Forecasting Realised Volatility Using ARFIMA and HAR Models

Marwan Izzeldin
Department of Economics,
Lancaster University LA1 4YX, UK

Mohammad Kabir Hassan[1]
Department of Economics & Finance,
University of New Orleans, New Orleans, USA

Vasileios Pappas
Kent Business School,
University of Kent ME4 4TE, UK

Mike Tsionas
Department of Economics,
Lancaster University LA1 4YX, UK

March 2019

Abstract

Recent literature provides mixed empirical evidence with respect to the forecasting performance of ARFIMA and HAR models. This paper compares the forecasting performance of both models using high frequency data of 100 stocks representing 10 business sectors for the period 2000-2010. We allow for different sectors, changing market conditions, variation in the sampling frequency and forecasting horizons. For the overall sample and using the 300 sec sampling frequency, the forecasting performance of both models is indistinguishable. However, differences arise under different market regimes, forecasting horizons and sampling frequencies. ARFIMA models are superior for the crisis and pre-crisis sub-samples. HAR forecasts are less sensitive to regime change and to longer forecasting horizons. Variations in forecasting performance could also be explained using differences in the levels of persistence underlying each model.

Keywords: High-Frequency data • Market conditions • Market Sectors • Realised Variance • HAR • ARFIMA

JEL Classification: C53 • C58 • G17

[1] *Corresponding author*: email: mhassan@uno.edu. We would like to thank the participants of INFINITI Conference on International Finance (2013), Aix-en-Provence, France and the 7th International Conference on Computational and Financial Econometrics (2013). We would like to thank John Maheu and Rodrigo Hizmeri for their valuable comments. The authors are grateful to the Gulf One Lab for computational and Economic Research (GOLCER) for support.

1. Introduction

Forecasting stock market volatility has long been and remains of interest to market traders and regulators. Given that financial risk is commonly assessed in terms of asset volatility, accurate volatility forecasts are desirable. Advances in computing and data technology make it possible to observe markets at very fine intervals of time. This has led to the introduction of so-called realised measures. Andersen et.al (2003) has shown that in the absence of microstructure noise, the realised variance calculated using high frequency data is a consistent estimator of the quadratic variation. Realised variance is now widely adopted owing to its desirable stylised facts and statistical properties which are superior to those parametric volatility measures generated from GARCH and Stochastic Volatility (SV) models. The superiority of the non-parametric realised variance is due to the utilization of information available at very fine intervals of time, which is lost at the lower frequencies at which GARCH and SV models operate.

Irrespective of the model generating the underlying volatility series, all volatility measures share a number of stylized facts and distributional assumptions that distinguish them from other processes. For instance actual realisations of return volatility are unobserved and are characterized by long memory; see Bollerslev & Mikkelsen (1996); Ding, Granger, & Engle (1993); Ray & Tsay (2000). As such, most forecasting models tend to exploit the long-memory property to generate in/out-of-sample forecasts. Traditionally this has been done via long-memory models such FIGARCH and ARFIMA, given that ARIMA processes are often found inadequate to capture the long memory feature in a parsimonious way Andersen et.al (2000, 2001,2003); Bandi & Perron (2006); Beine, Laurent, & Lecourt (2003); Caporin, Rossi, & Magistris (2014). But long-memory models have certain drawbacks: they are nontrivial to estimate, mainly univariate and require a large sample size to obtain accurate estimates of the fractional differencing parameter.

An alternative approach to long-memory modelling views the long memory feature of volatility the result of data aggregation, breaks and filtration: see Hyung et.al (2008); Wang and Yen, (2016) among others. This line of modelling has been stimulated by many factors. For instance, most observed processes are not pure fractional indicative of the presence of short-memory. Moreover, if the aggregation level is small relative to the lowest frequency of the model, then scaling laws do not apply, and short and long-memory features become hard to distinguish and model. LeBaron (2001) shows that the summation of short memory models with lags as low as three can generate memory patterns that are hyperbolic in nature. Building on these ideas, Müller et al., (1997) and subsequently Corsi, (2009) proposed the heterogeneous autoregressive model (HAR), which is capable of approximating the long memory features of the data and respond to short-term shocks; hence providing superior fitting and forecasting performance. The superior performance of the HAR in forecasting realised volatility is noted in Andersen et.al (2007); Andersen et.al (2011); Bollerslev et. al (2016) and Patton & Sheppard (2009) among others. Compared to ARFIMA, HAR models are more trivial to estimate and forecast from. Despite the numerous studies in volatility forecasting using a wide array of models and specifications, see for example Andersen et al., (2007, 2011); Brownlees, Engle, & Kelly (2011); Clements, Galvão, & Kim (2008); Corsi (2009); Fuertes, Izzeldin, & Kalotychou (2009); Scharth & Medeiros (2009). The question of how the HAR stacks up to the ARFIMA forecasts remains unaddressed.

The paper compares the forecasting performance of ARFIMA and HAR models allowing for variations in: i) business sector type; ii) market conditions; iii) forecasting horizons and iv) sampling frequencies. We use 100 stocks representing 10 distinctive business sectors for the period (2000-2010). Business sectors are heterogeneous in nature which leads to variations in stylised facts and responses to economic turmoil. This variation in our sample

As a preview to our results, we find the forecasting performance of both models is similar for the full sample and at sampling frequency of 300sec for the realised variance. Nevertheless, as the sampling frequency increases (60sec and above), ARFIMA forecasts take the lead. Crisis adversely affects the forecasting performance of HAR to a lesser extent. ARFIMA generates superior forecasts for both the pre-crisis and crisis regimes. Forecasts generated by both models are sensitive to the sampling frequency: at the benchmark sampling frequency of 300sec the HAR outperforms the ARFIMA. HAR forecasts are less sensitive to regime change and to longer forecasting horizons.

The paper is structured as follows. Data are presented in Section 2, Section 3 outlines the methodology, Section 4 presents the empirical results and Section 5 concludes.

2. Data Description

The sample period is 02/01/2000 to 12/31/2010 with a total of 2767 trading days observed at the tick level. Our data is obtained from Tick Data.[2] Data cleaning and filtration techniques are explained in the Tick Data website.[3] This makes our results easier to authenticate and replicate. We use trade data for 100 stocks from 10 business sectors: Consumer Discretionary (CD), Consumer Staples (CS), Energy (ENG), Financial (FIN), Health Care (HC), Industrials (IND), Information Technology (IT), Materials (MAR), Telecommunications (TEL) and Utilities (UTL). We consider stocks with the highest market capitalization within their representative sector. The sample range and sector coverage allow us to examine the sensitivity of the forecasting performance of our models across different market regimes and heterogeneous (in terms of volatility and liquidity) business sectors.

We use transaction prices sampled at the 300sec (5min) to construct returns and realized variance. The common use of this sampling frequency strikes a balance between information gain from high frequency data and microstructure effects; see Andersen et al. (2001, 2010). For robustness, we consider alternative sampling frequencies of 5, 15, 60, 150, 900 and 1800 seconds. Since persistence, microstructure noise and leverage are known to affect the forecasting performance, irrespective of the adopted model, and that the impact of these factors tends to vary with the sampling frequency. Hence, it is paramount to consider those sampling frequencies.

