4]	14.3088	5.8672	10.4671	12.7445	15.9307	30.1713
	$CXLT$	422	[0.8,1.2]	18.3476	7.0362	13.5811	16.5221	20.5653	2.3754
		629	[0.7,1.3]	18.6165	7.4680	13.6320	16.6063	20.7757	1.2482
		1239	[0.6,1.4]	18.7276	7.6217	13.6817	16.6578	20.8200	0.9822
	$CXRT$	422	[0.8,1.2]	16.3101	6.5466	11.9645	14.5647	18.1740	11.5773
		629	[0.7,1.3]	16.3463	6.6585	11.9645	14.5652	18.1789	11.0363
		1239	[0.6,1.4]	16.3535	6.6844	11.9645	14.5652	18.1789	10.9242
	VC	422	[0.8,1.2]	12.3426	4.8872	9.0802	11.0050	13.7968	58.9506
		629	[0.7,1.3]	12.3907	5.0354	9.0803	11.0066	13.8043	57.3536
		1239	[0.6,1.4]	12.4002	5.0696	9.0803	11.0066	13.8043	57.0195
	VP	422	[0.8,1.2]	15.5492	5.8333	11.5515	14.0358	17.4858	19.6655
		629	[0.7,1.3]	15.8618	6.3478	11.5960	14.1653	17.7415	15.3688
		1239	[0.6,1.4]	15.9815	6.5320	11.6480	14.2179	17.8116	13.9949

Table 2.4: Simulation study with $V_0 = 0.08$: out-of-sample forecast losses. This table reports the ratio of the losses (MSE and R^2) for different predictive regressions for future monthly realized volatility and logarithm of volatility, respectively. Different strike price increments and ranges of strikes are considered here. Data are obtained for every trading day and the forecasts are based on re-estimating the parameters of the different regressions each day with a fixed length Rolling Window (RW) made up of the previous 1000 days. Ranking is obtained for different cases of strike increments and represents the average volatility forecasting performances of implied volatilities across different strike ranges.

			MSE		Ranking		Out-of-sample R^2 (%)		Ranking	
			Vol	logVol	Vol	logVol	Vol	logVol	Vol	logVol
VIX_{Theo}			0.0387	0.0190			74.5090	76.8332		
rv_{t-1}			0.0558	0.0287			63.2640	65.1061		
$\Delta K = 5$	N_K	Moneyness Range								
VIX	86	[0.8,1.2]	0.0568	0.0269	5	5	62.6171	67.2937	5	5
	128	[0.7,1.3]	0.0553	0.0267			63.5527	67.5413		
	167	[0.6,1.4]	0.0544	0.0265			64.2093	67.7660		
$MFIV$	86	[0.8,1.2]	0.0564	0.0266	4	4	62.8468	67.6101	4	4
	128	[0.7,1.3]	0.0551	0.0265			63.7494	67.7624		
	167	[0.6,1.4]	0.0541	0.0264			64.3901	67.9136		
$CXNT$	86	[0.8,1.2]	0.0519	0.0258	3	3	65.8194	68.6326	3	3
	128	[0.7,1.3]	0.0519	0.0258			65.8194	68.6326		
	167	[0.6,1.4]	0.0519	0.0258			65.8194	68.6326		
$CXLT$	86	[0.8,1.2]	0.0577	0.0280	6	6	62.0262	65.9059	6	6
	128	[0.7,1.3]	0.0582	0.0283			61.6691	65.5402		
	167	[0.6,1.4]	0.0578	0.0283			61.9562	65.5705		
$CXRT$	86	[0.8,1.2]	0.0504	0.0244	2	2	66.8396	70.3366	2	2
	128	[0.7,1.3]	0.0489	0.0239			67.7939	70.9327		
	167	[0.6,1.4]	0.0483	0.0238			68.1703	71.0847		
VC	86	[0.8,1.2]	0.0496	0.0231	1	1	67.3493	71.8709	1	1
	128	[0.7,1.3]	0.0467	0.0224			69.2267	72.6933		
	167	[0.6,1.4]	0.0457	0.0223			69.9102	72.8904		
VP	86	[0.8,1.2]	0.0662	0.0311	7	7	56.3911	62.1576	7	7
	128	[0.7,1.3]	0.0651	0.0310			57.1374	62.2606		
	167	[0.6,1.4]	0.0635	0.0307			58.1738	62.6052		
$\Delta K = 1$										
VIX	422	[0.8,1.2]	0.0574	0.0270	5	5	62.1791	67.1571	5	5
	629	[0.7,1.3]	0.0556	0.0267			63.3723	67.4821		
	1239	[0.6,1.4]	0.0545	0.0265			64.1190	67.7361		
$MFIV$	422	[0.8,1.2]	0.0569	0.0267	4	4	62.5251	67.5176	4	4
	629	[0.7,1.3]	0.0553	0.0265			63.6167	67.7243		
	1239	[0.6,1.4]	0.0542	0.0264			64.3310	67.9012		
$CXNT$	422	[0.8,1.2]	0.0522	0.0259	3	3	65.6335	68.4597	3	3
	629	[0.7,1.3]	0.0522	0.0259			65.6335	68.4597		
	1239	[0.6,1.4]	0.0522	0.0259			65.6335	68.4597		
$CXLT$	422	[0.8,1.2]	0.0582	0.0281	6	6	61.6776	65.7394	6	6
	629	[0.7,1.3]	0.0586	0.0284			61.4355	65.4086		
	1239	[0.6,1.4]	0.0580	0.0284			61.8023	65.4760		
$CXRT$	422	[0.8,1.2]	0.0505	0.0244	2	2	66.7138	70.2520	2	2
	629	[0.7,1.3]	0.0490	0.0239			67.7127	70.8682		
	1239	[0.6,1.4]	0.0484	0.0238			68.1058	71.0259		
VC	422	[0.8,1.2]	0.0487	0.0227	1	1	67.9035	72.3489	1	1
	629	[0.7,1.3]	0.0460	0.0221			69.7165	73.1407		
	1239	[0.6,1.4]	0.0450	0.0219			70.3702	73.3265		
VP	422	[0.8,1.2]	0.0664	0.0310	7	7	56.2748	62.3124	7	7
	629	[0.7,1.3]	0.0649	0.0308			57.2576	62.4827		
	1239	[0.6,1.4]	0.0632	0.0305			58.3920	62.8674		

Table 2.5: Simulation study ($V_0 = 0.08$) with the use of interpolation-extrapolation method: out-of-sample forecast losses. This table reports the ratio of the losses (MSE and R^2) for different predictive regressions for future monthly realized volatility and logarithm of volatility, respectively. For values of MSE, the numbers in parenthesis are the relative differences, squared forecasting errors, between the raw implied volatilities and those with the method of interpolation and extrapolation; values in bold represent significant differences at 5% level, indicated by the Diebold-Mariano test. For values of R^2, the numbers in parenthesis are the percentage changes of the CP-measures relative to the corresponding measures based on the observed options only; numbers in bold highlight the gains of the CP-measures. Ranking is obtained for cases where measures are constructed by observed options only and by prices with the use of interpolation and extrapolation, respectively.

	MSE		Ranking		Out-of-sample R^2 (%)		Ranking	
	Vol	logVol	Vol	logVol	Vol	logVol	Vol	logVol
VIX_{Theo}	0.0387	0.0190			74.5090	76.8332		
rv_{t-1}	0.0558	0.0287			63.2640	65.1061		
VIX	0.0568	0.0269	5	5	62.6171	67.2937	5	5
$MFIV$	0.0564	0.0266	4	4	62.8468	67.6101	4	4
$CXNT$	0.0519	0.0258	3	3	65.8194	68.6326	3	3
$CXLT$	0.0577	0.0280	6	6	62.0262	65.9059	6	6
$CXRT$	0.0504	0.0244	2	2	66.8396	70.3366	2	2
VC	0.0496	0.0231	1	1	67.3493	71.8709	1	1
VP	0.0662	0.0311	7	7	56.3911	62.1576	7	7
$CP\text{-}MFIV$	0.0525 (**-0.39%**)	0.0259 (**-0.07%**)	4	4	65.3974 (**4.06%**)	68.4538 (**1.25%**)	4	4
$CP\text{-}CXNT$	0.0522 (0.03%)	0.0259 (0.01%)	3	3	65.6338 (-0.28%)	68.4592 (-0.25%)	3	3
$CP\text{-}CXLT$	0.0563 (-0.14%)	0.0279 (-0.01%)	5	5	62.9527 (**1.49%**)	66.0735 (**0.25%**)	5	5
$CP\text{-}CXRT$	0.0480 (**-0.24%**)	0.0236 (**-0.08%**)	2	2	68.4057 (**2.34%**)	71.2650 (**1.32%**)	2	2
$CP\text{-}VC$	0.0444 (**-0.52%**)	0.0217 (**-0.14%**)	1	1	70.7738 (**5.08%**)	73.6136 (**2.42%**)	1	1
$CP\text{-}VP$	0.0605 (**-0.58%**)	0.0298 (**-0.13%**)	6	6	60.1917 (**6.74%**)	63.7520 (**2.57%**)	6	6

Table 2.6: Simulation study with $V_0 = 0.02$: out-of-sample forecast losses. This table reports the ratio of the losses (MSE and R^2) for different predictive regressions for future monthly realized volatility and logarithm of volatility, respectively. Different strike price increments and ranges of strikes are considered here. Data are obtained for every trading day and the forecasts are based on re-estimating the parameters of the different regressions each day with a fixed length Rolling Window (RW) made up of the previous 1000 days. Ranking is obtained for different cases of strike increments and represents the average volatility forecasting performances of implied volatilities across different strike ranges.

			MSE		Ranking		Out-of-sample R^2 (%)		Ranking	
			Vol	logVol	Vol	logVol	Vol	logVol	Vol	logVol
VIX_{Theo}			0.0134	0.0267			66.7827	67.9234		
rv_{t-1}			0.0258	0.0477			36.0728	42.7877		
$\Delta K = 5$	N_K	Moneyness Range								
VIX	86	[0.8,1.2]	0.0140	0.0276	3	3	65.2932	66.8613	3	3
	128	[0.7,1.3]	0.0139	0.0276			65.6774	66.9107		
	167	[0.6,1.4]	0.0138	0.0276			65.7816	66.9276		
$MFIV$	86	[0.8,1.2]	0.0141	0.0278	4	4	65.0662	66.7200	4	4
	128	[0.7,1.3]	0.0139	0.0277			65.5327	66.7998		
	167	[0.6,1.4]	0.0138	0.0276			65.7326	66.9340		
$CXNT$	86	[0.8,1.2]	0.0136	0.0272	1	1	66.2040	67.4138	1	1
	128	[0.7,1.3]	0.0136	0.0272			66.2040	67.4138		
	167	[0.6,1.4]	0.0136	0.0272			66.2040	67.4138		
$CXLT$	86	[0.8,1.2]	0.0141	0.0278	5	5	65.1590	66.6712	5	5
	128	[0.7,1.3]	0.0140	0.0278			65.4379	66.7029		
	167	[0.6,1.4]	0.0138	0.0276			65.7044	66.9565		
$CXRT$	86	[0.8,1.2]	0.0137	0.0272	2	2	65.9595	67.3256	2	2
	128	[0.7,1.3]	0.0137	0.0272			66.1195	67.3665		
	167	[0.6,1.4]	0.0137	0.0272			66.1524	67.3752		
VC	86	[0.8,1.2]	0.0143	0.0284	6	6	64.4947	65.9123	6	6
	128	[0.7,1.3]	0.0142	0.0283			64.8982	66.0272		
	167	[0.6,1.4]	0.0141	0.0283			64.9679	66.0483		
VP	86	[0.8,1.2]	0.0146	0.0288	7	7	63.7524	65.4810	7	7
	128	[0.7,1.3]	0.0144	0.0287			64.3786	65.5817		
	167	[0.6,1.4]	0.0142	0.0284			64.7431	65.8882		
$\Delta K = 1$										
VIX	422	[0.8,1.2]	0.0141	0.0277	4	4	65.1813	66.8135	4	4
	629	[0.7,1.3]	0.0139	0.0276			65.6351	66.8826		
	1239	[0.6,1.4]	0.0138	0.0276			65.7564	66.9033		
$MFIV$	422	[0.8,1.2]	0.0141	0.0278	5	5	64.9976	66.7045	5	5
	629	[0.7,1.3]	0.0139	0.0277			65.5190	66.8078		
	1239	[0.6,1.4]	0.0138	0.0276			65.7180	66.9211		
$CXNT$	422	[0.8,1.2]	0.0137	0.0273	1	2	66.0310	67.2146	1	2
	629	[0.7,1.3]	0.0137	0.0273			66.0310	67.2146		
	1239	[0.6,1.4]	0.0137	0.0273			66.0310	67.2146		
$CXLT$	422	[0.8,1.2]	0.0141	0.0279	6	6	65.0340	66.6037	6	6
	629	[0.7,1.3]	0.0140	0.0278			65.3683	66.6738		
	1239	[0.6,1.4]	0.0139	0.0276			65.6063	66.8678		
$CXRT$	422	[0.8,1.2]	0.0138	0.0274	2	1	65.8512	67.1860	2	1
	629	[0.7,1.3]	0.0137	0.0273			66.0201	67.2307		
	1239	[0.6,1.4]	0.0137	0.0273			66.0544	67.2401		
VC	422	[0.8,1.2]	0.0139	0.0276	3	3	65.4487	66.9566	3	3
	629	[0.7,1.3]	0.0138	0.0275			65.8075	67.0437		
	1239	[0.6,1.4]	0.0138	0.0275			65.8692	67.0594		
VP	422	[0.8,1.2]	0.0145	0.0283	7	7	64.0916	66.0206	7	7
	629	[0.7,1.3]	0.0142	0.0282			64.7965	66.1591		
	1239	[0.6,1.4]	0.0141	0.0280			65.1224	66.3744		

Table 2.7: Simulation study ($V_0 = 0.02$) with the use of interpolation-extrapolation method: out-of-sample forecast losses. This table reports the ratio of the losses (MSE and R^2) for different predictive regressions for future monthly realized volatility and logarithm of volatility, respectively. For values of MSE, the numbers in parenthesis are the relative differences, squared forecasting errors, between the raw implied volatilities and those with the method of interpolation and extrapolation; values in bold represent significant differences at 5% level, indicated by the Diebold-Mariano test. For values of R^2, the numbers in parenthesis are the percentage changes of the CP-measures relative to the corresponding measures based on the observed options only; numbers in bold highlight the gains of the CP-measures. Ranking is obtained for cases where measures are constructed by observed options only and by prices with the use of interpolation and extrapolation, respectively.

	MSE		Ranking		Out-of-sample R^2 (%)		Ranking	
	Vol	logVol	Vol	logVol	Vol	logVol	Vol	logVol
VIX_{Theo}	0.0134	0.0267			66.7827	67.9234		
rv_{t-1}	0.0258	0.0477			36.0728	42.7877		
VIX	0.0140	0.0276	3	3	65.2932	66.8613	3	3
$MFIV$	0.0141	0.0278	5	4	65.0662	66.7200	5	4
$CXNT$	0.0136	0.0272	1	1	66.2040	67.4138	1	1
$CXLT$	0.0141	0.0278	4	5	65.1590	66.6712	4	5
$CXRT$	0.0137	0.0272	2	2	65.9595	67.3256	2	2
VC	0.0143	0.0284	6	6	64.4947	65.9123	6	6
VP	0.0146	0.0288	7	7	63.7524	65.4810	7	7
$CP\text{-}MFIV$	0.0138 (0.00%)	0.0276 (**-0.01%**)	4	4	65.8748 (**1.24%**)	66.8967 (**0.26%**)	4	4
$CP\text{-}CXNT$	0.0137 (0.00%)	0.0273 (**0.03%**)	2	2	66.0306 (-0.26%)	67.2154 (-0.29%)	2	2
$CP\text{-}CXLT$	0.0138 (0.00%)	0.0277 (**0.00%**)	5	5	65.7646 (**0.93%**)	66.7600 (**0.13%**)	5	5
$CP\text{-}CXRT$	0.0137 (0.01%)	0.0273 (**0.01%**)	1	1	66.0581 (**0.15%**)	67.2398 (-0.13%)	1	1
$CP\text{-}VC$	0.0138 (-0.02%)	0.0275 (**-0.05%**)	3	3	65.8772 (**2.14%**)	67.0599 (**1.74%**)	3	3
$CP\text{-}VP$	0.0140 (-0.02%)	0.0281 (**-0.05%**)	6	6	65.3964 (**2.58%**)	66.3051 (**1.26%**)	6	6

Table 2.8: Simulation study with $V_0 = 0.08$: multi-period return prediction. This table shows the adjusted R^2 from the daily regressions of the h-period returns on the current implied variance levels. Ranking is obtained for different cases of strike increments and represents the average ability of implied volatilities for predicting returns across different strike ranges.

				Return Predictability (Adj R^2 %)		Ranking	
				30-day	60-day	30-day	60-day
	VIX_{Theo}			0.7611	7.4725		
$\Delta K = 5$		N_K	Moneyness Range				
	VIX	86	[0.8,1.2]	1.4033	8.5224	4	4
		128	[0.7,1.3]	1.2048	8.3069		
		167	[0.6,1.4]	1.0217	8.0018		
	$MFIV$	86	[0.8,1.2]	1.4689	8.6001	2	3
		128	[0.7,1.3]	1.2625	8.3965		
		167	[0.6,1.4]	1.0648	8.0852		
	$CXNT$	86	[0.8,1.2]	0.7559	7.4092	7	7
		128	[0.7,1.3]	0.7559	7.4092		
		167	[0.6,1.4]	0.7559	7.4092		
	$CXLT$	86	[0.8,1.2]	1.1687	8.1426	5	5
		128	[0.7,1.3]	1.0992	8.0754		
		167	[0.6,1.4]	0.9731	7.8621		
	$CXRT$	86	[0.8,1.2]	1.0860	8.1113	6	6
		128	[0.7,1.3]	0.9967	7.9809		
		167	[0.6,1.4]	0.9164	7.8265		
	VC	86	[0.8,1.2]	1.6091	8.8549	1	1
		128	[0.7,1.3]	1.3245	8.5471		
		167	[0.6,1.4]	1.1319	8.2404		
	VP	86	[0.8,1.2]	1.4520	8.6069	3	2
		128	[0.7,1.3]	1.2745	8.4381		
		167	[0.6,1.4]	1.0611	8.1023		
$\Delta K = 1$	VIX	422	[0.8,1.2]	1.4875	8.6919	4	4
		629	[0.7,1.3]	1.2448	8.3953		
		1239	[0.6,1.4]	1.0541	8.0624		
	$MFIV$	422	[0.8,1.2]	1.5257	8.7259	2	2
		629	[0.7,1.3]	1.2881	8.4596		
		1239	[0.6,1.4]	1.0907	8.1320		
	$CXNT$	422	[0.8,1.2]	0.7798	7.4631	7	7
		629	[0.7,1.3]	0.7798	7.4631		
		1239	[0.6,1.4]	0.7798	7.4631		
	$CXLT$	422	[0.8,1.2]	1.2324	8.2971	5	5
		629	[0.7,1.3]	1.1336	8.1630		
		1239	[0.6,1.4]	1.0068	7.9322		
	$CXRT$	422	[0.8,1.2]	1.1023	8.1470	6	6
		629	[0.7,1.3]	1.0128	8.0122		
		1239	[0.6,1.4]	0.9320	7.8527		
	VC	422	[0.8,1.2]	1.4980	8.8190	3	1
		629	[0.7,1.3]	1.2473	8.5171		
		1239	[0.6,1.4]	1.0707	8.2103		
	VP	422	[0.8,1.2]	1.5611	8.6914	1	3
		629	[0.7,1.3]	1.3228	8.4411		
		1239	[0.6,1.4]	1.1071	8.0958		

Table 2.9: Simulation study ($V_0 = 0.08$) with the use of interpolation method: multi-period return prediction. This table shows the adjusted R^2 from the daily regressions of the h-period returns on the current implied variance levels. Numbers in parenthesis represent the percentage changes of the adjusted R^2 of the C-measures relative to their corresponding measures based on observed options only. Numbers in bold highlight the gains in the return predictability by the $C-$ measures. Ranking is obtained for cases where measures are constructed by observed options only and by prices with the use of interpolation, respectively.

	Return Predictability (Adj R^2 %)		Ranking	
	30-day	60-day	30-day	60-day
VIX_{Theo}	0.7611	7.4725		
VIX	1.4033	8.5224	4	4
$MFIV$	1.4689	8.6001	2	2
$CXNT$	0.7559	7.4092	7	7
$CXLT$	1.1687	8.1426	5	5
$CXRT$	1.0860	8.1113	6	6
VC	1.6091	8.8549	1	1
VP	1.4520	8.6069	3	3
C-$MFIV$	1.4611	8.5874	3	2
	(-0.53%)	(-0.15%)		
C-$CXNT$	0.7799	7.4629	6	6
	(3.17%)	**(0.72%)**		
C-$CXLT$	1.1828	8.1725	4	4
	(1.21%)	**(0.37%)**		
C-$CXRT$	1.1016	8.1354	5	5
	(1.44%)	**(0.30%)**		
C-VC	1.4937	8.7952	1	1
	(-7.18%)	(-0.67%)		
C-VP	1.4639	8.4910	2	3
	(0.82%)	(-1.35%)		

Table 2.10: Simulation study with $V_0 = 0.02$: multi-period return prediction. This table shows the adjusted R^2 from the daily regressions of the h-period returns on the current implied variance levels. Ranking is obtained for different cases of strike increments and represents the average ability of implied volatilities for predicting returns across different strike ranges.

				Return Predictability (Adj R^2 %)		Ranking	
				30-day	60-day	30-day	60-day
	VIX_{Theo}			0.4238	7.8759		
$\Delta K = 5$		N_K	Moneyness Range				
	VIX	86	[0.8,1.2]	0.6918	8.7715	4	3
		128	[0.7,1.3]	0.5330	8.2114		
		167	[0.6,1.4]	0.4655	7.9451		
	$MFIV$	86	[0.8,1.2]	0.7335	8.8962	2	2
		128	[0.7,1.3]	0.5546	8.2795		
		167	[0.6,1.4]	0.4723	7.9776		
	$CXNT$	86	[0.8,1.2]	0.4199	7.7976	7	7
		128	[0.7,1.3]	0.4199	7.7976		
		167	[0.6,1.4]	0.4199	7.7976		
	$CXLT$	86	[0.8,1.2]	0.6704	8.6732	3	4
		128	[0.7,1.3]	0.5506	8.2150		
		167	[0.6,1.4]	0.4718	7.9499		
	$CXRT$	86	[0.8,1.2]	0.5325	8.2695	6	6
		128	[0.7,1.3]	0.4519	7.9693		
		167	[0.6,1.4]	0.4275	7.8649		
	VC	86	[0.8,1.2]	0.5921	8.7234	5	5
		128	[0.7,1.3]	0.4394	8.1503		
		167	[0.6,1.4]	0.3970	7.9581		
	VP	86	[0.8,1.2]	0.8565	9.0860	1	1
		128	[0.7,1.3]	0.6413	8.3951		
		167	[0.6,1.4]	0.5217	8.0105		
$\Delta K = 1$	VIX	422	[0.8,1.2]	0.7290	8.8643	3	3
		629	[0.7,1.3]	0.5465	8.2584		
		1239	[0.6,1.4]	0.4706	7.9638		
	$MFIV$	422	[0.8,1.2]	0.7638	8.9634	2	2
		629	[0.7,1.3]	0.5665	8.3176		
		1239	[0.6,1.4]	0.4777	7.9995		
	$CXNT$	422	[0.8,1.2]	0.4237	7.8178	7	7
		629	[0.7,1.3]	0.4237	7.8178		
		1239	[0.6,1.4]	0.4237	7.8178		
	$CXLT$	422	[0.8,1.2]	0.6992	8.7464	4	4
		629	[0.7,1.3]	0.5616	8.2594		
		1239	[0.6,1.4]	0.4775	7.9851		
	$CXRT$	422	[0.8,1.2]	0.5430	8.3001	6	6
		629	[0.7,1.3]	0.4595	7.9904		
		1239	[0.6,1.4]	0.4330	7.8793		
	VC	422	[0.8,1.2]	0.6338	8.7119	5	5
		629	[0.7,1.3]	0.4740	8.1437		
		1239	[0.6,1.4]	0.4275	7.9471		
	VP	422	[0.8,1.2]	0.8629	9.1365	1	1
		629	[0.7,1.3]	0.6296	8.4181		
		1239	[0.6,1.4]	0.5068	8.0258		

Table 2.11: Simulation study ($V_0 = 0.02$) with the use of interpolation method: multi-period return prediction. This table shows the adjusted R^2 from the daily regressions of the h-period returns on the current implied variance levels. Numbers in parenthesis represent the percentage changes of the adjusted R^2 of the C-measures relative to their corresponding measures based on observed options only. Numbers in bold highlight the gains in the return predictability by the $C-$ measures. Ranking is obtained for cases where measures are constructed by observed options only and by prices with the use of interpolation, respectively.

	Return Predictability (Adj R^2 %)		Ranking	
	30-day	60-day	30-day	60-day
VIX_{Theo}	0.4238	7.8759		
VIX	0.6918	8.7715	3	3
$MFIV$	0.7335	8.8962	2	2
$CXNT$	0.4199	7.7976	7	7
$CXLT$	0.6704	8.6732	4	5
$CXRT$	0.5325	8.2695	6	6
VC	0.5921	8.7234	5	4
VP	0.8565	9.0860	1	1
C-$MFIV$	0.7315	8.8888	2	2
	(-0.28%)	(-0.08%)		
C-$CXNT$	0.4238	7.8195	6	6
	(**0.92%**)	(**0.28%**)		
C-$CXLT$	0.6702	8.6805	3	4
	(-0.03%)	(**0.08%**)		
C-$CXRT$	0.5397	8.2885	5	5
	(**1.36%**)	(**0.23%**)		
C-VC	0.6272	8.6878	4	3
	(**5.92%**)	(-0.41%)		
C-VP	0.8134	9.0368	1	1
	(-5.03%)	(-0.54%)		

Table 2.12: Empirical study: summary statistics. This table reports summary statistics for various volatility measures. All the numbers are percentages. The data under analysis ranges from Jan 02, 2003 to Dec 31, 2013, with a total of 2330 observations. After aggregating realized volatility to monthly level, the number of observations reduces to 2307.

	Mean	SD	25%	50%	75%	Skewness	Kurtosis	ρ_1	ρ_{21}	ρ_{63}
RV	11.1303	7.2109	6.9544	8.8190	12.3023	3.1585	15.8942	0.9981	0.7671	0.4531
RX	17.2084	7.6473	12.1256	15.2086	19.5409	2.1455	9.1302	0.9810	0.8023	0.5563
$MFIV$	17.2422	7.6435	12.1511	15.2488	19.5459	2.15415	9.18317	0.9809	0.8020	0.5563
$CXNT$	11.4029	5.2737	7.9047	9.7784	13.2237	2.0799	8.8063	0.9793	0.8016	0.5570
$CXLT$	21.0587	8.8156	15.4129	18.7203	23.7711	2.1243	8.8417	0.9679	0.7485	0.5007
$CXRT$	17.3546	8.2176	11.4495	15.3165	20.3142	1.6119	5.8851	0.9620	0.7543	0.5743
VC	12.7696	6.2396	8.3591	11.0583	15.0483	1.6374	6.0523	0.9528	0.7396	0.5644
VP	18.3542	8.0187	13.3445	16.1713	20.4744	2.2383	9.5957	0.9626	0.7341	0.4847
$CP\text{-}MFIV$	16.5469	7.6433	11.3773	14.4459	19.1326	2.1054	8.9193	0.9810	0.8001	0.5558
$CP\text{-}CXNT$	11.0137	5.1739	7.5377	9.4473	12.8460	2.1251	9.0571	0.9813	0.8038	0.5581
$CP\text{-}CXLT$	15.3803	7.1542	10.5873	13.4761	17.8154	2.1162	9.0209	0.9810	0.7891	0.5524
$CP\text{-}CXRT$	12.5612	5.8506	8.6225	10.7398	14.5407	2.0982	8.8405	0.9808	0.8074	0.5631
$CP\text{-}VC$	9.4145	4.3973	6.4795	8.0340	10.9034	2.1150	8.9629	0.9777	0.8109	0.5699
$CP\text{-}VP$	13.5061	6.3572	9.2902	11.8622	15.5371	2.1608	9.3899	0.9797	0.7910	0.5434

Table 2.13: Correlation Matrix. The sample period is from Jan 02, 2003 to 27 Nov, 2013, with a total of 2307 observations.

	RV	RX	$MFIV$	$CP\text{-}MFIV$	$CXNT$	$CP\text{-}CXNT$	$CXLT$	$CP\text{-}CXLT$	$CXRT$	$CP\text{-}CXRT$	VC	$CP\text{-}VC$	VP	$CP\text{-}VP$
RV	1.0000	0.7482	0.7487	0.7536	0.7585	0.7629	0.7389	0.7519	0.6500	0.7608	0.6299	0.7510	0.7248	0.7405
RX		1.0000	0.9999	0.9967	0.9895	0.9936	0.9714	0.9969	0.9290	0.9916	0.9041	0.9754	0.9488	0.9852
$MFIV$			1.0000	0.9967	0.9893	0.9936	0.9726	0.9968	0.9301	0.9916	0.9050	0.9754	0.9500	0.9850
$CP\text{-}MFIV$				1.0000	0.9937	0.9974	0.9677	0.9995	0.9206	0.9962	0.8968	0.9799	0.9451	0.9861
$CXNT$					1.0000	0.9973	0.9581	0.9915	0.9126	0.9972	0.8896	0.9844	0.9325	0.9773
$CP\text{-}CXNT$						1.0000	0.9657	0.9959	0.9154	0.9991	0.8930	0.9848	0.9419	0.9821
$CXLT$							1.0000	0.9682	0.9224	0.9628	0.8939	0.9422	0.9718	0.9539
$CP\text{-}CXLT$								1.0000	0.9191	0.9934	0.8947	0.9756	0.9467	0.9872
$CXRT$									1.0000	0.9169	0.9623	0.9050	0.8955	0.9075
$CP\text{-}CXRT$										1.0000	0.8951	0.9875	0.9377	0.9790
VC											1.0000	0.9008	0.9022	0.9017
$CP\text{-}VC$												1.0000	0.9397	0.9812
VP													1.0000	0.9618
$CP\text{-}VP$														1.0000

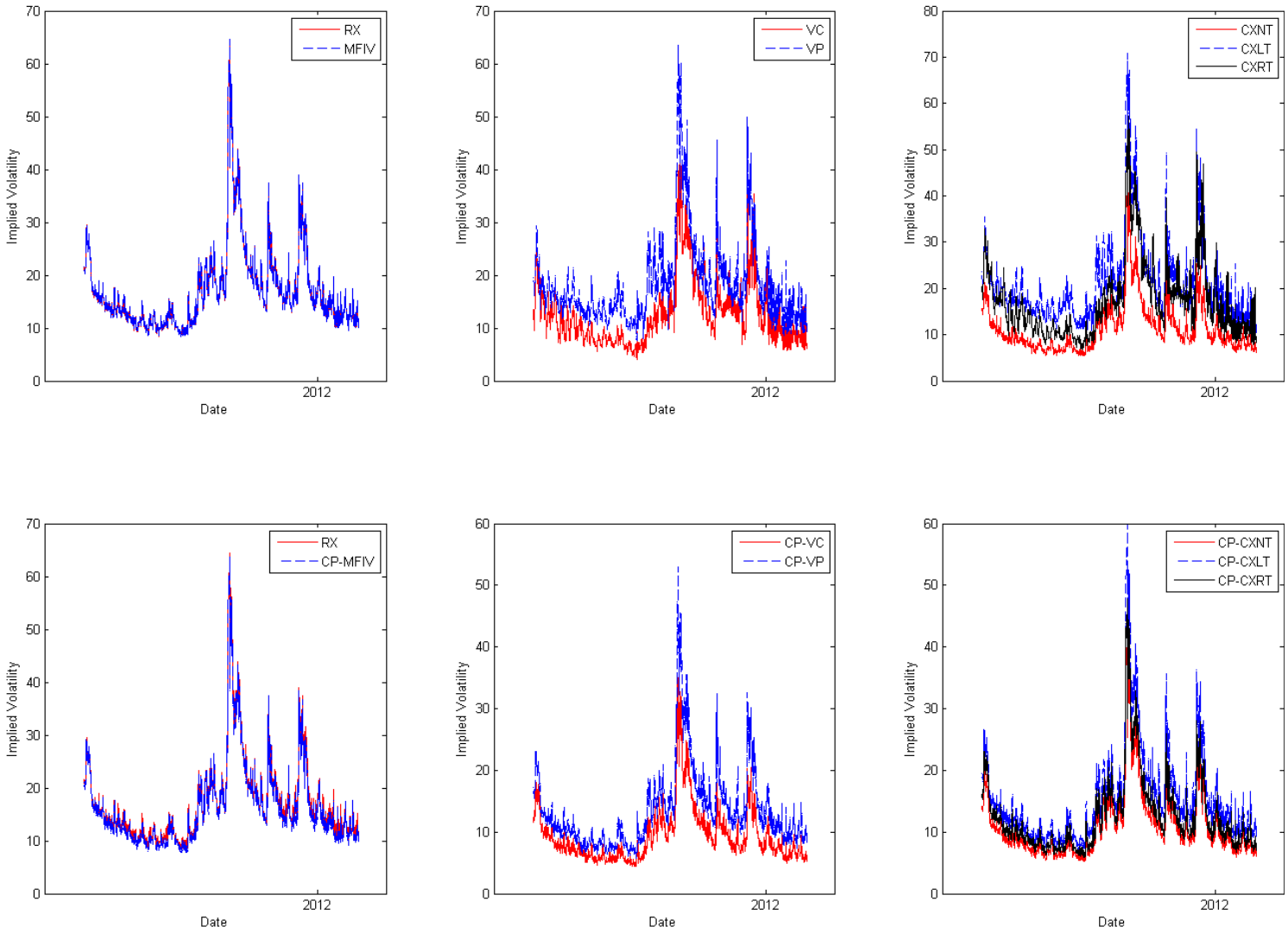
Table 2.14: Empirical study: out-of-sample forecast losses. This table reports the ratio of the losses (MSE and R^2) for different predictive regressions for future monthly realized volatility and logarithm of volatility, respectively. Data are obtained for every trading day and the forecasts are based on re-estimating the parameters of the different regressions each day with a fixed length Rolling Window (RW) made up of the previous 1000 days. For values of MSE, the numbers in parenthesis are the relative differences, squared forecasting errors, between the raw implied volatilities and those with the method of interpolation and extrapolation; values in bold represent significant differences at 5% level, indicated by the Diebold-Mariano test. For values of R^2, the numbers in parenthesis are the percentage changes of the CP-measures relative to the corresponding measures based on the observed options only; numbers in bold highlight the gains of the CP-measures. Ranking is obtained for cases where measures are constructed by observed options only and by prices with the use of interpolation and extrapolation, respectively. N_c denotes the number of OTM call options involved and N_p stands for the number of OTM put options.

	MSE		Ranking		Out-of-sample R^2 (%)		Ranking	
	Vol	logVol	Vol	logVol	Vol	logVol	Vol	logVol
	(N_C=34, N_P=63)				(N_C=34, N_P=63)			
rv_{t-1}	0.1943	0.1930			37.5615	29.6396		
RX	0.1556	0.1393	3	3	50.0027	49.2033	3	3
$MFIV$	0.1554	0.1390	2	2	50.0604	49.3357	2	2
$CXNT$	0.1552	0.1306	1	1	50.1026	52.3644	1	1
$CXLT$	0.1802	0.1485	4	4	42.0750	45.8654	4	4
$CXRT$	0.2315	0.1971	6	6	25.6019	28.1317	6	6
VC	0.2463	0.2114	7	7	20.8354	22.9354	7	7
VP	0.1831	0.1540	5	5	41.1424	43.8649	5	5
	(N_C=308, N_P=308)				(N_C=308, N_P=308)			
$CP\text{-}MFIV$	0.1545 (**-0.09%**)	0.1358 (-0.32%)	3	4	50.3542 (**0.59%**)	50.5030 (**2.37%**)	3	4
$CP\text{-}CXNT$	0.1502 (**-0.50%**)	0.1288 (**-0.18%**)	1	1	51.7218 (**3.23%**)	53.0330 (**1.28%**)	1	1
$CP\text{-}CXLT$	0.1549 (**-2.53%**)	0.1378 (**-1.07%**)	4	5	50.2225 (**19.36%**)	49.7529 (**8.48%**)	4	5
$CP\text{-}CXRT$	0.1524 (**-7.90%**)	0.1297 (**-6.74%**)	2	2	51.0072 (**99.23%**)	52.7195 (**87.40%**)	2	2
$CP\text{-}VC$	0.1557 (**-9.06%**)	0.1309 (**-8.05%**)	5	3	49.9474 (**139.72%**)	52.2884 (**127.98%**)	5	3
$CP\text{-}VP$	0.1576 (**-2.56%**)	0.1419 (**-1.21%**)	6	6	49.3625 (**19.98%**)	48.2794 (**10.06%**)	6	6

Table 2.15: Empirical study: multi-period return prediction. This table shows the adjusted R^2 from the daily regressions of the h-period returns on the current implied variance levels. Numbers in parenthesis represent the percentage changes of the adjusted R^2 of the C-measures relative to their corresponding measures based on observed options only. Numbers in bold highlight the gains in the return predictability by the $C-$ measures. Ranking is obtained for cases where measures are constructed by observed options only and by prices with the use of interpolation, respectively. N_c denotes the number of OTM call options involved and N_p stands for the number of OTM put options.

