Pontifícia Universidade Católica
do Rio de Janeiro

Iúri Honda Ferreira

Essays on Volatility and Returns Predictability

Tese de Doutorado

Thesis presented to the Programa de Pós–graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Economia.

Advisor : Prof. Marcelo Cunha Medeiros
Co-advisor: Prof. Ruy Monteiro Ribeiro

Rio de Janeiro
April 2022

Pontifícia Universidade Católica
do Rio de Janeiro

Iúri Honda Ferreira

Essays on Volatility and Returns Predictability

Thesis presented to the Programa de Pós–graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Economia. Approved by the Examination Committee.

Prof. Marcelo Cunha Medeiros
Advisor
Departamento de Economia – PUC-Rio

Prof. Ruy Monteiro Ribeiro
Co-advisor
Insper –

Prof. Alan De Genaro
Escola de Administração de Empresas – EAESP/FGV

Prof. Gabriel Vasconcelos
BOCOM – BBM

Prof. Marcelo Fernandes
Escola de Economia de São Paulo – EESP/FGV

Prof. Márcio Gomes Pinto Garcia
Departamento de Economia – PUC-Rio

Rio de Janeiro, April the 28th, 2022

Iúri Honda Ferreira

Bachelor of Arts (B.A.) degree in Economics - Universidade de Brasília (2016). Accepted into the Master of Science (M.Sc.) degree in Economics (2017); accepted into the Ph.D. Economics Program - Pontifical Catholic University of Rio de Janeiro (2018). Associate Visiting Scholar at Columbia University in the City of New York (2020).

Bibliographic data

Ferreira, Iúri H.

Essays on Volatility and Returns Predictability / Iúri Honda Ferreira; advisor: Marcelo Cunha Medeiros; co-advisor: Ruy Monteiro Ribeiro. – Rio de janeiro: PUC-Rio, Departamento de Economia, 2022.

v., 73 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia.

Inclui bibliografia

1. Economia – Teses. 2. Econometria e Finanças – Teses. 3. Previsibilidade de retornos;. 4. Dados em alta dimensão;. 5. Aprendizado de máquinas;. 6. Modelos não-lineares. I. Medeiros, Marcelo C.. II. Ribeiro, Ruy M.. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. IV. Título.

CDD: 620.11

Fear no more the heat o' the sun,
Nor the furious winter's rages;
Thou thy worldly task hast done,
Home art gone, and ta'en thy wages:
Golden lads and girls all must,
As chimney-sweepers, come to dust.
- W. S.

Acknowledgments

To my supervisors, Prof. Marcelo Medeiros and Prof. Ruy Ribeiro, for the support, friendship and patience along this path. In particular, I thank Professor Marcelo, who welcomed me not only as his advisee, but also as a research member of the COVID-19 Analytics project. This experience was fundamental in sharpening my critical thinking and research skills. More than that, it brought me great friendships and opportunities, for which I will always be grateful.

To my twin brother and best friend, Hugo, who I can always count on. To my sister Isabela, for her constant light. To my parents and eternal references, Olímpio and Luciane, for their unconditional love and commitment. I also thank my friends in Brasília, Rio de Janeiro and New York, for making my life lighter and happier, even in moments of tension. Without my family and friends, none of this would have been possible.

To Prof. José Scheinkman, who received me during the period as a visiting researcher at Columbia University in the City of New York. Finally, I would like to thank the Department of Economics at the Pontifical Catholic University of Rio de Janeiro for the support and partial funding of my scholarship. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

Ferreira, Iúri H.; Medeiros, Marcelo C. (Advisor); Ribeiro, Ruy M. (Co-Advisor). **Essays on Volatility and Returns Predictability**. Rio de Janeiro, 2022. 73p. Tese de doutorado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