The main quantity of interest is the daily realised variance. To define the daily realised variance, the time dimension is discretized and each day is divided into M equally-spaced subintervals of length δ. For instance, for $\delta = 5min$, we have $M = 78$ intraday returns obtained by dividing the number of seconds in the trading day by the sampling frequency (i.e. 23400/300). A few trading days consist of $M < 78$ owing to delayed openings and/or early closings of the NYSE. The price at the start of the j^{th} intraday interval is

[2] Tick Data is a data base provides data on a commercial basis for futures, Index and equity markets. Tick Data is sourced from NYSE's TAQ (Trade and Quote) database. Tick adjusts the TAQ database for ticker mapping, code filtering, price splits and dividend payments.
[3] https://www.tickdata.com/

computed as the average of the closing and opening prices of intervals $j-1$ and j, respectively. The j^{th} intraday return (on day t) is then computed as:

$$r_{t,j} = 100\left(\frac{\ln\left(p_{t,j}^{c}\right)+\ln\left(p_{t,j+1}^{o}\right)}{2}-\frac{\ln\left(p_{t,j-1}^{c}\right)+\ln\left(p_{t,j}^{o}\right)}{2}\right), j = 2, \dots, M-1 \qquad (1)$$

where each trading day [09:30am-16:00pm], and p_{tj}^{c}, p_{tj}^{o} are the closing and opening prices of the j^{th} intraday interval respectively. For instance, $j = 2$ corresponds to 09:35am-09:40am. The realized variance RV_t is defined as the sum of intraday returns and is given by:

$$RV_t = \sum_{j=1}^{M} r_{t,j}^{2}, t = 1, \dots, T \qquad (2)$$

As $M \to \infty$, the realized variance converges to quadratic variation of the process.

Table 1 outlines the tickers of the 100 stocks adopted in this study alongside their degree of market activity as measured by trading volume. Citigroup is the most active, and BT the least active, among all the 100 stocks considered. Table 2 provides descriptive statistics for daily returns, trading volume and realised variance by sectors. The IT and UTL sectors are the most/least active as measured by trading volume. Volatility as measured by the realised variance shows that the CS and FIN sectors are the least/most volatile sectors. Volatility across sectors features long memory as evident by the estimates of the fractional differencing parameter d.

[Table 1 here]

[Table 2 here]

3. Methodology

In this section, we outline our forecasting models by their underlying specifications. The two contenders are the long memory specification ARFIMA and the short memory specification HAR. In all cases the models are fitted to the natural log of realised volatility following Andersen et al., (2007).

Andersen et al., (2003); Areal & Taylor (2002); Bollerslev & Wright (2001); Deo, Hurvich, & Lu (2006); Granger & Joyeux (1980); Koopman, Jungbacker, & Hol (2005); Martens, De Pooter, & van Dijk (2009); Martens & Zein (2004); Oomen, (2001, 2004); Pong, Shackleton, Taylor, & Xu (2004); Thomakos & Wang (2003) highlight the long memory property of volatility and advocated that volatility persistence is better captured by ARFIMA type of models. An ARFIMA$(\mathrm{p}, \mathrm{d}, \mathrm{q})$ is given by:

$$\varphi(L)(1-L)^{d}(RV_t - \mu) = \theta(L)\varepsilon_t \qquad (3)$$

where $\varphi(L) = 1 - \sum_{i=1}^{p} \varphi_i L^i$ and $\theta(L) = 1 - \sum_{j=1}^{q} \theta_j L^j$ are the AR and MA lag polynomials accounting for the short-memory properties, whereas the long-memory properties are captured by the fractional differencing parameter d and ε_t is the error which is distributed as $N(0, \sigma_u^2)$. In our framework, the Autoregressive (AR) and Moving Average (MA) components are set to zero ($p = q = 0$), hence allowing only for the long memory feature to be in effect.

Equally there is evidence on the capacity of the HAR model to approximate the long memory property of the volatility, see Andersen et al., (2007); Bollerslev et al., (2016); Corsi, 2009) and references therein. Our HAR model specification follows that of Corsi (2009), which is given by:

$$RV_t = c + \beta^{(d)} RV_{t-1} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \varepsilon_t \tag{4}$$

where $\varepsilon_t \sim iid(0, \sigma^2)$, and weekly $RV_t^{(w)}$ and monthly $RV_t^{(m)}$ realised measures are respectively given by:

$$RV_t^{(w)} = \frac{1}{5}(RV_{t-1}^{(d)} + RV_{t-2}^{(d)} + RV_{t-3}^{(d)} + RV_{t-4}^{(d)} + RV_{t-5}^{(d)}) \tag{5}$$

$$RV_t^{(m)} = \frac{1}{22}(RV_{t-1}^{(d)} + RV_{t-2}^{(d)} + \cdots + RV_{t-21}^{(d)} + RV_{t-22}^{(d)}) \tag{6}$$

The notion behind the inclusion of weekly and monthly components in the HAR model is to accommodate market participants with different investment horizons, typically short (1 day), medium (1 week) and long (1 month), who may differ in how they perceive and react to volatility. The incorporation of the long lag structure is akin to a restricted AR (22), is capable of reproducing the long memory feature of realized volatility as evidenced in Corsi, (2009) and Andersen et al., (2007).[4]

3.1 Forecasting Calibration and Evaluation Criterion

The description of our forecasting exercise is based on our baseline set up; i.e., using the full sample with realized volatility sampled at the 300sec (5min) and the forecasting target being the 1d-ahead.

The total sample size is $T + 1$. We use the last $P (P = 500)$ observations as a holdout evaluation period. The first R ($R = 2267$) observations are used for the initial model estimation which generates a vector β of regression parameters. Under a rolling forecasting scheme, the βs are always estimated from a sample of size R. The first estimation window ranges from 1 to R, while volatility forecasts are generated for $R + 1$. The second estimation window ranges from 2 to $R + 1$, while forecasts are generated for $R + 2$. The last estimation window ranges from P to T, while forecasts are generated for $R + P = T + 1$.The initial, full sample estimation period is 2/1/2000 – 7/1/2009 while the forecasting covers the period 8/1/2009 – 31/12//2010. At every iteration, the 1-day-ahead volatility forecast h_t^2 is compared to the population volatility measure σ_t^2. The population volatility measure is the RV_t given its unbiased nature (Patton, 2011). The precision of the forecasts is assessed using two commonly used forecast evaluation criteria:

[4] Extensions of the HAR have appeared in the literature, see Andersen et al., (2011) and Corsi et al., (2010) for example. Also, papers using ARFIMA models have also modelled the short-memory process by setting $p > 0$, see for example Martens, De Pooter, & van Dijk (2009). Nevertheless, we feel that for our research question the baseline models are more appropriate as they model the long-memory process of the realised volatility.

the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE), see Andersen et al., (2011) and Patton (2011) among others.[5] Both criteria are expressed in percentage terms as follows.

$$MAE = \frac{1}{B}\sum_{b=1}^{B} |\hat{\sigma}_t^2 - h_t^2| \times 100 \tag{7}$$

$$RMSE = \sqrt{\frac{1}{B}\sum_{b=1}^{B} (\hat{\sigma}_t^2 - h_t^2)^2} \times 100 \tag{8}$$

where B denotes the number of the rolling forecasts.

To evaluate the impact of market conditions on forecasting performance we split the sample into two periods: pre-crisis and crisis. The period up to the end of 2005 is representative of pre-crisis and the forecasting is conducted as the baseline set up described earlier. As such, the pre-crisis forecasting is conducted in the period 2004-2005, while the crisis forecasting in the period 2007-2009. To evaluate the impact of forecasting horizons on both ARFIMA and HAR, we adjust our baseline set up for the cases of 5d-ahead and 22d-ahead forecasts as representative of 1-week and 1-month respectively. To gauge the sensitivity of the models to the sampling frequency, we used Realised variance sampled at various frequencies, namely 5, 15, 60, 150, 300, 900 and 1800 seconds.

4. Empirical Results

In this section, we present the results of our forecasting exercise. We begin by comparing the forecasting performance of the two models across the 10 sectors. We then compare the sensitivity of the models' forecasts to market conditions (pre-crisis/crisis). In a subsequent section, we evaluate the impact of forecasting over extended horizons. The impact of sampling frequency is reported thereafter.

A. Sectors

Figure 1 reports the forecast performance of the two competing models (ARFIMA and HAR) applied to the 100 stocks. Table 3 shows that the RMSE and MAE for the ARFIMA are 23.60% and 18.05% respectively, while for the HAR they are 23.63% and 18.04%, which shows that the performance of both models is at par.