	Return Predictability (Adj R^2 %)				Ranking			
	Pre-Crsis		Post-Crisis		Pre-Crsis		Post-Crisis	
	(N_C=19, N_P=35)		(N_C=35, N_P=82)					
	30-day	60-day	30-day	60-day	30-day	60-day	30-day	60-day
RX	8.2435	8.5540	10.7008	17.2237	5	5	3	3
$MFIV$	8.3097	8.5988	10.7418	17.2416	4	4	2	2
$CXNT$	9.1953	9.7213	12.5409	19.1597	3	3	1	1
$CXLT$	6.0534	5.5339	9.8895	15.4502	6	6	4	4
$CXRT$	10.5308	15.1420	4.5173	12.7485	1	1	7	6
VC	10.2042	14.3142	4.7760	12.5935	2	2	6	7
VP	3.7819	2.8295	8.9308	14.0298	7	7	5	5
	(N_C=136, N_P=371)		(N_C=397, N_P=216)					
C-$MFIV$	8.0393	7.9514	10.5657	16.8208	3	3	4	4
	(-3.25%)	(-7.53%)	(-1.64%)	(-2.44%)				
C-$CXNT$	7.8481	7.8074	13.0164	19.0621	4	4	3	3
	(-14.65%)	(-19.69%)	**(3.79%)**	(-0.51%)				
C-$CXLT$	7.3816	6.9832	9.7139	16.0005	5	5	5	5
	(21.94%)	**(26.19%)**	(-1.78%)	**(3.56%)**				
C-$CXRT$	8.7838	9.2190	13.8248	19.7243	2	2	2	2
	(-16.59%)	(-39.12%)	**(206.04%)**	**(54.72%)**				
C-VC	9.7937	10.8748	15.0924	20.7174	1	1	1	1
	(-4.02%)	(-24.03%)	**(216.01%)**	**(64.51%)**				
C-VP	6.8616	6.2286	8.4636	14.7009	6	6	6	6
	(81.43%)	**(120.14%)**	(-5.23%)	**(4.78%)**				

Figure 2.1: Time series plots. Data covers the period of Jan 02, 2003 to Dec 31, 2013.



Chapter 3

Volatility Forecasting Using the HAR and Lasso-based Models: an empirical investigation

3.1 Introduction

Modelling and forecasting volatility is of critical importance for asset and derivative pricing, asset allocation and risk management. The increasing availability of high-frequency data has inspired a new line of literature aimed at exploiting the intraday information in the estimation and forecast of return volatility. This literature begins with the work of Andersen and Bollerslev (1998) who first propose to use the cumulative sum of squared intraday returns over short time intervals during the trading day, as an alternative method of volatility estimation, and term this measure the realized variance (RV). Compared with the traditional volatilities built on daily, weekly and monthly frequency data, e.g. the parametric GARCH or stochastic volatility (SV) models, the RV provides model-free unbiased estimates of the ex post return variation under certain conditions specified by Barndorff-Nielsen and Shephard (2002). Although the GARCH and SV models can also be applied to intraday returns, empirical studies indicate that they fail to capture all information in high-frequency data; see Andersen and Bollerslev (1998), Blair, Poon, and Taylor (2001) and Hansen and Lunde (2005). In addition, the RV displays all the stylized facts of financial volatility documented in the case of latent model specifications, the long-memory property in particular.

A growing literature has sought to investigate the properties of the RV and suggest that reliable forecasts can be constructed by high-frequency time-series models; see Andersen et al. (2001b), Koopman, Jungbacker, and Hol (2005) and Ghysels and Sinko (2006), among others. A widely adopted model is the heterogeneous autoregressive, or HAR, model proposed by Corsi (2009). Although the HAR model is formally not a long-memory model, it is able to replicate the strong volatility persistence using the sum of RV components aggregated at different interval sizes. The HAR is easy-to-implement and provides an accurate fit of financial volatility. These features have made it the preferred specification for the forecast of RV. In spite of its substantial applications, the HAR model is rather restrictive in the way that the effect of a volatility shock which occurred two days

ago is identical to a shock which took place three, four or five days ago. However, little work has been conducted in terms of the appropriateness of the lag structure implied by the HAR model, e.g. the maximal autoregressive (AR) lag order and the arrangement of volatility components.

Applying the least absolute shrinkage and selection operator (Lasso) proposed by Tibshirani (1996) as a model selection device, Audrino and Knaus (2016) show that the HAR implied lag structure can be recovered by the Lasso estimator only if the HAR is the underlying data generating process (DGP). However, with their empirical application using nine stocks, the Lasso does not completely agree with the HAR with regard to the lag structure. They also find that the Lasso performs on a par with the HAR in the out-of-sample (OOS) volatility forecast. The Lasso approach produces estimated regression coefficients which are exactly zero, and therefore plays an important role in variable selections where only predictors with nonzero estimates are perceived to be relevant. Although the Lasso is important in determining a model for forecasting exercise, only a few studies have attempted to exploit it in the forecast of future RV; e.g. Audrino, Camponovo, and Roth (2015), Audrino, Huang, and Okhrin (2016) and Wilms, Rombouts, and Croux (2016). Wilms, Rombouts, and Croux (2016) are the pioneers in using the ordered Lasso approach together with the idea of forecast combination in an attempt to enhance the accuracy of forecasts of RV. Compared with the standard AR model estimated by least squares, the ordered Lasso method achieves a better performance in the OOS volatility forecast, where the forecast combination only slightly improves the results.

Against this background, this chapter contributes to the existing literature in two aspects. First, in the in-sample analysis, we adopt three different Lasso-based approaches-adaptive Lasso, group Lasso and cluster group Lasso-to verify the validity of the lag structure implied by the HAR model. In addition to the maximal AR lag order as analyzed in Audrino and Knaus (2016), we also investigate whether the choice of time horizons (daily, weekly and monthly time scales) in the HAR model, at which lags of RV are aggregated, is inherent to the real data. Second, we

are among the few to provide a comprehensive empirical study of the performance of the Lasso-based models in the forecast of future RV. We employ different Lasso-based models and examine the effectiveness of the forecast combination in improving the forecasting accuracy of the Lasso-based models. In addition, we also consider the standard HAR model and the two extensions proposed by Bollerslev et al. (2016) to draw comparisons with the Lasso-based models. The forecasts are based on different time horizons and constructed over both a rolling window and an increasing window. We also account for the impact of the financial crisis on the forecasting performance of various Lasso-based models. To examine the robustness of our results, we use the RV computed from intraday returns sampled at three different frequencies, i.e. 30, 300 and 600 seconds.

Our empirical analysis relies on the high-frequency data of the SPY and ten individual stocks with the same time span as in the work of Audrino and Knaus (2016). In summary, we show that the lag structure implied by the HAR model is not consistent with that given by the model selection devices, which brings into question whether the HAR model is appropriate for modelling and forecasting RV. In our OOS forecasting exercises using the full time period, the best performance is generally provided by the Lasso-based model, where the use of the forecast combination tends to deliver more accurate volatility forecasts. Among various models considered, the ordered Lasso AR with the forecast combination serves as the top performer more frequently than the others. In addition, its improvements over the benchmark HAR model in terms of the OOS volatility forecast are usually significant over monthly horizons. Relative to the pre-crisis period, there tends to be more situations in the post-crisis period where the ordered Lasso AR using the forecast combination dominates the HAR and produces the lowest forecast loss, especially over long forecasting horizons. A larger window size is found to result in a better performance of the Lasso-based models, which is in agreement with the work of Audrino and Knaus (2016). Moreover, the results summarized above are not affected by the variation in the sampling frequencies upon which the RV series are derived. However, as the sampling frequency grows, from 600 seconds to 30

seconds, the Lasso-based model in the OOS forecast is observed to be superior for more stocks in the full sample analysis.

The rest of chapter is organized as follows. Section 3.2 provides a review of relevant studies. Section 3.3 introduces various forecasting models considered in this chapter together with their estimation methods. Section 3.4 describes the data and outlines our empirical findings. Conclusion is presented in Section 3.5

3.2 Literature Review

In this section, we first provide an overview of the existing literature on the development of the RV. We then discuss the nature and construction of the standard HAR model with its various generalizations and extensions for modelling and forecasting RV. We also review studies using the Lasso-based methods to examine the appropriateness of the lag structure implied by the HAR and to forecast future RV from a model selection point of view.

3.2.1 Realized Variance

The modelling of financial volatility has played a central role in risk management and asset allocation. This could be attributed to the fact that, although the daily returns of financial assets are difficult to predict, the volatility of returns appears to be relatively easier to forecast. Since the conditional volatility is latent and thus not directly observable, some models are developed to capture the dynamics of volatility while accounting for its popular features, e.g. volatility clustering, slowly decaying autocorrelations and its non-linear behaviour in response to historical information. The most famous example is perhaps the (Generalized) Autoregressive Conditional Heteroskedasticity family of models, (G)ARCH, proposed by Engle (1982) and Bollerslev (1986). In (G)ARCH-type models, the volatility of a single day is considered unobservable, and so is computed as a function of the variance of daily returns over a given time period. In spite of different extensions and modifications of the GARCH model, e.g. Exponential GARCH, Threshold GARCH,

and GJR-GARCH among others, the basic GARCH(1,1) is found to perform well in an out-of-sample (OOS) forecast of future conditional volatility; see the work of Hansen and Lunde (2005). The latent conditional volatility can also be estimated by stochastic volatility (SV) models (Taylor (1996)) or exponentially weighted moving averages (EWMA) (Morgan (1996)). However, these models introduced above fail to simultaneously accommodate some of the stylized facts found in the time series of financial volatility, as suggested by Bollerslev (1987), Carnero, Peña, and Ruiz (2004) and Malmsten and Teräsvirta (2010). On the other hand, the estimation of these models generally depends on daily or coarser frequency data and therefore the important intraday information is omitted.

As shown by Merton (1980), Nelson (1992) and Foster and Nelson (1996), the volatility measurement tends to improve with the frequency of data. The use of high-frequency intraday data in estimating and predicting the conditional volatility of asset returns is first considered in the work of Andersen and Bollerslev (1998) who propose using realized volatility, computed from squared intraday returns sampled at five minutes, as a proxy for ex post daily foreign exchange volatility. Specifically, assume that an asset price, P_t, displays the dynamics $d\log(P_t) = \mu_t dt + \sigma_t dW_t$, where μ_t and σ_t represent the drift term and the instantaneous volatility, respectively; W_t is a standard Brownian motion which is supposed to be independent of σ_t. The one-day integrated variance is given by $IV_t = \int_{t-1}^{t} \sigma_s^2 ds$. In practice, σ_t^2 is not observable. However, the realized variance (RV) obtained by the sum of intraday returns, $RV_t = \sum_{i=1}^{M} r_{t,i}^2$, where $M = 1/\Delta$ and the Δ-period intraday return is calculated as $r_{t,i} = \log(P_{t-1+i\Delta}) - \log(P_{t-1+(i-1)\Delta})$, provides a consistent estimator of the IV_t as $1/\Delta \to \infty$.

Based upon the findings of Barndorff-Nielsen and Shephard (2002), Meddahi (2002) and Andersen et al. (2003) among others, further investigate the properties of the RV computed from high-frequency data. The superiority of the RV over the GARCH and SV models in terms of the OOS forecast is documented in the work of Andersen et al. (2003). Since then, the literature has increasingly concentrated on the RV estimator. Jacod and Shiryaev (2003) argue that the RV does not

converge as the sampling frequency grows. The impact of sampling frequency on the RV has been extensively examined; see Zhang, Mykland, and Aït-Sahalia (2005), Aït-Sahalia and Mancini (2008) and Bandi and Russell (2008), among others. The five-minute RV is widely considered an appropriate measure which avoids the microstructure noise arising from the use of high-frequency observations, e.g. bid-ask bounce, infrequent trading and price discreteness; see more details in Madhavan (2000) and Biais, Glosten, and Spatt (2005). To minimize the daily mean squared error (MSE) of the realized (co-)variance estimator, Bandi, Russell, and Zhu (2008) show that the optimal sampling frequency is 5 to 30 minutes. They also allow their optimal sampling frequency to vary across (co-)variances and over time. On the other hand, De Pooter, Martens, and Van Dijk (2008) achieve a different finding of around one hour, using a fixed frequency to directly optimize economic criteria. A general formula for the selection of the optimal sampling frequency based on the minimization of the MSE can be found in the work of Zhang, Mykland, and Aït-Sahalia (2005). In addition to the RV, there are many other approaches for estimating the integrated variance IV: for example, two-time scales estimator by Zhang, Mykland, and Aït-Sahalia (2005), kernel based estimator adopted in Zhou (1996) and generalized in Hansen and Lunde (2006), which are constructed by squared intraday returns, and estimators which rely on other ingredients of high-frequency data; see an extensive review in Pigorsch, Pigorsch, and Popov (2012).

3.2.2 HAR and its extensions

It is well established that the RV displays long-memory property, characterized by fractional integration and hyperbolic decaying rates of autocorrelation functions. The long-memory feature of the time series of the RV has been traditionally accommodated by an ARFIMA(p, d, q) process; see Andersen et al. (2003) for example. An alternative to the ARFIMA is the model which approximates long memory, i.e. high persistence, by the aggregation of the heterogenous components observed in the markets. Long-memory effects as a result of the aggregation of

dynamic heterogenous processes are discussed in detail in the work of Robinson (1978) and Granger (1980).

Corsi (2009) proposes the heterogenous autoregressive model of RV (HAR), which parameterizes the RV as a sum of lagged RVs aggregated at different interval sizes, i.e. daily, weekly and monthly averages of RVs. The HAR model is based on the heterogenous ARCH (HARCH) model of Müller et al. (1997) and inspired by the Heterogenous Market Hypothesis (HMH). The HMH suggests that a financial market is composed of agents, having different perspectives of their investment horizons, who react to and result in different types of volatility components. Corsi (2009) assumes that there are three primary trading activities based upon participants' trading duration preferences, i.e. short-, medium- and long-term investment, and thus each investment horizon may lead to a unique volatility. As a result, a financial market is made up of heterogenous market participants with a volatility cascade from low frequencies to high frequencies.

The cascade of heterogenous volatility components in the HAR has a simple AR structure in the RV with economically meaningful coefficient restrictions. Specifically, equality constraints are imposed on the AR coefficients representing a given time horizon. To evaluate whether the coefficient restrictions are valid, Corsi (2009) compares the HAR with the unrestricted AR model and finds that the former dominates the latter in terms of the in-sample fit. The HAR has emerged as a preferred specification for modelling and forecasting RV due to its advantages: first, it can be easily estimated by the OLS technique; and second, it reproduces the persistence properties of financial volatility and accommodates the observed long-memory behaviour. Although the HAR does not formally belong to the class of long-memory models, Corsi et al. (2008) argue that the HAR performs on a par with the long-memory model, ARFIMA, with regard to forecasting and model misspecifications and that the HAR might be more preferable in practice as a result of its straightforward estimation. The simplicity and flexibility of the HAR model allows for various extensions, which will be introduced in the remainder of this subsection. Applications of the HAR and its extensions can be found in a

wide range of areas in finance and economics, e.g. stock market, foreign exchange, bond market, commodity market and currency market.

Based on the bipower variation (BV) measure provided by Barndorff-Nielsen (2004), Andersen, Bollerslev, and Diebold (2007) propose the HAR-RV-J model by modifying the HAR to account for jumps and the HAR-RV-CJ model by incorporating jumps and continuous sample path variation. The HAR-RV-CJ model outperforms the HAR from a volatility forecasting perspective. Taking the analysis of Andersen, Bollerslev, and Diebold (2007) one step further, Andersen, Bollerslev, and Meddahi (2011) provide the HAR-CJN model by accounting, not only for the jumps and continuous component of volatility that occur during the trading day, but also for the overnight return variance using a discrete-time GARCH model. Results indicate that the popular discrete-time volatility models, including the GARCH(1, 1) and the HAR, are dominated by the HAR-CJN model in both in- and out-of-sample forecasts. In a departure from Andersen, Bollerslev, and Diebold (2007) and Andersen, Bollerslev, and Meddahi (2011), Bollerslev et al. (2009) adopt a more efficient maximum likelihood estimation method to model returns, BV and jumps using a coherent multivariate framework. The by-product of the study by Bollerslev et al. (2009) is the development of the HAR-GARCH-BV model. This model helps to explain the time-dependent conditional heteroskedasticity in the innovations of the BV measure by capturing the volatility of volatility using a separate GARCH type model.

In addition to the inclusion of jumps considered in the studies discussed above, Corsi and Renò (2012) extend the heterogenous structure by adding lagged negative returns over the previous day, week and month as explanatory variables for future volatility, in which clear evidence of the persistent leverage effect is delivered. From a different prospective, Corsi et al. (2008) propose two extensions of the HAR and ARFIMA models to accommodate the observed properties of the innovations, i.e. non-Gaussianity and volatility clustering. The extended models result in non-trivial gains in terms of the fit and predictive performance. Under conditions of continuous prices and the absence of measurement error, RV converges to

the true IV as the sampling frequency goes to zero. However, RV is subject to microstructure noise in any given finite sample and therefore ignoring the measurement error may be problematic. To deal with this issue, Bollerslev, Patton, and Quaedvlieg (2016) propose a HARQ model to allow the parameter of the HAR to vary with an estimate of the measurement error variance. The HARQ model is superior to other widely adopted models in terms of the in- and out-of-sample forecasts.

Other extensions of the HAR include (but are not restricted to): (1) the introduction of the nonlinear phenomenon of the RV by McAleer and Medeiros (2008) who combine the HAR with multiple regime smooth transition, and by Hillebrand and Medeiros (2010) who consider a log-linear specification based on the HAR model; (2) a vector heterogenous autoregressive (VecHAR) model proposed by Busch, Christensen, and Nielsen (2011b) for the joint modelling of implied volatility, continuous component of volatility and jumps; (3) Panel-based HAR employed in Bollerslev et al. (2016). Apart from the HAR-Free model where a more flexible lag structure is considered, Bollerslev et al. (2016) introduce two novel RV models in the spirit of the HAR. These are denoted as the heterogenous exponential RV model (HExp) and the Slope HAR model, which are designed to ensure that the predicted future RV relies on the lagged RVs in a way which is continuous and decreasing in the lag lengths. As shown by Bollerslev et al. (2016), the smoothness incorporated in these two models leads to better OOS volatility forecasts relative to the original unsmoothed HAR model.

3.2.3 Lasso applications in modelling and forecasting RV

In spite of the great popularity and extensive applications of the HAR model, few studies examine the validity of its heterogenous structure as a result of volatility cascade. Craioveanu and Hillebrand (2012) extend the standard lag structure corresponding to a daily, a weekly, and a monthly time horizon employed in the HAR by allowing for a flexible lag structure. The optimal lag specification in the HAR model is determined according to two different criteria: in-sample

fit (maximum likelihood) and OOS fit (MSE of the future volatility forecast). However, Craioveanu and Hillebrand (2012) suggest that the use of a more flexible lag structure delivers no gains in OOS volatility forecasts compared to the HAR by Corsi (2009).

In order to forecast a long-memory process containing structural breaks, an enhanced forecasting approach based on an AR approximation is proposed by Wang, Bauwens, and Hsiao (2013). In the context where the series of interest is a long-memory process subject to breaks, the AR-based method performs better than the traditional methods in the OOS forecasts. This provides an explanation, from an econometric perspective, for the empirical success of the HAR model, which can be seen as a special case of the AR approximation, i.e. an AR(22) model with only three lag coefficients. Another theoretical justification for the superior performance of the HAR is given by Hwang and Shin (2014). Motivated by the fact that the HAR (which is a restricted AR(22) process) has short memory, Hwang and Shin (2014) introduce an infinite-order HAR, i.e. HAR(∞), with exponentially decaying coefficients to obtain the genuine long-memory property. They show that the forecast errors are mainly due to the estimation of the unknown coefficients of a finite-order HAR(p) model rather than to errors induced by approximating the underlying HAR(∞) process by the HAR(p).

In light of the studies by Craioveanu and Hillebrand (2012) and Wang, Bauwens, and Hsiao (2013), Audrino and Knaus (2016) examine how much the daily, weekly and monthly frequencies of the HAR are inherent to the real data and whether the lag structure can be identified using a model selection method, i.e. the least absolute shrinkage and selection operator (Lasso) provided by Tibshirani (1996). Audrino and Knaus (2016) show that, if the HAR model is the underlying DGP, the Lasso performs well in recovering the lag structure implied by the HAR model. However, the HAR lag structure cannot be exactly recovered by the Lasso in their empirical application, which poses some doubts on whether the HAR is appropriate for the modelling of RV. In addition, the HAR model and the Lasso approach are found to exhibit indistinguishable performance with regard to the OOS volatility

forecasts in the work of Audrino and Knaus (2016).

As an extension of Audrino and Knaus (2016), Audrino, Camponovo, and Roth (2015) adopt the adaptive Lasso estimator and examine the significance of the estimated coefficients. The adaptive Lasso is introduced by Zou (2006) as a refined Lasso method, which helps to reduce the number of false positives, i.e. the scenario where too many variables are selected by the Lasso. In confirming the results of Audrino and Knaus (2016), the lags selected by the adaptive Lasso approach are generally inconsistent with those implied by the HAR model. Interestingly, the large lags given by the adaptive Lasso, i.e. lags far beyond the 22nd, are generally statistically insignificant, which, to some extent, explains the excellent empirical performance of the HAR model. Distinct from Audrino and Knaus (2016), who concentrate on lassoing the AR terms, Audrino, Huang, and Okhrin (2016) consider flexible HAR and HARQ specifications using the adaptive Lasso to investigate whether the lag structure implied by the HAR can be recovered. Again, no strong evidence for the hypothesis is observed in their application. In terms of the OOS forecasts, the proposed flexible HAR model only slightly outperforms other frequently employed specifications and the gains are not significant.

It is worth noting that the use of the (adaptive) Lasso may not be appropriate in a group of highly correlated or nearly linearly dependent variables, in which case the (adaptive) Lasso tends to select one or few variables even if many or all of them have important explanatory power in explaining the response variable. Such problem is the so-called false negatives. A possible solution to avoid false negatives is the use of the Group Lasso proposed by Yuan and Lin (2006) or the Cluster Lasso by Bühlmann et al. (2013). Group Lasso suggests that, if one group is active, then all the variables within the group will be active whereas the Cluster Lasso aims for small canonical correlations between groups. Audrino, Huang, and Okhrin (2016) divide the lags in AR(50) into four groups namely {1}, {2-5}, {6-22}, {23-50}, where the first three groups are implied by the lag structure of the HAR model. Applying the Group Lasso, Audrino, Huang, and Okhrin (2016) show that the hypothesis for the validity of the lag structure of the standard HAR is rejected in

most cases. They argue that the primary reason for rejection might be due to the inappropriate arrangement of the groups and that a minor reason is the equality restrictions imposed on the AR coefficients.

Lasso-based approaches are also employed to account for the dynamic nature of the volatility forecast. Two extensions of the Lasso method that are attempted in the work of Wilms, Rombouts, and Croux (2016) are termed the Hierarchical Lasso and the ordered Lasso. The Hierarchical Lasso forces the lower order lagged AR coefficient to be selected before its higher order lagged coefficients. Exhibiting a similar feature to the Hierarchical Lasso, the ordered Lasso strongly encourages, but does not ensure, that the absolute values of the lagged effects are monotonically non-increasing, which mimics the hierarchical structure assumed in the HAR model. Among four estimators (i.e. AR, Lasso AR, Hierarchical Lasso AR and ordered Lasso AR), Wilms, Rombouts, and Croux (2016) show that the ordered Lasso AR dominates the others in forecasting future volatility and that the forecast combination slightly improves the forecast accuracy of the Lasso-type models.

3.3 Methodology

In this section, we present the methods employed in our study. We start with an introduction of the HAR model and its two recent extensions. We then describe the Lasso-based methods considered in this chapter.

3.3.1 HAR

The HAR model proposed by Corsi (2009) is one of the most heavily adopted specifications for modelling and forecasting RV via the use of different volatility factors over daily, weekly and monthly horizons representing specific investment behaviors. To introduce the HAR model, let $RV_t^h = \frac{1}{h}\sum_{i=1}^{h} RV_{t-i+1}$ denote the average RV over the previous h days. Correspondingly, $RV_t^W = \frac{1}{5}\sum_{i=1}^{5} RV_{t-i+1}$ and $RV_t^M = \frac{1}{22}\sum_{i=1}^{22} RV_{t-i+1}$ are the weekly (5-day) and monthly (22-day) averages

of daily RV, respectively. The standard HAR is given by

$$RV_{t+1} = \beta_0 + \beta_D RV_t + \beta_W RV_t^W + \beta_M RV_t^M + \varepsilon_{t+1} \tag{3.1}$$

where $\{\varepsilon_t\}$ is a zero mean innovation process. Treating the average realized volatilities as directly observable, estimates of coefficients, β_0, β_D, β_W and β_M, can be consistently obtained by a standard OLS regression.

Highlighted by Corsi (2009), the HAR model in equation (3.1) is equivalent to an AR(22) model with imposed equality constraints on the AR coefficients as follows

$$RV_{t+1} = \beta_0 + \sum_{i=1}^{22} \phi_i^{HAR} RV_{t-i+1} + \varepsilon_{t+1} \tag{3.2}$$

The coefficient restrictions implied by the HAR are given by

$$\phi_i^{HAR} = \begin{cases} \beta_D + \frac{1}{5}\beta_W + \frac{1}{22}\beta_M & \text{for } i = 1 \\ \frac{1}{5}\beta_W + \frac{1}{22}\beta_M & \text{for } i = 2,\ldots, 5 \\ \frac{1}{22}\beta_M & \text{for } i = 6,\ldots, 22 \end{cases} \tag{3.3}$$

To evaluate whether the constraints are valid, Corsi (2009) compares the restricted HAR with the unrestricted AR(22) model. Corsi (2009) argues that the rejection of the joint F test for the restriction presented in equation (3.3) is to be expected due to the large number of restrictions and that the reason for the rejection is asset dependent. This result also provides an indication that the HAR may not be successful in fully capturing the observed effects in the real data.

The HAR in (3.1) can be used directly for forecasting the one-day-ahead RV. For longer-run forecasts, e.g. weekly or monthly horizons, we employ a simple forecasting approach by replacing the daily RV on the left-hand-side of equation (3.1) with the RV over h-day horizon as in the work of Bollerslev, Patton, and Quaedvlieg (2016) and Bollerslev et al. (2016), among others. The predictive regression is given by

$$RV_{t+h}^h = \beta_0^h + \beta_D^h RV_t + \beta_W^h RV_t^W + \beta_M^h RV_t^M + \varepsilon_{t+1}^h \tag{3.4}$$

which, again, can be estimated by a standard OLS technique. This forecasting method, the so-called direct forecasts in the forecasting literature, is directly applicable to other volatility forecasting models to be introduced below.

3.3.2 Two extensions of the HAR model

The key idea of this chapter is to examine whether parsimonious models produced by the Lasso-based estimators can result in superiority over the standard HAR model in terms of the forecast of future volatility. The parsimonious model indicates that many of the lagged RVs exert zero impact on the future RV. In other words, it selects only relevant volatility factors and thus achieves a more flexible lag structure than the standard HAR model, which may lead to a reduction in the estimation error. As such, by comparison with the Lasso-based forecast models, we consider some existing extensions of the HAR model which are designed to generalize the standard HAR model from the perspective of lag structure.

First, we adopt the HAR-Free[1] model proposed by Bollerslev et al. (2016) as follows

$$\begin{aligned} RV_{t+1} &= \beta_0 + \beta_1 RV_t + \beta_2 RV_{t-1} + \beta_3 RV_{t-2} + \beta_4 RV_{t-3} + \beta_5 RV_{t-4} \\ &\quad + \beta_6 RV_{t-5} + \beta_M RV_t^M + \varepsilon_{t+1} \end{aligned} \tag{3.5}$$

As an extension of the HAR model, the HAR-Free model allows for the coefficients of the first six daily lagged RVs to be freely estimated. Here, the RV_t^M is derived in the same way as that in equation (3.1).

As argued in Bollerslev et al. (2016), the step-wise nature of the RV_t^W and RV_t^M in the HAR model in equation (3.1) may result in sudden changes in the forecast of RV when a very large/small daily RV is removed from the sums for the longer-horizon lagged volatility factors with time passing. To overcome this issue,

[1] Unlike Bollerslev et al. (2016), we do not include the long-run volatility factor to "anchor" the HAR model since we do not conduct the panel-based estimation. In addition, we do not involve the annual volatility factor in order to make our analysis directly comparable to the standard HAR model.

Bollerslev et al. (2016) propose the slopeHAR model, which can be expressed as

$$RV_{t+1} = \beta_0 + \beta_D RV_t + \beta_W SlopeRV_t^W + \beta_M SlopeRV_t^M + \varepsilon_{t+1} \tag{3.6}$$

where $SlopeRV_t^k = \sum_{i=1}^{k} \left(\frac{k-i+1}{k+(k-1)+...+1} \right) RV_{t+1-i}$. The slopeHAR model is smooth in the sense that the effects on the forecast of future RV of the historical volatilities are continuous and decreasing with lags. In the same spirit as the slopeHAR model, Bollerslev et al. (2016) introduce another extended HAR-type model which depends on a mixture of Exponentially Weighted Moving Averages (EWMA) of the past RVs. Due to its similar performance to the slopeHAR model in the OOS volatility forecasts, we only consider the slopeHAR model in this chapter as an extended smooth HAR model.

3.3.3 Lasso-based Estimators

In the following subsection, we describe the Lasso-based approaches adopted in this chapter in an attempt to verify the validity of the lag structure implied by the HAR model and to enhance the performance in forecasting future RV by using a parsimonious model.

Lasso and adaptive Lasso

The Lasso method developed by Tibshirani (1996) has gained popularity in the recent econometric literature for model selection in linear and generalized linear models. As suggested by Friedman, Hastie, and Tibshirani (2010), the great interest in the Lasso arises from two aspects: first, it conducts model selection and estimation simultaneously; second, it adopts a highly efficient algorithm and is therefore computationally affordable. We describe below the Lasso and its refinement, i.e. adaptive Lasso in identifying the active lagged RVs in the forecast of future RV.

As discussed earlier, the HAR model can be expressed as an AR(22) process. In this subsection, we represent the daily RV_t by x_t and let $(x_t, \cdots, x_{t-p+1})'$ be

predictor variables in an $AR(p)$ process:

$$x_{t+1} = c + \sum_{j=1}^{p} \phi_j x_{t-j+1} + \varepsilon_t \tag{3.7}$$

where $t = p, \cdots, T$ and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) innovations with zero mean. The Lasso procedure employs an ℓ penalty to obtain a sparse solution as follows

$$\left(\widehat{c}^{Lasso}, \widehat{\phi}^{Lasso}\right) = \arg\min_{c,\, \phi} \left\{ \sum_{t=p}^{T} \left(x_{t+1} - c - \sum_{j=1}^{p} \phi_j x_{t-j+1} \right)^2 \right\} \text{ subject to } \sum_{j=1}^{p} |\phi_j| \leq \tau \tag{3.8}$$

where τ is a tuning parameter which controls the amount of shrinkage applied to the estimates. The solution for c is $\widehat{c} = \overline{x}$. The parameter c can be removed from the minimization once we let $\overline{x} = 0$ by demeaning the data. The minimization problem can be solved by

$$\widehat{\phi}^{Lasso} = \arg\min_{\phi} \left\{ \frac{1}{2} \sum_{t=p}^{T} \left(x_{t+1} - \sum_{j=1}^{p} \phi_j x_{t-j+1} \right)^2 + \lambda \sum_{j=1}^{p} |\phi_j| \right\} \tag{3.9}$$

where the parameter λ has a one-to-one relationship with τ in equation (3.8). It is clear that letting $\lambda = 0$ will lead to consistency between the Lasso estimator and the OLS estimator. However, the use of a strictly positive λ will penalize all AR coefficients which are not equal to zero. Moreover, a higher value of λ will result in an increasing number of coefficients being set exactly to zero. In short, the selection of the tuning parameter λ causes shrinkage of the solutions towards zero, where some of the AR coefficients may become exactly zero if λ is sufficiently large. The active set is defined as $S = \{j,\, \phi_j \neq 0\} \subset \{1, \ldots, p\}$ and the nonactive set is $S^c = \{1, \ldots, p\} \setminus S$. Nardi and Rinaldo (2011) establish that, under certain conditions, the Lasso is model selection consistent, estimation consistent and prediction consistent.

Assume that the underlying and unknown coefficient parameters to generate

the $AR(p)$ process in (3.7) are $\phi^* = (\phi_1^*, \ldots, \phi_p^*)'$. Model selection consistency is about precisely identifying S and S^c. Defined by Nardi and Rinaldo (2011), the sign function is given by

$$sgn(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{3.10}$$

and $sgn(\phi) = (sgn(\phi_1), \ldots, sgn(\phi_p))$. The estimator $\widehat{\phi}^{Lasso}$ is model selection consistent if

$$P(sgn(\widehat{\phi}^{Lasso}) = sgn(\phi^*)) \to 1 \text{ for } T \to \infty \tag{3.11}$$

In addition, $\widehat{\phi}^{Lasso}$ is said to be estimation consistent if $\left\|\widehat{\phi}^{Lasso} - \phi^*\right\|$ converges to zero as T goes to infinity. Prediction consistency holds if $\left\|X\widehat{\phi}^{Lasso} - X\phi^*\right\|$ converges to zero as $T \to \infty$, where $X = (x_t, \ldots, x_{t-p+1})$.

From the earlier discussion, the tuning parameter τ in equation (3.8), or λ in equation (3.9), is of critical importance for the success of the Lasso. To determine λ, cross-validation is commonly adopted although the information criteria, i.e. AIC or BIC, is also considered; see Audrino and Knaus (2016) for an example. As introduced in Tibshirani (1996), the prediction error for the Lasso approach is first estimated by K-fold cross-validation. The Lasso is then indexed with regard to the normalized parameter $s = \tau / \sum \widehat{\phi}_j^0$ where $\widehat{\phi}_j^0$ represents the full OLS estimates of equation (3.7). The prediction error is obtained over a grid of values of s ranging from zero to one. The $\widehat{s}$ resulting in the minimum error is finally selected.

In the original Lasso introduced above, all the AR coefficients are penalized equally. Zou (2006) provides a refined version of the Lasso allowing for a more flexible penalization, termed the adaptive Lasso, which helps to reduce the issue of false positives. The adaptive Lasso estimator is given by

$$\widehat{\phi}^{AL} = \arg\min_{\phi} \left\{ \frac{1}{2} \sum_{t=p}^{T} \left(x_{t+1} - \sum_{j=1}^{p} \phi_j x_{t-j+1} \right)^2 + \lambda \sum_{j=1}^{p} \lambda_j \left| \phi_j \right| \right\} \tag{3.12}$$

with λ_j being individual weights for each of the coefficients. When $\lambda_j = 1$ is set for all the coefficients, the adaptive Lasso becomes the standard Lasso in (3.9). As suggested by Zou (2006), a possible choice for the weights can be the inverse of the absolute values of the OLS coefficients. Under the condition that the innovations are i.i.d., Zou (2006) shows that the adaptive Lasso exhibits the oracle properties, i.e. it asymptotically detects the non-zero coefficients and displays the optimal estimation rate. Furthermore, in the context of linear time series processes, e.g. the AR process considered in the present chapter, the oracle properties are found for the adaptive Lasso in the work of Kock (2012) and Medeiros and Mendes (2012). One common problem of the Lasso and adaptive Lasso is that they are insensitive to highly correlated predictors and will tend to select one and omit the rest. The Lasso and adaptive Lasso will fail to produce reliable estimates in the extreme situation containing identical predictors only. To alleviate this problem, we consider other Lasso-type methods below.

Group Lasso

As discussed in section 3.2.2, the HAR is an additive cascade model of the RV aggregated at different time horizons. The cascade of heterogenous volatility components represents the behavior of market participants of different types. As such, it is natural to divide the lagged RVs into different groups and, if one group is active, then all the past RVs within this group should be active. This is the key idea of the so-called group Lasso proposed by Yuan and Lin (2006), who also argue that the Lasso can only be used to select individual variables rather than a group of correlated variables.

We keep the same notation used in the earlier subsection. As suggested by Yang and Zou (2015), the Group Lasso estimator can be obtained by solving the penalized least squares as follows

$$\widehat{\phi}^{Group} = \arg\min_{\phi} \left\{ \frac{1}{2} \sum_{t=p}^{T} \left(x_{t+1} - \sum_{j=1}^{p} \phi_j x_{t-j+1} \right)^2 + \lambda \sum_{k=1}^{K} \sqrt{p_k} \sqrt{\sum_{j \in I_k} \phi_j^2} \right\} \quad (3.13)$$

where the p lagged RVs are divided into K non-overlapping groups such that $(1, 2, \ldots, p) = \cup_{k=1}^{K} I_k$, and $I_k \cap I_{k'} = \varnothing$ for $k \neq k'$; the cardinality of index set I_k is p_k. We employ a simple unified algorithm, groupwise majorization descent (GMD), as proposed by Yang and Zou (2015). This can be used to solve the group Lasso learning problem if the loss function meets a quadratic majorization condition. Relative to some existing algorithms, e.g. LARS-type algorithm, coordinate descent algorithm and a block coordinate gradient descent algorithm, the GMD dominates primarily in two aspects: first, it does not require the group-wise orthonormal assumption and can therefore perform cross-validation or bootstrap analysis of the group Lasso; second, its computation is more efficient.

Cluster group Lasso

In the estimation of the group Lasso, the group membership is assumed known. However, this assumption usually does not hold in practice and therefore efforts need to be made to divide predictors into groups, i.e. homogenous clusters. Cluster Lasso is often adopted to split variables into groups so that elements in each group are strongly related to each other and contain similar information. In cases where predictors display high empirical correlations or near linear dependence, the invalidity of the Lasso estimator is also suggested by Bühlmann et al. (2013) who argue that the Lasso approach tends to choose only one variable from the group of variables and omit the others.

In this chapter, we employ the cluster Lasso proposed by Bühlmann et al. (2013), which differs from the existing methods of clustering the variables mainly by the use of canonical correlation. The arrangement of groups based on the cluster Lasso considered here guarantees identifiability and an oracle inequality for the group Lasso introduced earlier. Moreover, this procedure is intended to avoid false negatives, i.e. avoid omitting a variable from the active set. However, the trade-off is an increase in false positive selections. As argued in Bühlmann et al. (2013), the cluster Lasso is useful in practice and can be seen as a desirable screening method. We introduce this method as follows.