This thesis is composed of three papers on financial econometrics. The first two papers explore the relation between intraday equity market returns and implied volatility, represented by the CBOE Volatility Index (VIX). In both papers, we estimate one-minute-ahead forecasts using rolling windows within a day. In the first paper, the estimates indicate that our volatility factor models outperform traditional benchmarks at high frequency time-series analysis, even when excluding crisis periods. We also find that the model has a better out-of-sample performance at days without macroeconomic announcements. Interestingly, these results are amplified when we remove the crisis period. The second paper proposes a machine learning modeling approach to this forecasting exercise. We implement a minute-by-minute rolling window intraday estimation method using two nonlinear models: Long-Short-Term Memory (LSTM) neural networks and Random Forests (RF). Our estimations show that the VIX is the strongest candidate predictor for intraday market returns in our analysis, especially when implemented through the LSTM model. This model also improves significantly the performance of the lagged market return as predictive variable. Finally, the third paper explores a multivariate extension of the FarmPredict method, by combining factor-augmented vector autoregressive (FAVAR) and sparse models in a high-dimensional environment. Using a three-stage procedure, we estimate and forecast factors and its loadings, which can be observed, unobserved, or both, as well as a weakly sparse idiosyncratic structure. We provide an application of this methodology to a panel of daily realized volatilities. Finally, the accuracy of the stepwise method indicates improvements of this forecasting method when compared to consolidated benchmarks.

Keywords

Return predictability; High dimensional data; Machine learning; Nonlinear models

Resumo

Ferreira, Iúri H.; Medeiros, Marcelo C.; Ribeiro, Ruy M.. **Ensaios sobre Volatilidade e Previsibilidade de Retornos**. Rio de Janeiro, 2022. 73p. Tese de Doutorado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Essa tese é composta por três artigos em econometria financeira. Os dois primeiros artigos exploram a relação entre retornos intradiários do mercado de *equities* e a *implied volatility*, representada pelo Índice de Volatilidade da CBOE (VIX). Nos dois artigos, estimamos previsões um minuto à frente utilizando janelas rolantes para cada dia. No primeiro artigo, as estimativas indicam que nossos modelos de fatores de volatilidade têm uma performance superior a *benchmarks* tradicionais em uma análise de séries de tempo em alta frequência, mesmo aos excluirmos períodos de crise da amostra. Os resultados também indicam uma performance fora da amostra maior para dias em que não ocorrem anúncios macroeconômicos. A performance é ainda maior quando removemos períodos de crise. O segundo artigo propõe uma abordagem de aprendizado de máquinas para modelar esse exercício de previsão. Implementamos um método de estimação intradiário minuto a minuto com janelas móveis, utilizando dois tipos de modelos não lineares: redes neurais com *Long-Short-Term Memory* (LSTM) e *Random Forests* (RF). Nossas estimativas mostram que o VIX é o melhor previsor de retornos de mercado intradiários entre os candidatos na nossa análise, especialmente quando implementadas através do modelo LSTM. Esse modelo também melhora significativamente a performance quando utilizamos o retorno de mercado defasado como variável preditiva. Finalmente, o último artigo explora uma extensão multivariada do método FarmPredict, combinando modelos vetoriais autoregressivos aumentados em fatores (FAVAR) e modelos esparsos em um ambiente de alta dimensão. Utilizando um procedimento de três estágios, somos capazes de estimar e prever fatores e seus *loadings*, que podem ser observados, não observados ou ambos, assim como uma estrutura idiossincrática fracamente esparsa. Realizamos uma aplicação dessa metodologia em um painel de volatilidades realizadas e os resultados de performance do método em etapas indicam melhorias quando comparado a *benchmarks* consolidados.