[Figure 1 around here]

Table 3 also reports the average MAE and RMSE for both ARFIMA and HAR by sectors. Performance-wise, the sector analysis does not reveal a definite winner as ARFIMA is surpassed by HAR only for 4 out of the 10 sectors (ENG, IT, MAR and UTL).

[Table 3 around here]

Both ARFIMA and HAR provide similar ranking classification for all sectors, with the exception of IND and UTL sectors. Moreover, both models rank ENG and HC as the best/worst forecasted sectors. In

[5] We have used additional loss functions (MSE, QLIKE) in line with the arguments presented in Patton (2011), however these do not change the qualitative nature of our findings.

general, the ranking classification of both models is highly aligned as evidenced by the Spearman rank correlation coefficient which shows a strong positive correlation between the sectorial rankings.

B. Market Conditions

Table 4 reports the percentage MAE for each of the 10 business sectors and their corresponding ranks for both the pre-crisis and crisis regimes. The table delivers a number of important findings.

First, ARFIMA reports lower MAEs relative to HAR at both the sector level and for the average. This contrasts with the full sample results where the models performances were highly aligned. However, and unlike the case for the full sample, both models deliver different rankings of the forecasting performance across sectors.

Second, forecasting performance across all sectors deteriorates during crisis as both ARFIMA and HAR models, on average, show higher MAEs in the crisis period. Exceptions are the CD, ENG and UTL sectors with all featuring improved forecasting performance under both the ARFIMA and HAR models specification. The increase in the MAEs is more marked in the case of ARFIMA which records an average increase of around 2.7% as opposed to 0.7% for the HAR. This shows the HAR model specification to be less sensitive to regime change. The relative sensitivity of the ARFIMA forecasts to regime change can be attributed to the instability of the fractional differencing parameter d, since high volatility persistence can lead to pronounced spikes in future volatility which may in turn adversely impact the forecasts, see Maekawa & Xinhong, (2011) and Syczewska, (2011).[6]

Third, the d estimates are higher in the crisis as opposed to the pre-crisis regime and with variation across the two regimes are rather asymmetric across the sectors. For instance, the FIN sector features the worst forecasting performance and the highest negative percentage change in d (%Δ d = -18.24) whereas the ENG sector reports gains in forecasting performance and reports the highest positive change in d (%Δ d = 19.19). This finding matches those of the Earlier literature, see Andersen & Bollerslev, (1998) and Dufrenot, Guegan, & Peguin-Feissolle (2008), which highlighted the link between volatility changes and long memory during regime change. Our results here adds another dimension by outlining the association between the observed changes in the fractional parameter d's and changes in forecasting performance as measured by the MAEs.

Fourth, the standard deviation of forecast evaluation criteria is lower during the crisis than to the pre-crisis period. This is plausible given that during episodes of distress, stock price movements become more aligned as investors share common beliefs on the market's direction. Conversely during periods of tranquillity, investor sentiment is primarily driven by idiosyncratic information, leading to a wider spread of beliefs.

In summary, we find that the forecasting performance of ARFIMA models is more sensitive to regime change; however, ARFIMA reports lower MAEs when viewed at the sector level. This applies for both the pre-crisis and crisis regimes.

[6] Long-memory models have generally been found to be sensitive to structural breaks. Related to this, Granger & Joyeux, (1980) distinguish between genuine and spurious long memory processes where in the former case the property is inherent in the series, while in the latter it is caused by structural breaks. Several aspects of structural breaks and the long memory property of a series have been investigated by Granger & Terasvirta (1999) and Gourieroux & Jasiak (2001) among others.

[Table 4 around here]

[Figure 2 around here]

C. Forecasting Horizons

Table 5 reports the average MAEs for all sectors at 3 different horizons: the short-term (1d), the medium-term (5d) and the long-term (22d-ahead) forecasts. The RV measure adopted in this exercise is based on the 300 sec sampling frequency. Results obtained shows that ARFIMA (1d[18.05], 5d[18.71], and 22d[20.08]) and HAR (1d[18.04], 5d[18.08], and 22d[18.06]) compare well at the 1d ahead forecasts but deviate when the forecasting horizon is varied. ARFIMA forecasts drop by 3.5% for 5d-ahead and 10.6% for 22d-ahead forecasts compared to HAR forecasts which drop by 0.22% and 0.11% respectively for the 5d-ahead and 22d-ahead forecasts. The sensitivity of the ARFIMA model to changes in the forecasting horizon can also been seen in the variation of the ranking of sectors. In contrast, HAR based ranking of sectors features little or no change across the different horizons.

Figure 3 shows the relative gain/loss by sectors between the 1d and 22d forecasting horizons. ARFIMA forecasts deteriorate across all sectors where the maximum loss is observed in the FIN sector. HAR forecasting performance drops for all sectors with the exception of the IND and FIN sectors where gains are observed. The superiority of the HAR model in the medium and long horizons are in line with Corsi (2009) who highlighted the stability of the HAR model across different forecasting horizons.

[Table 5 around here]

[Figure 3 around here]

D. Sampling Frequency

The impact of varying the sampling frequencies is shown in Table 6, where there is clear tendency for the forecasting performance to vary with sampling frequency. This is true for both the ARFIMA and HAR model specifications. Also notable, is the differentiated impact of the sampling frequency on the two models. For instance, the ARFIMA outperforms the HAR at sampling frequencies higher than 150sec a result significant at the 5% level.[7] By contrast, the HAR significantly outperforms the ARFIMA for sampling frequencies lower than 150sec, including the literature benchmark of 300sec.Taking the average across the different frequencies shows that HAR outperforms ARFIMA. The fractional differencing estimator tends to increase with the sampling frequency, and might, in part, explain the superior performance of the ARFIMA at the higher end of the sampling frequencies.

The impact of the sampling frequency is further revealed by Figures 4a and 4b. Panel “a” shows an increasing tendency for the MAE to rise with the sampling frequency. Panel “b” shows the percentage gain in forecasting performance compared to the benchmark of 300sec for both ARFIMA and HAR.

[Figure 4 around here]

[Table 6 around here]

[7] The Kruskal-Wallis Singed Rank Test is suitable in this context forecasts of 100 stocks based on two models (ARFIMA and HAR) are compared against a common benchmark.

Our findings clearly highlight the impact of the sampling on the forecasting performance of the models under investigation. This is in with existing literature, which documents gains in forecasting from using realised volatility sampled at higher frequencies; see for example Martens, (2001) and Andersen et al., (2007). In specific, Patton & Sheppard (2009) conclude that forecasting performance of realised volatility peaks at a sampling frequency of 60 seconds. Our result here shows such peak is achieved at the 5 sec as evident by the lower MAEs achieved under both models.

Our results above applies for the 100 stocks' but given the multi-sector dimension of our data makes it a worthwhile exercise to examine the various sectors forecasting gains/losses with respect to the changes in the sampling frequency. Figures 5a and 5b outline the frequency gains/losses by sector. For example, relative to the widely adopted 300 sec benchmark, an increase of the sampling frequency from 300sec to 5 sec, leads to gains in all sectors, with highest forecasting gains attained by the IT sector (68.6% for HAR and 71.6% for ARFIMA) whereas the lowest forecasting gains are observed for the UTL sector (29.5% for HAR and 30.8% ARFIMA). Differences in forecasting gains across sectors could be attributed to factors such as the level of activity of the sector, market capitalization, as well as to variations in persistency levels. For instance, the IT (UTL) sectors have the highest (lowest) market capitalisation, the most (least) active by volume of traded shares and the most (least) persistent as measured by the fractional differencing parameter.