As in equation (3.7), we consider p lagged RVs. We represent the group of past RVs from a cluster $G \subseteq \{1, \ldots, p\}$ by $X^{\{G\}} = \{x_{t-j};\ j \in G\}$. The aim is to determine a partition G^* into non-overlapping clusters $G_1, \ldots, G_q : G^* = \{G_1, \ldots, G_q\}$ with $\cup_{r=1}^{q} G_r = \{1, \ldots, p\}$ and $G_r \cap G_\ell = \varnothing$ $(r \neq \ell)$. For a partition G^*, we define

$$\widehat{\rho}_{\max}(G^*) = \max \{\widehat{\rho}_{can}(G_r, G_\ell);\ r, \ell \in \{1, \ldots, q\},\ r \neq \ell \tag{3.14}$$

where $\widehat{\rho}_{can}(G_r, G_\ell)$ is the empirical canonical correlation between the lagged RVs from $X^{\{G_r\}}$ and $X^{\{G_\ell\}}$. In the work of Bühlmann et al. (2013), a structure with ξ-separation between clusters is given by

$$G^*(\xi) = \text{a partition } G^* \text{ of } \{1, \ldots, p\} \text{ such that } \widehat{\rho}_{\max}(G^*) \leq \xi \ (0 < \xi < 1) \tag{3.15}$$

To ensure that the clustering with ξ-separation has a unique solution, Bühlmann et al. (2013) further define the finest clustering with ξ-separation, $\widehat{G}^*_{finest}(\xi)$, as the one which has a strictly finer structure than any other clustering with the same separation ξ. In practice, we adopt the hierarchical bottom-up agglomerative clustering proposed by Bühlmann et al. (2013). Specifically, the procedure begins with the single variables, i.e. partition in p clusters; it then combines two clusters with the highest canonical correlation; this procedure is repeated until the criterion in (3.15) is satisfied. The parameter ξ is of critical importance in the procedure described above. Bühlmann et al. (2013) recommend the use of the minimal resulting ξ. To be more detailed, the maximal canonical correlation between clusters, termed $\widehat{\rho}_{\max}(b)$ where b is the number of iteration, is recorded in every iteration of the bottom-up agglomerative clustering algorithm. We adopt the partition which achieves the criterion $\arg\min_b \widehat{\rho}_{\max}(b)$. In this chapter, we follow the cluster group Lasso approach introduced in Bühlmann et al. (2013) to select the clusters: divide the predictors based on the identified clusters and then estimate the coefficients of predictors using the group Lasso by Yuan and Lin (2006).

Ordered Lasso

Finally, we consider a Lasso-type approach to account for the dynamic nature of the AR process, i.e. a higher order lag is considered only when its lower order lags are already selected. Two methods for accommodating such features are used in Wilms, Rombouts, and Croux (2016): the hierarchical Lasso and the ordered Lasso. As the latter is found to be superior to the former in terms of the volatility forecasts, we employ only the ordered Lasso in our analysis.

The ordered Lasso is proposed by Tibshirani and Suo (2016) to allow for a decreasing dependence on regressors in a time-lagged regression, by the introduction of an additional order restriction on the coefficients. The ordered Lasso estimator is given by

$$\widehat{\phi}^{order} = \arg\min_{\phi} \left\{ \frac{1}{2} \sum_{t=p}^{T} \left(x_{t+1} - \sum_{j=1}^{p} (\phi_j^+ - \phi_j^-) x_{t-j+1} \right)^2 + \lambda \sum_{j=1}^{p} (\phi_j^+ + \phi_j^-) \right\} \tag{3.16}$$

subject to $\phi_1^+ \geq \phi_2^+ \geq \ldots \geq \phi_p^+ \geq 0$ and $\phi_1^- \geq \phi_2^- \geq \ldots \geq \phi_p^- \geq 0$, where $\phi_j = \phi_j^+ - \phi_j^-$. The use of positive and negative terms in equation (3.16) instead of absolute values as in the case of Lasso and adaptive Lasso is to make it a convex problem. It is worth noting that ϕ_j^+ and ϕ_j^- can take positive values at the same time and therefore the $\left|\phi_j^+ - \phi_j^-\right|$ may display non-monotonicity. As shown in Tibshirani and Suo (2016), the minimization problem in equation (3.16) can be efficiently solved by the well-known Pool adjacent Violators algorithm as its proximal operator.

Forecast Combination

The concept of forecast combination is first introduced by Bates and Granger (1969). It constructs a new forecast using a linear combination of all forecasts obtained from individual models. Due to its superiority in terms of the OOS forecasting accuracy and stabilized forecasts errors, forecast combination has been applied in a wide range of topics in finance and economics; see Stock and Watson

(2004), Timmermann (2006), Corte and Tsiakas (2012), Rapach, Strauss, and Zhou (2013), Li and Chen (2014) and Wilms, Rombouts, and Croux (2016).

As mentioned earlier, the choice of the tuning parameter λ is crucial in the Lasso-type procedure. Following the work of Wilms, Rombouts, and Croux (2016), we attempt to combine forecasts produced by the Lasso-based estimators with several different tuning parameters.

Specifically, we use a logarithmic spaced grid of tuning parameters of length $L = 20$ in the interval $[\lambda_1, \lambda_L]$ where $\lambda_1 = 0$ and λ_L is an estimated parameter shrinking all the regression coefficients to zero. For each λ_m within the interval, $1 \leq m \leq L$, we estimate different Lasso-based models and construct their corresponding forecasts with the estimated coefficients. Finally, we obtain the weighted sum of the L forecasts of volatility given by L different tuning parameters. The calculation of the weights, w_m, is given by

$$w_m = \frac{\exp(-0.5BIC_{\lambda_m})}{\sum_{m=1}^{L} \exp(-0.5BIC_{\lambda_m})} \tag{3.17}$$

The widely adopted BIC criterion can be obtained by

$$BIC_{\lambda_m} = (T-p) \times \log(Loss_{\lambda_m}) + df_{\lambda_m} \times \log(T-p) \tag{3.18}$$

where the $Loss_{\lambda_m}$ is the average sum of squared prediction errors using the tuning parameter λ_m and df_{λ_m} is the number of nonzero regression coefficients. This forecast combination approach is considered more robust to an inappropriate selection of tuning parameter.

3.3.4 Models

We now list models to be adopted in our subsequent empirical study.

- HAR: the model proposed by Corsi (2009) as in equation (3.1).
- slopeHAR: the slope model proposed by Bollerslev et al. (2016) as in equation (3.6).

- freeHAR: the model proposed by Bollerslev et al. (2016) as in equation (3.5).

- adaptive Lasso AR: a model that is similar to that considered by Audrino and Knaus (2016). In a departure, we apply the adaptive Lasso method rather than the Lasso to equation (3.7) to determine the lag terms. In addition, we choose the tuning parameter based on the one standard error rule via cross-validation. Specifically, we partition the whole sample $\{1, \cdots, T\}$ into K folds randomly where K is often set to five. Among the K subsamples, a single subsample is treated as validation observations for examining the model and the other $K-1$ subsamples are employed as training observations. For each value of λ, we derive the estimate of regression coefficients on the training set and record the squared error on the validation set, denoted as $e_k(\lambda)$. This procedure is then replicated K folds. We compute the averaged validation errors in each fold as $CV_k(\lambda) = \frac{1}{T_k} e_k(\lambda)$ where T_k is the number of observations in the kth fold. The sample standard deviation of $CV_1(\lambda), \cdots, CV_K(\lambda)$ is obtained as $SD(\lambda) = \sqrt{var(CV_1(\lambda), \cdots, CV_K(\lambda))}$. The standard error of $CV(\lambda)$ is given by $SE(\lambda) = SD(\lambda)/\sqrt{K}$. The average error over all folds is given by

$$CV(\lambda) = \frac{1}{T} \sum_{k=1}^{K} e_k(\lambda) \tag{3.19}$$

We first choose a tuning parameter which minimizes the CV error, i.e. $\widehat{\lambda}$=arg $\min_{\lambda \in \{\lambda_1, \cdots, \lambda_m\}} CV(\lambda)$. However, this usually results in insufficient regularization for the purpose of recovering the underlying model. In our practice, we apply the one standard error rule: change λ in order to increase regularization until the following is true

$$CV(\lambda) \leq CV(\widehat{\lambda}) + SE(\widehat{\lambda}) \tag{3.20}$$

In so doing, we obtain the most regularized model which produces the prediction error within one standard error of the minimal error.

- adaptive Lasso HAR: a model that is similar to the flexible HAR model considered by Audrino, Huang, and Okhrin (2016). Assuming that the number of volatility factors to be added in the model is unknown, we use the adaptive Lasso procedure to select the active terms in the equation below

$$RV_{t+1} = \beta_0 + \sum_{i=1}^{p} \beta_i \left(\frac{1}{i} \sum_{j=1}^{i} RV_{t-j+1} \right) + \varepsilon_{t+1} \tag{3.21}$$

- adaptive Lasso slopeHAR: a model which applies the adaptive Lasso method to the additive model of slope RV defined in section 3.3.2

$$RV_{t+1} = \beta_0 + \sum_{i=1}^{p} \beta_i \left(\sum_{j=1}^{i} \left(\frac{i-j+1}{i+(i-1)+\ldots+1} \right) RV_{t-j+1} \right) + \varepsilon_{t+1} \tag{3.22}$$

- adaptive Lasso freeHAR: a model in which we replace the first six terms in equation (3.21) with six daily lagged RV and then apply the adaptive Lasso approach

$$RV_{t+1} = \beta_0 + \sum_{i=1}^{6} \beta_i RV_{t-j+1} + \sum_{i=7}^{p} \beta_i \left(\frac{1}{i} \sum_{j=1}^{i} RV_{t-j+1} \right) + \varepsilon_{t+1} \tag{3.23}$$

- ordered Lasso AR: a model in which we first implement the variable screening and then apply the ordered Lasso to the selected lagged RVs. Specifically, we use the adaptive Lasso technique to select the active terms of equation (3.7), where we term this active set S_1. Since the Lasso method tends to choose only one or few variables from a group of highly correlated variables, we also consider the lagged RVs not in S_1 but which are correlated with the variables in S_1 where the coefficient is greater than 0.65. The new active set is denoted S_2. Section 3.4 shows that the S_2 tends to be consistent with the selection of variables using the cluster group Lasso, which is perceived as a good screening method in Bühlmann et al. (2013). Finally, we employ the ordered Lasso method to enforce an order restriction on the coefficients of the variables in S_2.

- Model-FC: the models described above with the suffix FC are those using the idea of forecast combination introduced in section 3.3.3.

In the following empirical analysis, we employ the Diebold-Mariano (DM) test to evaluate the significance of the forecasting differences. The DM test is considered since the HAR and the Lasso-based approaches are non-nested. First, the difference of the squared forecasting errors is given by

$$d_t = e_{1,t}^2 - e_{2,t}^2,\ t = N+1, \cdots, M \tag{3.24}$$

The DM test statistic is derived as

$$DM = \frac{\frac{1}{M-N}\sum_{t=N+1}^{M} d_t}{\sqrt{\frac{1}{M-N}\left(\widehat{\gamma}_0 + 2\sum_{k=1}^{h-1}\widehat{\gamma}_k\right)}} \tag{3.25}$$

where $M-N$ is the number of OOS forecasts, $\widehat{\gamma}_k$ is the estimated kth autocovariance of the series d_t, and h represents the time horizon. To account for heteroskedastic autocorrelated errors, the DM test statistic can be estimated using the Newey and West (1987) corrected standard errors. Under the null of equal forecasting performance between the HAR and the Lasso-based model, the DM follows an asymptotic standard normal distribution.

3.4 Empirical Application

In this section, we first examine the appropriateness of the lag structure implied by the HAR model in an in-sample analysis using the Lasso-type approaches described in section 3.3.3. We then compare the performance of the HAR model, its two extensions introduced in section 3.3.2 and the Lasso-based model[2] in the OOS volatility forecasts. We consider the daily realized variance measure proposed by Barndorff-Nielsen and Shephard (2002), i.e. the daily RV is equal to the sum of

[2]Throughout our analysis, the Lasso-based model indicates any model using the Lasso-based technique, namely, adaptive Lasso AR, adaptive Lasso HAR, adaptive Lasso slopeHAR, adaptive Lasso freeHAR, ordered Lasso AR and each of these models using the forecast combination.

intraday squared returns. Throughout our analysis, we employ the SPY index and ten individual stocks from different sectors sourced from Tick Data Inc..

Our forecasts are based on direct projection in equation (3.4) over different horizons, i.e. daily, weekly and monthly horizons. Three different sampling frequencies, 30, 300 and 600 seconds, upon which the RV is based, are used in our analysis. We construct the forecasts by re-estimating the parameters of various models every day with a fixed length Rolling Window (RW) and an Increasing Window (IW).

3.4.1 Data Description

In order to obtain the RV, we employ the intraday data of SPY, an exchange traded fund (ETF) which tracks the S&P 500 index closely, Microsoft (MSFT), Citigroup Inc. (C), Pfizer (PFE), General Electric (GE), The Home Depot (HD), Sprint Nextel Corp (S), ExxonMobil (XOM), Alcoa (AA), Wal-Mart (WMT) and Duke Energy (DUK). We consider the time span from Jan 02, 2001 to Nov 15, 2010 with a total of 2483 observations, which is the same as that of the study by Audrino and Knaus (2016). As in the work of Audrino, Camponovo, and Roth (2015), Audrino and Knaus (2016) and Wilms, Rombouts, and Croux (2016), we use the realized variance in logarithmic form, i.e. log RV. In the rest of this chapter, we assume the use of log RV when mentioning RV unless otherwise noted.

Table 3.1 provides summary statistics of the RV series based upon three different sampling frequencies for each of the stocks considered. As the sampling frequency increases, from 600 seconds to 30 seconds, the mean of the RV tends to increase while the standard deviation generally decreases. Almost all series exhibit positive skewness and excess kurtosis. The time series movements of the 300-second RV are presented in Figure 3.1. For the SPY and all individual stocks, an evident increase occurs in the magnitude of the RV during the financial crisis starting from 2007.

3.4.2 In-Sample Analysis

We start by investigating the memory property of the RV series under different market conditions by employing the exact local Whittle estimator of Shimotsu and Phillips (2005) with a commonly adopted bandwidth parameter $m = T^{0.6}$, where T=2483 is the sample size; see Nielsen and Shimotsu (2007), Garvey and Gallagher (2012) and Caporin, Ranaldo, and Santucci de Magistris (2013), among others, for the same choice of the bandwidth parameter. In line with the work of Audrino and Knaus (2016), we set the beginning of the financial crisis to Sep 01, 2007 and split the sample into the pre- and post-crisis periods. In Table 3.2, we show that the long memory estimates, $\widehat{d}$, tend to increase with the sampling frequency, i.e. from 600 seconds to 30 seconds. In addition, the memory property of each series generally strengthens in the post-crisis period, except the RV of C and WMT based on 300- and 600-second frequencies. Values of memory estimates range from 0.6 to 0.8, suggesting the RV series here are non-stationary long-memory processes. Evidence for the long memory and non-stationarity in volatility is also found by Bandi and Perron (2006) and Kellard, Dunis, and Sarantis (2010). To summarize, our results are indicative of substantially different properties of the RV series in pre- and post-crisis periods, which motivates us to examine separately the forecasting performance of various models in these two sub-periods. In the work of Audrino and Knaus (2016), the clear difference in the autocorrelation function (ACF) of the RV for sub-periods is explained by a structural break in the memory of the process occurring around the beginning of 2007.

In the following plot analysis, we choose the 300-second RV of each stock as an illustrative example. Results for other frequencies remain virtually unchanged. First, we provide plots of the partial autocorrelation function (PACF) of the RV during the whole sample, and the pre-crisis and post-crisis periods, in Figures 3.2, 3.3 and 3.4, respectively. We find that, for most stocks, there are some lags beyond 22, with the PACF significantly different from zero, suggesting a possible AR(p^0) for the series of RV where $p^0 > 22$. In the rest of our in-sample evaluation, we make

use of the full sample and adopt various Lasso-based techniques to examine the appropriateness of the lag structure implied by the standard HAR model proposed by Corsi (2009). We let x_t represent the RV_t and $p = 100$ in equation (3.7). In each of the Lasso-based models, the dependent variable is x_{t+1}, i.e. our analysis concentrates on the daily horizons. We then define the HAR implied active set as $S_{HAR} = \{x_{t-1},\ldots, x_{t-22}\}$ and the nonactive set as $S^c_{HAR} = \{x_{t-23},\ldots, x_{t-100}\}$. In Figure 3.5, we present the AR coefficients implied by the HAR model in equation (3.3) as well as those given by the adaptive Lasso approach. In line with the work of Audrino and Knaus (2016), we find that not all of the predictors in S_{HAR} are selected by the adaptive Lasso. Furthermore, for most stocks, a few lags beyond x_{t-23}, i.e. terms in S^c_{HAR}, are selected by the adaptive Lasso as relevant. The results outlined above indicate that the adaptive Lasso does not completely agree with the lag structure implied by the HAR model. To gain a closer look at the lag structure implied by each of the models, in Tables 3.12 and 3.13, we provide the estimates of the AR coefficients for $\{x_{t-1},\ldots,x_{t-100}\}$ of the SPY and MSFT as two examples.

As mentioned in section 3.3.2, Bollerslev et al. (2016) propose two extensions of the HAR model: the freeHAR and slopeHAR respectively allow for a more flexible lag structure and avoid the step-wise abrupt changes associated with the HAR model. In Figure 3.6, we plot their implied AR coefficients for all the stocks considered. First, in contrast to the HAR model in Figure 3.5, the freeHAR does not impose equality restrictions on the coefficients of $\{x_{t-2},\ldots,x_{t-6}\}$ but allow those coefficients to be estimated individually by the OLS technique. Second, the implied AR coefficients of the slopeHAR are continuous and decreasing with the lag length. The enforced smoothness in the dependence of the future RV upon the historical RVs removes predictable jumps and non-monotonicities in the volatility forecasts. Such a feature helps to improve the accuracy of the forecasts, as argued in the work of Bollerslev et al. (2016).

In Figure 3.5, we observe that the lags beyond x_{t-4} are much less frequently selected by the adaptive Lasso method. This result may be attributed to the

invalidity of the adaptive Lasso approach in the context where regressors exhibit non-trivial correlations or almost linear dependence, as noted in section 3.3.3. To deal with the false negatives induced by the use of the adaptive Lasso in selecting relevant predictors, we employ two different Lasso-based methods below.

First, following the work of Audrino, Huang, and Okhrin (2016), we group the lags in AR(100) as {1}, {2-5}, {6-22}, {23-50}, {51,75} and {76-100} and adopt the group Lasso proposed by Yuan and Lin (2006) to further investigate the appropriateness of the lag structure implied by the HAR model. Specifically, the selection of the groups {1}, {2-5} and {6-22} as active lags can be seen as an indication of the validity of the HAR lag structure. We demonstrate the results for the selection of active predictors in Figure 3.7. Since $\{x_{t-1},\ldots, x_{t-5}\}$ is identified as relevant predictors for all stocks considered, we only report the coefficients for lags beyond x_{t-6}. As for the group Lasso AR(1, 5, 22, 50, 75, 100), this is found to be consistent with the HAR lag structure for SPY, PFE, AA, WMT and DUK. Only lags in groups {1} and {2-5} get selected for XOM. The other stocks all have nonzero coefficients for lags after 22. This evidence suggests that the standard HAR model may not be appropriate for the modelling and forecasting of volatility. This is because lags beyond x_{t-22}, which are omitted by the HAR model, may still carry some predictive power for the future RV.

The group Lasso method discussed above adopts the arrangement of groups implied by the HAR model. However, it remains unclear whether such specification for the heterogenous volatility components is inherent to the underlying data or not. In order to answer this question, we apply the cluster Lasso method introduced by Bühlmann et al. (2013) to partition $\{x_{t-1},\ldots,x_{t-100}\}$ into different groups and then select groups as relevant predictors for the future RV using the group Lasso by Yuan and Lin (2006). In Figure 3.7, it is evident that the cluster group Lasso AR is inconsistent with the HAR with regard to the selection of the active lags of the RV for all stocks considered. Furthermore, the cluster group Lasso AR results in the arrangement of groups, which contradicts that implied by the standard HAR model, i.e. {1}, {2-5} and {6-22}, in all cases; see SPY and

MSFT in Tables 3.12 and 3.13 for instance.

Next, we implement the ordered Lasso AR as discussed in section 3.3.4. This estimator accounts for the dynamic nature of volatility by the tendency for the absolute values of the estimated AR coefficients to be monotonically non-increasing with the lag length. We demonstrate the comparison between the cluster group Lasso AR and the ordered Lasso AR in terms of their estimated coefficients in Figure 3.8. The lags of the RV included in the ordered Lasso AR approach are, in most cases, similar to those selected by the cluster group Lasso AR. Bühlmann et al. (2013) consider the latter to be a desirable screening method in a group of highly correlated variables which helps to avoid the problem of false negatives present in the adaptive Lasso estimator. The similarity between the ordered Lasso AR and the cluster group Lasso AR with regard to the selection of relevant predictors can also be found in Tables 3.12 and 3.13. In addition, the ordered Lasso method results in decreasing effects of the past RVs on the future RV, in line with the HAR and slopeHAR models.

Finally, we report the BIC in-sample fit for each of the models using the full sample 300-second RV in Table 3.3. For most stocks over daily and weekly horizons, the slopeHAR clearly dominates the other candidates by having the lowest BIC value. However, the superiority of the slopeHAR no long exists over the monthly horizons, where the adaptive Lasso HAR and the adaptive Lasso HAR-Free perform best. Furthermore, the ordered Lasso AR achieves better in-sample fit than the cluster Group Lasso AR and the group Lasso AR(1, 5, 22, 50, 75, 100), especially over weekly and monthly horizons, where the last two result in rather poor in-sample fit. In the following OOS forecasting exercises, we do not consider the group Lasso AR(1, 5, 22, 50, 75, 100) and the cluster Group Lasso AR. This is because the former specifies the arrangement of volatility groups which may not be appropriate for the underlying data and the latter results in a selection of variables generally captured by the ordered Lasso AR approach, as mentioned earlier.

3.4.3 Out-of-Sample Forecast

In the in-sample analysis above, we provide evidence for the inappropriateness of the lag structure of the standard HAR model in forecasting future RV. In the following subsection, we concentrate on the performance of various models outlined in section 3.3.4 in the OOS volatility forecast. Following Audrino and Knaus (2016), we make rolling window (RW) forecasts of 1000 and 2000 daily observations, respectively, and account for the impact of the financial crisis on the forecasts of future RV. In a departure from the work of Audrino and Knaus (2016), we consider various Lasso-based models and conduct comprehensive empirical exercises to evaluate their forecasting ability for future RV.

First, our analysis focuses on the forecast of RV over daily, weekly and monthly horizons. The forecast is based on direct projection as presented in equation (3.4) where $h = 1$, 5 or 22. Here, longer horizons are of more interest since the HAR model is proposed to accommodate the long-memory properties of the series. Second, our empirical study involves the RV constructed from 30-second, 300-second and 600-second intraday returns. In doing so, we are able to ascertain the effect of the sampling frequency on the forecasting performance of the model selection devices and examine the robustness of our results. Third, we also construct the forecasts by re-estimating the parameters of the regressions each day with an increasing window (IW). Specifically, the first training set includes the first 1000 observations with each subsequent training set containing one more observation. The OOS forecasting performance is measured by the mean square error (MSE). We standardize the MSE of each of the models by the MSE of the HAR model in order to highlight the relative gains.

To obtain the general performance of each of the models in forecasting future RV, we report the results of the SPY in various scenarios in Table 3.4 and present the average loss ratios across all of the 10 individual stocks in Table 3.5. In line with Audrino and Knaus (2016), we set the evaluation window for the case of RW=1000 to May 12, 2009 to Nov 15, 2010, the same as that implied by the case

where RW=2000, so that the results of the RW=1000 and RW=2000 are directly comparable. With the crisis, we concentrate on RW=1000 only and compute the MSE using the maximal evaluation windows in both pre- and post-crisis periods. In each case, the model with the best forecasting performance, i.e. lowest relative MSE, is highlighted in bold blue. As for the SPY in Table 3.4, it is evident that the ordered Lasso AR using the forecast combination provides the best OOS forecasting performance in most cases considered. However, such superior performance is not always observed during the pre- and post-crisis periods. Similar findings in terms of the averaged performance of the models across 10 stocks are shown in Table 3.5. Compared with the case of RW=1000, the averaged performance of the Lasso-based models relative to the HAR tends to improve when RW=2000. For the SPY and 10 individual stocks, the forecast gains of the ordered Lasso AR-FC over the standard HAR generally increase as the forecasting horizons increase, i.e. from $h = 1$ to $h = 22$, except in the pre-crisis period. In most scenarios, the use of the forecast combination results in a better performance of the Lasso-based models. A thorough investigation of the forecasting ability of the Lasso-based models in various situations is presented below.

First, we provide a summary of the forecasting performance of different models in Tables 3.6 and 3.7. Table 3.6 reports the number of cases where each of the models achieves the best forecast and Table 3.7 records the number of cases where the benchmark HAR model is dominated by other candidate. Ordered Lasso AR and its FC version are highlighted due to their superior performance over the other Lasso-based models. Values of the MSE of each of the models against the HAR model are reported in Tables 8(a) to 11(c). We also employ the Diebold-Mariano test using Newey-West standard errors to assess the statistical significance of differences in squared forecasting errors between the HAR and each of the other models. As suggested by Andersen, Bollerslev, and Diebold (2007), the Newey-West heteroskedasticity consistent covariance matrix estimator with 5, 10, and 44 lags are used for the daily ($h = 1$), weekly ($h = 5$) and monthly ($h = 22$) forecasts, respectively.

We start with the forecast of future RV based upon 30-second returns using the full sample period and RW=1000. As can be seen from Table 3.6, the best performance in forecasting future RV is delivered by the Lasso-based model for 9/11 stocks over daily horizons, and for 8/11 stocks over weekly and monthly horizons. Table 3.7 shows that the Lasso-based model using the concept of forecast combination outperforms the HAR model in many cases and that, over long forecasting horizons, this occurs much more frequently than the situation where the extended HAR dominates the standard HAR model. As detailed in Table 8(a), the forecast combination tends to result in more accurate forecasts of future RV for the Lasso-based models. However, there is no clear evidence for which Lasso-based model is consistently the top performer. The superiority of the ordered Lasso AR-FC over the others is observed for 5/11 stocks over daily horizons and for 6/11 stocks over weekly horizons although such gains relative to the HAR model are insignificant in most cases. Over monthly horizons, the adaptive Lasso HAR-FC provides the most accurate forecasts for 4/11 stocks; the ordered Lasso AR and that using the forecasting combination perform best for 2/11 stocks, where the improvements in the OOS volatility forecasts relative to the HAR model are all significant.

We now move to the results with the full sample period and RW=2000. Relative to the case with RW=1000, the Lasso-based model produces the best forecasting performance for more stocks (see Table 3.6) and there tends to be more Lasso-based models which dominate the benchmark HAR model (see Table 3.7). The ordered Lasso AR-FC dominates the other candidates and serves as the top performer in many cases, i.e. 7/11 stocks over daily horizons and 4/11 stocks over weekly horizons. Over monthly horizons, the ordered Lasso AR plays the leading role for 4/11 stocks and its FC version ranks highest for 3/11 stocks. In all these 7 cases, the gains with regard to the forecasting performance are significant.

Given the distinct properties of the RV series in the pre- and post-crisis periods mentioned earlier, we then examine the forecasts made by various models in each of these two sub-periods. In Tables 3.6 and 3.7, we find that the Lasso-based models

lose the forecasting superiority during the pre-crisis period as the forecasting horizons grow. In the case of $h = 22$, for 9/11 stocks the best performance is achieved by either the HAR or the freeHAR. As for the post-crisis period, the ordered Lasso AR-FC model regains its advantages in the forecast of future RV over different horizons. Specifically, it is superior to the other models for 5/11 stocks over daily horizons and for 6/11 stocks over weekly as well as monthly horizons. However, the performance of other Lasso-based models remains disappointing, as evident from Table 3.7.

Our analysis above focuses on the case where the RV series is constructed from the 30-second intraday returns. To examine the impact of the sampling frequency on our results, we next consider the RV computed from the 300-second returns. Looking at the 300-second RV with RW=1000 in Table 3.6, we find that, over different horizons, the best forecasting performance for most stocks is offered by the Lasso-based model. Again, the number of times that the ordered Lasso AR-FC serves as the top performer is generally greater than the other candidates and its improvements over the HAR model are all significant over monthly horizons; see details in Table 9(a). In terms of the results for RW=2000, the conclusions are the same. Compared with the 30-second RV in RW=1000 and RW=2000, there tends to be slightly fewer cases where the most accurate RV forecasts are produced by the Lasso-based model when the RV is based on the 300-second returns.

Next, we split the full sample into the pre- and post-crisis periods. As can be observed from Table 3.6, in contrast to the pre-crisis output based upon the 30-second RV, the ordered Lasso AR-FC model displays the top performance for more stocks during the pre-crisis period in forecasting future 300-second RV over short horizons. It dominates all the other models for 4/11 stocks over daily horizons and for 6/11 stocks over weekly horizons. Consistent with the corresponding case of the 30-second RV, during the pre-crisis period, few of the other Lasso-based models outperform the benchmark HAR model (see Table 3.7) and the superiority of the ordered Lasso-AR-FC is no longer evident over monthly horizons (see Table 3.6). When looking at the results for the 300-second RV in post-crisis period

(see Table 3.6), we again find that the ordered Lasso AR-FC model has the best forecasting performance for most of the stocks considered, i.e. 7/11 over daily horizons, 8/11 over weekly as well as monthly horizons.

Similar investigations are conducted for the case where the RV series are derived from 600-second returns. Starting with the 600-second RV using the full sample and RW=1000, we show that, compared with the situations of 30-second and 300-second RV, the Lasso-based model displays the best forecasting performance for fewer stocks (see Table 3.6). Among the cases where the Lasso-based model achieves the best performance, the most frequent top performer is still the ordered Lasso AR-FC model. Evidence for the role of the Lasso-based model is more clear for the case of the full sample with RW=2000. As can be observed from Table 3.7, relative to the case of RW=1000, there are generally more situations where the Lasso-based model dominates the benchmark HAR model when RW=2000. However, such situations are fewer than those for the case of 30-second and 300-second RV with RW=2000. In Table 3.6, the best forecasting performance for the 600-second RV with RW=2000 is given by the Lasso-based model for 7/11 stocks over daily and monthly horizons, and for 9/11 stocks over weekly horizons. Among these cases, the ordered Lasso AR-FC model most frequently provides the top performance.

As for the impact of the financial crisis on the results of the 600-second RV, we make two observations. First, during the pre-crisis period, generally more Lasso-based models are found (see Table 3.7) to perform better than the benchmark HAR model compared with the case of 300-second RV. In Table 3.6, for the 600-second RV during the pre-crisis period, the superiority of the ordered Lasso AR-FC model is evident for 7/11 stocks over daily horizons, 6/11 stocks over weekly horizons, and only 2/11 stocks over monthly horizons. Second, as can be seen from Table 3.6, during the post-crisis period, the ordered Lasso AR-FC model regains its leading role in forecasting future 600-second RV over monthly horizons by serving as the top performer for 8/11 stocks. However, relative to the pre-crisis period, there are more cases where the slopeHAR provides the best performance

over daily and weekly horizons. In addition, Table 3.7 shows no clear evidence for the enhanced performance of the other Lasso-based models as we move from the pre-crisis period to the post-crisis period.

To sum up, the results from Tables 8(a) to 10(d) show that: first, the ordered Lasso AR-FC model is superior to the other Lasso-based models in forecasting future RV by serving as the top performer for a large fraction of the stocks/index. When the analysis is based on the full sample, for most stocks the gains of the ordered Lasso AR-FC model relative to the HAR model are significant at 1% level over monthly horizons. Second, using the idea of the forecast combination, there tends to be more cases where each of the Lasso-based models dominates the benchmark HAR model. Third, moving from the case of RW=1000 to RW=2000, there are more cases where the Lasso-based model provides the best forecasting performance, which indicates that the Lasso-based model may need a sufficiently large window size to display the advantages in terms of the OOS forecast. Here, our finding is in line with the work of Audrino and Knaus (2016) who show that, for longer training windows, the Lasso outperforms the HAR in most cases.

Fourth, conclusions outlined above are not affected by the variation in the sampling frequency upon which the RV series are based. However, as the sampling frequency increases, i.e. from 600 seconds to 30 seconds, the superiority of the Lasso-based model is generally apparent for more cases using the full sample, whichever window size is used. Fifth, considering three sampling frequencies, most of the Lasso-based models perform rather poorly during the pre-crisis period. The only exception is the ordered Lasso AR-FC model in the case of 300-second and 600-second RV, where it still dominates the others in many cases over daily and weekly horizons. During the post-crisis period, the ordered Lasso AR-FC model regains its superiority by serving as the top performer for far more stocks, especially when $h = 22$, compared with its corresponding performance during the pre-crisis period. However, in the case of 600-second RV, this finding only holds for monthly horizons.

Finally, as a robustness check, we evaluate the forecasts given by each of the

models using an IW approach. The earlier findings suggest that the Lasso-based models display better performance in a larger window size and therefore more clear evidence for the excellent performance of the Lasso-based models are to be expected in the following IW practice. Results of the forecasts across different sampling frequencies are provided in Tables 11(a) to 11(c) and summarized in Tables 3.6 and 3.7. First, in most cases the concept of the forecast combination clearly improves the forecasting performance of the Lasso-based models. As a result (see Table 3.7), compared with the Lasso-based models without using the forecast combination, there is an increasing number of cases where the Lasso-based model outperforms the HAR model once the forecast combination is implemented. In addition, significant gains of the Lasso-based models using the forecasting combination over the benchmark HAR are more frequently observed over monthly horizons; see Tables 11(a), 11(b) and 11(c). Second, the ordered Lasso AR-FC is undoubtedly the most desirable model among all the Lasso-based models, achieving the best performance in most cases as shown in Table 3.6. Third, the impact of the sampling frequency is in line with our earlier results. In Table 3.6, as the sampling frequency increases, i.e. from 600 seconds to 30 seconds, we observe more scenarios where the Lasso-based model, or the ordered Lasso AR-FC model specifically, serves as the top performer.

3.5 Conclusion

We have reviewed the performance of various Lasso-based models in forecasting future (log) realized variance (RV) over different horizons. We employ the RV series of the SPY and ten individual stocks representing ten different sectors using the same sample period as that used by Audrino and Knaus (2016), i,e, from 2001 to 2010 with a total of 2483 observations. We consider three different sampling frequencies upon which the RV is based and construct the forecasts of the RV by both a rolling window of 1000 and 2000 observations, and an increasing window. With the financial crisis, the RV exhibits different memory properties in the pre-

and post-crisis periods. The popular HAR is treated as the benchmark model in our analysis, which reproduces the documented long-memory behaviour of financial volatility. Two recent extensions of the HAR are also adopted for the purpose of comparison.

In our in-sample analysis, we provide evidence, from the perspective of model selection, for the invalidity of the lag structure implied by the HAR model. Neither the maximal AR lag order nor the specification of the heterogenous volatility components of the HAR are in line with those produced by the Lasso-based approaches. In the out-of-sample forecast of future RV using the full time span, the top performance is mostly produced by the Lasso-based model. Forecast combination tends to improve the accuracy of forecasts made by Lasso-based models and result in more cases where the Lasso-based model dominates the standard HAR model.

Compared with the other candidates, we note a generally greater number of occasions in which the ordered Lasso AR using the forecast combination displays the best forecasting performance. Its gains over the HAR model are, in most cases, significant over monthly horizons. For the pre- and post-crisis periods, there are more cases in the post-crisis period where the ordered Lasso AR using the forecast combination outperforms the HAR and delivers the most superior performance, especially over longer forecasting horizons. In addition, with a larger forecasting window size, there tends to be more situations where the Lasso-based model exhibits the top performance. Finally, our conclusions are unaffected by different sampling frequencies. However, as the frequency grows, from 600 seconds to 30 seconds, the superiority of the Lasso-based model becomes more evident over the full time period.

Table 3.1: Summary Statistics of log *RV* Series. This table reports the descriptive statistics of the log realized variance for the SPY and 10 individual stocks, namely: Microsoft (MSFT), Citigroup Inc. (C), Pfizer (PFE), General Electric (GE), The Home Depot (HD), Sprint Nextel Corp (S), ExxonMobil (XOM), Alcoa (AA), Wal-Mart (WMT), Duke Energy (DUK). We consider the same time period as in the work of Audrino and Knaus (2016), from Jan 02, 2001 to Nov 15, 2010 for a total of 2483 daily observations.

	SPY	MSFT	C	PFE	GE	HD	S	XOM	AA	WMT	DUK
						30 sec					
Mean	-0.255	0.896	1.386	0.936	0.868	1.050	1.947	0.594	1.477	0.522	0.987
SD	0.996	0.770	1.479	0.694	1.028	0.852	1.179	0.810	0.863	0.802	0.826
Kurtosis	3.399	3.304	2.520	4.218	3.278	3.437	2.709	4.370	3.993	3.307	4.689
Skewness	0.603	0.751	0.504	0.954	0.703	0.780	0.552	0.919	0.918	0.768	0.994
Median	-0.330	0.770	1.241	0.840	0.783	0.908	1.749	0.463	1.375	0.337	0.847
25%-quantile	-1.013	0.288	0.075	0.428	-0.014	0.399	0.967	0.020	0.822	-0.086	0.417
75%-quantile	0.362	1.344	2.523	1.308	1.520	1.575	2.845	1.066	1.927	1.078	1.424
						300 sec					
Mean	-0.415	0.615	0.986	0.580	0.597	0.880	1.550	0.412	1.300	0.365	0.589
SD	1.027	0.925	1.372	0.807	1.110	0.899	1.142	0.831	0.876	0.863	0.960
Kurtosis	3.596	3.009	3.417	3.725	3.347	3.237	3.146	4.620	4.271	3.368	3.924
Skewness	0.598	0.478	0.666	0.652	0.612	0.604	0.466	0.850	0.891	0.603	0.745
Median	-0.482	0.517	0.897	0.509	0.521	0.756	1.487	0.308	1.192	0.231	0.464
25%-quantile	-1.195	-0.081	-0.152	0.020	-0.290	0.217	0.705	-0.143	0.685	-0.242	-0.099
75%-quantile	0.180	1.209	1.831	1.046	1.291	1.449	2.280	0.865	1.735	0.925	1.140
						600 sec					
Mean	-0.469	0.526	0.892	0.491	0.532	0.825	1.465	0.341	1.240	0.303	0.489
SD	1.044	0.976	1.385	0.845	1.135	0.933	1.179	0.841	0.906	0.892	0.995
Kurtosis	3.644	2.982	3.543	3.669	3.348	3.233	3.136	4.710	3.860	3.374	3.626
Skewness	0.559	0.381	0.682	0.570	0.566	0.513	0.414	0.834	0.745	0.525	0.639
Median	-0.544	0.436	0.795	0.412	0.445	0.709	1.388	0.249	1.140	0.197	0.372
25%-quantile	-1.226	-0.203	-0.188	-0.091	-0.342	0.156	0.629	-0.212	0.617	-0.333	-0.222
75%-quantile	0.143	1.190	1.729	0.977	1.246	1.430	2.199	0.789	1.722	0.862	1.086

Table 3.2: Estimates of Long Memory Parameters. For each log RV series, we obtain the estimate of long memory parameter, $\widehat{d}$, using the univariate exact local Whittle likelihood procedure of Shimotsu and Phillips (2005) with the bandwidth parameter $m = T^{0.6}$ where T is the sample size. The full time span is from Jan 02, 2001 to Nov 15, 2010, with a total of 2483 daily observations. We set the date for the beginning of the financial crisis to Sep 1, 2007.