Palavras-chave

Previsibilidade de retornos; Dados em alta dimensão; Aprendizado de máquinas; Modelos não-lineares

Table of contents

List of figures

List of tables

List of Abreviations

VIX – *Chicago Board Options Exchange Volatility Index*
RV – *Realized Volatility*
VRP – *Variance Risk Premium*
RNNs – *Recurrent Neural Networks*
LSTM – *Long-Short-Term Memory*
RF – *Random Forests*
RMSE – *Root-mean-square error*

1
Forecasting intraday returns using VIX

1.1
Introduction

The prediction of equity market returns an inexhaustible subject of interest for both practitioners and researchers in finance. With advances in high-frequency data processing capability, the challenge of predicting returns at higher frequencies has been gradually overcome. Ait-Sahalia and Jacod (2014) emphasize the undeniable progress of mathematical tools to analyze these data, which has become increasingly more accessible, allowing the development of new fields of study, as well as the boost of high-frequency trading strategies.

This paper presents a one-minute-ahead rolling window experiment to forecast intraday equity returns. We show that the Chicago Board Options Exchange (CBOE) VIX index succeeds as a strong predictor for market returns in high-frequency, outperforming traditional benchmarks. The outputs of minute-by-minute estimations indicate that our volatility factor models employing the intraday VIX in level as the main predictor present a much higher *out-of-sample* performance. These results hold, even when we remove crisis periods from the analysis.

Additional empirical findings suggest that our method has a better forecasting performance on days without macroeconomic announcements, while it seems to perform poorly when analyzing only the dates on which the announcements take place. These results do not find any definitive pattern of forecasting throughout the day, but highlight the aggregate *out-of-sample* forecasting capability in each day in the sample. Finally, we calculate similar performance metrics for each minute across the days, as a way to verify this pattern in different hours of the day. In general, the results point in the same direction as the ones obtained through rolling-windows.

The choice of VIX as a predictive variable is due to its importance in the literature and practice in finance. Also known as *"fear gauge"*, this index stands out as a measure of the market's implied volatility, based on S&P 500 Index option prices. This metric provides real-time updates on market sentiment and expectations of future 30-day volatility (Moran and Liu, 2020).

As pointed out by Martin (2011), it is generally interpreted as a measure of risk-neutral variance, and under certain circumstances may be thought as a proxy for the risk-neutral expectation of the quadratic variation of log returns.

The relation between market returns and VIX is widely explored in the literature through multiple approaches. Fernandes et al. (2014) examine time-series properties of the index at daily frequency, which holds a very strong negative relationship with the S&P 500 index returns as well as a positive link with the contemporary S&P 500 volume change. Bollerslev et al. (2009) explore the characteristics of a innovattive risk measure to explain the aggregate stock market returns. They propose the *variance risk premium* (VRP) variable, defined as the difference between implied (IV_t) and realized variances (RV_t), an useful predictor for observed return variation. The authors use the VIX index to quantify IV_t, while the RV_t is defined as the sum of the intraday non-overlapping 5-minute returns, in a similar framework to what is done in many works as Liu et al. (2015) and Bollerslev et al. (2018).

Bekaert and Hoerova (2014) suggest that the well-known results obtained in Bollerslev et al. (2009) exaggerate the predictive power of the VRP for stock returns. The authors strategy is to decompose the squared VIX into a conditional variance of the stock market (CV) and the equity variance premium (VP). In their work, they find that the the VP is a significant predictor of stock returns, while the CV mostly is not, recognizing that this decomposition critically depends on the accuracy of the model for the conditional component.

Martin (2017) introduces a volatility index named SVIX and claims that, under determinate circumstances, it provides a bound on the equity premium perceived by the investor. In the paper, the author infers that the equity premium is extremely volatile (more than implied by valuation-ratio predictors) and approximately tight, so that the SVIX index provides a direct measure of the equity market premium. Martin and Wagner (2019) explore a similar approach for stock returns.

The challenge of predicting equity returns in within a day is not a novel issue. Chinco et al. (2019) use intraday data to identify simultaneously unexpected, short-lived, and sparse predictors. The authors apply shrinkage methods in high-frequency intraday data, for the cross-section of a randomly chosen subset of NYSE-listed stocks. They implement a thirty-minute window framework, also using rolling estimations throughout the day. In a similar fashion, our work evaluates one-minute-ahead market return forecasts using the VIX as the main predictive variable.