[Figure 5 around here]

5. Conclusion

Forecasting return volatility has always been of interest to policy makers and practitioners. Such interest has increased, especially amidst the recent global financial crisis. The advent of high frequency data has spurred the development of realised volatility measures that have dominated the use of parametric models. Forecasting models utilising realised volatility aims to exploit the observed persistence in volatility either via long-memory formulations or, more recently, with short-memory models capable of approximating the hyperbolic decay in the autocorrelation function. Perhaps the most commonly referenced model of this sort is the HAR model, popularised by Corsi (2009). Although the HAR has been widely adopted due to its convenience in estimation, a formal comparative study with the ARFIMA is lacking.

In this paper we compare the forecasting performance ARFIMA and HAR across a variety of scenarios. In specific, we assess the impact of market conditions, forecasting horizons and sampling frequency on the models forecasting performance. Our inferences are drawn from high frequency stock data comprising 100 stocks from 10 business sectors over the period 2000 – 2010.

Forecasting using both models is sensitive to sector type, regime change, the degree of persistence and the sampling frequency. The Energy sector offers the best forecasts while Health care sector ranks worst. The ARFIMA model is more sensitive to variations in market regimes. The long memory differencing parameter is sensitive to changes in both the market regime and the sampling frequency and affects the forecasting performance of both models. HAR forecasts are more stable across forecasting horizons as evident by the little or no change of the sector rankings. Higher sampling frequency leads to improvements in forecasting performance and this is verified by both ARFIMA and HAR models for all business sectors. ARFIMA outperforms the HAR for the sampling frequencies below 150sec. Conversely, sampling frequencies of at least 300sec are better suited for a HAR model. Both models generate the best forecasts using realised variance based on the 5 sec sampling frequency. Our findings refute the notation of a definite winner and highlight the merits underlying both models.

References

Andersen, T. G., & Bollerslev, T. (1998). Deutsche Mark – Dollar Volatility : Intraday Activity Patterns , Macroeconomic Announcements. *Journal of Finance*, *53*(1), 219–265. http://doi.org/10.1111/0022-1082.85732

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing It Up : Including Jump Components in the Return Volatility, *89*(November), 701–720. http://doi.org/10.1162/rest.89.4.701

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). *The distribution of realized stock return volatility*. *Journal of Financial Economics* (Vol. 61). http://doi.org/10.1016/S0304-405X(01)00055-1

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2000). Exchange rate returns standardized by realized volatility are (nearly) Gaussian. *Multinational Finance Journal*, *4*(September), 159–179. http://doi.org/10.3386/w7488

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association*, *96*(August), 42–55. http://doi.org/10.1198/016214501750332965

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modelling and Forecasting Realised Volatility. *Econometrica*, *71*, 529–626.

Andersen, T. G., Bollerslev, T., & Huang, X. (2011). A reduced form framework for modeling volatility of speculative prices based on realized variation measures. *Journal of Econometrics*, *160*(1), 176–189. http://doi.org/10.1016/j.jeconom.2010.03.029

Areal, N. M. P. C., & Taylor, S. J. (2002). The Realized Volatility of FTSE-100 Futures Prices. *Journal of Futures Markets*, *22*(7), 627–648. http://doi.org/10.1002/fut.10018

Bandi, F. M., & Perron, B. (2006). Long memory and the relation between implied and realized volatility. *Journal of Financial Econometrics*, *4*(4), 636–670. http://doi.org/10.1093/jjfinec/nbl003

Beine, M., Laurent, S., & Lecourt, C. (2003). Official central bank interventions and exchange rate volatility: Evidence from a regime-switching analysis. *European Economic Review*, *47*(5), 891–911. http://doi.org/10.1016/S0014-2921(02)00306-9

Bollerslev, T., & Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, *73*(1), 151–184. http://doi.org/10.1016/0304-4076(95)01736-4

Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, *192*(1), 1–18. http://doi.org/10.1016/j.jeconom.2015.10.007

Bollerslev, T., & Wright, J. H. (2001). High-Frequency Data, Frequency Domain Inference, and Volatility Forecasting. *Review of Economics and Statistics*, *83*(4), 596–602. http://doi.org/10.1162/003465301753237687

Brownlees, C. T., Engle, R. F., & Kelly, B. T. (2011). A Practical Guide to Volatility Forecasting through Calm and Storm. *SSRN Electronic Journal*, *14*(2), 1–37. http://doi.org/10.2139/ssrn.1502915

Caporin, M., Rossi, E., & Magistris, P. S. De. (2014). Chasing volatility - A persistent multiplicative error model with jumps Chasing volatility A persistent multiplicative error model with jumps.

Clements, M. P., Galvão, A. B., & Kim, J. H. (2008). Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance*, *15*(4), 729–750. http://doi.org/10.1016/j.jempfin.2007.12.001

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*, 174–196. http://doi.org/10.1093/jjfinec/nbp001

Corsi, F., Pirino, D., & Reno, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, *159*(2), 276-288.

Deo, R., Hurvich, C., & Lu, Y. (2006). Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment. *Journal of Econometrics*, *131*, 29–58. http://doi.org/10.1016/j.jeconom.2005.01.003

Ding, Z., Granger, C. W. J., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*(1), 83–106. http://doi.org/10.1016/0927-5398(93)90006-D

Dufrenot, G., Guegan, D., & Peguin-Feissolle, A. (2008). Changing-regime volatility: a fractionally integrated SETAR model. *Applied Financial Economics*, *18*(7), 519–526. http://doi.org/10.1080/09603100600993778

Fuertes, A. M., Izzeldin, M., & Kalotychou, E. (2009). On forecasting daily stock volatility: The role of intraday information and market conditions. *International Journal of Forecasting*, *25*(2), 259–281. http://doi.org/10.1016/j.ijforecast.2009.01.006

Gourieroux, C., & Jasiak, J. (2001). *Financial Econometrics*. Princeton NJ: Princeton University Press.

Granger, C. W. J., & Joyeux, R. (1980). An Introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, *1*(1), 15–29.

Granger, C. W. J., & Terasvirta, T. (1999). A simple nonlinear time series model with misleading linear properties. *Economics Letters*, *62*(2), 161–165. http://doi.org/10.1016/S0165-1765(98)00228-6

Hyung, N., Poon, S. H., & Granger, C. W. J. (2008). A source of long memory in volatility. *Frontiers of Economics and Globalization*, *3*, 329–380. http://doi.org/10.1016/S1574-8715(07)00209-6

Koopman, S. J., Jungbacker, B., & Hol, E. (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. http://doi.org/10.1016/j.jempfin.2004.04.009

LeBaron, B. (2001). Stochastic volatility as a simple generator of financial power laws and long memory. *Quantitative Finance*, *1*, 621–631.

Maekawa, K., & Xinhong, L. (2011). Long Memory in the Realized Volatility of Returns on the Yen / US $ Exchange Rate during the Three Financial Crises. *Working Paper, Hiroshima University of Economics*, 1–24.

Martens, M. (2001). Forecasting daily exchange rate volatility using intraday returns. *Journal of International Money and Finance*, *20*, 1–23. http://doi.org/10.1016/S0261-5606(00)00047-4

Martens, M., & Zein, J. (2004). Predicting financial volatility: High-frequency timeseries forecasts vis-`a-vis implied volatility, *Journal of Futures Markets, 24*(11), 1005-1028.

Martens, M., Van Dijk, D., & De Pooter, M. (2009). Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of forecasting*, *25*(2), 282-303.

Müller, U. a., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V., & von Weizsäcker, J. E. (1997). Volatilities of different time resolutions -- Analyzing the dynamics of market components. *Journal of Empirical Finance*, *4*(2-3), 213–239. http://doi.org/10.1016/S0927-5398(97)00007-8

Oomen, R. C. A. (2001). Using High Frequency Stock Market Index Data to Calculate, Model and Forecast Realised Return Variance. *EUI Working Paper, 6*.