		SPY	MSFT	C	PFE	GE	HD	S	XOM	AA	WMT	DUK
							30 sec					
$\widehat{d}$	Full	0.703	0.668	0.747	0.680	0.687	0.723	0.679	0.698	0.692	0.691	0.688
	Pre	0.664	0.658	0.678	0.668	0.660	0.681	0.634	0.676	0.647	0.654	0.696
	Post	0.765	0.759	0.723	0.773	0.749	0.758	0.785	0.806	0.697	0.782	0.727
							300 sec					
$\widehat{d}$	Full	0.687	0.647	0.703	0.630	0.638	0.698	0.629	0.646	0.647	0.668	0.614
	Pre	0.661	0.654	0.668	0.623	0.632	0.658	0.623	0.648	0.635	0.636	0.606
	Post	0.747	0.729	0.654	0.715	0.695	0.771	0.716	0.744	0.648	0.647	0.725
							600 sec					
$\widehat{d}$	Full	0.674	0.642	0.682	0.617	0.636	0.681	0.629	0.621	0.625	0.662	0.621
	Pre	0.653	0.651	0.658	0.603	0.633	0.640	0.624	0.620	0.626	0.635	0.602
	Post	0.744	0.703	0.642	0.716	0.680	0.743	0.683	0.726	0.626	0.626	0.674

Table 3.3: BIC In-sample Fit. This table provides the values of the BIC for each of the models considered using the 300-sec log RV over the whole sample period and three forecasting horizons. We concentrate on the daily horizon where h=1. The model with the best in-sample fit is highlighted in bold blue.

horizon=1	SPY	MSFT	C	PFE	GE	HD	S	XOM	AA	WMT	DUK
HAR	-3548	-3707	-3332	-3281	-3276	-3605	-2675	-3744	-3503	-3556	-3153
slopeHAR	**-3558**	**-3719**	**-3346**	**-3292**	**-3300**	**-3618**	**-2690**	**-3750**	**-3523**	**-3569**	**-3157**
freeHAR	-3538	-3689	-3325	-3273	-3267	-3595	-2677	-3722	-3496	-3552	-3135
adaptive Lasso AR	-3487	-3623	-3267	-3226	-3201	-3555	-2609	-3703	-3411	-3477	-3068
adaptive Lasso HAR	-3487	-3643	-3252	-3197	-3210	-3516	-2663	-3636	-3448	-3506	-3040
adaptive Lasso slopeHAR	-3243	-3505	-2274	-3081	-3023	-3379	-2258	-3537	-3256	-3340	-2855
adaptive Lasso freeHAR	-3429	-3645	-3222	-3194	-3194	-3535	-2625	-3678	-3453	-3469	-3044
ordered Lasso AR	-3210	-3380	-2614	-3080	-3103	-3301	-2081	-3415	-3408	-3218	-2988
cluster Group Lasso AR	-3091	-3487	-2789	-2832	-2773	-3223	-2369	-3156	-3070	-3255	-2723
group Lasso AR(1,5,22,50,75,100)	-3286	-3255	-2871	-3017	-2634	-3129	-2215	-3594	-3262	-3319	-2888
horizon=5											
HAR	-4337	-4995	-4053	-4770	-4376	-4939	**-4108**	-4795	-4838	-4972	-4295
slopeHAR	**-4358**	**-5018**	**-4070**	**-4787**	-4412	**-4959**	-4107	**-4823**	**-4858**	**-4993**	**-4298**
freeHAR	-4326	-4985	-4043	-4755	-4372	-4928	-4092	-4792	-4835	-4960	-4278
adaptive Lasso AR	-4274	-4973	-4025	-4733	-4355	-4934	-4035	-4764	-4808	-4894	-4196
adaptive Lasso HAR	-4310	-5014	-4062	-4765	-4383	-4940	-4051	-4770	-4856	-4934	-4211
adaptive Lasso slopeHAR	-4082	-4864	-3139	-4637	-4260	-4813	-3964	-4598	-4628	-4775	-4091
adaptive Lasso freeHAR	-4314	-5009	-4046	-4732	**-4414**	-4950	-4071	-4741	-4838	-4942	-4211
ordered Lasso AR	-3896	-4494	-3396	-4510	-3941	-4538	-3837	-4371	-4602	-4533	-4102
cluster Group Lasso AR	-3955	-4757	-3315	-4376	-3833	-4439	-3663	-4009	-4078	-4598	-3729
group Lasso AR(1,5,22,50,75,100)	-3903	-4565	-3241	-4295	-3764	-4276	-3603	-4119	-4176	-4492	-3803
horizon=22											
HAR	-3719	-4638	-3664	-4862	-4123	-4574	**-4292**	**-4111**	-4739	-4939	**-4165**
slopeHAR	-3712	-4631	-3621	-4852	-4099	-4556	-4244	-4109	-4719	-4928	-4153
freeHAR	-3696	-4617	-3643	-4838	-4103	-4552	-4269	-4092	-4719	-4917	-4141
adaptive Lasso AR	-3731	-4662	-3687	-4801	-4151	-4575	-4180	-4063	-4695	-4865	-4049
adaptive Lasso HAR	-3771	**-4764**	**-3753**	**-4920**	-4228	-4659	-4288	-4074	-4736	**-4985**	-4150
adaptive Lasso slopeHAR	-3623	-4717	-3352	-4737	-4084	-4574	-4199	-4013	-4667	-4889	-4071
adaptive Lasso freeHAR	**-3785**	-4758	-3705	-4896	**-4232**	**-4664**	-4266	-4083	**-4778**	-4973	-4092
ordered Lasso AR	-3431	-4254	-3106	-4692	-3774	-4162	-4049	-3973	-4496	-4558	-3904
cluster Group Lasso AR	-3320	-3963	-2957	-4511	-3598	-4114	-3953	-3685	-4124	-4489	-3584
group Lasso AR(1,5,22,50,75,100)	-3141	-3901	-2910	-4241	-3566	-3983	-3782	-3502	-4102	-4500	-3492

Table 3.4: Summary of the OOS Forecasts: SPY. This table reports the results of the SPY with regard to the ratio of the losses for the different models relative to the losses of the HAR model. For each horizon, the model with the best performance is highlighted in bold blue.

	30 sec					300 sec					600 sec				
	Full Sample			Pre-crisis	Post-crisis	Full Sample			Pre-crisis	Post-crisis	Full Sample			Pre-crisis	Post-crisis
horizon=1	RW=1000	RW=2000	IW			RW=1000	RW=2000	IW			RW=1000	RW=2000	IW		
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
slopeHAR	0.998	0.999	0.999	1.001	0.998	0.998	0.999	0.997	0.996	0.999	0.995	0.995	0.996	0.995	**0.999**
freeHAR	0.994	0.990	0.996	0.997	0.998	1.000	0.996	0.996	0.995	1.003	1.006	1.004	0.997	0.993	1.006
adaptive Lasso AR	1.077	1.021	1.022	1.468	1.183	1.061	1.006	1.010	1.085	1.056	1.052	1.026	1.016	1.037	1.063
adaptive Lasso HAR	1.035	1.046	1.117	1.649	1.348	1.049	1.047	1.080	1.353	1.186	1.015	1.032	1.084	1.200	1.138
adaptive Lasso slopeHAR	1.114	1.106	1.079	1.063	1.116	1.118	1.059	1.030	1.019	1.105	1.121	1.075	1.027	1.002	1.118
adaptive Lasso freeHAR	1.029	1.045	1.094	1.645	1.358	1.044	1.027	1.047	1.326	1.175	1.025	1.026	1.054	1.199	1.130
ordered Lasso AR	0.990	0.989	1.057	1.137	1.042	0.996	0.993	1.013	1.059	1.011	0.995	0.996	1.008	1.037	1.006
adaptive Lasso AR-FC	1.013	0.994	**0.988**	1.323	1.091	1.018	0.995	0.995	1.034	1.022	1.036	1.015	0.998	0.996	1.038
adaptive Lasso HAR-FC	0.995	0.990	1.015	1.583	1.241	1.002	0.995	1.015	1.248	1.098	0.998	1.005	1.019	1.100	1.062
adaptive Lasso slopeHAR-FC	1.063	1.053	1.014	1.004	1.052	1.066	1.054	1.005	1.007	1.078	1.078	1.019	1.008	0.990	1.075
adaptive Lasso freeHAR-FC	1.002	0.997	1.008	1.584	1.252	1.007	0.999	1.003	1.240	1.085	1.006	0.998	1.008	1.099	1.058
ordered Lasso AR-FC	**0.986**	**0.986**	0.996	**0.992**	**0.992**	**0.990**	**0.990**	**0.993**	**0.986**	**0.994**	**0.991**	**0.992**	**0.995**	**0.985**	1.000
horizon=5															
HAR	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
slopeHAR	1.002	1.004	0.998	1.003	0.994	1.002	1.004	0.995	0.995	0.994	0.996	0.999	0.992	0.992	0.992
freeHAR	1.002	1.000	1.002	1.006	1.003	1.002	1.002	0.998	0.998	1.002	0.997	0.998	0.996	0.997	0.999
adaptive Lasso AR	1.047	1.039	1.032	1.637	1.179	1.056	1.021	1.009	1.269	1.112	1.057	1.014	1.011	1.198	1.108
adaptive Lasso HAR	1.066	1.047	1.051	1.642	1.428	1.056	1.045	1.028	1.303	1.234	1.046	1.039	1.034	1.297	1.196
adaptive Lasso slopeHAR	1.151	1.143	1.075	1.146	1.153	1.133	1.128	1.026	1.070	1.154	1.187	1.127	1.027	1.047	1.151
adaptive Lasso freeHAR	1.061	1.045	1.037	1.652	1.447	1.068	1.046	1.016	1.298	1.234	1.038	1.031	1.021	1.288	1.187
ordered Lasso AR	0.986	0.988	1.053	1.186	1.028	0.975	0.998	1.016	1.060	1.009	0.982	0.987	1.002	1.033	0.997
adaptive Lasso AR-FC	0.998	0.993	0.991	1.488	1.107	1.007	1.005	0.993	1.179	1.068	1.026	1.002	0.995	1.134	1.070
adaptive Lasso HAR-FC	1.002	0.986	0.988	1.549	1.298	1.003	0.992	0.990	1.289	1.149	0.993	0.991	0.993	1.312	1.116
adaptive Lasso slopeHAR-FC	1.095	1.075	1.029	1.058	1.102	1.112	1.114	0.994	1.029	1.121	1.152	1.040	0.993	1.014	1.120
adaptive Lasso freeHAR-FC	1.008	0.986	0.982	1.552	1.307	1.007	1.005	0.987	1.283	1.152	0.998	0.990	0.989	1.308	1.124
ordered Lasso AR-FC	**0.978**	**0.982**	**0.980**	1.006	**0.976**	**0.969**	**0.991**	**0.984**	**0.977**	**0.985**	**0.973**	**0.980**	**0.983**	**0.970**	**0.980**
horizon=22															
HAR	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	**1.000**	1.000
slopeHAR	1.004	1.004	1.005	1.029	**0.988**	1.006	1.006	1.007	1.030	0.990	1.001	1.000	1.004	1.026	**0.987**
freeHAR	1.004	1.003	1.004	1.006	1.006	1.004	1.002	1.001	1.003	1.005	1.002	1.001	1.000	1.003	1.003
adaptive Lasso AR	0.975	0.963	0.991	1.472	1.015	0.959	0.946	1.003	1.248	1.007	0.963	0.952	1.007	1.232	1.032
adaptive Lasso HAR	1.029	0.979	0.991	1.431	1.132	1.013	1.003	0.995	1.117	1.144	1.033	1.007	1.008	1.098	1.148
adaptive Lasso slopeHAR	1.024	0.999	0.991	1.187	1.021	1.014	0.959	0.988	1.124	1.031	1.023	0.977	0.990	1.097	1.059
adaptive Lasso freeHAR	1.010	0.975	0.991	1.444	1.139	1.011	0.998	0.983	1.117	1.137	1.026	0.992	0.991	1.094	1.157
ordered Lasso AR	0.954	0.941	1.001	1.263	1.054	0.936	0.929	0.972	1.118	1.013	0.935	0.942	0.966	1.119	1.019
adaptive Lasso AR-FC	0.968	0.952	0.980	1.359	1.002	0.940	0.936	0.991	1.190	1.004	0.963	0.951	0.997	1.174	1.026
adaptive Lasso HAR-FC	0.965	0.953	0.957	1.287	1.097	0.961	0.944	0.960	1.043	1.069	0.979	0.971	0.963	1.045	1.079
adaptive Lasso slopeHAR-FC	1.007	0.996	0.991	1.118	1.017	1.007	0.984	0.976	1.080	1.037	1.026	0.993	0.972	1.067	1.054
adaptive Lasso freeHAR-FC	0.969	0.956	0.960	1.302	1.098	0.969	0.958	0.959	1.050	1.072	0.978	0.968	0.961	1.046	1.080
ordered Lasso AR-FC	**0.937**	**0.929**	**0.948**	1.087	0.992	**0.921**	**0.918**	**0.941**	1.021	**0.972**	**0.919**	**0.930**	**0.944**	1.046	0.988

Table 3.5: Summary of the OOS Forecasts: individual stocks. This table reports the average ratio of the losses for different models relative to the losses of the HAR model across all of the 10 stocks. For each horizon, the model with the best performance is highlighted in bold blue.

	30 sec					300 sec					600 sec				
	Full Sample			Pre-crisis	Post-crisis	Full Sample			Pre-crisis	Post-crisis	Full Sample			Pre-crisis	Post-crisis
horizon=1	RW=1000	RW=2000	IW			RW=1000	RW=2000	IW			RW=1000	RW=2000	IW		
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
slopeHAR	0.994	0.994	0.995	**0.997**	0.993	**0.994**	0.993	0.995	**0.997**	0.993	**0.995**	0.994	**0.995**	**0.997**	**0.993**
freeHAR	0.993	0.991	0.993	0.998	**0.992**	0.997	0.993	0.996	1.003	0.994	0.998	0.996	0.997	1.003	0.996
adaptive Lasso AR	1.588	1.016	1.382	1.239	1.595	1.045	1.020	1.066	1.092	1.137	1.050	1.027	1.045	1.068	1.102
adaptive Lasso HAR	1.217	1.067	1.492	1.382	1.813	1.037	1.026	1.095	1.139	1.231	1.022	1.021	1.072	1.103	1.163
adaptive Lasso slopeHAR	1.121	1.066	1.082	1.067	1.116	1.066	1.039	1.061	1.068	1.089	1.070	1.039	1.060	1.061	1.098
adaptive Lasso freeHAR	1.211	1.055	1.486	1.385	1.811	1.033	1.014	1.090	1.142	1.231	1.020	1.017	1.072	1.108	1.167
ordered Lasso AR	1.079	1.046	1.181	1.121	1.381	1.009	1.003	1.049	1.036	1.106	1.008	1.000	1.030	1.020	1.072
adaptive Lasso AR-FC	1.405	0.984	1.268	1.147	1.410	1.013	1.002	1.033	1.054	1.071	1.022	1.008	1.017	1.032	1.048
adaptive Lasso HAR-FC	1.090	0.994	1.326	1.285	1.595	1.005	0.995	1.035	1.094	1.148	0.999	0.997	1.024	1.063	1.093
adaptive Lasso slopeHAR-FC	1.051	1.024	1.036	1.010	1.055	1.022	1.007	1.027	1.029	1.035	1.025	1.007	1.022	1.022	1.045
adaptive Lasso freeHAR-FC	1.093	0.992	1.323	1.287	1.601	1.003	0.990	1.038	1.098	1.153	1.003	0.997	1.029	1.063	1.098
ordered Lasso AR-FC	**0.991**	**0.983**	**0.988**	1.016	1.045	0.997	**0.988**	**0.994**	1.000	**0.991**	1.003	**0.992**	0.998	1.006	1.001
horizon=5															
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
slopeHAR	0.998	1.000	0.995	**0.996**	**0.992**	1.000	1.000	0.995	0.995	0.993	0.998	0.999	0.994	0.996	0.992
freeHAR	0.997	0.995	0.996	0.999	0.997	1.000	0.997	0.995	0.997	0.996	0.999	0.996	0.994	0.997	0.995
adaptive Lasso AR	1.771	1.025	1.590	1.434	1.675	1.079	1.027	1.106	1.253	1.205	1.093	1.037	1.071	1.187	1.168
adaptive Lasso HAR	1.283	1.038	1.666	1.547	1.938	1.023	1.013	1.124	1.273	1.295	1.008	1.015	1.089	1.196	1.232
adaptive Lasso slopeHAR	1.181	1.097	1.126	1.119	1.155	1.108	1.065	1.075	1.078	1.122	1.108	1.061	1.053	1.078	1.124
adaptive Lasso freeHAR	1.292	1.043	1.659	1.549	1.934	1.025	1.006	1.115	1.269	1.296	1.011	1.010	1.079	1.192	1.230
ordered Lasso AR	1.107	1.056	1.204	1.204	1.423	1.010	0.997	1.058	1.036	1.136	1.014	0.996	1.029	1.013	1.089
adaptive Lasso AR-FC	1.559	0.997	1.454	1.329	1.503	1.058	1.006	1.065	1.187	1.139	1.061	1.016	1.038	1.136	1.116
adaptive Lasso HAR-FC	1.151	0.983	1.507	1.454	1.739	1.001	0.987	1.073	1.222	1.222	**0.989**	0.996	1.042	1.152	1.167
adaptive Lasso slopeHAR-FC	1.096	1.045	1.079	1.058	1.087	1.069	1.029	1.035	1.034	1.078	1.064	1.008	1.017	1.033	1.083
adaptive Lasso freeHAR-FC	1.159	0.987	1.506	1.454	1.734	1.004	0.987	1.070	1.222	1.226	0.992	0.997	1.041	1.151	1.171
ordered Lasso AR-FC	**0.986**	**0.977**	**0.980**	1.060	1.039	**0.993**	**0.976**	**0.979**	**0.982**	**0.990**	1.000	**0.984**	**0.981**	**0.989**	**0.991**
horizon=22															
HAR	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	**1.000**	1.000	**1.000**	1.000	1.000	**1.000**	1.000
slopeHAR	1.011	1.012	1.009	1.027	**0.997**	1.014	1.015	1.013	1.022	1.006	1.010	1.013	1.012	1.023	1.004
freeHAR	1.001	1.000	1.000	1.002	1.002	1.003	1.001	0.999	1.001	1.001	1.001	1.000	0.999	1.001	1.000
adaptive Lasso AR	1.504	0.997	1.392	1.404	1.334	1.106	1.042	1.097	1.348	1.152	1.117	1.050	1.076	1.287	1.159
adaptive Lasso HAR	1.199	1.022	1.441	1.396	1.516	1.075	1.051	1.077	1.294	1.231	1.075	1.060	1.063	1.230	1.212
adaptive Lasso slopeHAR	1.100	1.015	1.088	1.176	1.081	1.110	1.064	1.050	1.149	1.093	1.124	1.086	1.054	1.117	1.116
adaptive Lasso freeHAR	1.204	1.019	1.433	1.397	1.510	1.080	1.050	1.068	1.293	1.235	1.081	1.059	1.055	1.227	1.217
ordered Lasso AR	1.059	0.978	1.095	1.318	1.200	1.022	0.990	1.014	1.113	1.072	1.028	1.007	1.008	1.075	1.045
adaptive Lasso AR-FC	1.372	0.977	1.312	1.320	1.243	1.086	1.029	1.064	1.277	1.107	1.093	1.037	1.048	1.227	1.115
adaptive Lasso HAR-FC	1.071	0.965	1.313	1.293	1.379	1.037	1.011	1.021	1.210	1.144	1.040	1.025	1.009	1.147	1.135
adaptive Lasso slopeHAR-FC	1.087	1.019	1.059	1.118	1.047	1.091	1.044	1.018	1.093	1.064	1.108	1.055	1.026	1.065	1.084
adaptive Lasso freeHAR-FC	1.076	0.968	1.313	1.295	1.380	1.044	1.012	1.019	1.210	1.147	1.043	1.026	1.008	1.146	1.139
ordered Lasso AR-FC	**0.978**	**0.955**	**0.958**	1.185	1.010	**0.996**	**0.980**	**0.957**	1.074	**0.969**	1.006	**0.997**	**0.964**	1.059	**0.965**

Table 3.6: This table provides the number of times that the HAR, slopeHAR, freeHAR and the Lasso-based model serve as the top performer for the SPY and 10 individual stocks (11 log RV series). In addition, the performance of the ordered Lasso AR and that using the forecast combination are also summarized here, which clearly dominate the other Lasso-based models.

	30 sec					300 sec					600 sec				
	Full Sample			Pre-crisis	Post-crisis	Full Sample			Pre-crisis	Post-crisis	Full Sample			Pre-crisis	Post-crisis
	RW=1000	RW=2000	IW			RW=1000	RW=2000	IW			RW=1000	RW=2000	IW		
horizon=1															
HAR	0	0	0	1	1	1	0	1	1	1	2	0	0	1	1
slopeHAR	0	0	0	0	3	3	0	3	5	1	2	2	5	2	5
freeHAR	2	0	0	1	2	1	3	1	1	2	1	2	0	1	2
Lasso-based model	9	11	11	9	5	6	8	6	4	7	6	7	6	7	3
ordered Lasso AR	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
ordered Lasso AR-FC	5	7	7	2	5	2	3	6	4	7	3	4	6	7	3
horizon=5															
HAR	1	0	0	1	0	1	1	0	0	0	2	1	0	0	0
slopeHAR	1	0	0	4	3	0	0	1	2	2	2	1	2	1	4
freeHAR	1	1	0	1	2	0	0	0	0	1	0	0	0	1	1
Lasso-based model	8	10	11	5	6	10	10	10	9	8	7	9	9	9	6
ordered Lasso AR	0	0	0	1	0	2	2	0	1	0	0	3	0	2	0
ordered Lasso AR-FC	6	4	8	2	6	4	4	9	6	8	3	4	7	6	6
horizon=22															
HAR	3	1	0	8	1	2	2	1	7	1	3	1	1	3	0
slopeHAR	0	0	0	0	2	1	0	0	2	0	2	1	0	1	1
freeHAR	0	0	1	1	1	0	0	0	0	0	1	2	0	0	0
Lasso-based model	8	10	10	2	7	8	9	10	2	10	5	7	10	7	10
ordered Lasso AR	2	4	1	1	0	2	4	0	0	1	2	3	0	2	1
ordered Lasso AR-FC	2	3	9	0	6	4	4	8	0	8	2	3	6	2	8

Table 3.7: This table provides the number of times that each of the models outperforms the benchmark HAR for the SPY and 10 individual stocks (11 log RV series).

	30 sec					300 sec					600 sec				
	Full Sample			Pre-crisis	Post-crisis	Full Sample			Pre-crisis	Post-crisis	Full Sample			Pre-crisis	Post-crisis
	RW=1000	RW=2000	IW			RW=1000	RW=2000	IW			RW=1000	RW=2000	IW		
horizon=1															
slopeHAR	8	8	10	7	10	10	10	10	7	10	8	9	11	8	10
freeHAR	9	10	11	6	8	6	9	9	3	7	6	7	9	4	7
adaptive Lasso AR	0	3	0	2	0	0	1	0	0	0	0	1	0	0	0
adaptive Lasso HAR	1	0	0	0	0	2	1	0	0	0	2	1	0	0	0
adaptive Lasso slopeHAR	0	1	0	0	0	0	2	0	0	0	0	2	0	0	0
adaptive Lasso freeHAR	3	1	0	0	0	3	2	0	0	0	0	1	0	0	0
ordered Lasso AR	3	4	0	0	0	7	8	3	1	2	4	5	3	3	2
adaptive Lasso AR-FC	2	10	8	2	1	2	4	3	0	1	1	3	4	1	0
adaptive Lasso HAR-FC	8	7	1	0	0	4	7	1	0	0	5	6	0	0	0
adaptive Lasso slopeHAR-FC	0	2	1	5	0	2	3	0	0	0	1	1	1	1	0
adaptive Lasso freeHAR-FC	5	8	1	0	0	5	9	1	0	0	5	7	0	0	0
ordered Lasso AR-FC	9	11	10	5	7	7	9	8	5	8	5	6	6	7	4
horizon=5															
slopeHAR	6	4	9	8	10	7	6	10	8	9	6	6	8	8	9
freeHAR	6	7	10	5	7	7	7	11	8	8	8	9	10	10	10
adaptive Lasso AR	1	3	3	0	0	1	2	1	1	0	0	1	3	1	0
adaptive Lasso HAR	1	3	1	0	0	2	2	1	0	0	4	2	1	1	0
adaptive Lasso slopeHAR	0	0	0	0	0	1	1	0	2	0	1	2	0	1	0
adaptive Lasso freeHAR	1	2	1	0	0	2	2	1	0	0	3	2	3	0	0
ordered Lasso AR	5	7	0	1	0	5	7	2	4	1	5	8	3	5	2
adaptive Lasso AR-FC	4	4	6	1	1	2	3	6	1	0	2	4	7	1	0
adaptive Lasso HAR-FC	6	10	5	0	0	9	9	7	2	1	7	5	5	1	0
adaptive Lasso slopeHAR-FC	1	1	1	3	0	2	4	4	2	0	2	3	5	2	0
adaptive Lasso freeHAR-FC	5	10	5	0	0	5	7	7	2	2	6	5	6	2	1
ordered Lasso AR-FC	7	10	11	3	7	9	9	11	11	8	7	9	10	8	8
horizon=22															
slopeHAR	2	2	1	0	7	3	4	1	2	5	3	3	2	2	6
freeHAR	2	5	4	1	1	0	4	8	2	0	3	6	9	1	4
adaptive Lasso AR	4	7	6	1	1	3	6	4	0	0	2	3	3	0	0
adaptive Lasso HAR	4	8	5	0	0	2	3	4	0	0	3	3	3	0	0
adaptive Lasso slopeHAR	2	6	4	0	0	2	3	4	1	0	1	2	5	0	0
adaptive Lasso freeHAR	3	8	5	0	0	2	4	4	0	0	3	4	4	0	0
ordered Lasso AR	5	8	2	1	0	6	8	5	1	2	5	7	4	2	2
adaptive Lasso AR-FC	6	7	7	1	3	4	5	7	0	2	3	4	6	0	1
adaptive Lasso HAR-FC	8	9	6	0	2	7	7	9	0	4	5	5	9	1	4
adaptive Lasso slopeHAR-FC	2	7	5	2	2	3	6	7	1	2	2	5	4	3	1
adaptive Lasso freeHAR-FC	8	9	6	0	2	6	6	9	0	3	5	5	9	1	3
ordered Lasso AR-FC	7	9	10	0	8	6	7	10	1	10	5	6	10	3	10

Table 8(a): OOS Forecasts based on Full Sample 30-sec log RV and RW=1000. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with a fixed length Rolling Window (RW). Consistent with the work of Audrino and Knaus (2016), we set the evaluation window to (May 12, 2009 to Nov 15, 2010), in order to make the results comparable to those obtained from the RW=2000. We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.998	1.000	0.993	0.987	0.994	0.989	0.994	1.002	1.005	0.989	0.991	8/11
freeHAR	0.994	1.010	**0.991**	0.997	0.997	0.966	**0.985**	1.003	0.999	0.992	0.993	9/11
adaptive Lasso AR	1.077	1.042	5.151***	1.036	1.097	1.014	2.260***	1.068	1.052	1.076	1.079	0/11
adaptive Lasso HAR	1.035	1.078	2.454***	1.086	1.009	0.992	1.382***	1.032	1.058	1.016	1.062	1/11
adaptive Lasso slopeHAR	1.114	1.097	1.177**	1.064	1.154	1.101	1.201**	1.137	1.068	1.092	1.121	0/11
adaptive Lasso freeHAR	1.029	1.049	2.439***	1.077	1.012	1.006	1.360***	1.033	1.016	1.030	1.082	3/11
ordered Lasso AR	0.990	1.029	1.319	1.033	0.992	0.965	1.091	1.040	1.014	1.161	1.150	3/11
adaptive Lasso AR-FC	1.013	1.017	4.131***	0.998	1.010	0.975	1.872***	1.010	1.018	1.017	1.007	2/11
adaptive Lasso HAR-FC	0.995	1.021	1.812***	**0.978**	0.989**	**0.953**	1.180*	0.997	0.990	**0.985**	0.996	8/11
adaptive Lasso slopeHAR-FC	1.063	1.051	1.124	1.005	1.067	1.027	1.091	1.088**	1.026	1.010	1.018	0/11
adaptive Lasso freeHAR-FC	1.002	1.009	1.814***	0.987	0.993	0.964	1.173	1.013	**0.977***	0.999	1.005	5/11
ordered Lasso AR-FC	**0.986**	**0.996**	1.005	0.981	**0.986**	0.964	0.998	**0.996**	1.006	0.993	**0.986**	9/11
horizon=5												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	
slopeHAR	1.002	1.001	1.001	0.993	0.994	1.000	**0.992**	0.998	1.016	0.998	0.991	6/11
freeHAR	1.002	1.005	**0.998**	0.991	0.994	0.995	1.000	1.000	1.001	0.993	0.992	6/11
adaptive Lasso AR	1.047	1.029	5.866***	0.980	1.116	1.060	3.285***	1.034	1.150*	1.153	1.035	1/11
adaptive Lasso HAR	1.066	1.038	3.159***	1.001	1.007	0.983	1.500***	1.028	1.013	1.029	1.068*	1/11
adaptive Lasso slopeHAR	1.151	1.111	1.340***	1.042**	1.228	1.151**	1.412***	1.212	1.163**	1.099	1.052*	0/11
adaptive Lasso freeHAR	1.061**	1.019	3.195***	1.024	1.009	0.999	1.505***	1.023	1.017	1.048	1.082*	1/11
ordered Lasso AR	0.986	0.987	1.446***	0.957**	0.997	0.972	1.167***	1.039	1.022	1.293	1.187	5/11
adaptive Lasso AR-FC	0.998	1.007	4.793***	0.942	1.043	1.010	2.693***	0.989	1.072	1.070	0.967	4/11
adaptive Lasso HAR-FC	1.002	1.004	2.353***	0.956**	0.989	**0.937**	1.301***	0.974	1.023	**0.984**	0.994	6/11
adaptive Lasso slopeHAR-FC	1.095	1.064	1.293***	1.021	1.107	1.045	1.238***	1.096	1.071	1.033	0.997	1/11
adaptive Lasso freeHAR-FC	1.008	1.002	2.357***	0.964**	0.993	0.947	1.306***	0.997	1.029	0.998	1.000	5/11
ordered Lasso AR-FC	**0.978**	**0.973**	1.042	**0.934****	**0.976**	0.973	1.021	**0.966**	1.006	1.009	**0.963**	7/11
horizon=22												
HAR	1.000	1.000	**1.000**	1.000	**1.000**	1.000	**1.000**	1.000	1.000	1.000	1.000	
slopeHAR	1.004	1.002	1.029	0.997	1.021	1.000	1.021***	1.003	1.037***	0.998	1.004*	2/11
freeHAR	1.004	1.006	1.000	0.998	1.002	1.002	1.001	1.002	1.005	1.000	0.998	2/11
adaptive Lasso AR	0.975**	0.954	3.959***	0.888***	1.118	0.972	2.923***	1.009	1.136**	1.083**	1.002	4/11
adaptive Lasso HAR	1.029	1.045***	2.515***	0.993***	1.126	0.999	1.260***	0.958**	0.995***	1.054	1.044	4/11
adaptive Lasso slopeHAR	1.024	1.020	1.180***	0.962***	1.173*	1.059	1.379***	1.046	1.100	1.083**	0.998*	2/11
adaptive Lasso freeHAR	1.010	1.043**	2.527***	0.996***	1.164	1.008	1.255***	0.950***	0.990***	1.066	1.043	3/11
ordered Lasso AR	0.954***	**0.873*****	1.440***	**0.783*****	1.090	0.937***	1.199***	0.947***	1.036	1.233***	1.049	5/11
adaptive Lasso AR-FC	0.968***	0.968*	3.394***	0.888***	1.084	0.972	2.367***	0.952***	1.084	1.044	0.965	6/11
adaptive Lasso HAR-FC	0.965***	0.958***	1.919***	0.947***	1.064*	**0.915*****	1.051	**0.915*****	**0.988*****	**0.963****	0.992*	8/11
adaptive Lasso slopeHAR-FC	1.007	1.011	1.344***	1.005***	1.154	1.015	1.289***	0.984	1.046*	1.035	0.990*	2/11
adaptive Lasso freeHAR-FC	0.969***	0.960***	1.914***	0.950***	1.069*	0.924***	1.054	0.926***	0.993***	0.977	0.992**	8/11
ordered Lasso AR-FC	**0.937*****	0.943***	1.076*	0.873***	1.039*	0.950**	1.034	0.923***	1.000**	0.990*	**0.953*****	7/11

Table 8(b): OOS Forecasts based on Full Sample 30-sec log RV and RW=2000. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with a fixed length Rolling Window (RW). Consistent with the work of Audrino and Knaus (2016), we set the evaluation window to (May 12, 2009 to Nov 15, 2010). We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.999	1.000	0.989	0.983	0.994	0.987	0.995	1.002	1.006	0.989	0.993	8/11
freeHAR	0.990	1.009	0.985	0.996	0.994	0.968	0.988	0.995	0.993	0.988	0.994	10/11
adaptive Lasso AR	1.021	1.027	1.044	0.974	1.034	0.983	1.042	1.015	0.993	1.022	1.022	3/11
adaptive Lasso HAR	1.046	1.050	1.082	1.189	1.007	1.002	1.088	1.065	1.058	1.046	1.080	0/11
adaptive Lasso slopeHAR	1.106	1.067	1.093	1.027	1.101	1.025	1.071	1.060	0.994	1.101	1.123	1/11
adaptive Lasso freeHAR	1.045	1.029	1.099	1.185	0.998	1.001	1.061	1.070	1.001	1.021	1.087	1/11
ordered Lasso AR	0.989	1.053	1.032	1.037	0.988	0.981	0.982**	1.075	1.000	1.245	1.065	4/11
adaptive Lasso AR-FC	0.994	1.007	**0.972**	0.975	0.988	0.973	0.985*	**0.982**	**0.980**	0.988	0.991	10/11
adaptive Lasso HAR-FC	0.990	1.013	1.005	1.017	0.985	0.956	0.985	1.001	0.996	0.994	0.989	7/11
adaptive Lasso slopeHAR-FC	1.053	1.014	1.097	1.001	1.054	0.986	1.003	1.010	0.987	1.003	1.083	2/11
adaptive Lasso freeHAR-FC	0.997	0.998	1.006	1.018	0.987	0.957	**0.976**	1.018	0.983	0.987	0.996	8/11
ordered Lasso AR-FC	**0.986**	**0.997**	0.992*	**0.967**	**0.983**	**0.959**	0.984	0.995	0.986	**0.986**	**0.978**	11/11
horizon=5												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	1.004	1.002	1.000	0.987*	0.996	1.003	0.997	1.002	1.022	1.000	0.990	4/11
freeHAR	1.000	1.000	**0.993**	0.983*	0.992	0.994	1.003	0.995	1.001	0.992	0.992	7/11
adaptive Lasso AR	1.039	1.016	1.011	**0.908****	1.062	1.036	1.120**	0.994	1.038	1.076	0.993	3/11
adaptive Lasso HAR	1.047	1.019	1.212***	1.060*	1.005	0.997	0.978**	1.032*	0.991	1.045*	1.043*	3/11
adaptive Lasso slopeHAR	1.143	1.100	1.131**	1.137***	1.198	1.078*	1.080	1.001	1.047	1.092	1.110	0/11
adaptive Lasso freeHAR	1.045*	1.009	1.298***	1.043	1.014	1.006	0.982***	1.011	0.982	1.050*	1.040	2/11
ordered Lasso AR	0.988	0.996	1.120**	0.949**	0.988	0.987	0.987**	1.066	0.975	1.421	1.068	7/11
adaptive Lasso AR-FC	0.993	1.007	1.038	0.923**	1.010	1.001	1.028	0.967	1.006	1.028	0.961	4/11
adaptive Lasso HAR-FC	0.986	0.983	1.053	0.978**	**0.973***	0.965	**0.970*****	0.965*	0.975**	**0.992**	0.972	10/11
adaptive Lasso slopeHAR-FC	1.075	1.051	1.234***	1.003	1.077	**0.963**	1.007	1.009	1.020	1.008	1.077*	1/11
adaptive Lasso freeHAR-FC	0.986	0.975	1.069	0.979**	0.976**	0.971	0.972***	0.987	**0.972****	0.995	0.974	10/11
ordered Lasso AR-FC	**0.982**	**0.974**	1.032	0.939**	0.974	0.967	0.988**	**0.965***	0.983	0.996	**0.951***	10/11
horizon=22												
HAR	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	1.004	1.003	1.026	0.997*	1.023	1.004	1.026***	1.006	1.043***	0.993	1.001**	2/11
freeHAR	1.003	1.003	0.997	0.996	1.001	1.001	1.002	0.999	1.005	0.997	0.997	5/11
adaptive Lasso AR	0.963***	0.949	**0.970**	0.871***	1.076	0.938***	1.125***	0.968*	1.093**	1.012	0.967	7/11
adaptive Lasso HAR	0.979**	0.988***	1.346***	0.998***	1.079	0.979*	0.964**	0.942***	0.952***	1.024	0.946***	8/11
adaptive Lasso slopeHAR	0.999	0.988	1.004	0.930***	1.151	1.054	0.968**	0.987	1.047	1.057**	0.962	6/11
adaptive Lasso freeHAR	0.975**	0.970***	1.381***	0.979***	1.101	0.975	0.962**	0.909***	0.957***	1.023	0.937***	8/11
ordered Lasso AR	0.941***	**0.841*****	1.219***	**0.811*****	1.039**	**0.901*****	0.943***	0.960***	**0.896*****	1.228	0.946***	8/11
adaptive Lasso AR-FC	0.952***	0.966**	1.025*	0.883***	1.044*	0.941***	1.007	0.929***	1.055	0.995	0.924***	7/11
adaptive Lasso HAR-FC	0.953***	0.949***	1.138**	0.947***	1.028**	0.931***	**0.941*****	0.902***	0.974***	**0.929****	0.907***	9/11
adaptive Lasso slopeHAR-FC	0.996	0.996**	1.217***	1.010**	1.172	1.029	0.957***	0.926***	0.986***	0.969	0.927***	7/11
adaptive Lasso freeHAR-FC	0.956***	0.948***	1.139**	0.961***	1.028**	0.933***	0.942***	0.907***	0.974***	0.937**	0.909***	9/11
ordered Lasso AR-FC	**0.929*****	0.935***	1.078**	0.896***	1.008**	0.934***	0.950***	**0.900*****	0.987***	0.958**	**0.902*****	9/11