Motivated by the results presented by Lucca and Moench (2015) and

Faust and Wright (2018), we also investigate if the market return predictability varies accordingly to macroeconomic announcement dates. Through a resourceful approach, Lucca and Moench (2015) find the existence of large average excess returns in anticipation of U.S. monetary policy actions. Faust and Wright (2018) analyze intraday data to compute the expected excess returns earned in short windows around the exact times of macroeconomic announcements, compared to the other times of the day. We select major macroeconomic announcement dates, and we run the model for two groups of days: one in which the announcements happen and another for the days without announcements, in which our model performs better.

For all the estimations in this paper, we set multiple regression benchmarks to contrast the predictive power of our outputs. Through *out-of-sample* performance metrics as the ones developed by Welch and Goyal (2008), Campbell and Thompson (2008), and Chinco et al. (2019), we are able to assess and attest the predictive power our models in a high-frequency environment. We combine these statistics with robustness exercises for crisis periods and different samples of days.

The paper is organized as follows. Section 1.2 presents the model definition, as well as the methodology adopted to forecast intradaily returns. It also introduces the data, its summary statistics, and the model implementation framework for estimation and prediction. Section 1.3 gathers the main results and robustness for the estimations. Finally, Section and 1.4 concludes the paper.

1.2 Methodology and Estimation Setup

For each asset i and day t, define $m = 1, 2, \ldots, M$ as a rolling one minute interval. Define $Y_{it,m}$ as the variable we want to predict. We want to capture short-lived effects from the predictors $\boldsymbol{X}_{it,m}$

$$Y_{it,m} = \boldsymbol{\beta}'_m \cdot \boldsymbol{X}_{it,m} + U_{it,m}. \tag{1-1}$$

Using a very short-term estimation framework, we can estimate the one-minute-ahead forecast for each window

$$\hat{Y}_{it,m+1|m} = \hat{\boldsymbol{\beta}}'_{m+1|m} \cdot \boldsymbol{X}_{it,m}. \tag{1-2}$$

1.2.1 Data

We use available intraday data for the SPDR S&P 500 ETF Trust (SPY) and the Cboe Volatility Index (VIX), comprising all business days between February 2005 and February 2017, totaling 3032 days. We perform a standard data pre-processing, described in Appendix A.1.1. For each day in the sample, we filter the data to include observations during the regular market trade hours, between 09:40 and 15:50 (inclusive).

In order to forecast market returns, we use five-minute log returns in minute-by-minute frequency as our main estimation framework. To compute rolling five-minute log-return of SPY for each minute of day t, we calculate the logarithmic difference of prices at minute m and minute $(m-4)$

$$r_{t,m-4:m} = \log(P_{t,m}) - \log(P_{t,m-4}), \tag{1-3}$$

where $P_{t,m}$ is the close price of SPY at day t and minute m, and $r_{t,m_2:m_1}$ represents the log-return computed between minutes m_1 and m_2. The summary statistics for intraday SPY log-returns and daily VIX are given in Table A.1. Both stock market returns (SPY) and expectation of volatility (VIX) are skewed right and leptokurtic. Compared to a normal distribution, its tails are longer and fatter, presenting a higher and sharper central peak, as illustrated by the Q-Q plot in Figure A.1. As expected, market returns and volatility are negatively correlated.

Finally, to better interpret the outputs of the estimates presented in the following section, we transform the VIX, defined as an annualized implied volatility, in an intraday minute measure[1]. This transformation allows us to analyze directly the estimated coefficients for each model, as described below.