Oomen, R. C. A. (2004). Modelling realized variance when returns are serially correlated. *Discussion Papers, The Open Access Publication Server of the ZBW*. Retrieved from http://hdl.handle.net/10419/51076

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1980), 246–256. http://doi.org/10.1016/j.jeconom.2010.03.034

Patton, A. J., & Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, *25*(2), 218–238. http://doi.org/10.1016/j.ijforecast.2009.01.011

Pong, S., Shackleton, M. B., Taylor, S. J., & Xu, X. (2004). Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models. *Journal of Banking and Finance*, *28*, 2541–2563. http://doi.org/10.1016/j.jbankfin.2003.10.015

Ray, B., & Tsay, R. (2000). Long-range dependence in daily stock volatilities. *Journal of Business & Economic Statistics*, *18*(January 2013), 254–262. http://doi.org/10.2307/1392562

Scharth, M., & Medeiros, M. C. (2009). Asymmetric effects and long memory in the volatility of Dow Jones stocks. *International Journal of Forecasting*, *25*(2), 304–327. http://doi.org/10.1016/j.ijforecast.2009.01.008

Syczewska, E. M. (2008). Financial crisis influence on the BUX index of Hungarian Stock Exchange. Long memory measures: 1991-2008. *Working Paper, Warsaw School of Economics*.

Thomakos, D. D., & Wang, T. (2003). Realized volatility in the futures markets. *Journal of Empirical Finance*, *10*, 321–353. http://doi.org/10.1016/S0927-5398(02)00052-X

Wang, Y.-H., & Yen, K.-C. (2016). The information content of the implied volatility term structure on future returns. *SSRN Electronic Journal*, 1–44. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2483222

Table 1. Sector-wise list of 100 stocks

Stock Name	Ticker	Volume	Stock Name	Ticker	Volume	Stock Name	Ticker	Volume	Stock Name	Ticker	Volume	Stock Name	Ticker	Volume
Consumer Discretionary (CD)			**Consumer Staples (CS)**			**Energy (ENG)**			**Financials (FIN)**			**Health Care (HC)**		
Amazon.com Inc	AMZN	7.895	Wal-Mart	WMT	12.014	ExxonMobil	XOM	16.606	Wells Fargo	WFC	22.247	Johnson & Johnson	JNJ	8.806
McDonald's	MCD	6.186	Procter & Gamble	PG	8.590	Chevron Corporation	CVX	7.521	JPMorgan Chase	JPM	20.185	Pfizer	PFE	28.251
Walt Disney	DIS	8.887	Coca-Cola	KO	6.657	Occidental Petroleum	OXY	4.465	Citigroup Inc.	C	111.497	Merck	MRK	9.839
The Home Depot	HD	11.387	PepsiCo Inc.	PEP	4.852	Halliburton Co.	HAL	11.727	Bank of America	BAC	59.480	Abbott Laboratories	ABT	4.972
Time Warner Inc.	TWX	6.749	Unilever	UL	0.610	Devon Energy Corp	DVN	3.314	Goldman Sachs Group	GS	6.802	United Health Group	UNH	8.145
Marriott Int'l.	MAR	2.710	Costco	COST	4.268	Baker Huges	BHI	3.541	American Express	AXP	7.345	Amgen Inc	AMGN	8.867
Gap (The)	GPS	6.397	Kimberly-Clark	KMB	1.851	Chesapeaks Energy	CHK	7.251	Morgan Stanley	MS	10.399	Medtronic Inc	MDT	4.817
News Corporation	NWSA	6.753	Estee Lauder Cos.	EL	1.240	Williams Cos.	WMB	4.572	The Bank of New York Mellon	BK	4.656	Gilead Sciences	GILD	10.475
Interpublic Group	IPG	3.750	Brown-Forman Corp.	BFB	0.309	Sunoco Inc.	SUN	2.184	Travelers	TRV	2.514	Humana Inc.	HUM	1.628
Best Buy Co. Inc.	BBY	6.025	Avon Products	AVP	2.858	TECO Energy	TE	1.311	Allstate Corp	ALL	3.155	Boston Scientific	BSX	8.832
Average Trading Volume		**6.670**	**Average**		**4.320**	**Average**		**6.250**	**Average**		**24.830**	**Average**		**9.460**
Industrials (IND)			**Information Technology (IT)**			**Materials (MAR)**			**Telecommunications Services (TEL)**			**Utilities (UTL)**		
General Electric	GE	38.335	Apple Inc.	AAPL	20.871	DuPont	DD	4.445	AT&T	T	14.646	The Southern Company	SO	2.661
United Technologies	UTX	4.413	Microsoft	MSFT	58.939	Freeport-McMoran	FCX	6.307	Vodafone Group Plc (ADR)	VOD	3.451	Exelon Corp.	EXC	2.892
United Parcel Service	UPS	3.010	IBM	IBM	6.775	Newmont Mining	NEM	5.509	Verizon Communications	VZ	9.583	Duke Energy	DUK	5.114
Minnesota Mining & Mfg Co	MMM	3.683	Oracle Corp.	ORCL	39.187	Dow Chemical	DOW	5.883	Telefonica S.A. (ADR)	TEF	0.453	American Electric Power	AEP	1.964
Caterpillar	CAT	5.833	Intel Corp.	INTC	57.647	Alcoa	AA	11.115	American Tower Corp A	AMT	2.480	Public Serv. Enterprise Inc.	PEG	2.220
Boeing	BA	4.429	Cisco Systems	CSCO	56.347	International Paper	IP	3.626	BT Group plc (ADR)	BT	0.119	PG&E Corp.	PCG	1.818
Honeywell Int'l Inc.	HON	4.298	Hewlett-Packard	HPQ	12.321	Nucor Corp.	NUE	4.186	CenturyTel Inc	CTL	1.201	Progress Energy, Inc.	PGN	1.020
General Dynamics	GD	2.088	EMC Corp.	EMC	19.749	Weyerhauser Co	WY	4.152	Sprint Nextel Corp	S	18.150	Entergy Corp.	ETR	1.134
Cummins Inc.	CMI	2.700	Dell Inc.	DELL	22.782	United States Steel Corp.	X	5.022	Frontier Communications	FTR	2.466	Constellation Energy Group	CEG	1.250
Southwest Airlines	LUV	5.084	Xerox Corp.	XRX	5.908	AK Steel Holding Corp	AKS	3.230	Qwest Communication Int	Q	13.300	ONEOK	OKE	0.545
Average Trading Volume		**7.390**	**Average**		**30.050**	**Average**		**5.350**	**Average**		**6.580**	**Average**		**2.060**

Notes: The table reports the 100 stocks traded in the US equity market, along with tickers and volume that are featured in our analysis.

Table 2. Summary Statistics by Sectors

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL
					% Daily Returns (R_t)					
Mean	0.032	0.050	-0.037	-0.029	-0.001	-0.001	0.015	-0.088	-0.027	0.011
S.D.	2.297	1.449	2.301	2.488	1.932	1.842	2.328	2.510	2.158	1.722
Skew	0.166	0.132	-0.697	-0.192	0.030	-0.053	0.175	-0.186	-0.030	-0.785
Kurt	12.506	8.964	20.780	16.455	7.999	7.633	8.485	8.019	13.283	27.297
					Trading Volume ($vol_t \times 10^{-7}$)					
Mean	6.674	4.325	6.249	24.828	9.462	7.387	30.564	5.347	6.585	2.062
S.D.	2.678	1.806	3.296	38.161	3.984	5.375	8.471	4.005	4.711	1.018
Skew	1.18	1.877	1.617	2.641	1.519	3.663	1.108	1.615	1.684	1.636
Kurt	4.893	8.953	8.348	11.024	6.788	28.916	6.087	5.965	7.358	7.672
Ranking	**5**	**9**	**7**	**2**	**3**	**4**	**1**	**8**	**6**	**10**
					Realized Variance (RV_t)					
Mean	5.455	2.453	5.836	6.339	4.245	3.696	5.722	6.703	5.637	3.419
S.D.	7.100	2.913	9.250	17.207	4.660	5.111	6.787	9.499	9.721	13.239
Skew	7.440	5.183	8.681	11.436	4.918	8.352	3.648	8.620	8.961	31.670
Kurt	111.077	50.947	118.145	198.683	56.216	139.260	24.865	123.691	140.262	1267.273
$d_{300\ sec}$	0.329	0.330	0.433	0.424	0.352	0.394	0.401	0.424	0.317	0.354
Ranking	**6**	**10**	**3**	**2**	**7**	**8**	**5**	**1**	**4**	**9**

Notes: The table reports mean, standard deviation (S.D.), skewness (Skew), Kurtosis (Kurt) for Daily returns, Trading volume, Realized variance. The Robinson's d for long-memory of the volatility series is denoted as d. The statistics are for the full sample 2001-2010.