Table 8(c): OOS Forecasts based on Pre-Crisis 30-sec log RV and RW=1000. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with a fixed length Rolling Window (RW). Consistent with the work of Audrino and Knaus (2016), here we set the evaluation window to (May 20, 2005 to Aug 31, 2007), with a total of 575 observations. We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	1.001	0.993	0.996	1.002	0.998	0.995	0.997	1.006	0.992	0.996	1.000	7/11
freeHAR	0.997	**0.992**	0.988	1.005	1.005	0.996	1.005	1.003	0.991	1.001	0.998	6/11
adaptive Lasso AR	1.468	1.490**	1.590	1.138	1.726	1.139	1.102	0.992	0.997	1.080	1.139	2/11
adaptive Lasso HAR	1.649*	1.486**	1.737*	1.254***	1.845*	1.309**	1.262*	1.187**	1.191	1.248**	1.299	0/11
adaptive Lasso slopeHAR	1.063	1.133	1.056	1.107	1.061	1.032	1.069	1.048	1.010	1.009	1.141	0/11
adaptive Lasso freeHAR	1.645*	1.493**	1.740*	1.256**	1.845*	1.315**	1.270*	1.192***	1.184*	1.263***	1.296	0/11
ordered Lasso AR	1.137	1.324**	1.057	1.190	1.316	1.079	1.031	1.016	1.093	1.075	1.034	0/11
adaptive Lasso AR-FC	1.323	1.344	1.406	1.031	1.553	1.076	1.014	**0.982**	**0.973**	1.016	1.076	2/11
adaptive Lasso HAR-FC	1.583*	1.416*	1.656*	1.040*	1.722	1.263**	1.181***	1.048**	1.038	1.225**	1.261	0/11
adaptive Lasso slopeHAR-FC	1.004	1.066	**0.981**	1.012	**0.993**	**0.992**	**0.992**	1.037	1.000	**0.965**	1.060	5/11
adaptive Lasso freeHAR-FC	1.584*	1.425*	1.658*	1.043**	1.726	1.273**	1.172***	1.059**	1.018	1.230**	1.264	0/11
ordered Lasso AR-FC	**0.992**	1.047*	1.000	1.003	1.120	1.002	0.995	0.995	0.998	1.007	**0.995**	5/11
horizon=5												
HAR	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	1.003	0.993	**0.995***	1.010**	**0.998***	**0.986**	1.005	0.999	**0.988***	0.991	0.991	8/11
freeHAR	1.006	**0.991**	0.997	1.001	1.001	1.001	0.998	0.996	1.002	0.997	1.002	5/11
adaptive Lasso AR	1.637	1.877***	1.770	1.311	2.027	1.386	1.270	1.003	1.079	1.287	1.326*	0/11
adaptive Lasso HAR	1.642	1.976***	2.006	1.264	2.122	1.395	1.327	1.387***	1.362	1.345	1.288**	0/11
adaptive Lasso slopeHAR	1.146	1.189*	1.123	1.064	1.170	1.103	1.085	1.025	1.109	1.013	1.311**	0/11
adaptive Lasso freeHAR	1.652	1.971***	1.999	1.272	2.130	1.391	1.314	1.403***	1.364	1.349	1.293**	0/11
ordered Lasso AR	1.186	1.564***	1.170**	1.285	1.522***	1.140*	1.015	**0.939*** **	1.164	1.197***	1.041	1/11
adaptive Lasso AR-FC	1.488	1.709***	1.579	1.199	1.810	1.315	1.200	0.967*	1.040	1.220	1.252	1/11
adaptive Lasso HAR-FC	1.549	1.813***	1.885	1.195	1.925	1.331	1.276	1.299**	1.297	1.290	1.232**	0/11
adaptive Lasso slopeHAR-FC	1.058	1.112	1.043	**0.992**	1.084	1.048	1.025	0.995	1.068	**0.987**	1.221*	3/11
adaptive Lasso freeHAR-FC	1.552	1.829***	1.881	1.195	1.928	1.330	1.267	1.298**	1.291	1.292	1.226**	0/11
ordered Lasso AR-FC	1.006	1.105***	1.083**	1.028	1.250***	1.030*	**0.982**	0.987	1.064	1.080***	**0.990**	3/11
horizon=22												
HAR	**1.000**	1.000	**1.000**	**1.000**	**1.000**	1.000	**1.000**	1.000	**1.000**	**1.000**	**1.000**	
slopeHAR	1.029**	1.016***	1.017**	1.037**	1.037**	1.038**	1.014	1.015*	1.020	1.068	1.004	0/11
freeHAR	1.006	**0.998**	1.002	1.001	1.006	1.003	1.002	1.003	1.004	1.000	1.001	1/11
adaptive Lasso AR	1.472	1.716***	1.478	1.635***	1.997***	1.349	1.276	0.957***	1.085**	1.238	1.309**	1/11
adaptive Lasso HAR	1.431	1.839***	1.542***	1.492***	2.025***	1.175	1.208*	1.035	1.242***	1.068**	1.333***	0/11
adaptive Lasso slopeHAR	1.187	1.214	1.149	1.290***	1.245*	1.073	1.147	1.002	1.171*	1.050	1.416***	0/11
adaptive Lasso freeHAR	1.444	1.850***	1.546***	1.505***	2.026***	1.177	1.219*	1.030	1.229***	1.067**	1.324***	0/11
ordered Lasso AR	1.263***	1.649***	1.245***	1.518***	1.798***	1.237***	1.074	**0.935*** **	1.190*	1.453***	1.085	1/11
adaptive Lasso AR-FC	1.359	1.560*	1.340	1.534***	1.802***	1.264	1.223	0.961**	1.066**	1.177	1.277**	1/11
adaptive Lasso HAR-FC	1.287	1.688***	1.428**	1.360***	1.799***	1.111*	1.143**	1.033*	1.179***	1.005***	1.188***	0/11
adaptive Lasso slopeHAR-FC	1.118	1.141	1.095	1.198***	1.182*	**0.981**	1.082	0.980	1.153**	1.005	1.361***	2/11
adaptive Lasso freeHAR-FC	1.302	1.690***	1.430**	1.376***	1.798***	1.108*	1.141**	1.033*	1.177***	1.004***	1.192***	0/11
ordered Lasso AR-FC	1.087***	1.216***	1.170***	1.235***	1.519***	1.127***	1.094	1.009**	1.131**	1.327***	1.021*	0/11

Table 8(d): OOS Forecasts based on Post-Crisis 30-sec log RV and RW=1000. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with a fixed length Rolling Window (RW). Consistent with the work of Audrino and Knaus (2016), here we set the evaluation window to (Sep 04, 2007 to Nov 15, 2010), with a total of 808 observations. We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	
slopeHAR	0.998	0.996	**0.998**	0.989**	**0.993**	0.993*	**0.989**	1.005	0.990	0.990	0.986**	10/11
freeHAR	0.998	1.001	1.000	0.994	0.998	**0.977**	0.990	1.001	0.984	**0.988***	0.988	8/11
adaptive Lasso AR	1.183	1.094	3.813***	1.231**	1.760***	1.083	2.634***	1.098	1.100**	1.096	1.043	0/11
adaptive Lasso HAR	1.348***	1.268***	4.218***	1.260**	2.295***	1.231***	3.071***	1.123**	1.362***	1.134**	1.174**	0/11
adaptive Lasso slopeHAR	1.116	1.112	1.071	1.160**	1.101	1.101	1.188	1.121	1.096**	1.130**	1.077	0/11
adaptive Lasso freeHAR	1.358***	1.274***	4.220***	1.271**	2.281***	1.230***	3.056***	1.116*	1.345***	1.133**	1.187**	0/11
ordered Lasso AR	1.042	1.442***	1.645***	1.666***	1.125**	1.295***	1.212	1.434**	1.163***	1.546***	1.287**	0/11
adaptive Lasso AR-FC	1.091	1.026	3.174***	1.057	1.547***	1.022	2.196***	1.024	1.031	1.041	0.981	1/11
adaptive Lasso HAR-FC	1.241***	1.186***	3.708***	1.050	2.012***	1.094***	2.629***	1.027*	1.187***	1.035	1.023	0/11
adaptive Lasso slopeHAR-FC	1.052	1.046	1.026	1.042	1.058	1.054*	1.122	1.055	1.059*	1.059*	1.027	0/11
adaptive Lasso freeHAR-FC	1.252***	1.189**	3.704***	1.059	2.011***	1.100***	2.626***	1.046	1.190***	1.040	1.043	0/11
ordered Lasso AR-FC	**0.992**	**0.995**	1.341***	**0.977****	1.068	0.984	1.151	1.011	**0.968**	0.990	**0.970****	7/11
horizon=5												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.994	0.993	1.003**	0.985**	**0.993**	0.997	**0.988**	**0.995**	0.992	0.994	0.981**	10/11
freeHAR	1.003	0.999	**0.998**	0.993	0.999	0.997	0.998	1.000	0.995	**0.992**	0.995	7/11
adaptive Lasso AR	1.179***	1.106	3.568***	1.083*	1.812***	1.272***	3.614***	1.096***	1.106*	1.088	1.003	0/11
adaptive Lasso HAR	1.428***	1.254***	3.911***	1.280***	2.384***	1.470***	4.423***	1.081***	1.310***	1.073**	1.191***	0/11
adaptive Lasso slopeHAR	1.153***	1.131***	1.093	1.108***	1.156	1.165***	1.369**	1.134***	1.145**	1.171***	1.084***	0/11
adaptive Lasso freeHAR	1.447***	1.231***	3.884***	1.296***	2.391***	1.476***	4.449***	1.064***	1.286***	1.074**	1.189***	0/11
ordered Lasso AR	1.028**	1.527***	1.489**	1.775***	1.126***	1.350***	1.265	1.520***	1.189***	1.704***	1.280***	0/11
adaptive Lasso AR-FC	1.107***	1.058	3.021***	1.014	1.614***	1.175***	3.033***	1.007	1.088**	1.054**	0.971	1/11
adaptive Lasso HAR-FC	1.298***	1.181***	3.459***	1.159***	2.138***	1.330***	3.805***	1.003***	1.193***	1.028*	1.094***	0/11
adaptive Lasso slopeHAR-FC	1.102***	1.081***	1.041	1.030**	1.124	1.095***	1.266**	1.038**	1.080*	1.080**	1.035**	0/11
adaptive Lasso freeHAR-FC	1.307***	1.189***	3.427***	1.159***	2.136***	1.336***	3.803***	1.013***	1.185***	1.029	1.064***	0/11
ordered Lasso AR-FC	**0.976****	**0.969****	1.273**	**0.954*****	1.059	**0.991**	1.215*	1.002	**0.977**	0.993	**0.960****	7/11
horizon=22												
HAR	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	**0.988****	0.996	1.040***	0.985**	1.002	1.004	**0.973**	1.004	0.990*	0.999	0.979***	7/11
freeHAR	1.006	1.004	**0.999**	1.000	1.003	1.002	1.001	1.003	1.001	1.001	1.002	1/11
adaptive Lasso AR	1.015	1.018	2.184***	1.010***	1.232***	1.082***	2.571***	1.057***	1.111**	1.081**	0.994	1/11
adaptive Lasso HAR	1.132***	1.092**	2.312***	1.165***	1.642***	1.298***	3.401***	1.022**	1.114***	1.078	1.032***	0/11
adaptive Lasso slopeHAR	1.021**	1.039*	1.090***	1.037***	1.031	1.026	1.338*	1.069***	1.142**	1.018	1.022*	0/11
adaptive Lasso freeHAR	1.139***	1.081**	2.271***	1.167***	1.643***	1.286***	3.409***	1.030**	1.100***	1.077	1.034***	0/11
ordered Lasso AR	1.054***	1.235***	1.094	1.442***	1.068***	1.233***	1.242	1.171***	1.138***	1.299***	1.080***	0/11
adaptive Lasso AR-FC	1.002	1.006	1.917***	0.975**	1.117***	1.065*	2.207***	0.987	1.109***	1.058*	0.985*	3/11
adaptive Lasso HAR-FC	1.097**	1.016	2.060***	1.055***	1.507***	1.207***	2.974***	**0.950*****	1.034	1.022	0.962**	2/11
adaptive Lasso slopeHAR-FC	1.017**	1.037	1.035	1.009**	1.007	1.020	1.258	0.987	1.110***	1.017	0.991**	2/11
adaptive Lasso freeHAR-FC	1.098***	1.020	2.041***	1.060***	1.515***	1.211***	2.983***	0.964	1.026	1.022	0.963**	2/11
ordered Lasso AR-FC	0.992	**0.982****	1.071*	**0.956*****	1.009	**0.988****	1.261**	0.962	**0.958**	**0.966**	**0.949**	8/11

Table 9(a): OOS Forecasts based on Full Sample 300-sec log RV and RW=1000. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with a fixed length Rolling Window (RW). Consistent with the work of Audrino and Knaus (2016), we set the evaluation window to (May 12, 2009 to Nov 15, 2010), in order to make the results comparable to those obtained from the RW=2000. We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	
slopeHAR	0.998	0.997	0.996	0.991	0.991	**0.988**	0.990	1.002	0.998	**0.996**	**0.991***	10/11
freeHAR	1.000	1.007	**0.991**	1.006	0.994	1.001	0.972	1.006	0.993	0.999	0.997	6/11
adaptive Lasso AR	1.061	1.072	1.012	1.034	1.038	1.018	1.024**	1.080	1.001	1.082	1.085	0/11
adaptive Lasso HAR	1.049	1.012	1.189*	0.996	1.030	1.027	0.986*	1.027	1.043	1.030	1.028*	2/11
adaptive Lasso slopeHAR	1.118	1.097	1.072	1.048	1.095	1.034	1.045	1.142	1.005	1.049	1.070	0/11
adaptive Lasso freeHAR	1.044	1.018	1.189*	0.997	1.013	1.029	0.991	1.029	0.999	1.028**	1.031*	3/11
ordered Lasso AR	0.996	**0.995**	1.026	0.987	0.995	0.998	0.982*	1.008	0.996	1.053	1.047	7/11
adaptive Lasso AR-FC	1.018	1.039	1.005	1.022	1.004	1.007	0.985	1.025	**0.980**	1.024	1.037	2/11
adaptive Lasso HAR-FC	1.002	0.995	1.080	0.983	1.001	1.003	**0.968*** **	1.005	1.005	1.007	0.999	4/11
adaptive Lasso slopeHAR-FC	1.066	1.032	1.058	1.018	1.023	1.008	0.986	1.084	0.990	1.008	1.013	2/11
adaptive Lasso freeHAR-FC	1.007	1.004	1.083	**0.978**	0.997	0.990	0.971**	1.017	0.986	1.001	1.005	5/11
ordered Lasso AR-FC	**0.990**	0.998	1.008	0.994	**0.989**	0.994	0.975*	1.004	0.987	1.014	1.010	7/11
horizon=5												
HAR	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	1.002	0.996	1.010	0.990	0.997	1.005	0.995	0.996	1.012	0.999	0.997	7/11
freeHAR	1.002	1.007	1.015	0.992	0.997	1.002	0.999	0.998	0.999	0.999	0.993	7/11
adaptive Lasso AR	1.056	1.066*	1.207***	0.986	1.066	1.044	1.131***	1.045	1.045	1.130**	1.069	1/11
adaptive Lasso HAR	1.056**	1.025	1.191*	0.941***	1.027	1.006	1.000	1.026	0.961*	1.018	1.031**	2/11
adaptive Lasso slopeHAR	1.133	1.085	1.298***	0.987	1.130*	1.146**	1.062*	1.189	1.052	1.065	1.070*	1/11
adaptive Lasso freeHAR	1.068*	1.040	1.180	0.961**	1.029	1.001	1.005	1.022	**0.949*** **	1.025	1.034*	2/11
ordered Lasso AR	0.975	**0.974**	1.099***	**0.909*** **	1.000	1.000	0.999	0.977	1.011	1.097*	1.036	5/11
adaptive Lasso AR-FC	1.007	1.047	1.254***	0.981	1.042	1.024	1.087**	0.999	1.029	1.092**	1.022	2/11
adaptive Lasso HAR-FC	1.003	0.999	1.158	0.935***	0.996*	**0.991**	0.983	0.990	0.968**	**0.993**	0.997	9/11
adaptive Lasso slopeHAR-FC	1.112*	1.067	1.293***	0.979*	1.090	1.084**	0.987	1.157**	1.000	1.014	1.018	2/11
adaptive Lasso freeHAR-FC	1.007	1.007	1.161	0.938**	0.998*	1.000	**0.977**	0.995	0.964**	1.000	1.003	5/11
ordered Lasso AR-FC	**0.969***	0.983	1.062***	0.926**	**0.983**	0.992	0.986	**0.967***	0.996	1.038*	**0.992**	9/11
horizon=22												
HAR	1.000	1.000	**1.000**	1.000	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	1.006	0.992*	1.011	0.989	1.031	**0.992***	1.054***	1.004	1.051**	1.002	1.010	3/11
freeHAR	1.004	1.004	1.008	1.000	1.003	1.004	1.003	1.002	1.005	1.000	1.000	0/11
adaptive Lasso AR	0.959***	0.967	1.501***	0.901***	1.093	1.008	1.241***	1.009	1.162	1.111***	1.061	3/11
adaptive Lasso HAR	1.013	1.060**	1.440***	0.971***	1.106	1.096	1.017**	0.956**	1.004***	1.070*	1.034	2/11
adaptive Lasso slopeHAR	1.014	1.076	1.447***	0.966***	1.182**	1.119	1.044	1.079	1.097	1.093***	0.995	2/11
adaptive Lasso freeHAR	1.011*	1.078*	1.450***	0.967***	1.107	1.110	1.028	0.967***	1.003***	1.077**	1.016*	2/11
ordered Lasso AR	0.936***	**0.954*** **	1.291***	**0.814*** **	1.076	1.032	0.936***	0.912***	1.101	1.100***	0.999**	6/11
adaptive Lasso AR-FC	0.940***	0.999	1.540***	0.902***	1.102	1.030	1.119*	0.952***	1.128	1.081***	1.012	4/11
adaptive Lasso HAR-FC	0.961***	0.990***	1.418***	0.944***	1.059*	1.028*	1.019***	0.933***	0.999**	**0.980**	0.999*	7/11
adaptive Lasso slopeHAR-FC	1.007	1.104	1.501***	0.946***	1.190	1.135*	0.994	1.000	1.013**	1.035*	0.994	3/11
adaptive Lasso freeHAR-FC	0.969***	0.998***	1.440***	0.946***	1.063	1.044	1.022**	0.935***	**0.993*** **	1.001	0.997*	6/11
ordered Lasso AR-FC	**0.921*** **	0.976***	1.207***	0.867***	1.037*	1.017*	**0.912*** **	**0.903*** **	1.059	1.017	**0.969*** **	6/11

Table 9(b): OOS Forecasts based on Full Sample 300-sec log RV and RW=2000. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with a fixed length Rolling Window (RW). Consistent with the work of Audrino and Knaus (2016), we set the evaluation window to (May 12, 2009 to Nov 15, 2010). We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.999	0.997	0.991	0.989	0.990	0.987	0.995	1.003	0.994	0.996	0.992	10/11
freeHAR	0.996	1.007	**0.984**	0.999	0.991	1.000	0.973*	**0.999**	0.989	**0.996**	0.992	9/11
adaptive Lasso AR	1.006	1.031	1.026	1.010	1.012	1.020	0.994	1.029	1.000	1.048	1.034	1/11
adaptive Lasso HAR	1.047	1.009	1.033	1.039	1.031	1.043	0.978**	1.031	1.039	1.027	1.034*	1/11
adaptive Lasso slopeHAR	1.059	1.093	1.011	1.061	1.045	1.036	0.988	1.059	0.988	1.051**	1.063*	2/11
adaptive Lasso freeHAR	1.027	1.008	1.020	1.043	1.012	1.011	0.975**	1.024	0.985	1.019*	1.041	2/11
ordered Lasso AR	0.993	**0.990**	0.999	0.988	0.986	0.990	0.967***	1.022	0.979*	1.096*	1.014	8/11
adaptive Lasso AR-FC	0.995	1.019	1.007	1.000	1.000	1.010	0.972	1.003	0.985	1.024	0.998	4/11
adaptive Lasso HAR-FC	0.995	0.994	1.005	0.988	0.999	1.004	0.959***	1.003	0.995	1.008	0.995	7/11
adaptive Lasso slopeHAR-FC	1.054	1.056	1.019	1.006	1.040	**0.975**	0.968**	1.018	0.983	1.005	1.003	3/11
adaptive Lasso freeHAR-FC	0.999	0.999	0.993*	**0.987**	0.992	0.988	**0.958*** **	1.010	**0.967**	0.999	1.003	9/11
ordered Lasso AR-FC	**0.990**	0.996	0.990	0.988	**0.982**	0.986	0.964**	1.000	0.980	1.007	**0.990**	9/11
horizon=5												
HAR	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	1.004	0.996	1.007	0.982	0.999	1.001	1.002	0.999	1.013	0.999	0.998	6/11
freeHAR	1.002	1.003	1.003	0.988	0.994	0.996	1.001	0.996	0.998	0.994	0.992	7/11
adaptive Lasso AR	1.021	1.023	1.104***	0.959**	1.037	1.020	1.045*	1.012	1.017	1.054	0.996	2/11
adaptive Lasso HAR	1.045**	1.019	1.074**	1.002	1.016	1.031*	**0.932****	1.022	0.953**	1.036*	1.042*	2/11
adaptive Lasso slopeHAR	1.128	1.225*	1.132***	1.026	1.079	1.024*	0.983	1.026	1.037	1.055	1.065*	1/11
adaptive Lasso freeHAR	1.046*	1.016	1.048	1.000	1.012	1.027*	0.932**	1.004	**0.949****	1.033*	1.038*	2/11
ordered Lasso AR	0.998	**0.941**	1.039*	**0.902*****	1.000	0.992	0.965**	1.016	0.952**	1.174**	0.994	7/11
adaptive Lasso AR-FC	1.005	1.008	1.089***	0.958**	1.019	1.003	1.001	0.994	1.004	1.012	0.976**	3/11
adaptive Lasso HAR-FC	0.992	0.995	1.042*	0.965***	1.001	0.996	0.952**	0.988	0.959***	**0.989**	0.983	9/11
adaptive Lasso slopeHAR-FC	1.114	1.222*	1.125***	1.002	0.994	**0.987**	0.958**	0.988	1.002	1.007	1.001	4/11
adaptive Lasso freeHAR-FC	1.005	0.992	1.035	0.963***	0.998	1.007	0.951**	0.982	0.953***	1.001	0.990	7/11
ordered Lasso AR-FC	**0.991**	0.964*	1.022	0.926***	**0.988**	0.991	0.959**	**0.976**	0.968*	1.008	**0.962****	9/11
horizon=22												
HAR	1.000	1.000	**1.000**	1.000	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	1.006	0.988*	1.015	0.988	1.033	0.998	1.059***	1.007	1.053**	0.999	1.013*	4/11
freeHAR	1.002	1.002	1.003	0.996	1.001	0.999	1.006	1.000	1.005	0.997	0.997	4/11
adaptive Lasso AR	0.946***	0.955	1.293***	0.891***	1.073	0.993**	1.075	0.982*	1.125	1.038**	0.995	6/11
adaptive Lasso HAR	1.003*	1.051**	1.340***	1.006***	1.123	1.082	0.935***	0.943***	1.014***	1.028	0.985**	3/11
adaptive Lasso slopeHAR	0.959*	1.041	1.249***	0.994**	1.154	1.089	1.025	1.002	1.049	1.041**	0.997	3/11
adaptive Lasso freeHAR	0.998*	1.052***	1.340***	1.001***	1.125	1.079	0.943***	0.947***	1.019***	1.027*	0.972***	4/11
ordered Lasso AR	0.929***	**0.905*****	1.204***	**0.832*****	1.073***	**0.973*****	0.942***	0.937***	**0.936*****	1.143***	0.951***	8/11
adaptive Lasso AR-FC	0.936***	0.972**	1.305***	0.904***	1.083	1.003**	1.012	0.935***	1.099	1.015	0.964***	5/11
adaptive Lasso HAR-FC	0.944***	0.986***	1.238***	0.961***	1.084	1.034	0.945***	0.928***	1.043**	**0.947**	0.943***	7/11
adaptive Lasso slopeHAR-FC	0.984	1.075	1.355***	0.969***	1.126	1.041	0.950***	0.961	1.012***	0.971	0.979**	6/11
adaptive Lasso freeHAR-FC	0.958***	1.001***	1.231***	0.957***	1.083	1.044	0.945***	0.925***	1.037**	0.949	0.945***	6/11
ordered Lasso AR-FC	**0.918*****	0.951***	1.162**	0.900***	1.046**	1.003**	**0.928*****	**0.908*****	1.000**	0.981	**0.923*****	7/11

Table 9(c): OOS Forecasts based on Pre-Crisis 300-sec log RV and RW=1000. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with a fixed length Rolling Window (RW). Consistent with the work of Audrino and Knaus (2016), here we set the evaluation window to (May 20, 2005 to Aug 31, 2007), with a total of 575 observations. We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	
slopeHAR	0.996	**0.996**	0.995	1.001	**0.995**	**0.994**	1.000	**0.997**	**0.992**	1.001	1.000	7/11
freeHAR	0.995	0.998	1.001	**0.995**	1.006	1.001	1.011	1.003	1.007	1.003	1.000	3/11
adaptive Lasso AR	1.085	1.137**	1.213*	1.105	1.160**	1.046	1.065*	1.024	1.046	1.022	1.103	0/11
adaptive Lasso HAR	1.353***	1.197***	1.319	1.075	1.238	1.059	1.073	1.060	1.079	1.068	1.220	0/11
adaptive Lasso slopeHAR	1.019	1.062	1.141**	1.092	1.065***	1.042***	1.060	1.062**	1.058**	1.018	1.078***	0/11
adaptive Lasso freeHAR	1.326***	1.198**	1.319**	1.082	1.231***	1.059**	1.070	1.069**	1.087**	1.083	1.227***	0/11
ordered Lasso AR	1.059	1.048	0.991	1.113	1.026	1.040	1.008	1.019	1.091	1.022	1.001	1/11
adaptive Lasso AR-FC	1.034	1.093**	1.135	1.042	1.104	1.032	1.039	1.013	1.024	1.015	1.044	0/11
adaptive Lasso HAR-FC	1.248***	1.176***	1.330**	1.002	1.206**	1.006*	1.028	1.028**	1.033	1.003	1.132**	0/11
adaptive Lasso slopeHAR-FC	1.007	1.019	1.064	1.006	1.017	1.031	1.016	1.034**	1.035	1.014	1.055	0/11
adaptive Lasso freeHAR-FC	1.240***	1.176***	1.333**	1.008	1.202**	1.012	1.022	1.040**	1.036	1.011	1.142**	0/11
ordered Lasso AR-FC	**0.986**	1.000	**0.991**	1.009	1.002	0.995	1.007	1.001	1.015	**0.986**	**0.992**	5/11
horizon=5												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.995	0.998*	0.985	1.008	**0.981**	0.993	1.008	0.984	**0.982****	1.015	0.999	8/11
freeHAR	0.998	0.998	0.992	1.003	0.998	0.996	1.001	0.987	0.998	1.001	0.995	8/11
adaptive Lasso AR	1.269	1.509***	1.534***	1.095**	1.636***	1.233***	1.115	0.972	1.043	1.170***	1.222**	1/11
adaptive Lasso HAR	1.303**	1.508***	1.639***	1.027	1.571***	1.306***	1.174**	1.025	1.039	1.174***	1.263***	0/11
adaptive Lasso slopeHAR	1.070	1.188*	1.191	0.935***	1.167**	1.058	1.055	1.048*	1.027	0.990	1.119	2/11
adaptive Lasso freeHAR	1.298**	1.525***	1.650***	1.028	1.566***	1.303***	1.158**	1.010	1.020	1.169***	1.262***	0/11
ordered Lasso AR	1.060	1.086**	0.977	1.121**	1.050	1.040	0.999*	**0.941****	1.134	1.029	0.985**	4/11
adaptive Lasso AR-FC	1.179	1.401**	1.409**	1.028	1.533***	1.161***	1.064	0.970	1.015	1.130**	1.162*	1/11
adaptive Lasso HAR-FC	1.289**	1.419***	1.550***	1.037	1.460***	1.259***	1.168***	0.991	0.996	1.139**	1.200**	2/11
adaptive Lasso slopeHAR-FC	1.029	1.125	1.122	**0.923*****	1.093	1.028	1.001*	1.009	1.001	**0.978**	1.057	2/11
adaptive Lasso freeHAR-FC	1.283**	1.424***	1.556***	1.035	1.457***	1.259***	1.166***	0.982	0.989	1.139**	1.208**	2/11
ordered Lasso AR-FC	**0.977**	**0.998**	**0.966**	0.977	0.996	**0.954***	**0.984**	0.976	0.995	0.999	**0.980**	11/11
horizon=22												
HAR	**1.000**	**1.000**	1.000	**1.000**	**1.000**	1.000	1.000	1.000	**1.000**	**1.000**	**1.000**	
slopeHAR	1.030**	1.008	**0.997**	1.044***	1.038	1.042**	**0.999**	1.005	1.003	1.077	1.007	2/11
freeHAR	1.003	1.002	0.999	1.003	1.003	1.000	1.003	0.998	1.002	1.004	1.000	2/11
adaptive Lasso AR	1.248	1.551***	1.347	1.344***	1.915***	1.354	1.203	1.031***	1.098***	1.421	1.211	0/11
adaptive Lasso HAR	1.117	1.593***	1.446***	1.201***	1.790***	1.181	1.097***	1.058***	1.094***	1.161	1.316***	0/11
adaptive Lasso slopeHAR	1.124	1.230*	1.185	1.174***	1.283***	0.979	1.183	1.012*	1.113**	1.096	1.230**	1/11
adaptive Lasso freeHAR	1.117	1.596***	1.451***	1.206***	1.791***	1.178	1.103***	1.043***	1.090***	1.166	1.304***	0/11
ordered Lasso AR	1.118*	1.191***	1.077***	1.243***	1.270***	1.064***	1.067**	**0.911*****	1.112**	1.166	1.033	1/11
adaptive Lasso AR-FC	1.190	1.429*	1.248	1.241***	1.772***	1.252	1.166	1.038***	1.106***	1.337	1.177	0/11
adaptive Lasso HAR-FC	1.043	1.477**	1.380***	1.111**	1.607***	1.131	1.042***	1.029***	1.022***	1.118	1.184	0/11
adaptive Lasso slopeHAR-FC	1.080	1.163	1.125	1.075**	1.172***	**0.934**	1.094	1.049***	1.071***	1.036	1.208**	1/11
adaptive Lasso freeHAR-FC	1.050	1.484**	1.379***	1.116**	1.608***	1.126	1.042***	1.018***	1.016***	1.113	1.196	0/11
ordered Lasso AR-FC	1.021**	1.092***	1.090***	1.038*	1.185***	0.999**	1.062	1.031***	1.084***	1.151	1.010	1/11

Table 9(d): OOS Forecasts based on Post-Crisis 300-sec log RV and RW=1000. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with a fixed length Rolling Window (RW). Consistent with the work of Audrino and Knaus (2016), here we set the evaluation window to (Sep 04, 2007 to Nov 15, 2010), with a total of 808 observations. We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	
slopeHAR	0.999	0.998	0.997	0.991*	0.988*	0.992**	0.991**	1.002	0.991*	0.990*	**0.989**	10/11
freeHAR	1.003	1.007	**0.992**	1.002	0.993**	0.988	0.988**	1.007	0.986**	**0.988****	0.993	7/11
adaptive Lasso AR	1.056	1.062	1.505***	1.131*	1.092*	1.054	1.190***	1.097*	1.027	1.110	1.097	0/11
adaptive Lasso HAR	1.186***	1.074*	1.904***	1.118***	1.362***	1.080***	1.374***	1.088**	1.107***	1.110***	1.097**	0/11
adaptive Lasso slopeHAR	1.105	1.096*	1.045	1.130***	1.067	1.091	1.110	1.086*	1.047	1.124***	1.091**	0/11
adaptive Lasso freeHAR	1.175***	1.088*	1.892***	1.121***	1.356***	1.090***	1.381***	1.079**	1.089***	1.105***	1.106**	0/11
ordered Lasso AR	1.011	1.094*	1.075	1.161***	0.994	1.121***	0.979**	1.240***	1.047*	1.222***	1.127**	2/11
adaptive Lasso AR-FC	1.022	1.020	1.380***	1.055	1.031	1.005	1.110***	1.027	0.995	1.054	1.034	1/11
adaptive Lasso HAR-FC	1.098***	1.023	1.744***	1.030	1.265***	1.020	1.286***	1.025*	1.041**	1.030*	1.018***	0/11
adaptive Lasso slopeHAR-FC	1.078	1.037	1.029	1.054***	1.022	1.043*	1.034	1.043	1.013	1.047***	1.025	0/11
adaptive Lasso freeHAR-FC	1.085*	1.037	1.749***	1.028	1.261***	1.026	1.288***	1.041**	1.044**	1.033**	1.022**	0/11
ordered Lasso AR-FC	**0.994**	**0.996**	1.006	**0.986***	**0.981***	**0.981***	**0.970****	1.015*	**0.984**	0.992	1.003	8/11
horizon=5												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.994*	0.992	1.004**	0.982*	0.988	0.995	0.997	0.993	1.002	**0.986**	**0.994**	9/11
freeHAR	1.002	1.001	0.999	0.991	0.993	0.997	1.000	0.999	**0.993**	0.995	0.994	8/11
adaptive Lasso AR	1.112***	1.071**	1.710***	1.048	1.267***	1.093*	1.570***	1.091***	1.044	1.074	1.083***	0/11
adaptive Lasso HAR	1.234***	1.171***	1.952***	1.068***	1.680***	1.207***	1.695***	1.040***	1.025*	1.042	1.074***	0/11
adaptive Lasso slopeHAR	1.154***	1.110***	1.089	1.061***	1.105	1.157***	1.309***	1.079**	1.094**	1.111**	1.103***	0/11
adaptive Lasso freeHAR	1.234***	1.173***	1.937***	1.087***	1.692***	1.207***	1.696***	1.038***	1.028*	1.040	1.065***	0/11
ordered Lasso AR	1.009*	1.092***	1.059	1.218***	1.009	1.204***	0.972**	1.294***	1.081**	1.289***	1.137***	1/11
adaptive Lasso AR-FC	1.068**	1.045*	1.559***	1.009	1.164***	1.039	1.452***	1.009	1.020	1.053	1.042**	0/11
adaptive Lasso HAR-FC	1.149***	1.124***	1.783***	1.000	1.537***	1.142***	1.590***	0.999**	1.000	1.019	1.022***	1/11
adaptive Lasso slopeHAR-FC	1.121***	1.064**	1.064	1.042**	1.083	1.109***	1.244***	1.044**	1.039	1.059*	1.034***	0/11
adaptive Lasso freeHAR-FC	1.152***	1.130***	1.781***	1.011**	1.550***	1.154***	1.588***	0.999**	0.999	1.023	1.024***	2/11
ordered Lasso AR-FC	**0.985**	**0.985**	**0.996***	**0.964**	**0.980****	**0.975**	**0.968***	**0.982**	1.009	1.004	1.039*	8/11
horizon=22												
HAR	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.990**	0.994	1.035***	0.987	1.006*	1.009	1.021	1.005	1.009*	0.996	0.992**	5/11
freeHAR	1.005	1.002	1.003	1.000	1.000	1.001	1.002	1.002	1.000	1.000	1.001	0/11
adaptive Lasso AR	1.007*	1.035	1.538***	1.011**	1.073***	1.072**	1.541***	1.048***	1.091	1.059	1.055**	0/11
adaptive Lasso HAR	1.144***	1.103**	1.683***	1.071***	1.429***	1.256***	1.621***	1.027**	1.048**	1.038	1.040***	0/11
adaptive Lasso slopeHAR	1.031**	1.065**	1.093*	1.039*	1.037	1.061**	1.342***	1.068**	1.122**	1.054	1.051***	0/11
adaptive Lasso freeHAR	1.137***	1.108***	1.677***	1.085***	1.443***	1.256***	1.634***	1.050***	1.027*	1.038	1.035***	0/11
ordered Lasso AR	1.013**	1.111***	0.943**	1.044***	1.059**	1.172***	**0.956*****	1.082***	1.124***	1.187***	1.048***	2/11
adaptive Lasso AR-FC	1.004*	1.022	1.442***	0.990	1.008	1.049	1.471***	0.978	1.054	1.039	1.022**	2/11
adaptive Lasso HAR-FC	1.069**	1.019	1.536***	0.994	1.313***	1.156***	1.497***	0.947***	0.977	0.998	1.001**	4/11
adaptive Lasso slopeHAR-FC	1.037**	1.077**	1.069	1.029	1.029	1.070*	1.319***	0.983	1.050*	1.017	0.998**	2/11
adaptive Lasso freeHAR-FC	1.072**	1.023	1.528***	1.004	1.320***	1.164***	1.500***	0.955	**0.975**	1.005	0.997**	3/11
ordered Lasso AR-FC	**0.972**	**0.970****	**0.939****	**0.924*****	1.004*	**0.986**	0.974**	**0.923*****	0.991**	**0.994***	**0.983**	10/11