1.2.2 Implementation

For each day t we run 340 intradaily minute estimations, using 30 minute rolling windows as *in-sample* data, described by equation 1-1. As portrayed below, the set of predictions at day t starts at 10:11 (09:40 + 30 minutes window) and goes until 15:50, estimated minute by minute in a rolling scheme

[1]For 1440 minutes a day and 252 days a year: $VIX_{\text{Intraday}} = VIX_{\text{Annual}} \times (1/\sqrt{1440}) \times (1/\sqrt{252})$

$$\ldots, \quad \underbrace{m-29, \quad m-28, \quad m-27, \quad \ldots, \quad m-2, \quad m-1, \quad m,}_{\text{Estimate coefficients using data from previous 30 minutes}} \quad \underbrace{m+1}_{\text{Forecast}}$$

Once we obtain the window coefficients from the past 30 minutes, we are able to forecast the 31^{st} minute $(m+1)$. Following the methodological framework presented in the begin of this section, the general forecasting equation will be given by

$$\hat{r}_{t,m-4:m} = \hat{\boldsymbol{\beta}}_{m-4:m} \cdot \boldsymbol{X}_{t,m-5} \tag{1-4}$$

where $\boldsymbol{X}_{t,m-5}$ may represent a range of setups for the lagged intraday VIX and potential benchmarks. The high frequency estimation allows us not only to the derive a relation between the market returns and volatility, but also to predict the one minute ahead return, by using an extremely simple and easy to implement framework.

1.2.3 Intraday Models

The minute by minute models are based on *non-overlapping* dependent and independent variables. For example, in order to run a regression of the 5-min SPY log-return on the lagged VIX^2 difference at day t, compute both variables in the following way

$$\ldots, \quad \underbrace{VIX^2_{t,m-6}, \quad VIX^2_{t,m-5},}_{\text{Compute } \Delta VIX^2_{t,m-6:m-5}} \quad \underbrace{P_{t,m-4}, \quad P_{t,m-3}, \quad P_{t,m-2}, \quad P_{t,m-1}, \quad P_{t,m}}_{\text{Compute } r_{t,m-4:m}}$$

This strategy allows us to predict intraday cumulative returns coherently, without using non-disclosed information from future minutes. After calculating the properly lagged variables at each m, one can run the equation

$$r_{t,m-4:m} = \alpha_m + \beta_m \Delta VIX^2_{t,m-6:m-5} + \varepsilon_{t,m}.$$

The same rationale is applied to the model benchmarks, described below. In the interest of identifying each regression, we define volatility factor models

VIX 1 to VIX 5 as

Model	**Equation**
VIX 1	$r_{t,m-4:m} = \alpha_m + \beta_m \cdot VIX_{t,m-5} + \varepsilon_{t,m}$
VIX 2	$r_{t,m-4:m} = \alpha_m + \beta_m \cdot VIX^2_{t,m-5} + \varepsilon_{t,m}$
VIX 3	$r_{t,m-4:m} = \alpha_m + \beta_m \cdot \Delta VIX_{t,m-6:m-5} + \varepsilon_{t,m}$
VIX 4	$r_{t,m-4:m} = \alpha_m + \beta_m \cdot \Delta VIX^2_{t,m-6:m-5} + \varepsilon_{t,m}$
VIX 5	$r_{t,m-4:m} = \alpha_m + \beta_m \cdot \overline{\Delta VIX^2}_{t,m-6:m-5} + \varepsilon_{t,m}$

where models 1 and 2 use the intraday VIX and squared-VIX (VIX^2) as independent variables in level, respectively. Model 3 and 4 use the first differences of these same variables. Finally, model 5 applies a rolling five-minute moving average over the first difference squared-VIX. These transformations aim to capture how the market's expectation of future volatility relates to intraday returns, one-period-ahead.

As a strategy to compare *out-of-sample* performance of volatility models, we use standard time-series analysis benchmarks, composed by autoregressive (AR) models with lags $p = 1, 2, \ldots, 5$:

$$\text{AR}(p): \quad r_{t,m-4:m} = \alpha_m + \sum_{k=1}^{p} \beta_{k,m} \cdot r_{t,m-8-k:m-4-k} + \varepsilon_{t,m}$$

where the past log-return is also *non-overlapping*, as described above. The following section presents the outputs obtained for both volatility factor and benchmarks models.