Table 3. Full Sample

Sector	MAE				RMSE			
	ARFIMA	rank	HAR	rank	ARFIMA	rank	HAR	rank
CD	17.26	2	17.31	2	22.35	2	22.44	2
CS	18.38	8	18.42	8	24.56	8	24.61	8
ENG	16.96	1	16.84	1	22.17	1	22.04	1
FIN	19.16	9	19.12	9	24.82	9	24.94	9
HC	19.16	10	19.16	10	25.36	10	25.40	10
IND	18.02	6	18.13	7	23.54	7	23.69	7
IT	17.77	3	17.74	3	23.32	4	23.34	4
MAR	17.84	4	17.82	4	22.97	3	22.94	3
TEL	17.86	5	17.84	5	23.36	5	23.39	5
UTL	18.06	7	18.05	6	23.51	6	23.46	6
Average	18.05		18.04		23.60		23.63	
S.D.	1.35		1.39		1.99		2.05	
$\boldsymbol{\rho}$	0.99^{***}				0.99^{***}			

Notes: Table reports Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) in percentage terms for the 1-day-ahead volatility forecasts. A 500-day rolling window has been used and the statistics reported here are averages for the 10 sectors. The ARFIMA (0,d,0) is used where the order of differencing is estimated. S.D. denotes the Standard Deviation of the forecast evaluation measures across the 100 stocks. ρ is the bi-variate Spearman rank correlation of ARFIMA vis-à-vis HAR. ***, **, * denote the 1, 5 and 10% significance level respectively. 1(10): best (worst) forecast.

Table 4. Forecasting and Market Conditions.

	Panel A: Pre-Crisis					Panel B: Crisis					Panel C: Percentage Changes (%Δ)		
	ARFIMA	**rank**	**HAR**	**rank**	$d_{300\ sec}$	**ARFIMA**	**rank**	**HAR**	**rank**	$d_{300\ sec}$	**ARFIMA**	**HAR**	$d_{300\ sec}$
CD	11.47	8	19.57	9	0.36	11.30	6	18.75	4	0.42	-1.50	-4.37	14.42
CS	11.27	6	18.98	4	0.41	12.01	9	19.86	8	0.37	6.16	4.43	-10.48
ENG	11.42	7	19.34	7	0.36	10.85	2	18.09	1	0.44	-5.25	-6.91	19.19
FIN	10.05	2	18.22	1	0.40	11.62	8	21.07	10	0.34	13.51	13.53	-18.24
HC	12.32	10	20.56	10	0.42	12.51	10	20.39	9	0.39	1.52	-0.83	-6.12
IND	10.70	4	18.7	3	0.40	10.90	3	18.68	3	0.41	1.84	-0.11	2.44
IT	9.84	1	18.62	2	0.44	10.26	1	18.56	2	0.42	4.09	-0.32	-5.00
MAR	11.70	9	19.06	5	0.38	11.62	7	18.99	5	0.42	-0.69	-0.37	10.14
TEL	11.01	5	19.46	8	0.37	11.04	4	19.29	6	0.41	0.27	-0.88	8.80
UTL	10.44	3	19.24	6	0.38	11.27	5	19.70	7	0.36	7.37	2.34	-5.51
Average	11.02		19.17		0.39	11.34		19.34		0.40	2.73	0.65	0.96
S.D.	1.63		1.83			1.16		1.52					
$\boldsymbol{\rho}$	-0.35		-0.23			-0.47		-0.89					
Z	2.497^{**}		1.27			—		—					

Notes: The table reports Mean Absolute Error (MAE) in percentage terms for the 1-day-ahead volatility forecasts. A 500-day rolling window has been used and the statistics reported here are averages for the 10 sectors. The ARFIMA (0,d,0) is used where the order of differencing is estimated. S.D. denotes the Standard Deviation of the forecast evaluation measures across the 100 stocks. ρ is the bi-variate Spearman rank correlation of Volatility (see Table 1) vis-à-vis the rankings of the models. Z denotes the Wilcoxon Signed Rank test statistic for the equality of ranks test between pre-crisis (Panel A) and crisis (Panel B) periods. Panel C shows percentage changes in the MAEs of the ARFIMA and HAR models as well as the long-memory parameter (d) between the pre-crisis and crisis periods. ***, **, * denote the 1, 5 and 10% significance level respectively. 1(10): best (worst) forecast.

Table 5. Forecasting and Extended Horizons.

	1-day ahead		5-day ahead		22-day ahead	
	MAE	**Rank**	**MAE**	**Rank**	**MAE**	**Rank**
			ARFIMA			
CD	17.26	2	17.97	2	19.59	3
CS	18.38	8	18.89	8	20.16	7
ENG	16.96	1	17.73	1	19.09	1
FIN	19.16	9	20.34	10	22.00	10
HC	19.16	10	19.66	9	20.93	9
IND	18.02	6	18.79	7	20.21	8
IT	17.77	3	18.28	3	19.52	2
MAR	17.84	4	18.50	6	19.95	6
TEL	17.86	5	18.46	4	19.73	5
UTL	18.06	7	18.47	5	19.64	4
Average	18.05		18.71		20.08	
S.D.	1.35		1.46		1.62	
ρ	—		0.95^{***}		0.88^{***}	
			HAR			
CD	17.31	2	17.35	2	17.36	2
CS	18.42	8	18.48	8	18.50	8
ENG	16.84	1	16.87	1	16.92	1
FIN	19.12	9	19.14	9	18.89	9
HC	19.16	10	19.22	10	19.25	10
IND	18.13	7	18.18	7	18.03	6
IT	17.74	3	17.77	3	17.80	4
MAR	17.82	4	17.86	4	17.79	3
TEL	17.84	5	17.91	5	17.94	5
UTL	18.05	6	18.04	6	18.09	7
Average	18.04		18.08		18.06	
S.D.	1.39		1.40		1.38	
ρ	—		1.00^{***}		0.99^{***}	