Table 10(a): OOS Forecasts based on Full Sample 600-sec log RV and RW=1000. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with a fixed length Rolling Window (RW). Consistent with the work of Audrino and Knaus (2016), we set the evaluation window to (May 12, 2009 to Nov 15, 2010), in order to make the results comparable to those obtained from the RW=2000. We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	**1.000**	
slopeHAR	0.995	**0.999**	0.995	0.995	0.984*	**0.985****	0.989	1.002	1.003	0.994	1.002	8/11
freeHAR	1.006	1.008	**0.991**	1.013	0.977*	0.994	0.987	1.002	0.994*	0.995	1.013	6/11
adaptive Lasso AR	1.052	1.077	1.006	1.033	1.023	1.037	1.058**	1.069	1.013	1.081	1.101	0/11
adaptive Lasso HAR	1.015	1.008	1.071	1.035	0.983	1.023	1.020	1.011	1.024	0.986	1.060*	2/11
adaptive Lasso slopeHAR	1.121	1.157**	1.048	1.050	1.112	1.050	1.010	1.148**	1.027	1.044	1.051	0/11
adaptive Lasso freeHAR	1.025	1.018	1.053	1.024	1.019	1.003	1.010	1.023	1.022	1.001	1.030*	0/11
ordered Lasso AR	0.995	1.006	1.015	1.020	0.974*	0.989	0.985*	1.002	1.003	1.030	1.057	4/11
adaptive Lasso AR-FC	1.036	1.047	1.009	1.027	0.992	1.007	1.026	1.027	1.000	1.032	1.054	1/11
adaptive Lasso HAR-FC	0.998	1.002	1.027	1.006	0.970	1.004	0.989	1.000	0.999	**0.976**	1.022	5/11
adaptive Lasso slopeHAR-FC	1.078	1.100**	1.012	1.007	1.049	1.010	**0.971****	1.082	1.006	1.007	1.010	1/11
adaptive Lasso freeHAR-FC	1.006	1.012	1.025	**0.993**	0.998	0.994	0.986	1.013	1.000	0.993	1.010	5/11
ordered Lasso AR-FC	**0.991**	1.004	1.003	1.028	**0.968***	0.986	0.982*	1.001	**0.994**	1.020	1.040	5/11
horizon=5												
HAR	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	
slopeHAR	0.996	**0.988**	1.010	0.979	1.002	**0.988**	0.979	0.996	1.028	1.004	1.005	6/11
freeHAR	0.997	0.998	1.018*	0.996	0.987	0.996	0.997	0.997	1.001	0.997	1.002	8/11
adaptive Lasso AR	1.057	1.065	1.248***	1.001	1.044	1.062	1.141***	1.041	1.080	1.150***	1.102	0/11
adaptive Lasso HAR	1.046*	1.019	1.108	0.964**	1.030	1.009	**0.905*****	1.026	0.979	0.996	1.044	4/11
adaptive Lasso slopeHAR	1.187**	1.186***	1.296***	0.974	1.113**	1.098**	1.010	1.232	1.042	1.042	1.083**	1/11
adaptive Lasso freeHAR	1.038*	1.024	1.101	0.965	1.018	1.020	0.912***	1.029	**0.970***	1.013	1.054**	3/11
ordered Lasso AR	0.982	1.003	1.081***	0.956**	1.001	0.997	0.983	0.976	1.041	1.048	1.052	5/11
adaptive Lasso AR-FC	1.026	1.048	1.245***	0.990	1.034	1.041*	1.083***	0.996	1.052	1.075*	1.050	2/11
adaptive Lasso HAR-FC	0.993	1.008	1.042	**0.946****	0.993	1.002	0.929**	0.996	0.977*	**0.991**	1.001	7/11
adaptive Lasso slopeHAR-FC	1.152**	1.135**	1.241***	0.965**	1.110**	1.051*	0.963	1.123***	1.009	1.012	1.034*	2/11
adaptive Lasso freeHAR-FC	0.998	1.013	1.041	0.948*	1.002	1.010	0.920***	0.998	0.978*	0.997	1.012	6/11
ordered Lasso AR-FC	**0.973**	0.995	1.053***	0.973	**0.984**	0.990	0.974	**0.972**	1.025	1.016	1.017	7/11
horizon=22												
HAR	1.000	1.000	**1.000**	1.000	**1.000**	1.000	1.000	1.000	**1.000**	1.000	1.000	
slopeHAR	1.001	**0.992**	1.008	0.976	1.038*	**0.973*****	1.016	1.003	1.075***	1.010	1.010	3/11
freeHAR	1.002	1.002	1.008	0.998	1.000	0.999	1.001	1.002	1.004	**0.999**	1.001	3/11
adaptive Lasso AR	0.963**	1.030	1.408***	0.875***	1.112	1.029	1.230***	1.021	1.199	1.200***	1.071	2/11
adaptive Lasso HAR	1.033	1.099**	1.307***	0.976***	1.100	1.176**	0.973***	0.971*	1.023	1.086**	1.042	3/11
adaptive Lasso slopeHAR	1.023	1.092	1.373***	0.943***	1.227***	1.201***	1.090**	1.067	1.113	1.127***	1.012	1/11
adaptive Lasso freeHAR	1.026	1.112*	1.327***	0.974***	1.111	1.183***	0.982**	0.989**	1.011**	1.085**	1.031	3/11
ordered Lasso AR	0.935***	1.029***	1.180***	**0.830*****	1.097	1.056	0.973***	**0.926*****	1.109	1.099**	0.982**	5/11
adaptive Lasso AR-FC	0.963***	1.058	1.394***	0.883***	1.127	1.051	1.121***	0.994	1.153	1.135***	1.013	3/11
adaptive Lasso HAR-FC	0.979***	1.032***	1.328**	0.952***	1.053	1.095	**0.947*****	0.949***	1.031**	1.025	0.985	5/11
adaptive Lasso slopeHAR-FC	1.026	1.126	1.381***	0.936***	1.237**	1.164***	1.117**	1.024	1.034*	1.072**	0.989	2/11
adaptive Lasso freeHAR-FC	0.978***	1.040***	1.328***	0.956***	1.053	1.103	0.949***	0.949***	1.029**	1.032	0.988	5/11
ordered Lasso AR-FC	**0.919*****	1.016***	1.123**	0.860***	1.057	1.038	0.951***	0.930***	1.070	1.065	**0.952*****	5/11

Table 10(b): OOS Forecasts based on Full Sample 600-sec log RV and RW=2000. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with a fixed length Rolling Window (RW). Consistent with the work of Audrino and Knaus (2016), we set the evaluation window to (May 12, 2009 to Nov 15, 2010). We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.995	**0.998**	0.988	**0.991**	0.985	0.985**	0.992	1.003	1.002	0.995	0.999	9/11
freeHAR	1.004	1.007	**0.987**	1.015	0.983*	0.992	0.982*	**0.998**	1.001	0.996	0.998	7/11
adaptive Lasso AR	1.026	1.049	1.007	1.023	0.998	1.024	1.017	1.036	1.010	1.060	1.045	1/11
adaptive Lasso HAR	1.032**	1.021	1.011	1.051	1.020	1.012	0.997	1.018	1.043	1.002	1.032*	1/11
adaptive Lasso slopeHAR	1.075	1.087	0.999	1.061	1.062	1.073	0.976	1.048	1.010	1.024	1.046	2/11
adaptive Lasso freeHAR	1.026	1.012	1.005	1.056	1.015	1.018	0.992	1.022	1.018	1.002*	1.033*	1/11
ordered Lasso AR	0.996	1.001	0.995	1.001	0.977*	0.983	0.971**	1.015	1.001	1.047	1.007	5/11
adaptive Lasso AR-FC	1.015	1.029	0.993	1.017	0.988	1.004	0.996	1.004	1.006	1.038	1.008	3/11
adaptive Lasso HAR-FC	1.005	1.008	0.992	1.004	0.993	0.993	0.976	1.000	1.013	0.987	0.998	6/11
adaptive Lasso slopeHAR-FC	1.019	1.019	1.000	1.009	1.037	1.030	**0.964****	1.003	1.002	1.002	1.005	1/11
adaptive Lasso freeHAR-FC	0.998	1.010	0.991	1.001	0.995	0.998	0.972**	1.010	**0.997**	**0.986**	1.003	7/11
ordered Lasso AR-FC	**0.992**	1.002	0.988	1.003	**0.972****	**0.981***	0.969**	1.000	1.002	1.012	**0.993**	6/11
horizon=5												
HAR	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.999	0.989	1.006	0.971	1.003	**0.985**	0.987	0.999	1.031	1.009*	1.007	6/11
freeHAR	0.998	0.995	1.005	0.989*	0.987	0.993	0.999	0.996	1.001	0.995	0.997	9/11
adaptive Lasso AR	1.014	1.008	1.109***	0.974*	1.026	1.029	1.068***	1.017	1.047	1.077	1.015	1/11
adaptive Lasso HAR	1.039**	1.019	1.034	1.018	1.013	1.019	0.923***	1.045	0.995	1.023	1.055**	2/11
adaptive Lasso slopeHAR	1.127**	1.085*	1.103***	1.091**	1.080*	1.055**	0.959	1.137	0.989	1.033	1.072**	2/11
adaptive Lasso freeHAR	1.031	1.013	1.028	1.005	1.013	1.022	**0.917*****	1.026	0.986	1.030	1.058**	2/11
ordered Lasso AR	0.987	**0.969**	1.024**	**0.931*****	0.996	0.989	0.964	1.006	**0.980****	1.106*	0.990	8/11
adaptive Lasso AR-FC	1.002	0.999	1.088***	0.964**	1.026	1.019	1.020**	0.994	1.030	1.033	0.990*	4/11
adaptive Lasso HAR-FC	0.991	0.988	1.041*	0.972***	1.002	1.000	0.951*	1.009	0.990*	1.000	1.008	5/11
adaptive Lasso slopeHAR-FC	1.040	1.004	1.078**	0.975*	1.043	1.019	0.937**	1.027	1.001	**0.991**	1.003	3/11
adaptive Lasso freeHAR-FC	0.990	0.993	1.027*	0.979	1.000	1.009	0.947*	1.004	0.986*	1.008	1.013	5/11
ordered Lasso AR-FC	**0.980**	0.973	1.011*	0.960***	**0.984**	0.986	0.959*	**0.993**	0.992*	1.009	**0.970****	9/11
horizon=22												
HAR	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	1.000	0.991	1.014	0.972	1.039*	**0.979*****	1.023	1.006	1.081***	1.010	1.014*	3/11
freeHAR	1.001	0.999	1.004	0.995	**0.998**	0.995	1.005	1.000	1.004	**0.997**	0.998	6/11
adaptive Lasso AR	0.952***	1.003	1.194***	0.868***	1.095	0.985**	1.097***	1.006	1.153	1.099***	1.003*	3/11
adaptive Lasso HAR	1.007	1.060***	1.208***	0.992***	1.141	1.133	0.952***	0.968**	1.043**	1.068**	1.032	3/11
adaptive Lasso slopeHAR	0.977	1.055	1.220***	0.988**	1.200*	1.140***	1.043	1.015	1.087	1.102***	1.005	2/11
adaptive Lasso freeHAR	0.992**	1.065***	1.211***	0.992***	1.132	1.141	0.956**	0.964**	1.028**	1.076***	1.026	4/11
ordered Lasso AR	0.942***	**0.989*****	1.127**	**0.839*****	1.093	1.016**	0.970***	0.960**	**0.992*****	1.131**	0.952***	7/11
adaptive Lasso AR-FC	0.951***	1.017	1.185***	0.883***	1.110	1.019*	1.035	0.972*	1.113	1.062*	0.977***	4/11
adaptive Lasso HAR-FC	0.971***	1.020***	1.150**	0.958***	1.101	1.081	0.968***	0.936***	1.069*	1.007	0.964***	5/11
adaptive Lasso slopeHAR-FC	0.993	1.108	1.213***	0.958***	1.160	1.089	0.994	0.999	1.033**	1.015	0.982	5/11
adaptive Lasso freeHAR-FC	0.968***	1.022***	1.141**	0.963***	1.096	1.090	0.968**	**0.935*****	1.062*	1.013	0.972**	5/11
ordered Lasso AR-FC	**0.930*****	0.994***	1.097**	0.890***	1.065	1.029*	**0.950*****	0.949***	1.029**	1.034	**0.932*****	6/11

Table 10(c): OOS Forecasts based on Pre-Crisis 600-sec log RV and RW=1000. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with a fixed length Rolling Window (RW). Consistent with the work of Audrino and Knaus (2016), here we set the evaluation window to (May 20, 2005 to Aug 31, 2007), with a total of 575 observations. We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	
slopeHAR	0.995	0.990	1.000	1.002	0.996	0.997	1.002	0.994	**0.996**	**0.998**	0.999	8/11
freeHAR	0.993	0.992	1.003	**0.995**	1.007**	1.003	1.012	0.999	1.007	1.001	1.008	4/11
adaptive Lasso AR	1.037	1.063*	1.117**	1.093**	1.081	1.046	1.080**	1.016	1.066*	1.029	1.084	0/11
adaptive Lasso HAR	1.200***	1.145*	1.307***	1.060	1.096**	1.047***	1.068	1.053**	1.080***	1.040*	1.130***	0/11
adaptive Lasso slopeHAR	1.002	1.047*	1.091	1.095**	1.035	1.042	1.068*	1.043	1.077*	1.025	1.084	0/11
adaptive Lasso freeHAR	1.199***	1.135*	1.305**	1.075	1.095*	1.057**	1.081	1.060**	1.090***	1.051***	1.130**	0/11
ordered Lasso AR	1.037	0.998	0.995	1.080**	1.004	1.022	1.010	1.001	1.074*	1.018	0.997	3/11
adaptive Lasso AR-FC	0.996	1.029*	1.060	1.043*	1.038	1.012	1.039	1.005	1.035	1.011	1.049	1/11
adaptive Lasso HAR-FC	1.100**	1.127**	1.273***	1.012	1.074*	1.015*	1.030	1.024*	1.037**	1.005	1.033***	0/11
adaptive Lasso slopeHAR-FC	0.990	1.010	1.039	1.007	1.005	1.031	1.013	1.023	1.039	1.007	1.050	1/11
adaptive Lasso freeHAR-FC	1.099**	1.120**	1.274***	1.013	1.071	1.017	1.027	1.019	1.044	1.007	1.042	0/11
ordered Lasso AR-FC	**0.985**	**0.987**	**0.995**	1.052***	**0.990**	**0.988**	1.009	**0.990**	1.050*	1.004	**0.990**	7/11
horizon=5												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.992	0.993	0.989	1.012	0.984	0.999	1.005	0.988	**0.982**	1.013	0.995	8/11
freeHAR	0.997	0.996	0.994	1.008	0.998	0.996	0.999	0.990	0.996	0.997	**0.992**	10/11
adaptive Lasso AR	1.198	1.335***	1.406***	1.080***	1.500***	1.132**	1.111	0.981	1.048*	1.096**	1.185***	1/11
adaptive Lasso HAR	1.297***	1.343***	1.455***	0.995	1.405***	1.193***	1.141**	1.031*	1.057**	1.090**	1.246***	1/11
adaptive Lasso slopeHAR	1.047	1.145**	1.172**	0.971	1.172**	1.036	1.077	1.042*	1.053*	1.008	1.103	1/11
adaptive Lasso freeHAR	1.288***	1.339***	1.451***	1.006	1.410***	1.178***	1.132**	1.018	1.059*	1.088**	1.244***	0/11
ordered Lasso AR	1.033	1.014	**0.948*****	1.094***	1.017	0.992	0.991	**0.942****	1.116**	1.019	0.996	5/11
adaptive Lasso AR-FC	1.134	1.263***	1.307***	1.038**	1.412***	1.073	1.053	0.986	1.037	1.057	1.139**	1/11
adaptive Lasso HAR-FC	1.312***	1.284***	1.405***	**0.962****	1.325***	1.146***	1.142***	1.004	1.015	1.053*	1.186**	1/11
adaptive Lasso slopeHAR-FC	1.014	1.095	1.101*	0.965	1.077	1.004	1.002	1.009	1.040*	0.992	1.046	2/11
adaptive Lasso freeHAR-FC	1.308***	1.283***	1.404***	0.965*	1.322***	1.144***	1.142***	0.991	1.014	1.050	1.190**	2/11
ordered Lasso AR-FC	**0.970**	**0.979**	0.964	1.027*	**0.974**	**0.924****	**0.984***	0.981	1.056*	**0.995**	1.006	8/11
horizon=22												
HAR	**1.000**	1.000	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	**1.000**	
slopeHAR	1.026*	1.004	1.006	1.056***	1.041	1.050***	1.005	0.999	**0.994**	1.073	1.003	2/11
freeHAR	1.003	1.000	0.999	1.004	1.002	1.001	1.004	1.000	1.001	1.002	1.001	1/11
adaptive Lasso AR	1.232	1.365**	1.265	1.251***	1.829***	1.353	1.211	1.002***	1.073	1.312	1.208	0/11
adaptive Lasso HAR	1.098	1.374**	1.310**	1.048	1.641***	1.201	1.077***	1.080***	1.120***	1.153	1.294**	0/11
adaptive Lasso slopeHAR	1.097	1.149**	1.175	1.077**	1.248***	1.013	1.154	1.016	1.110**	1.025	1.203*	0/11
adaptive Lasso freeHAR	1.094	1.378***	1.301**	1.047	1.648***	1.189	1.084***	1.061***	1.127***	1.153	1.287**	0/11
ordered Lasso AR	1.119*	1.014**	**0.979**	1.249***	1.027***	1.139*	1.087	**0.892****	1.130**	1.135	1.103*	2/11
adaptive Lasso AR-FC	1.174	1.285*	1.191	1.168***	1.685***	1.257	1.174	1.018***	1.078**	1.246	1.172	0/11
adaptive Lasso HAR-FC	1.045	1.300*	1.229**	**0.971**	1.484***	1.148	1.017***	1.042***	1.022***	1.105	1.152	1/11
adaptive Lasso slopeHAR-FC	1.067	1.087	1.118	0.978	1.128***	**0.967**	1.080	1.027***	1.094**	**0.991**	1.184*	3/11
adaptive Lasso freeHAR-FC	1.046	1.301*	1.224**	0.974	1.486***	1.142	1.016***	1.029***	1.023***	1.105	1.162	1/11
ordered Lasso AR-FC	1.046*	**0.989****	0.989**	1.092***	**0.975****	1.060	1.095	1.025***	1.126***	1.141	1.101*	3/11

Table 10(d): OOS Forecasts based on Post-Crisis 600-sec log RV and RW=1000. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with a fixed length Rolling Window (RW). Consistent with the work of Audrino and Knaus (2016), here we set the evaluation window to (Sep 04, 2007 to Nov 15, 2010), with a total of 808 observations. We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	
slopeHAR	**0.999**	**0.998**	0.997	**0.993**	0.985*	0.986***	0.992**	1.001	0.997	**0.992****	**0.992**	10/11
freeHAR	1.006	1.006	**0.994**	1.006	0.985**	0.987**	0.992	1.005	**0.991*****	0.995	0.997	7/11
adaptive Lasso AR	1.063	1.065	1.266***	1.140***	1.023	1.055	1.104***	1.094*	1.048	1.109*	1.116*	0/11
adaptive Lasso HAR	1.138***	1.058	1.541***	1.122**	1.219***	1.067***	1.278***	1.099**	1.082***	1.052**	1.108***	0/11
adaptive Lasso slopeHAR	1.118**	1.140**	1.086	1.135***	1.083	1.122***	1.067	1.103***	1.058	1.112***	1.076**	0/11
adaptive Lasso freeHAR	1.130***	1.076*	1.533***	1.116**	1.228***	1.068***	1.276***	1.114**	1.098***	1.065**	1.098***	0/11
ordered Lasso AR	1.006	1.057	1.053	1.129***	0.982**	1.055***	0.985**	1.180***	1.025	1.132***	1.121**	2/11
adaptive Lasso AR-FC	1.038	1.021	1.194***	1.071***	1.000	1.005	1.046	1.030	1.018	1.055	1.042	0/11
adaptive Lasso HAR-FC	1.062**	1.021	1.395***	1.038	1.150***	1.017	1.187***	1.030**	1.026	1.029**	1.033***	0/11
adaptive Lasso slopeHAR-FC	1.075**	1.084***	1.054	1.061**	1.053*	1.048*	1.006	1.048*	1.028	1.040**	1.028	0/11
adaptive Lasso freeHAR-FC	1.058*	1.033	1.401***	1.034	1.167***	1.019	1.184***	1.039**	1.042***	1.037**	1.024**	0/11
ordered Lasso AR-FC	1.000	1.008	1.001	1.010	**0.972****	**0.982****	**0.976****	1.013	0.996	1.017	1.038	4/11
horizon=5												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.992*	0.991	1.008**	**0.978****	0.992	0.981*	0.994	**0.991**	1.009	**0.986**	**0.993**	9/11
freeHAR	0.999	0.998	0.996	0.993	0.990*	0.991	1.000	0.998	**0.993**	0.995	0.994	10/11
adaptive Lasso AR	1.108***	1.057	1.564***	1.061**	1.175***	1.049*	1.452***	1.075***	1.065	1.086	1.092***	0/11
adaptive Lasso HAR	1.196***	1.113***	1.755***	1.089***	1.567***	1.118***	1.513***	1.043***	1.039**	1.007	1.075***	0/11
adaptive Lasso slopeHAR	1.151***	1.206***	1.144**	1.054***	1.115	1.158***	1.309***	1.070*	1.061	1.038	1.086***	0/11
adaptive Lasso freeHAR	1.187***	1.110***	1.728***	1.108***	1.571***	1.117***	1.524***	1.037***	1.024*	1.014	1.070***	0/11
ordered Lasso AR	0.997	1.058**	1.040	1.133***	1.002	1.093***	0.990	1.249***	1.066**	1.180***	1.081***	2/11
adaptive Lasso AR-FC	1.070**	1.031	1.456***	1.043**	1.091	1.013	1.367***	1.018**	1.031	1.065	1.050***	0/11
adaptive Lasso HAR-FC	1.116***	1.077**	1.618***	1.016**	1.422***	1.056***	1.440***	1.001*	1.015	1.009	1.016***	0/11
adaptive Lasso slopeHAR-FC	1.120***	1.154***	1.110	1.053***	1.099	1.061***	1.252***	1.033**	1.014	1.019	1.038***	0/11
adaptive Lasso freeHAR-FC	1.124***	1.084**	1.603***	1.035***	1.434***	1.065***	1.448***	0.993*	1.015	1.018	1.018***	1/11
ordered Lasso AR-FC	**0.980**	**0.975**	**0.985**	0.991	**0.977*****	**0.955****	**0.980**	0.999	1.006	1.015	1.030*	8/11
horizon=22												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	**0.987****	0.998	1.038***	0.980	1.007**	0.999	1.002	1.004	1.019	0.999	0.990**	6/11
freeHAR	1.003	1.001	1.002	0.998	0.999	0.998	1.001	1.001	1.000	0.999	1.000	4/11
adaptive Lasso AR	1.032**	1.049	1.546***	1.012***	1.084***	1.067**	1.508***	1.060***	1.143***	1.057	1.060***	0/11
adaptive Lasso HAR	1.148***	1.075	1.592***	1.100***	1.461***	1.232***	1.567***	1.023	1.035	1.013*	1.025***	0/11
adaptive Lasso slopeHAR	1.059***	1.076*	1.170***	1.032	1.076	1.110***	1.384***	1.096***	1.119**	1.043	1.050***	0/11
adaptive Lasso freeHAR	1.157***	1.083*	1.583***	1.109***	1.460***	1.235***	1.583***	1.044**	1.030	1.014*	1.032***	0/11
ordered Lasso AR	1.019**	1.084***	0.946**	1.040***	1.029	1.107***	**0.977*****	1.050***	1.102***	1.098***	1.016***	2/11
adaptive Lasso AR-FC	1.026**	1.035	1.470***	1.006**	1.029	1.041	1.444***	0.987	1.093**	1.030	1.016**	1/11
adaptive Lasso HAR-FC	1.079***	1.029	1.450***	1.017**	1.331***	1.129***	1.477***	0.947***	0.996	0.996**	0.982	4/11
adaptive Lasso slopeHAR-FC	1.054***	1.074***	1.146**	1.037*	1.069	1.081***	1.382***	1.003	1.048	1.011	0.991**	1/11
adaptive Lasso freeHAR-FC	1.080***	1.032	1.442***	1.036***	1.335***	1.133***	1.479***	0.952***	**0.992**	1.001**	0.983	3/11
ordered Lasso AR-FC	0.988	**0.965*****	**0.938****	**0.959**	**0.980*****	**0.959*****	0.990*	**0.934*****	1.002*	**0.967*****	**0.960**	10/11

Table 11(a): OOS Forecasts based on Full Sample 30-sec log RV: an IW approach. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with an increasing training set. The first training set includes the first 1000 observations and each subsequent training set contains one more observation. The evaluation window is set to (Dec 28, 2004 to Nov 15, 2010). We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.999	0.996	0.998	0.992**	0.994	0.993	0.996	1.005	0.990**	0.992	0.994	10/11
freeHAR	0.996	0.996	0.991	0.995	0.996	0.985*	0.994	0.999	0.986	0.992*	0.993*	11/11
adaptive Lasso AR	1.022	1.031	3.971***	1.066	1.037	1.002	1.680***	1.000	1.000	1.007*	1.021	0/11
adaptive Lasso HAR	1.117**	1.068***	3.906***	1.211***	1.427***	1.048**	1.786***	1.117***	1.163**	1.094***	1.103	0/11
adaptive Lasso slopeHAR	1.079	1.083*	1.120	1.092	1.066	1.022	1.220	1.050	1.016	1.063	1.088	0/11
adaptive Lasso freeHAR	1.094***	1.052	3.916***	1.213***	1.420***	1.031**	1.795***	1.119***	1.125*	1.076**	1.112	0/11
ordered Lasso AR	1.057	1.327**	1.334**	1.416**	1.039	1.144	1.010	1.128	1.230	1.164	1.016	0/11
adaptive Lasso AR-FC	**0.988**	0.998	3.341***	1.000	0.998	0.983	1.421	**0.982***	**0.970**	**0.986***	0.998	8/11
adaptive Lasso HAR-FC	1.015	1.006	3.319***	1.022	1.293***	0.990	1.548***	1.027**	1.032	1.015*	1.004	1/11
adaptive Lasso slopeHAR-FC	1.014	1.057**	1.065	1.024	1.022	1.002	1.099	1.016	0.992	1.014	1.070	1/11
adaptive Lasso freeHAR-FC	1.008	1.012	3.322***	1.029	1.268***	0.989	1.555***	1.036*	1.002	1.010	1.010	1/11
ordered Lasso AR-FC	0.996	**0.994**	**0.985**	**0.985****	**0.980**	**0.981**	**0.991**	1.000	0.980	0.993	**0.990**	10/11
horizon=5												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.998	0.995	1.004***	0.992	0.995	0.994	1.005	0.995	0.993	0.993	0.986	9/11
freeHAR	1.002	0.995	0.996	0.994	0.996	0.997	0.997	0.995	0.997	0.993	0.996	10/11
adaptive Lasso AR	1.032	0.999	4.553***	1.000	1.166	1.014	2.988***	0.992	1.171	0.994	1.024	3/11
adaptive Lasso HAR	1.051***	1.016*	4.447***	1.100***	1.534***	0.991*	3.128***	1.038***	1.289***	1.043***	1.072***	1/11
adaptive Lasso slopeHAR	1.075**	1.057**	1.168	1.078***	1.125	1.008**	1.458***	1.055***	1.142	1.058**	1.109***	0/11
adaptive Lasso freeHAR	1.037	1.004	4.462***	1.075***	1.567***	0.985	3.109***	1.026**	1.267**	1.030***	1.060**	1/11
ordered Lasso AR	1.053	1.380***	1.351**	1.521***	1.020	1.172***	1.030	1.104**	1.244**	1.214**	1.008	0/11
adaptive Lasso AR-FC	0.991	0.984	3.834***	0.980	1.099	0.980	2.579***	**0.974****	1.128	**0.977**	1.006	6/11
adaptive Lasso HAR-FC	0.988	0.993	3.821***	1.008**	1.415***	**0.959**	2.691***	0.998	1.198**	0.987*	1.004	5/11
adaptive Lasso slopeHAR-FC	1.029*	1.010	1.112	1.018***	1.078	0.970	1.346***	1.003	1.097	1.067***	1.087***	1/11
adaptive Lasso freeHAR-FC	0.982	0.984	3.804***	1.002	1.432***	0.961	2.708***	1.002	1.184**	0.990*	0.998	5/11
ordered Lasso AR-FC	**0.980**	**0.977**	**0.986**	**0.969**	**0.965***	0.971	**0.995**	0.987	**0.987**	0.980	**0.986**	11/11
horizon=22												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	1.005	1.007	1.040***	1.002*	1.014	1.016	1.002	1.004	1.004*	1.013	0.989**	1/11
freeHAR	1.004	1.000	0.999	0.998	1.001	1.000	**0.999**	1.000	1.000	0.998	1.000	4/11
adaptive Lasso AR	0.991	1.000	3.135***	0.973	1.066	0.978	2.636***	0.999	1.133*	0.996	0.999	6/11
adaptive Lasso HAR	0.991**	0.970	3.053***	1.030***	1.136***	0.940**	3.125***	0.996	1.084***	0.977	1.096***	5/11
adaptive Lasso slopeHAR	0.991	1.026	1.195***	1.028***	1.073	0.994	1.453***	0.991	1.134**	0.975	1.008*	4/11
adaptive Lasso freeHAR	0.991*	0.968	3.045***	1.021***	1.139***	0.933**	3.126***	0.985	1.075***	0.967	1.070***	5/11
ordered Lasso AR	1.001	1.098**	1.199**	1.309***	1.011*	1.074***	1.050	0.996***	1.158***	1.089***	**0.967**	2/11
adaptive Lasso AR-FC	0.980	0.995	2.745***	0.970	1.078	0.949**	2.321***	0.984	1.107**	0.978	0.997	7/11
adaptive Lasso HAR-FC	0.957	0.960	2.615***	0.985	1.020***	0.912***	2.674***	0.970*	1.031	0.949	1.010**	6/11
adaptive Lasso slopeHAR-FC	0.991	0.996	1.152***	1.013**	1.071	0.951	1.362***	0.994	1.064*	0.960	1.028**	5/11
adaptive Lasso freeHAR-FC	0.960	0.957	2.611***	0.993**	1.016***	0.914***	2.681***	0.968	1.029	0.948	1.014**	6/11
ordered Lasso AR-FC	**0.948**	**0.940**	**0.945*****	**0.966**	**0.951*****	**0.912*****	1.011*	**0.956****	**0.975****	**0.945**	0.974	10/11

Table 11(b): OOS Forecasts based on Full Sample 300-sec log RV: an IW approach. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with an increasing training set. The first training set includes the first 1000 observations and each subsequent training set contains one more observation. The evaluation window is set to (Dec 28, 2004 to Nov 15, 2010). We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	
slopeHAR	0.997	0.997	0.995	**0.993****	0.991	0.994*	0.998	1.000	**0.989****	0.995	**0.994**	10/11
freeHAR	0.996	1.000	0.994	0.996*	0.993*	0.992	0.998	1.002	0.993	**0.994****	0.996	9/11
adaptive Lasso AR	1.010	1.025	1.375***	1.055	1.002	1.027	1.058*	1.017	1.038	1.028	1.041	0/11
adaptive Lasso HAR	1.080***	1.040	1.416***	1.081*	1.075***	1.030***	1.051**	1.058***	1.071***	1.048***	1.084**	0/11
adaptive Lasso slopeHAR	1.030	1.060	1.118**	1.080**	1.058	1.039	1.057**	1.038*	1.017	1.065***	1.077	0/11
adaptive Lasso freeHAR	1.047***	1.037	1.409***	1.085**	1.059***	1.036**	1.040*	1.037**	1.045**	1.047***	1.101***	0/11
ordered Lasso AR	1.013	1.013	1.058	1.101	0.985*	1.046*	0.992	1.085**	1.134***	1.075*	0.999	3/11
adaptive Lasso AR-FC	0.995	1.002	1.284***	1.016	0.990	0.999	1.017	1.000	1.005	1.008	1.012	3/11
adaptive Lasso HAR-FC	1.015*	1.002	1.289***	1.005	1.010*	0.994	1.004	1.015**	1.013*	1.005	1.014	1/11
adaptive Lasso slopeHAR-FC	1.005	1.034*	1.091**	1.014*	1.037**	1.006	1.006	1.018*	1.001	1.021**	1.042	0/11
adaptive Lasso freeHAR-FC	1.003	1.011	1.295***	1.008	1.009*	1.003	0.994	1.024**	1.003	1.011*	1.025	1/11
ordered Lasso AR-FC	**0.993**	**0.993**	**0.990**	1.005	**0.980***	**0.985***	**0.989**	1.001	0.997	1.001	0.995*	8/11
horizon=5												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.995	0.995	0.997**	0.989	0.989	0.997	1.008	0.990	0.996	0.997	**0.989**	10/11
freeHAR	0.998	0.996	0.994	0.995	0.992	0.994	0.999	0.992	0.993	0.995	0.994	11/11
adaptive Lasso AR	1.009	1.006	1.767***	1.019	1.017	1.002	1.162***	0.999	1.057	1.017	1.010	1/11
adaptive Lasso HAR	1.028***	0.998	1.735***	1.036**	1.121***	1.007***	1.199***	1.033***	1.025**	1.005**	1.077***	1/11
adaptive Lasso slopeHAR	1.026**	1.102***	1.168**	1.037**	1.046*	1.024***	1.171***	1.015**	1.035*	1.062**	1.093***	0/11
adaptive Lasso freeHAR	1.016***	0.995	1.722***	1.025*	1.112***	1.000**	1.192***	1.004**	1.016*	1.009**	1.073***	1/11
ordered Lasso AR	1.016	1.003	1.050	1.101**	0.977	1.075***	1.003	1.081***	1.197***	1.097*	0.997*	2/11
adaptive Lasso AR-FC	0.993	0.987	1.594***	0.996	0.995	0.978	1.091*	0.986	1.021	1.000	1.004	6/11
adaptive Lasso HAR-FC	0.990	0.984	1.599***	0.990	1.052***	0.979	1.137***	0.998	0.983	0.988	1.021***	7/11
adaptive Lasso slopeHAR-FC	0.994	1.097***	1.130	1.015	0.995*	0.991*	1.071	0.996	1.004	1.020	1.029***	4/11
adaptive Lasso freeHAR-FC	0.987	0.985	1.580***	0.985	1.043***	0.985	1.143***	0.989	**0.974**	0.998*	1.021***	7/11
ordered Lasso AR-FC	**0.984**	**0.971**	**0.966*****	**0.979**	**0.966****	**0.970**	**0.996**	**0.980**	0.985	**0.984**	0.991	11/11
horizon=22												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	
slopeHAR	1.007	1.004	1.034***	1.004	1.018*	1.021	1.031	1.000	1.014	1.014	0.991*	1/11
freeHAR	1.001	1.000	0.999	0.999	0.998	0.999	1.000	0.999	0.999	0.999	0.999	8/11
adaptive Lasso AR	1.003	0.991	1.616***	0.991	1.054	1.019	1.252***	0.989	1.058*	1.022	0.973*	4/11
adaptive Lasso HAR	0.995***	0.975	1.452***	1.020***	1.027**	0.965	1.241***	1.016**	1.046***	0.977	1.054***	4/11
adaptive Lasso slopeHAR	0.988	1.006	1.153***	1.032***	1.034	0.975	1.246***	0.993	1.031	0.983	1.044***	4/11
adaptive Lasso freeHAR	0.983**	0.974	1.459***	1.011***	1.017	0.948	1.235***	1.004	1.025**	0.972	1.033***	4/11
ordered Lasso AR	0.972	0.945	1.012	1.048***	0.983***	1.019	1.025**	0.976**	1.119***	1.050**	0.965	5/11
adaptive Lasso AR-FC	0.991	0.985	1.487***	0.978	1.036	0.989	1.197**	0.969	1.023	0.998	0.977	7/11
adaptive Lasso HAR-FC	0.960	0.947	1.320***	0.978	0.983	**0.932*****	1.149**	0.971	0.975	**0.953**	0.998***	9/11
adaptive Lasso slopeHAR-FC	0.976	0.998	1.106	0.996**	1.016	0.948	1.145*	0.984	0.990	0.974	1.026***	7/11
adaptive Lasso freeHAR-FC	0.959	0.950	1.309*	0.984*	0.980	0.935**	1.145**	0.965	0.969	0.955	0.998***	9/11
ordered Lasso AR-FC	**0.941**	**0.932**	**0.933*****	**0.956**	**0.967*****	0.933***	1.014**	**0.951**	**0.960*****	0.963	**0.959**	10/11

Table 11(c): OOS Forecasts based on Full Sample 600-sec log RV: an IW approach. This table reports the values of the mean squared error (MSE) for different models considered relative to the MSE of the HAR model. Forecasts are constructed by re-estimating the parameters of the regressions each day with an increasing training set. The first training set includes the first 1000 observations and each subsequent training set contains one more observation. The evaluation window is set to (Dec 28, 2004 to Nov 15, 2010). We use the Diebold-Mariano test to compare the predictive ability between the benchmark HAR model and each of the other models (two extensions of the HAR and Lasso-based models). *, **, *** indicate that the differences in squared forecasting errors are significant at 10%, 5% and 1% level, respectively. For each horizon, the model with the best performance is highlighted in bold blue. The last column reports the number of times that each model outperforms the benchmark HAR model.

horizon=1	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.996	0.995	0.996	**0.995***	0.990	0.992**	0.999	**0.998**	**0.995**	**0.995**	**0.995**	11/11
freeHAR	0.997	0.997	0.995*	0.998	0.992*	0.992*	1.000	0.999	0.999	0.996	1.000	9/11
adaptive Lasso AR	1.016	1.023	1.142***	1.052**	1.009	1.031	1.059	1.018	1.053**	1.034	1.034	0/11
adaptive Lasso HAR	1.084***	1.043	1.202***	1.082**	1.062***	1.029**	1.062*	1.066***	1.056***	1.045***	1.072***	0/11
adaptive Lasso slopeHAR	1.027	1.052***	1.084	1.085***	1.049	1.066***	1.055***	1.036**	1.040	1.049**	1.086	0/11
adaptive Lasso freeHAR	1.054***	1.040**	1.200***	1.089**	1.062***	1.036**	1.050*	1.065***	1.056***	1.035***	1.088***	0/11
ordered Lasso AR	1.008	0.997	1.032	1.075***	0.983**	1.015	0.998	1.067*	1.084***	1.042	1.002	3/11
adaptive Lasso AR-FC	0.998	1.004	1.094**	1.022	0.993	0.998	1.020	0.998	1.017	1.012	1.012	4/11
adaptive Lasso HAR-FC	1.019***	1.008	1.144***	1.010	1.010**	1.000	1.010	1.022***	1.008	1.009*	1.019**	0/11
adaptive Lasso slopeHAR-FC	1.008	1.011	1.055	1.020	1.035**	1.028*	0.997	1.005	1.013	1.017**	1.038	1/11
adaptive Lasso freeHAR-FC	1.008	1.012	1.152***	1.016	1.015**	1.008	1.002	1.025**	1.018	1.008*	1.031**	0/11
ordered Lasso AR-FC	**0.995**	**0.993**	**0.991**	1.029**	**0.979****	**0.986****	**0.992**	1.001	1.004	1.006	1.001	6/11
horizon=5	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
slopeHAR	0.992	0.993	1.001**	**0.988**	0.993	0.992	1.006	0.990	1.001	0.994	**0.987**	8/11
freeHAR	0.996	0.994	0.992	0.997	0.991**	0.991	1.000	0.993	0.993	0.995	0.993	10/11
adaptive Lasso AR	1.011	0.991	1.519***	1.034**	0.997	1.003	1.086**	1.010	1.056**	1.016	0.994	3/11
adaptive Lasso HAR	1.034***	0.995	1.488***	1.054***	1.075***	1.003***	1.120***	1.041***	1.031**	1.006**	1.076***	1/11
adaptive Lasso slopeHAR	1.027***	1.010	1.175**	1.027*	1.035	1.025***	1.106***	1.017**	1.010	1.040**	1.081***	0/11
adaptive Lasso freeHAR	1.021***	0.996	1.460***	1.030**	1.071***	0.996	1.127***	1.025***	1.011*	0.999	1.072***	3/11
ordered Lasso AR	1.002	0.971	1.027	1.085***	0.966***	1.022**	1.000	1.082***	1.104***	1.046	0.990	3/11
adaptive Lasso AR-FC	0.995	0.976	1.400***	1.009	0.979	0.980	1.036	**0.988**	1.019	0.998	0.990	7/11
adaptive Lasso HAR-FC	0.993	0.976	1.385***	1.001	1.022	0.972	1.071***	1.004*	0.991	0.988	1.008**	5/11
adaptive Lasso slopeHAR-FC	0.993	0.990	1.116	1.001	1.011	0.994	1.040**	0.991	0.996	1.007*	1.020***	5/11
adaptive Lasso freeHAR-FC	0.989	0.978	1.368***	1.000	1.023*	0.979	1.075***	0.995	**0.984**	0.995	1.012***	6/11
ordered Lasso AR-FC	**0.983**	**0.959**	**0.968****	1.021*	**0.958*****	**0.961**	**0.986**	0.993	0.992	**0.982**	0.988	10/11
horizon=22	**SPY**	**MSFT**	**C**	**PFE**	**GE**	**HD**	**S**	**XOM**	**AA**	**WMT**	**DUK**	
HAR	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	1.000	
slopeHAR	1.004	1.006	1.039***	1.002	1.019**	1.016	1.019	0.998	1.018*	1.014**	0.988*	2/11
freeHAR	1.000	0.998	0.998	0.998	0.998	0.998	1.001	0.999	0.999	0.999	0.998	9/11
adaptive Lasso AR	1.007*	0.980	1.489***	1.007	1.044	1.033*	1.167	0.999*	1.067*	1.023	0.947**	3/11
adaptive Lasso HAR	1.008***	0.974	1.324***	1.041***	1.037**	0.983	1.199***	1.026**	1.019**	0.981	1.044***	3/11
adaptive Lasso slopeHAR	0.990*	0.994	1.174**	1.050***	1.037	0.973	1.215***	1.004	1.107**	0.984	0.999*	5/11
adaptive Lasso freeHAR	0.991**	0.964	1.317**	1.027***	1.034*	0.974	1.190***	1.019**	1.016**	0.978	1.033***	4/11
ordered Lasso AR	0.966	0.939	1.004	1.052***	0.979***	1.010	1.039	1.005***	1.098***	1.017	0.935**	4/11
adaptive Lasso AR-FC	0.997*	0.975	1.391***	0.994	1.033	0.999	1.118	0.981	1.035	1.003	0.948	6/11
adaptive Lasso HAR-FC	0.963	0.945	1.232*	0.985**	0.996	0.946	1.114**	0.974	0.972	**0.958**	0.966	9/11
adaptive Lasso slopeHAR-FC	0.972	1.003	1.108	1.025***	1.017	0.949	1.133**	0.974*	1.046**	0.980	1.027***	4/11
adaptive Lasso freeHAR-FC	0.961	0.945	1.229	**0.982****	0.995	**0.945**	1.114**	0.970	**0.968**	0.962	0.969	9/11
ordered Lasso AR-FC	**0.944**	**0.931**	**0.941****	0.996**	**0.964*****	0.951	1.018*	**0.969**	0.973**	0.966	**0.931****	10/11

Table 3.12: Estimates of AR Coefficients for the SPY. This table reports the AR coefficients implied by the HAR, slopeHAR, freeHAR and those given by the adaptive Lasso AR, Ordered Lasso AR, cluster group Lasso AR as well as by the group Lasso AR (1, 5, 22, 50, 75, 100). Cluster corresponds to the arrangements of groups of the lagged RVs identified by the cluster group Lasso method. Models are estimated on the full sample from Jan 02, 2001 to Nov 15, 2010.

	SPY							
Lag	HAR	slopeHAR	freeHAR	LassoAR	orderedLassoAR	Cluster	ClusterGroupLassoAR	GroupLassoAR
1	0.49499	0.47752	0.47122	0.55597	0.46458	1	0.26892	0.47562
2	0.09021	0.13314	0.18840	0.24514	0.18062	1	0.16339	0.10224
3	0.09021	0.10339	0.03674	0	0.05839	1	0.09033	0.08036
4	0.09021	0.07365	0.07820	0.03429	0.05839	1	0.07977	0.08404
5	0.09021	0.04390	0.07687	0.06059	0.05840	1	0.06858	0.08356
6	0.00612	0.01415	0.00572	0	0.01489	1	0.03766	0.00248
7	0.00612	0.01332	0.00647	0	0.01489	1	0.02581	0.00253
8	0.00612	0.01249	0.00647	0	0.01489	1	0.01362	0.00251
9	0.00612	0.01166	0.00647	0	0.01489	1	0.03391	0.00290
10	0.00612	0.01082	0.00647	0	0.01489	1	0.03091	0.00282
11	0.00612	0.00999	0.00647	0	0.01489	1	0.02399	0.00273
12	0.00612	0.00916	0.00647	0	0.00217	1	0.00873	0.00250
13	0.00612	0.00833	0.00647	0	0.00217	1	0.01655	0.00261
14	0.00612	0.00749	0.00647	0	0.00217	1	0.01157	0.00248
15	0.00612	0.00666	0.00647	0	0.00216	1	0.01287	0.00246
16	0.00612	0.00583	0.00647	0	0.00125	1	-0.01814	0.00200
17	0.00612	0.00500	0.00647	0	0.00125	1	-0.00808	0.00221
18	0.00612	0.00416	0.00647	0	0.00125	1	-0.01029	0.00218
19	0.00612	0.00333	0.00647	0.01031	0.00125	1	0.02214	0.00267
20	0.00612	0.00250	0.00647	0	0.00125	1	0.01594	0.00250
21	0.00612	0.00167	0.00647	0	0.00125	1	0.00113	0.00231
22	0.00612	0.00083	0.00647	0	0.00125	1	0.01102	0.00245
23	0	0	0	0	0.00125	1	-0.00128	0
24	0	0	0	0	0.00125	1	-0.00015	0
25	0	0	0	0	0.00125	1	-0.01679	0
26	0	0	0	0	0.00125	1	0.00169	0
27	0	0	0	0	0.00125	1	-0.00457	0
28	0	0	0	0	0.00125	1	0.00283	0
29	0	0	0	0	0.00125	1	0.00302	0
30	0	0	0	0	0.00125	1	-0.00128	0
31	0	0	0	0	0.00125	1	0.00383	0
32	0	0	0	0	0.00125	1	0.00009	0
33	0	0	0	0	0.00125	1	0.00778	0
34	0	0	0	0	0.00125	1	0.00127	0
35	0	0	0	0	0.00125	1	-0.00158	0
36	0	0	0	0	0.00125	1	0.00313	0
37	0	0	0	0	0.00125	1	0.01030	0
38	0	0	0	0.01373	0.00124	1	0.02844	0
39	0	0	0	0	0.00124	1	0.01048	0
40	0	0	0	0	0.00124	1	-0.00739	0
41	0	0	0	0	0.00124	1	-0.01723	0
42	0	0	0	0	0.00124	1	-0.00307	0
43	0	0	0	0	0.00124	1	0.00899	0
44	0	0	0	0	0.00124	1	0.00183	0
45	0	0	0	0	0.00124	1	-0.00484	0
46	0	0	0	0	0.00124	1	-0.00434	0
47	0	0	0	0	0.00124	1	0.00493	0
48	0	0	0	0	0.00124	1	0.00950	0
49	0	0	0	0	0.00124	1	0.01391	0
50	0	0	0	0	0.00075	1	0.00002	0
51	0	0	0	0	0	2	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	0	0	0	0	0	2	0	0

Table 3.13: Estimates of AR Coefficients for the MSFT. This table reports the AR coefficients implied by the HAR, slopeHAR, freeHAR and those given by the adaptive Lasso AR, Ordered Lasso AR, cluster group Lasso AR as well as by the group Lasso AR (1, 5, 22, 50, 75, 100). Cluster corresponds to the arrangements of groups of the lagged RVs identified by the cluster group Lasso method. Models are estimated on the full sample from Jan 02, 2001 to Nov 15, 2010.

	MSFT							
Lag	HAR	slopeHAR	freeHAR	LassoAR	orderedLassoAR	Cluster	ClusterGroupLassoAR	GroupLassoAR
1	0.40269	0.38889	0.38794	0.49017	0.37446	1	0.21233	0.38573
2	0.09883	0.13804	0.15218	0.19320	0.14251	1	0.13173	0.10470
3	0.09883	0.10885	0.08608	0.04982	0.08002	1	0.09663	0.09280
4	0.09883	0.07966	0.09061	0.12025	0.08003	1	0.08629	0.09189
5	0.09883	0.05047	0.05026	0	0.04187	1	0.06726	0.08614
6	0.00943	0.02128	0.05188	0	0.03769	1	0.05902	0.00582
7	0.00943	0.02002	0.00873	0	0.02136	1	0.05002	0.00578
8	0.00943	0.01877	0.00873	0	0.01920	1	0.03039	0.00520
9	0.00943	0.01752	0.00873	0	0.01920	1	0.04879	0.00601
10	0.00943	0.01627	0.00873	0	0.01920	1	0.03827	0.00552
11	0.00943	0.01502	0.00873	0.03859	0.01920	1	0.05675	0.00608
12	0.00943	0.01377	0.00873	0	0.00104	2	0	0.00496
13	0.00943	0.01252	0.00873	0	0.00104	2	0	0.00515
14	0.00943	0.01126	0.00873	0	0.00104	2	0	0.00524
15	0.00943	0.01001	0.00873	0	0.00104	2	0	0.00505
16	0.00943	0.00876	0.00873	0	0.00104	2	0	0.00522
17	0.00943	0.00751	0.00873	0	0.00104	2	0	0.00442
18	0.00943	0.00626	0.00873	0	0.00104	2	0	0.00447
19	0.00943	0.00501	0.00873	0	0.00104	2	0	0.00459
20	0.00943	0.00375	0.00873	0	0.00104	2	0	0.00489
21	0.00943	0.00250	0.00873	0	0	2	0	0.00513
22	0.00943	0.00125	0.00873	0	0	2	0	0.00517
23	0	0	0	0	0	2	0	0
24	0	0	0	0	0	2	0	0
25	0	0	0	0	0	2	0	0
26	0	0	0	0	0	3	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
35	0	0	0	0	0	3	0	0
36	0	0	0	0	0	4	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
47	0	0	0	0	0	4	0	0
48	0	0	0	0	0	5	0	0
49	0	0	0	0	0	5	0	0
50	0	0	0	0	0.00103	5	0	0
51	0	0	0	0	0.00102	5	0	0.00025
52	0	0	0	0	0.00102	5	0	0.00023
53	0	0	0	0	0.00102	5	0	0.00026
54	0	0	0	0	0.00102	5	0	0.00024
55	0	0	0	0	0.00102	5	0	0.00024
56	0	0	0	0	0.00102	6	0.00283	0.00024
57	0	0	0	0	0.00102	6	0.00381	0.00027
58	0	0	0	0	0.00102	6	0.00380	0.00027
59	0	0	0	0	0.00102	6	0.00348	0.00026
60	0	0	0	0	0.00102	6	0.00393	0.00027
61	0	0	0	0.00850	0.00102	6	0.00509	0.00030
62	0	0	0	0	0.00102	6	0.00459	0.00028
63	0	0	0	0	0.00102	6	0.00406	0.00027
64	0	0	0	0	0.00102	6	0.00380	0.00026
65	0	0	0	0	0.00101	6	0.00368	0.00026
66	0	0	0	0	0.00101	6	0.00377	0.00026
67	0	0	0	0	0.00101	6	0.00364	0.00026
68	0	0	0	0	0.00101	7	0	0.00028
69	0	0	0	0	-2.27E-11	7	0	0.00024
70	0	0	0	0	-2.27E-11	7	0	0.00026
71	0	0	0	0	-2.27E-11	7	0	0.00023
72	0	0	0	0	-5.97E-12	7	0	0.00022
73	0	0	0	0	0	7	0	0.00022
74	0	0	0	0	0	7	0	0.00022
75	0	0	0	0	0	7	0	0.00023
76	0	0	0	0	0	7	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
83	0	0	0	0	0	7	0	0
84	0	0	0	0	0	8	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	0	0	0	0	0	8	0	0

Figure 3.1: Time series plots. Figures depict the time series of log realized variance (RV) of the SPY and 10 individual stocks based upon 300-second frequency. Data covers the period of Jan 02, 2001 to Nov 15, 2010.

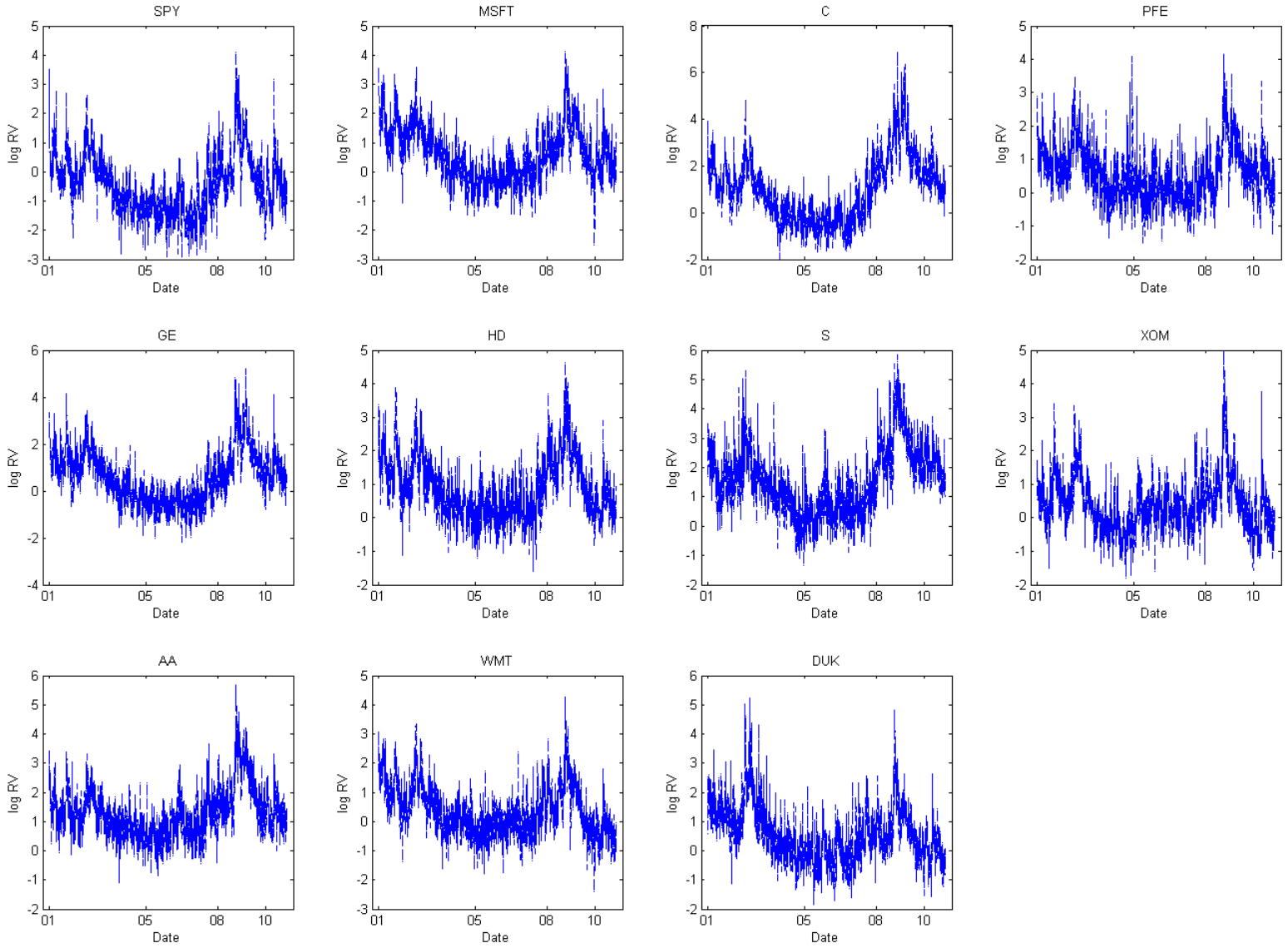


Figure 3.2: Plots of Partial Autocorrelation Function (PACF) for the Full Sample Period. The PACF of log 300-second RV for each stock is provided for the whole period considered.

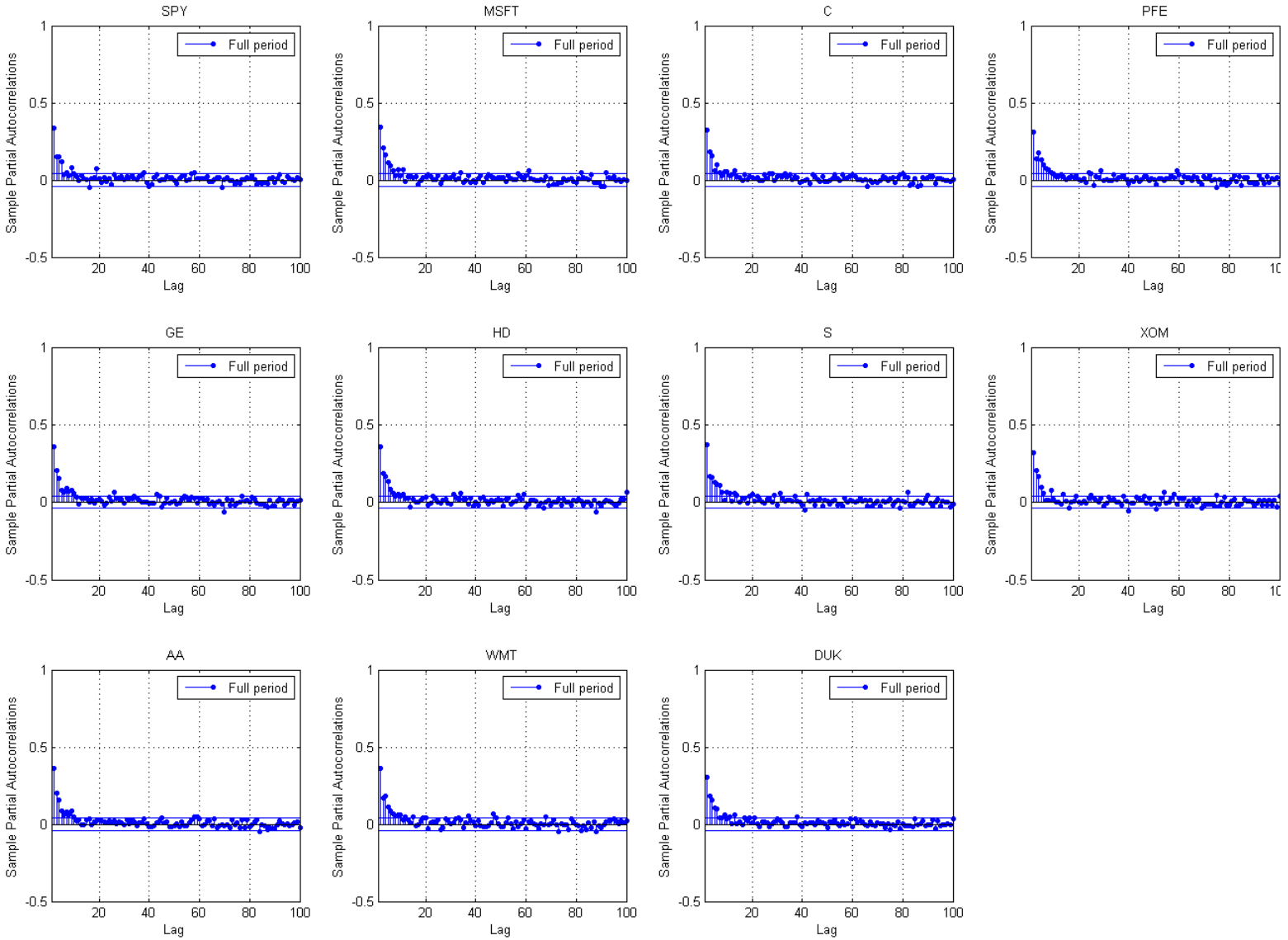


Figure 3.3: Plots of Partial Autocorrelation Function (PACF) for the Pre-crisis Period. The PACF of log 300-second RV for each stock is provided for the pre-crisis period.

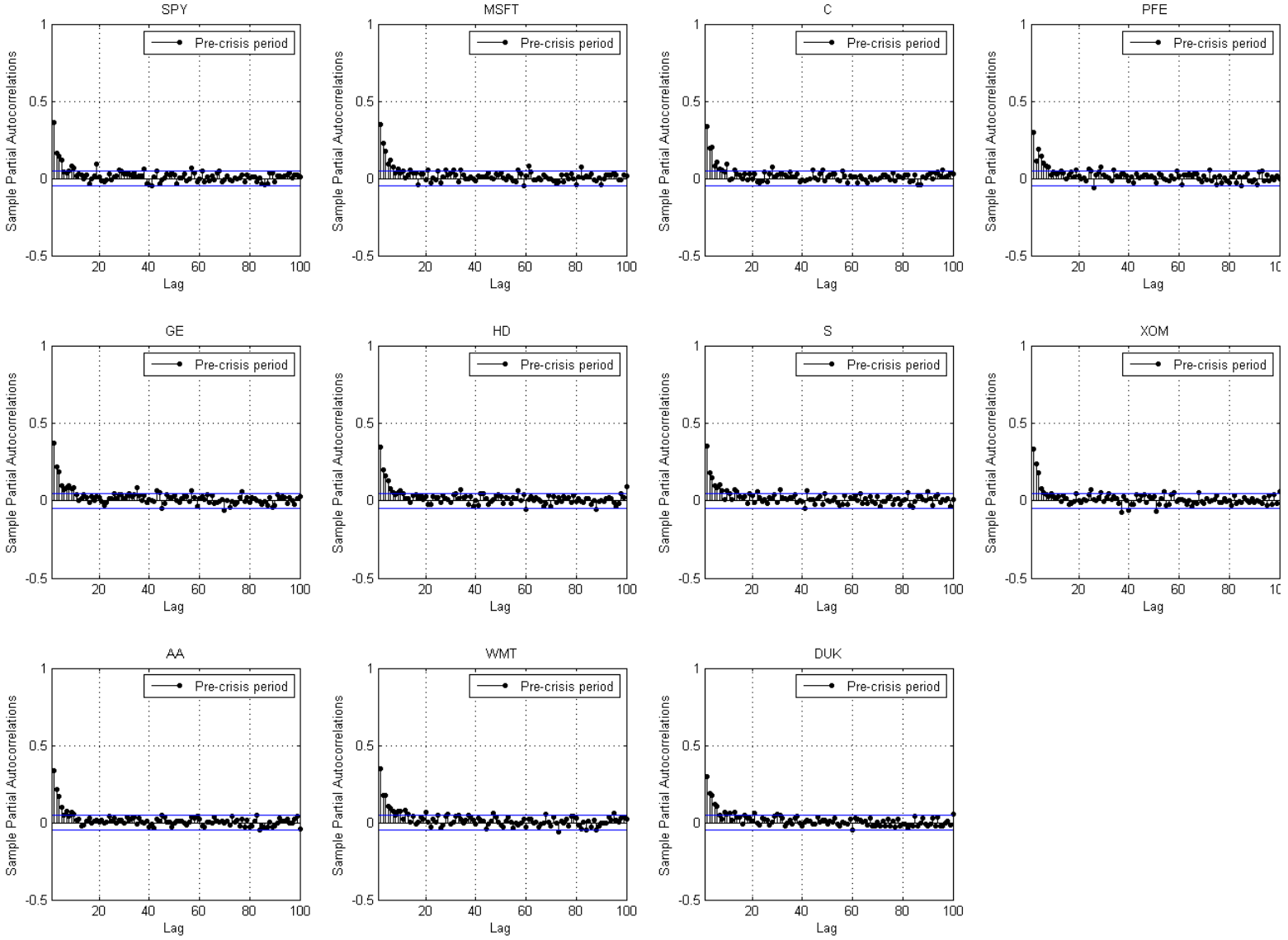


Figure 3.4: Plots of Partial Autocorrelation Function (PACF) for the Post-crisis Period. The PACF of log 300-second RV for each stock is provided for the post-crisis period.

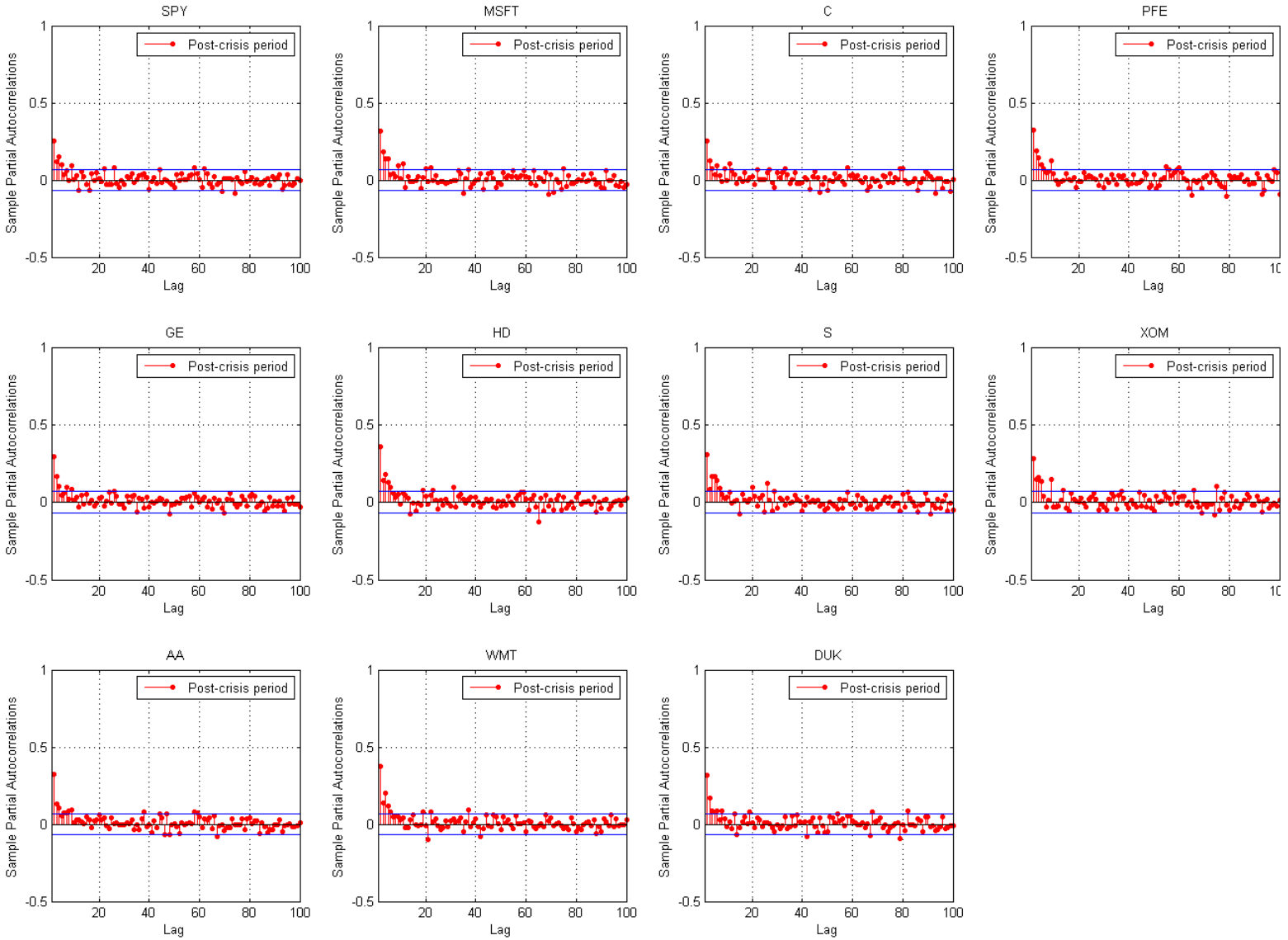


Figure 3.5: HAR versus adaptive Lasso AR Coefficients. The graph plots the estimated coefficients of the HAR and adaptive Lasso AR for different stocks. RV is constructed from 300-second intraday returns. The time span is from Jan 02, 2001 to Nov 15, 2010.

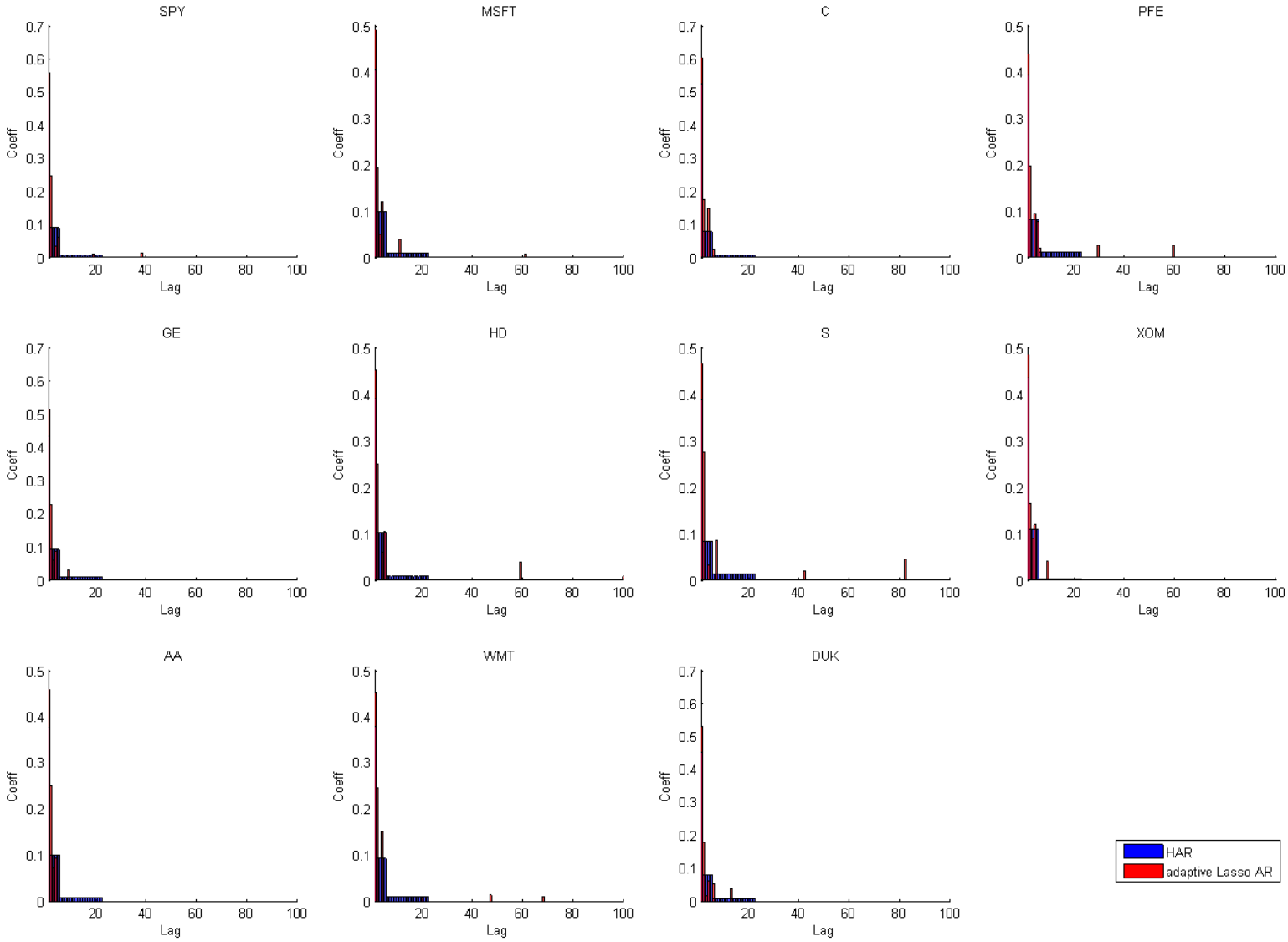


Figure 3.6: slopeHAR versus freeHAR AR Coefficients. The graph plots the estimated coefficients of the slopeHAR and freeHAR for different stocks. RV is constructed from 300-second intraday returns. The time span is from Jan 02, 2001 to Nov 15, 2010.

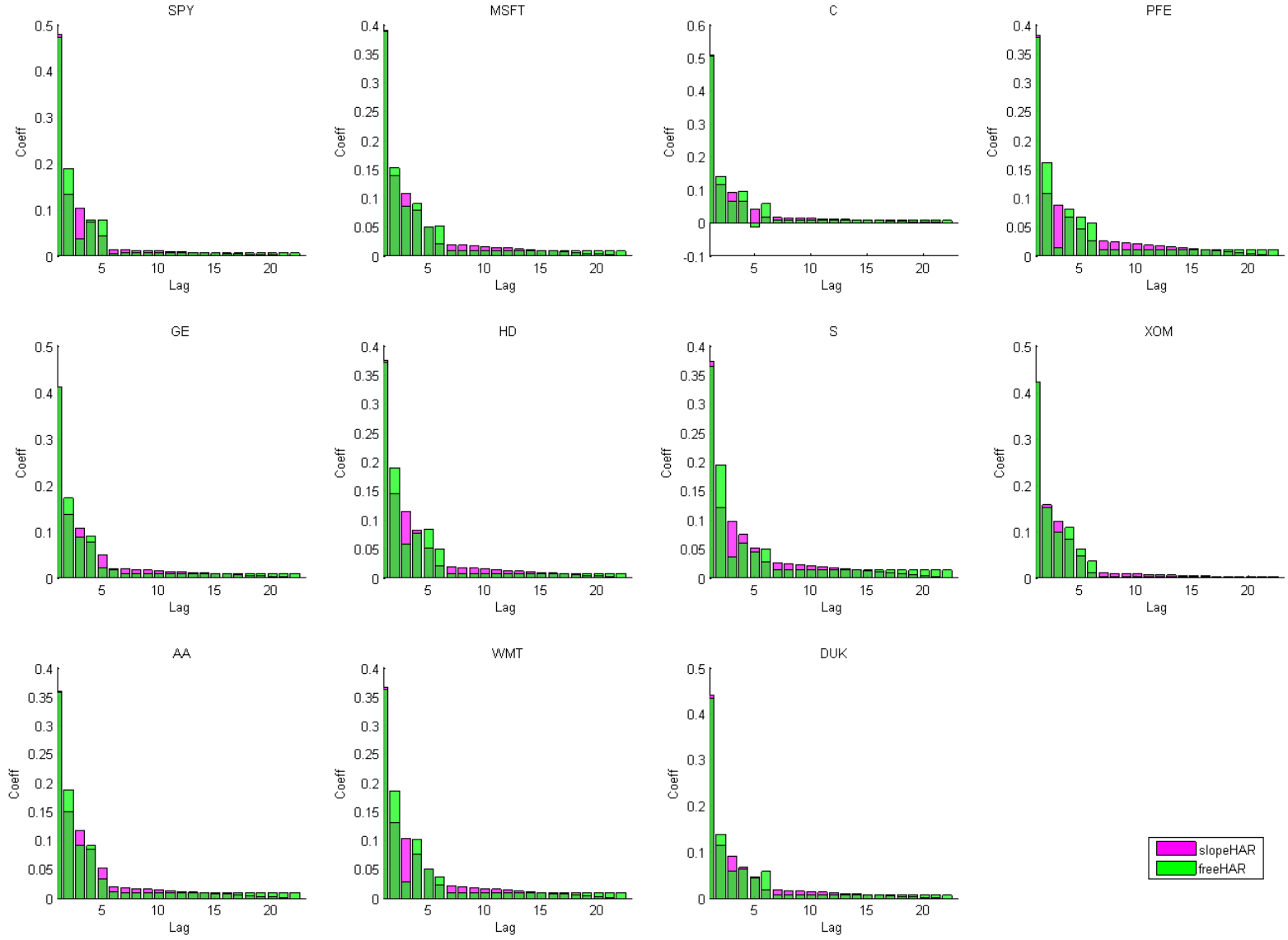


Figure 3.7: HAR, cluster group Lasso AR and group Lasso AR (1, 5, 22, 50, 75, 100) Coefficients. The graph plots the estimated coefficients of the HAR, cluster group Lasso AR and group Lasso AR (1, 5, 22, 50, 75, 100) models for different stocks. For better exposition of the plots, we report the coefficients for lags beyond $logRV_{t-6}$ only. For the group Lasso AR (1, 5, 22, 50, 75, 100), we group the lags in AR(100) as 1, 2-5, 6-22, 23-50, 51-75 and 76-100 in order to examine the appropriateness of the traditional HAR lag structure. RV is constructed from 300-second intraday returns. The time span is from Jan 02, 2001 to Nov 15, 2010.