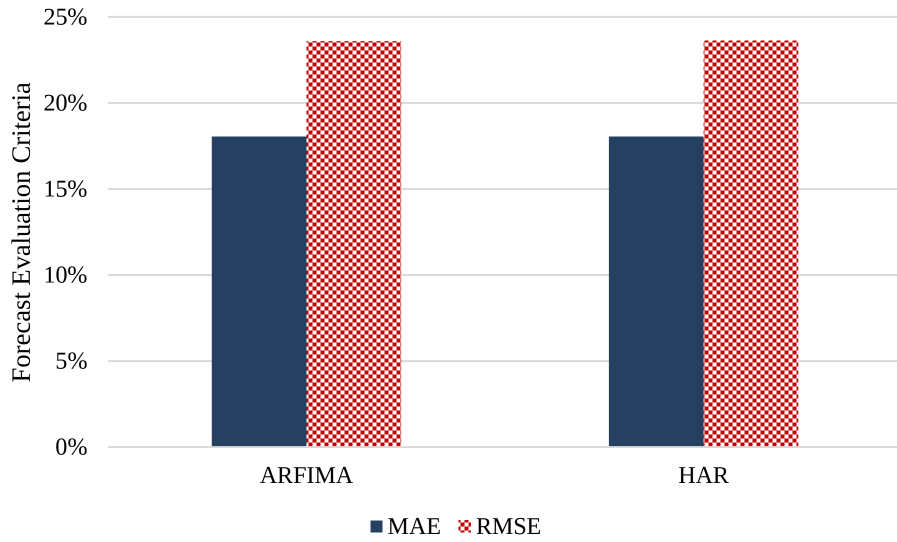
Notes: Table reports Mean Absolute Error (MAE) in percentage terms for the 1d-ahead, 5d-ahead and 22d-ahead volatility forecasts. A 500-day rolling window has been used and the statistics reported here are averages for the 10 sectors. The ARFIMA(0,d,0) is used where the order of differencing is estimated. S.D. denotes the Standard Deviation of the forecast evaluation measures across the 100 stocks. ρ is the bi-variate Spearman rank correlation where 1d-ahead is the baseline. ***, **, * denote the 1, 5 and 10% significance level respectively. 1(10): best (worst) forecast.

Table 6: Forecasting and Sampling Frequency.

Sampling Frequency	d	HAR (MAE)	ARFIMA (MAE)	KW	Best performer
5 sec	0.391	11.190	11.160	2.441**	ARFIMA
15 sec	0.389	12.380	12.340	3.128***	ARFIMA
60 sec	0.381	14.380	14.350	2.568***	ARFIMA
150 sec	0.370	15.920	15.920	0.155	—
300 sec	0.377	18.040	18.050	6.628***	HAR
900 sec	0.345	22.750	22.880	6.066***	HAR
1800 sec	0.326	27.590	27.860	7.495***	HAR
Average	**0.368**	**17.464**	**17.509**		**HAR**

Notes: The table shows the impact of the sampling frequency on the forecasting performance of ARFIMA and HAR models. KW is the Kruskal-Wallis Signed Rank test statistic to test for significant differences between the models across the various sampling frequencies. ***, **, * denote the 1, 5 and 10% significance level respectively.

Figure 1. Models Performance for Full Sample.



Notes: Figure 2 displays the overall average forecasting performance for the 100 stocks for ARFIMA and HAR using the MAE and RMSE loss criterion functions.

Figure 2. Forecasting Performance and Market Conditions.

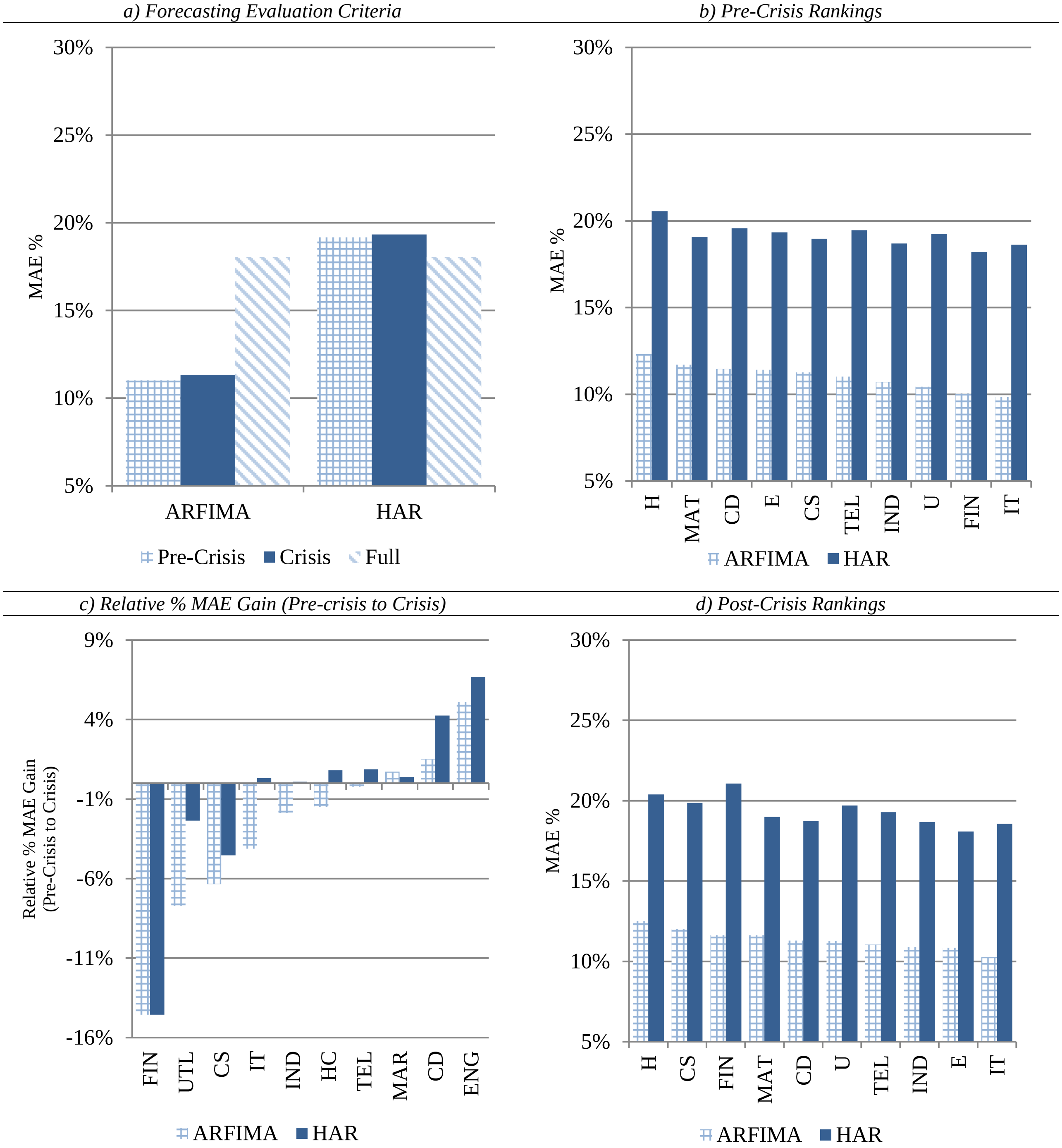


Figure 3. Relative % MAE Gain (1d to 22d).