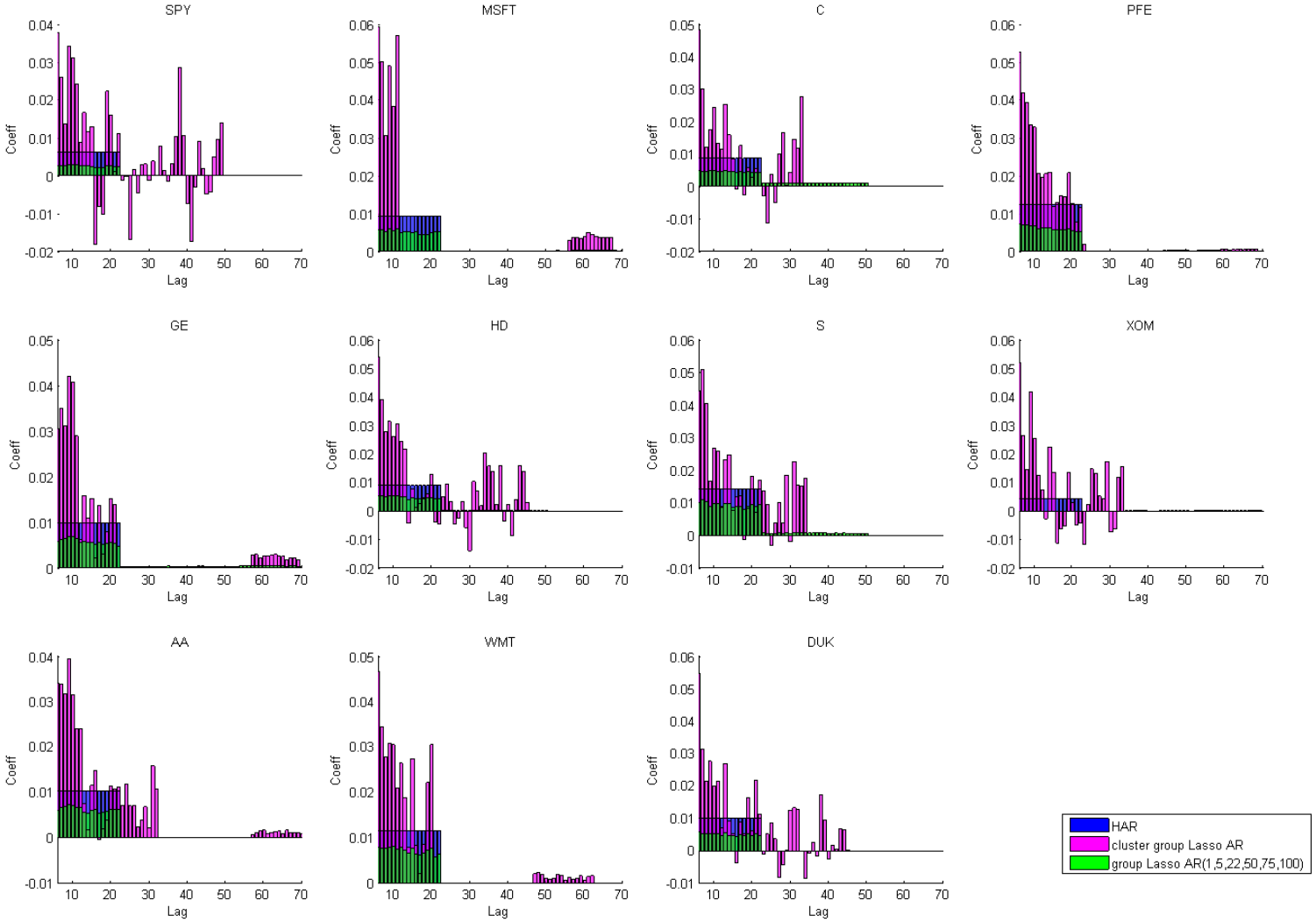
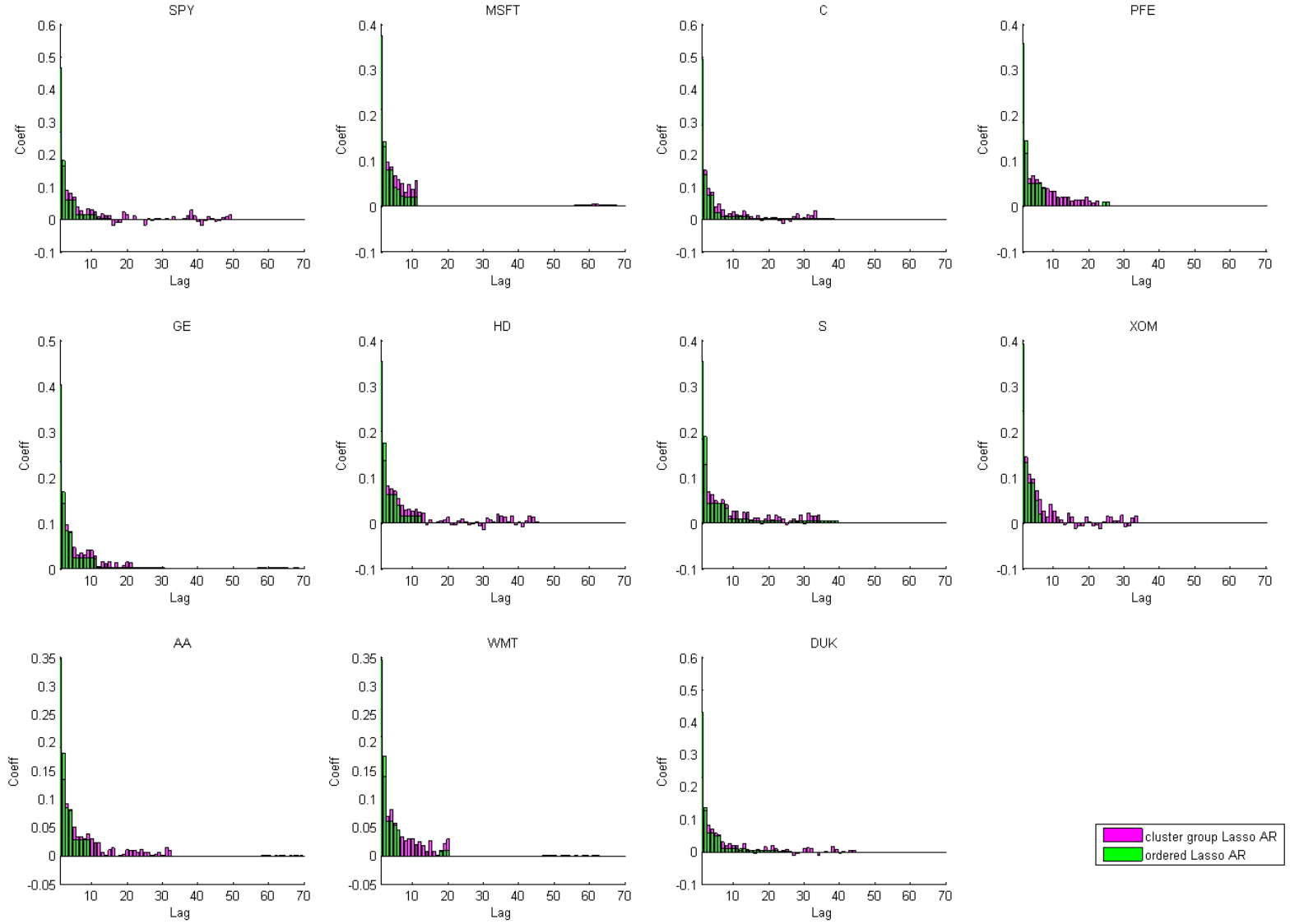


Figure 3.8: Ordered Lasso AR and cluster group Lasso AR Coefficients. The graph plots the estimated coefficients of the Ordered Lasso AR and cluster group Lasso AR models for different stocks. RV is constructed from 300-second intraday returns. The time span is from Jan 02, 2001 to Nov 15, 2010.



Concluding Remarks

This thesis provides different methods of enhancing the modelling and forecasting of future returns and realized variance (RV). We adopt both time series models (Chapter 1 and 3) and options-based approach (Chapter 2) in order to achieve better volatility forecast (Chapter 2 and 3) and return predictions (Chapter 1 and 2). Among the time series techniques considered, both univariate (Chapter 3) and multivariate (Chapter 1) models are employed.

In Chapter 1, we propose modifications to the fractionally co-integrated vector autoregressive (FCVAR) model developed by Johansen (2008) to accommodate systems containing $I(d)$ and $I(0)$ variables under the presence of long memory in the co-integrating residuals. The proposed model is termed the M-FCVAR. In the simulation study and empirical application, we show that the M-FCVAR delivers better inference than the FCVAR. In addition, we investigate the impact of the shock to the $I(0)$ variable on the $I(d)$ variables within the system, which could either be permanent or transitory. Particular equation specifications are outlined to restrict the shock arising from the $I(0)$ variable, so that it exerts only transitory effect on the $I(d)$ variables. The simulation evidence indicates that failure in restricting the shock associated with the $I(0)$ variable when its impact on the $I(d)$ variables is transitory may result in biased model estimates and low in-sample fit. In the empirical study, the FCVAR and M-FCVAR are used for a joint modelling of the dynamic dependencies in stock market returns, RV and option-implied volatility. With less biased estimates of the fractional integration order, degree of co-integration and the co-integrating relationship, the M-FCVAR dominates the FCVAR in terms of the return predictions over long horizons.

In Chapter 2, we evaluate the performance of different measures of model-free implied volatility in forecasting future returns and RV. The Monte Carlo simulations suggest that: first, the out-of-sample volatility forecast can be improved by adopting an interpolation and extrapolation technique; second, only an interpolation method is needed in attempt to enhance the predictive power of implied volatilities for future returns. In the empirical study using SPX options, the aforementioned procedure, i.e. interpolation/extrapolation approach, is found to work well for most measures considered. In addition, with the use of this procedure, the SPX OTM call options outperform the OTM put options in terms of their forecasting performance for future RV and returns, which is consistent with the simulation results. However, the advantages of the SPX OTM put options are evident when implied volatilities are derived from the observed options only.

In Chapter 3, we examine the usefulness of least absolute shrinkage and selection operator (Lasso) based models in the forecast of future RV using a comprehensive empirical study containing the SPY and ten individual stocks from different sectors. The in-sample analysis implies that the popular heterogeneous autoregressive (HAR) model is not fully consistent with the Lasso-type models with regard to the lag structure, which casts doubt on the appropriateness of the HAR in modelling the dynamics of the financial volatility. In the out-of-sample analysis, we find that, in most cases, the Lasso-based model dominates the other candidates including the HAR and its extensions and that the forecast combination tends to improve the accuracy of volatility forecast delivered by the Lasso-based models. The ordered Lasso AR with the forecast combination provides the top forecast most frequently and its improvements over the HAR model are generally significant over monthly forecasting horizons. The global financial crisis is shown to produce non-trivial impact on the performance of the Lasso-based models. However, the order Lasso AR using the forecast combination still plays a leading role in the post crisis period, especially over long horizons. Furthermore, a larger window size helps the Lasso-based models to display their advantages in the volatility forecast. Finally, although the variation in the sampling frequency upon which the RV is

based does not alter the conclusions outlined above, the superior performance of the Lasso-based model becomes more evident in the full sample as the sampling frequency increases.

References

Aït-Sahalia, Y., and L. Mancini (2008): "Out of sample forecasts of quadratic variation," *Journal of Econometrics*, 147(1), 17–33.

Andersen, T., T. Bollersley, F. X. Diebold, and H. Ebens (2001a): "The distribution of realized stock return volatility," *Journal of Financial Economics*, 61(1), 43–76.

Andersen, T. G., and T. Bollerslev (1997): "Intraday periodicity and volatility persistence in financial markets," *Journal of Empirical Finance*, 4(2-3), 115–158.

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

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

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001b): "The distribution of realized exchange rate volatility," *Journal of the American Statistical Association*, 96(453), 42–55.

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

Andersen, T. G., T. Bollerslev, and N. Meddahi (2011): "Realized

volatility forecasting and market microstructure noise," *Journal of Econometrics*, 160(1), 220–234.

Andersen, T. G., and O. Bondarenko (2007): "Construction and interpretation of model-free implied volatility," in *Volatility as an Asset Class*, ed. by I. Nelken, pp. 141–181. Risk Books, London, UK.

Andersen, T. G., O. Bondarenko, and M. T. Gonzalez-Perez (2015): "Exploring Return Dynamics via Corridor Implied Volatility," *Review of Financial Studies*, 28(10), 2902–2945.

Andersen, T. G., N. Fusari, and V. Todorov (2015): "The risk premia embedded in index options," *Journal of Financial Economics*, 117(3), 558–584.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006): "The cross-section of volatility and expected returns," *Journal of Finance*, 61(1), 259–299.

Audrino, F., L. Camponovo, and C. Roth (2015): "Testing the lag structure of assets' realized volatility dynamics," *Working Paper*.

Audrino, F., C. Huang, and O. Okhrin (2016): "Flexible HAR Model for Realized Volatility," *Working Paper*, pp. 1–25.

Audrino, F., and S. D. Knaus (2016): "Lassoing the HAR model: A Model Selection Perspective on Realized Volatility Dynamics," *Econometric Reviews*, 35(8-10), 1485–1521.

Avarucci, M., and C. Velasco (2009): "A Wald test for the cointegration rank in nonstationary fractional systems," *Journal of Econometrics*, 151(2), 178–189.

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

Baillie, R. T., G. Geoffrey Booth, Y. Tse, and T. Zabotina (2002): "Price discovery and common factor models," *Journal of Financial Markets*, 5(3), 309–321.

Bandi, F. M., and B. Perron (2006): "Long memory and the relation between implied and realized volatility," *Journal of Financial Econometrics*, 4(4), 636–670.

Bandi, F. M., and J. R. Russell (2008): "Microstructure noise , realized volatility , and optimal sampling," *Review of Economic Studies*, 75(2), 339–369.

Bandi, F. M., J. R. Russell, and Y. Zhu (2008): "Using High-Frequency Data in Dynamic Portfolio Choice," *Econometric Reviews*, 27(June 2009), 163–198.

Banerjee, A., M. Marcellino, and C. Osbat (2004): "Some cautions on the use of panel methods for integrated series of macroeconomic data," *Econometrics Journal*, 7(2), 322–340.

Banerjee, P. S., J. S. Doran, and D. R. Peterson (2007): "Implied volatility and future portfolio returns," *Journal of Banking and Finance*, 31, 3183–3199.

Barndorff-Nielsen, O. E. (2004): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2(1), 1–37.

Barndorff-Nielsen, O. E., and N. Shephard (2002): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 64, 253–280.

Bates, D. S. (2000): "Post-'87 crash fears in the S{&}P 500 futures option market," *Journal of Econometrics*, 94(1-2), 181–238.

Bates, D. S. (2008): "The market for crash risk," *Journal of Economic Dynamics and Control*, 32, 2291–2321.

Bates, D. S. (2012): "U.S. stock market crash risk, 1926-2010," *Journal of Financial Economics*, 105(2), 229–259.

Bates, J. M., and C. W. J. Granger (1969): "The Combination of Forecasts," *OR*, 20(4), 451.

Biais, B., L. Glosten, and C. Spatt (2005): "Market microstructure: A survey of microfoundations, empirical results, and policy implications," *Journal of Financial Markets*, 8(2), 217–264.

Blair, B. J., S.-h. Poon, and S. J. Taylor (2001): "Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns," *Journal of Econometrics*, 105(1), 5–26.

Blanco, R., S. Brennan, and I. W. Marsh (2005): "An empirical analysis of the dynamic relationship between investment-grade bonds and credit default swaps," *The Journal of Finance*, 60(5).

Bliss, R. R., and N. Panigirtzoglou (2002): "Testing the stability of implied probability density functions," *Journal of Banking and Finance*, 26(2-3), 381–422.

Bliss, R. R., and N. Panigirtzoglou (2004): "Option-Implied Risk Aversion Estimates by Option-Implied Risk Aversion Estimates," *The Journal of Finance*, 59(1), 407–446.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, 21, 307–327.

Bollerslev, T. (1987): "A conditionally heteroskedastic time series model for speculative prices and rates of return," *The review of economics and statistics*, 69, 542–547.

Bollerslev, T., B. Hood, J. Huss, and L. H. Pedersen (2016): "Risk Everywhere: Modeling and Managing Volatility," *Working Paper*.

Bollerslev, T., U. Kretschmer, C. Pigorsch, and G. Tauchen (2009): "A discrete-time model for daily S & P500 returns and realized variations: Jumps and leverage effects," *Journal of Econometrics*, 150(2), 151–166.

Bollerslev, T., J. Marrone, L. Xu, and H. Zhou (2014): "Stock Return Predictability and Variance Risk Premia: Statistical Inference and International Evidence," *Journal of Financial and Quantitative Analysis*, 49(03), 633–661.

Bollerslev, T., D. Osterrieder, N. Sizova, and G. Tauchen (2013): "Risk and return: Long-run relations, fractional cointegration, and return predictability," *Journal of Financial Economics*, 108(2), 409–424.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2016): "Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting," *Journal of Econometrics*, 192(1), 1–18.

Bollerslev, T., N. Sizova, and G. Tauchen (2012): "Volatility in equilibrium: Asymmetries and dynamic dependencies," *Review of Finance*, 16(1), 31–80.

Bollerslev, T., G. Tauchen, and H. Zhou (2009): "Expected stock returns and variance risk premia," *Review of Financial Studies*, 22(11), 4463–4492.

Bondarenko, O. (2003): "Estimation of risk-neutral densities using positive convolution approximation," *Journal of Econometrics*, 116(1-2), 85–112.

Breitung, J., and U. Hassler (2002): "Inference on the cointegration rank in fractionally integrated processes," *Journal of Econometrics*, 110(2), 167–185.

Britten-Jones, M., and A. Neuberger (2000): "Option prices, implied price processes, and stochastic volatility," *Journal of Finance*, 55, 839–866.

Bühlmann, P., P. Rütimann, S. van de Geer, and C. H. Zhang (2013): "Correlated variables in regression: Clustering and sparse estimation," *Journal of Statistical Planning and Inference*, 143(11), 1835–1858.

Busch, T., B. J. Christensen, and M. Ø. Nielsen (2011a): "The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets," *Journal of Econometrics*, 160(1), 48–57.

——— (2011b): "The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets," *Journal of Econometrics*, 160(1), 48–57.

Campbell, J. Y. (1991): "A variance decomposition for stock returns," *Economic Journal*, 101, 151–179.

Caporin, M., A. Ranaldo, and P. Santucci de Magistris (2013): "On the predictability of stock prices: A case for high and low prices," *Journal of Banking and Finance*, 37(12), 5132–5146.

Carlini, F., and P. Santucci de Magistris (2017): "On the Identification of Fractionally Cointegrated VAR Models with the F(d) Condition," *Journal of Business and Economic Statistics*, (just-accepted).

Carnero, M. a., D. Peña, and E. Ruiz (2004): "Persistence and Kurtosis in GARCH and Stochastic Volatility Models," *Journal of Financial Econometrics*, 2(2), 319–342.

Carr, P., and L. Wu (2006): "A Tale of Two Indices," *The Journal of Derivatives*, 13, 13–29.

Chen, W. W., and C. M. Hurvich (2003a): "Estimating fractional cointegration in the presence of polynomial trends," *Journal of Econometrics*, 117(1), 95–121.

Chen, W. W., and C. M. Hurvich (2003b): "Semiparametric estimation of multivariate fractional cointegration," *Journal of the American Statistical Association*, 98(463), 629–642.

Chen, W. W., and C. M. Hurvich (2006): "Semiparametric estimation of fractional cointegrating subspaces," *The Annals of Statistics*, 34(6), 2939–2979.

Choi, K., W. Yu, and E. Zivot (2010): "Long memory versus structural breaks in modeling and forecasting realized volatility," *Journal of International Money and Finance*.

Christensen, B., and N. Prabhala (1998): "The relation between implied and realized volatility," *Journal of Financial Economics*, 50(2), 125–150.

Christensen, B. J., and M. Ø. Nielsen (2006): "Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting," *Journal of Econometrics*, 133(1), 343–371.

Comte, F., L. Coutin, and E. Renault (2012): "Affine fractional stochastic volatility models," *Annals of Finance*, 8(2-3), 337–378.

Comte, F., and E. Renault (1998): "Long memory in continuous-time stochastic volatility models," *Mathematical Finance*, 8(4), 291–323.

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

Corsi, F., S. Mittnik, C. Pigorsch, and U. Pigorsch (2008): "The Volatility of Realized Volatility," *Econometric Reviews*, 27(1-3), 46–78.

Corsi, F., and R. Renò (2012): "Discrete-Time Volatility Forecasting With Persistent Leverage Effect and the Link With Continuous-Time Volatility Modeling," *Journal of Business and Economic Statistics*, 30(3), 368–380.

Corte, P. D., and I. Tsiakas (2012): "Statistical and Economic Methods for Evaluating Exchange Rate Predictability," in *Handbook of Exchange Rates*, ed. by J. James, L. Sarno, and I. Marsh. Wiley, London.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985): "A theory of the term structure of interest rates," *Econometrica*, 53, 385–407.

Craioveanu, M., and E. Hillebrand (2012): "Why It Is OK to Use the HAR-RV (1, 5, 21) Model," *Working Paper*.

De Pooter, M., M. Martens, and D. Van Dijk (2008): "Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data-But Which Frequency to Use?," *Econometric Reviews*, 27(13), 199–229.

Degiannakis, S., and C. Floros (2013): "Modeling CAC40 volatility using ultra-high frequency data," *Research in International Business and Finance*, 28(1), 68–81.

Demeterfi, K., E. Derman, M. Kamal, and J. Zou (1999): "More Than You Ever Wanted To Know About Volatility Swaps," *Working Paper*.

Dolatabadi, S., M. Nielsen, and K. Xu (2015): "A fractionally cointegrated VAR analysis of price discovery in commodity futures markets," *Journal of Futures Markets*, 35(4), 339–356.

Dolatabadi, S., M. Ø. Nielsen, and K. Xu (2016): "A fractionally cointegrated VAR model with deterministic trends and application to commodity futures markets," *Journal of Empirical Finance*, 38, 623–639.

Dotsis, G., and N. Vlastakis (2016): "Corridor volatility risk and expected returns," *Journal of Futures Markets*, 36, 488–505.

Drechsler, I., and A. Yaron (2010): "What's Vol got to do with it," *Review of Financial Studies*, 24(1), 1–45.

Duan, J.-C., and J.-G. Simonato (1998): "Empirical martingale simulation for asset prices," *Management Science*, 44, 1218–1233.

Duan, J.-C., and C.-Y. Yeh (2010): "Jump and volatility risk premiums implied by VIX," *Journal of Economic Dynamics and Control*, 34(11), 2232–2244.

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

Engle, R. F., and C. W. J. Granger (1987): "Co-Integration and Error Correction: representation, estimation, and testing," *Econometrica: Journal of the Econometric Society*, 55(2), 251–276.

Engle, R. F., D. M. Lilien, and R. P. Robins (1987): "Estimating time varying risk premia in the term structure: The ARCH-M model," *Econometrica: Journal of the Econometric Society*, pp. 391–407.

Figlewski, S. (2008): "Estimating the Implied Risk-Neutral Density for the US Market Portfolio," in *Volatility and Time Series Econometrics: essays in honor of Robert F. Engle*, ed. by T. Bollersley, J. R. Russell, and M. Watson. Oxford University Press, Oxford, UK.

Fisher, L. A., H.-S. Huh, and A. R. Pagan (2016): "Econometric Methods for Modelling Systems With a Mixture of I(1) and I(0) Variables," *Journal of Applied Econometrics*, 31(5), 892–911.

Foster, D., and D. Nelson (1996): "Continuous Record Asymptotics for Rolling Sample Variance Estimators," *Econometrica*, 64(1), 139–174.

Franchi, M. (2010): "A representation theory for polynomial cofractionality in vector autoregressive models," *Econometric Theory*, 26(4), 1201–1217.

Friedman, J., T. Hastie, and R. Tibshirani (2010): "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal Of Statistical Software*, 33(1), 1–22.

Garvey, J. F., and L. A. Gallagher (2012): "the Realised-Implied volatility relationship : Recent empirical evidence from FTSE-100 stocks," *Journal of Forecasting*, 31(7), 639–660.

Geweke, J., and S. Porter-Hudak (1983): "The estimation and application of long memory time series models," *Journal of time series analysis*, 4(4), 221–238.

Ghysels, E., and A. Sinko (2006): "Comment," *Journal of Business and Economical Statistics*, 24(2), 192–194.

Gil-Alana, L. a. (2003): "Testing of Fractional Cointegration in Macroeconomic Time Series," *Oxford Bulletin of Economics and Statistics*, 65(4), 517–529.

Gil-Alana, L. A., and J. Hualde (2009): "Fractional integration and cointegration. An overview and an empirical application," in *Palgrave handbook of econometrics*, pp. 434–469. MacMillan Publishers UK.

Giot, P. (2005): "Relationships between implied volatility indexes and stock index returns," *The Journal of Portfolio Management*, 31, 92–100.

Gonzalo, J., and C. W. Granger (1995): "Estimation of Common Long-Memory Components in Cointegrated Systems," *Journal of Business and Economic Statistics*, 13(1), 27–35.

Gonzalo, J., and S. Ng (2001): "A systematic framework for analyzing the dynamic effects of permanent and transitory shocks," *Journal of Economic Dynamics and Control*, 25(10), 1527–1546.

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

Granger, C. W. J. (1986): "Development in the Study of Cointegrated Economic Variables," *Oxford Bulletin of Economics and Statistics*, 48(3), 213–228.

Granger, C. W. J., and R. Joyeux (1980): "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1(1), 15–29.

Guo, H., and R. F. Whitelaw (2006): "Uncovering the Risk-Return Relation in the Stock Market," *The Journal of Finance*, 61(3), 1433–1463.

Hansen, P. R., and A. Lunde (2005): "A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?," *Journal of applied econometrics*, 20(7), 873–889.

Hansen, P. R., and A. Lunde (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24(2), 127–161.

Hilal, S., S.-h. Poon, and J. Tawn (2011): "Hedging the black swan : Conditional heteroskedasticity and tail dependence in S&P500 and VIX," *Journal of Banking and Finance*, 35(9), 2374–2387.

Hillebrand, E., and M. C. Medeiros (2010): "The Benefits of Bagging for Forecast Models of Realized Volatility," *Econometric Reviews*, 29(5-6), 571–593.

Hosking, J. R. M. (1981): "Fractional differencing," *Biometrika*, 68, 165–176.

Hualde, J., and P. Robinson (2010): "Semiparametric inference in multivariate fractionally cointegrated systems," *Journal of Econometrics*, 157(2), 492–511.

Hwang, E., and D. W. Shin (2014): "Infinite-order, long-memory heterogeneous autoregressive models," *Computational Statistics and Data Analysis*, 76, 339–358.

Jackwerth, J. C. (2000): "Recovering risk aversion from option prices and realized returns," *The Review of Financial Studies*, 13, 433–451.

Jacod, J., and A. N. Shiryaev (2003): *Limit theorems for stochastic processes*. Springer-Verlag, New York.

Jiang, G. J., and Y. S. Tian (2005): "The model-free implied volatility and its information content," *The Review of Financial Studies*, 18, 1305–1342.

Jiang, G. J., and Y. S. Tian (2007): "Extracting Model-Free Volatility from Option Prices: An examination of the VIX index," *Journal of Derivatives*, 14(3), 35–60.

Johansen, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.

Johansen, S. (2008): "A representation theory for a class of vector autoregressive models for fractional processes," *Econometric Theory*, 24(03), 651–676.

JOHANSEN, S., AND M. Ø. NIELSEN (2012): "Likelihood inference for a fractionally cointegrated vector autoregressive model," *Econometrica*, 80, 2667–2732.

JOHANSEN, S., AND M. Ø. NIELSEN (2015): "The Role of Initial Values in Conditional Sum-of-Squares Estimation of Nonstationary Fractional Time Series Models," *Econometric Theory*, 32(5), 1–45.

JONES, C. S. (2006): "A nonlinear factor analysis of S&P 500 index option returns," *The Journal of Finance*, 61, 2325–2363.

JONES, M. E. C., M. Ã. NIELSEN, AND M. K. POPIEL (2014): "A fractionally cointegrated VAR analysis of economic voting and political support," *Canadian Journal of Economics*, 47(4).

KELLARD, N., C. DUNIS, AND N. SARANTIS (2010): "Foreign exchange, fractional cointegration and the implied-realized volatility relation," *Journal of Banking and Finance*, 34(4), 882–891.

KIM, C. S., AND P. C. PHILLIPS (2006): *Log periodogram regression: the nonstationary case*. Cowles Foundation Discussion Paper.

KOCK, A. B. (2012): "On the Oracle Property of the Adaptive LASSO in Stationary and Nonstationary Autoregressions," *CREATES Research Paper*, 5.

KOOPMAN, S. J., B. JUNGBACKER, AND E. HOL (2005): "Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements," *Journal of Empirical Finance*, 12(3), 445–475.

KUENSCH, H. R. (1987): "Statistical aspects of self-similar processes," in *Proceedings of the first World Congress of the Bernoulli Society*, vol. 1, pp. 67–74.

ŁASAK, K. (2010): "Likelihood based testing for no fractional cointegration," *Journal of Econometrics*, 158(1), 67–77.

Łasak, K., and C. Velasco (2015): "Fractional Cointegration Rank Estimation," *Journal of Business and Economic Statistics*, 33(2), 241–254.

Lettau, M., and S. Ludvigson (2001): "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 56(3), 815–849.

Li, J., and W. Chen (2014): "Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models," *International Journal of Forecasting*, 30(4), 996–1015.

MacKinnon, J., and M. Nielsen (2014): "Numerical distribution functions of fractional unit root and cointegration tests," *Journal of Applied Econometrics*, 29(1), 161–171.

Madhavan, A. (2000): "Market microstructure: A survey," *Journal of Financial Markets*, 3(3), 205–258.

Malmsten, H., and T. Teräsvirta (2010): "Stylized facts of financial time series and three popular models of volatility," *European Journal of pure and applied mathematics*, 3(3), 443–477.

Malz, A. M. (2014): "A Simple and Reliable Way to Compute Option-Based Risk-Neutral Distributions," *FRB of New York Staff Report*, 677.

Mandelbrot, B. B., and J. W. Van Ness (1968): "Fractional Brownian motions, fractional noises and applications," *SIAM review*, 10(4), 422–437.

Marinucci, D., and P. Robinson (2001): "Semiparametric fractional cointegration analysis," *Journal of Econometrics*, 105(1), 225–247.

McAleer, M., and M. C. Medeiros (2008): "A multiple regime smooth transition Heterogeneous Autoregressive model for long memory and asymmetries," *Journal of Econometrics*, 147(1), 104–119.

Meddahi, N. (2002): "A theoretical comparison between integrated and realized volatility," *Journal of applied econometrics*, 17(5).

Medeiros, M., and E. Mendes (2012): "Estimating high-dimensional time series models," *CREATES Research Paper*, 37.

Merton, R. C. (1980): "On estimating the expected return on the market: An exploratory investigation," *Journal of Financial Economics*, 8(4), 323–361.

Morgan, J. P. (1996): "JP Morgan/Reuters Riskmetrics-Technical Document," *JP Morgan, New York*.

Müller, U. A., M. M. Dacorogna, R. D. Davé, R. B. Olsen, O. V. Pictet, and J. E. von Weizsäcker (1997): "Volatilities of different time resolutions -Analyzing the dynamics of market components," *Journal of Empirical Finance*, 4(2-3), 213–239.

Nardi, Y., and A. Rinaldo (2011): "Autoregressive process modeling via the Lasso procedure," *Journal of Multivariate Analysis*, 102(3), 528–549.

Nelson, D. (1992): "Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model," *Journal of Econometrics*, 52(1-2), 61–90.

Neuberger, A. (2012): "Realized skewness," *Review of Financial Studies*, 25(11), 3424–3455.

Newey, W. K., and K. D. West (1987): "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Nielsen, M. (2010): "Nonparametric cointegration analysis of fractional systems with unknown integration orders," *Journal of Econometrics*, 155(2), 170–187.

Nielsen, M. Ø. (2005): "Multivariate Lagrange Multiplier Tests for Fractional Integration," *Journal of Financial Econometrics*, 3(3), 372–398.

——— (2007): "Local Whittle Analysis of Stationary Fractional Cointegration and the Implied-Realized Volatility Relation," *Journal of Business and Economic Statistics*, 25(4), 427–446.

Nielsen, M. Ø., and P. Frederiksen (2011): "Fully modified narrow-band least squares estimation of weak fractional cointegration," *The Econometrics Journal*, 14(1), 77–120.

Nielsen, M. Ø., and K. Shimotsu (2007): "Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach," *Journal of Econometrics*, 141(2), 574–596.

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

Pigorsch, C., U. Pigorsch, and I. Popov (2012): "Volatility Estimation Based on High-Frequency Data," in *Handbook of Computational Finance*, pp. 335–369. Springer Berlin Heidelberg, Berlin, Heidelberg.

Poon, S.-H., and C. Granger (2003): "Forecasting volatility in financial markets: a review," *Journal of Economic Literature*, 41, 478–539.

Poteshman, A. M. (2000): "Forecasting Future Volatility from Option Prices," *Working Paper*.

Rapach, D. E., J. K. Strauss, and G. Zhou (2013): "International Stock Return Predictability: What Is the Role of the United States?," *The Journal of Finance*, 68(4), 1633–1662.

Ray, B. K., and R. S. Tsay (2000): "Long-range Dependence in Daily Stock Volatilities," *Journal of Business and Economic Statistics*, 18(2), 254–262.

Robinson, P. (1994): "Semiparametric analysis of long-memory time series," *The annals of Statistics*, 22(1), 515–539.

Robinson, P., and Y. Yajima (2002): "Determination of cointegrating rank in fractional systems," *Journal of Econometrics*, 106(2), 217–241.

Robinson, P. M. (1978): "Statistical Inference for a Random Coefficient Autoregressive Model," *Scandinavian Journal of Statistics*, 5, 163–168.

Robinson, P. M. (1995): "Log-periodogram regression of time series with long range dependence," *The annals of Statistics*, pp. 1048–1072.

Robinson, P. M. (1997): "Large-sample inference for nonparametric regression with dependent errors," *The annals of Statistics*, 25(5), 2054–2083.

Robinson, P. M. (2003): *Time Series with Long Memory*. Oxford University Press, Oxford.

Robinson, P. M., and D. Marinucci (2001): "Narrow-Band Analysis of Nonstationary Processes," *Annals of Statistics*, 29(4), 947–986.

Robinson, P. M., and D. Marinucci (2003): "Semiparametric frequency domain analysis of fractional cointegration," in *Time Series with Long Memory*, ed. by P. M. Robinson, pp. 334–373. Oxford University Press, Oxford, UK.

Rossi, A., and A. Timmermann (2010): "What is the shape of the risk-return relation?," *Working Paper*.

Rossi, E., and P. Santucci de Magistris (2013): "A No-Arbitrage Fractional Cointegration Model for Futures and Spot Daily Ranges," *Journal of Futures Markets*, 33(1), 77–102.

Shimko, D. (1993): "Bounds of Probability," *Risk*, 6(4), 33–37.

Shimotsu, K. (2010): "Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend," *Econometric Theory*, 26(02), 501–540.

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

Stock, J. H., and M. W. Watson (2004): "Combination forecasts of output growth in a seven-country data set," *Journal of Forecasting*, 23(6), 405–430.

Taylor, S. J. (1996): *Modelling Financial Time Series*. Wiley, Chichester.

Taylor, S. J., P. K. Yadav, and Y. Zhang (2010): "The information content of implied volatilities and model-free volatility expectations: evidence from options written on individual stocks," *Journal of Banking and Finance*, 34, 871–881.

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

Tibshirani, R., and X. Suo (2016): "An ordered lasso and sparse time-lagged regression," *Technometrics*, 58(4), 415–423.

Timmermann, A. (2006): *Forecast Combinations*, vol. 1. Elsevier.

Wang, C. S.-H., L. Bauwens, and C. Hsiao (2013): "Forecasting a long memory process subject to structural breaks," *Journal of Econometrics*, 177(2), 171–184.

Wilms, I., J. Rombouts, and C. Croux (2016): "Lasso-based forecast combinations for forecasting realized variances," *Working Paper*.

Yang, Y., and H. Zou (2015): "A Fast Unified Algorithm for Solving Group-Lasso Penalized Learning Problems," *Statistics and Computing*, 25(6), 1129–1141.

Yuan, M., and Y. Lin (2006): "Model Selection and Estimation in Regression With Grouped Variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1), 49–67.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100(472), 1394.

Zhang, Y., S. J. Taylor, and L. Wang (2013): "Investigating the Information Content of the Model-Free Volatility Expectation by Monte Carlo Methods," *Journal of Futures Markets*, 33(11), 1071–1095.

ZHOU, B. (1996): "High-frequency data and volatility in foreign-exchange rates," *Journal of Business and Economic Statistics*, 14(1), 45–52.

ZOU, H. (2006): "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101(476), 1418–1429.