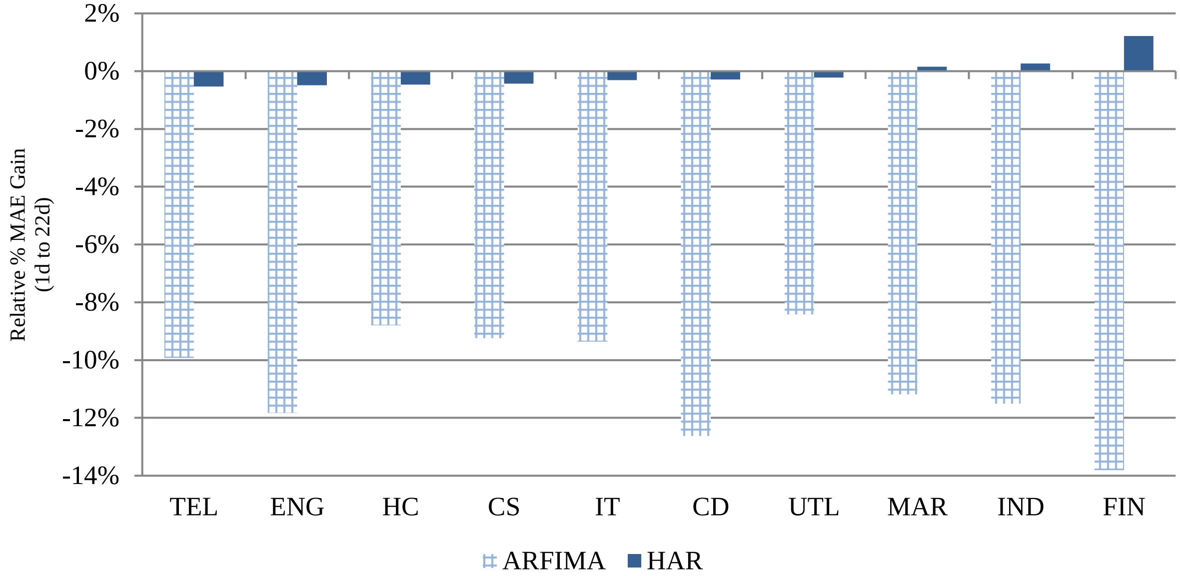
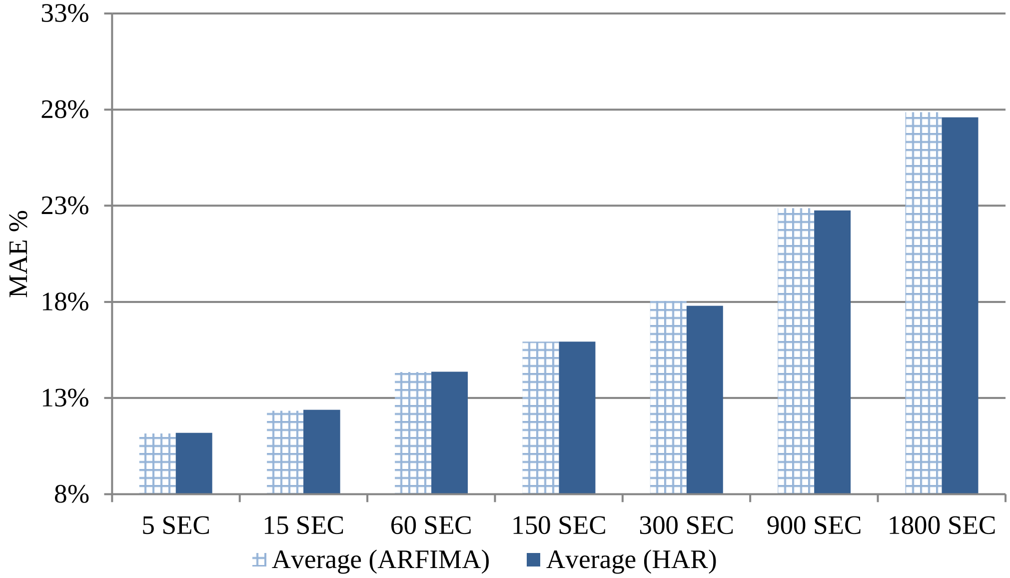


Figure 4. Forecasting Performance and Sampling Frequency.

a) Forecasting Evaluation Criteria



b) Relative % MAE Gain (baseline: 300sec)

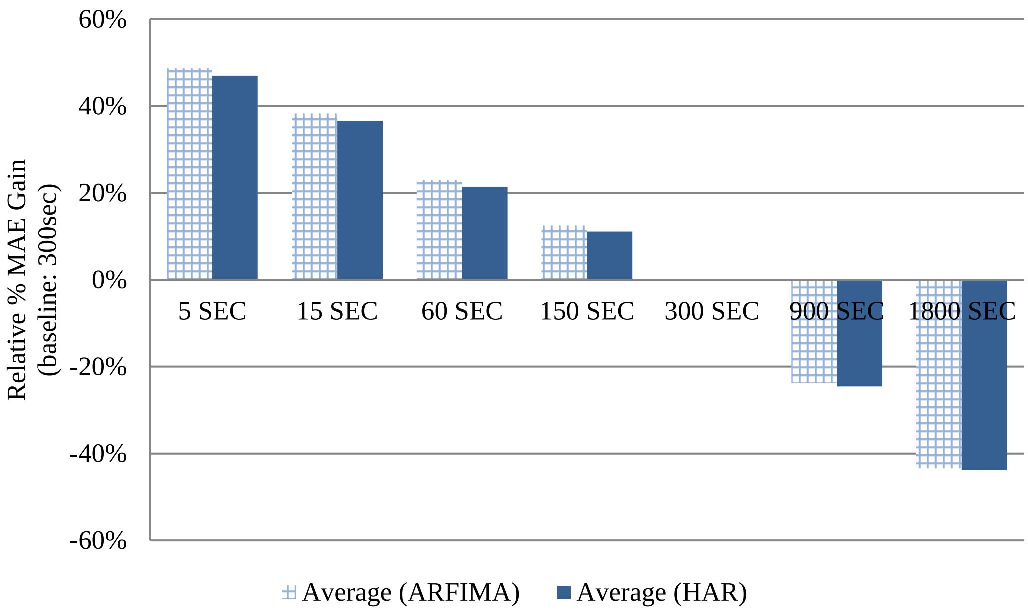
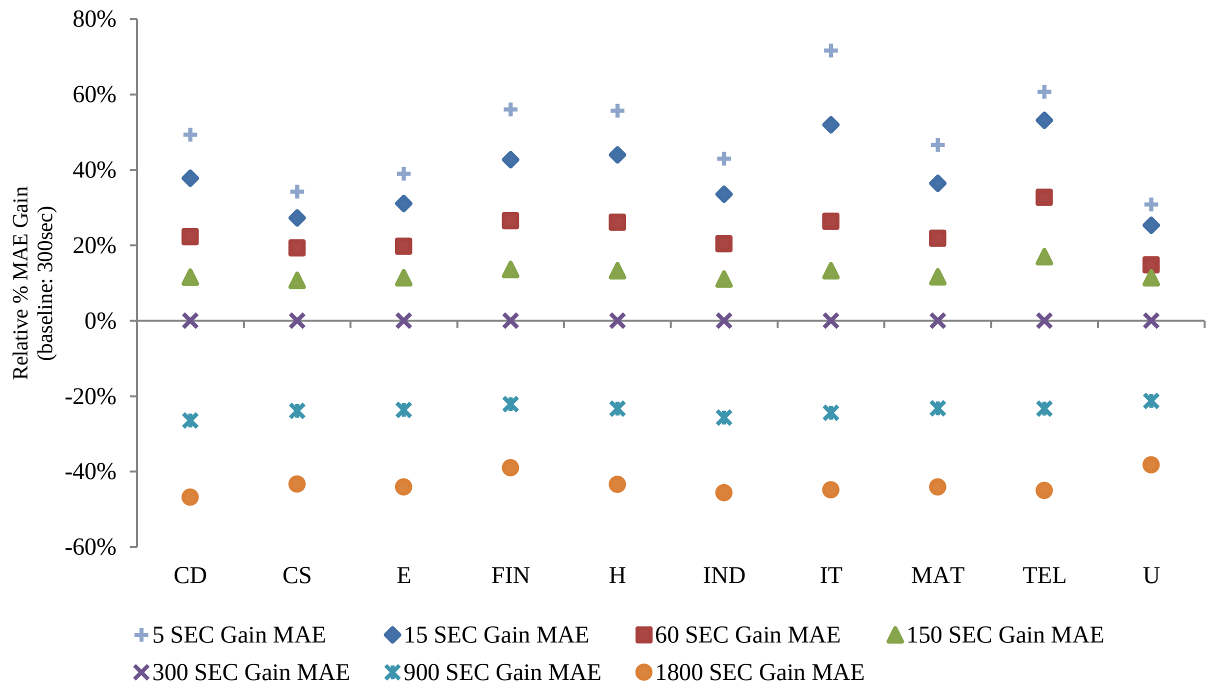


Figure 5. Forecasting Gains by Frequency and Sector.

a) ARFIMA



b) HAR