1.3 Results

Table A.2 evaluates the performance of the forecasts through the average *out-of-sample* R^2 statistic that can be directly compared to the *in-sample* R^2 in daily frequency. As done in Welch and Goyal (2008) and Campbell and Thompson (2008), we calculate the average daily $R^2_{OOS,t}$

$$R^2_{OOS} = 1 - \frac{\sum_{t=1}^{T}\sum_{m=1}^{M}(r_{t,m-4:m} - \hat{r}_{t,m-4:m})^2}{\sum_{t=1}^{T}\sum_{m=1}^{M}(r_{t,m-4:m} - \overline{r}_{t,m-4:m})^2} \tag{1-5}$$

where $\overline{r}_{t,m-4:m}$ is the *in-sample* window average of the realized log-return $r_{t,m-4:m}$, and $\hat{r}_{t,m-4:m}$ is the one-minute-ahead return forecast at minute m. This measure compares the average error assigned to the model prediction and the one of the past window average. When the R^2_{OOS} is positive, it means that our model outperforms on average the naive approach. If this metric is negative, it indicates that the model was not able to overcome the accuracy of the prediction based solely on the *in-sample* window average.

In line with the performance evaluation framework presented by Chinco et al. (2019), we also estimate a set of predictive regressions as an alternative measure of *out-of-sample* fit for each day t

$$r_{t,m-4:m} = \overline{a}_t + \overline{b}_t \cdot \hat{r}_{t,m-4:m} + e_{t,m} \tag{1-6}$$

$$r_{t,m-4:m} = \tilde{b}_t \cdot \hat{r}_{t,m-4:m} + e_{t,m}. \tag{1-7}$$

We aim to capture the ability of the predicted return ($\hat{r}_{t,m-4:m}$) to explain the observed return ($r_{t,m-4:m}$) at minute m, by computing the ordinary least squares coefficients and adjusted fit measure as additional *out-of-sample* performance metrics. The first two columns of Table A.3 report the average metrics and confidence intervals from equation 1-6 estimation outputs, where $\overline{b} = \frac{1}{T}\sum_{\tau=1}^{T} \overline{b}_\tau$ and $\text{Adj.}R^2 = \frac{1}{T}\sum_{\tau=1}^{T} \text{Adj.}R^2_\tau$. Analogously, the last two columns report $\tilde{b}$ and Adj. $\tilde{R}^2$ retrieved from the estimation of equation 1-7, in which we force the intercept to be zero ($\overline{a}_t = 0$). Additionally, Figure A.2 presents the box plots of the daily *out-of-sample* metrics for each model.

As carried out by Martin (2017), we run the same regressions removing the crisis period between 2008 and 2009, characterized by high volatility and the stock market crash. The argument for this robustness exercise is to check if the results obtained above are entirely driven by these dates, that comprise the days between August 1, 2008 and July 31, 2009. Table A.4 reports the summary statistics for the restrained data set, while Tables A.5 and A.6 display the *out-of-sample* metrics for the model estimations stated above. The results show that suppressing this period does not change significantly the predictability of intraday returns by our models, in which VIX and VIX^2 still outperform as independent variables.

To illustrate *out-of-sample* performance over time, Figure A.3 plots each metric described above at daily frequency. We choose to represent only models VIX 1, VIX 2 and AR(1) (the most accurate benchmark on average) for the sake of organization. As shown in Tables A.2 and A.3, these are best

performance volatility factor and benchmark models, respectively. As the figure suggests, the R^2_{OOS}, $\bar{b}$, and adjusted R^2 are consistently higher for VIX 1 and VIX 2 models across the days, when compared to the AR(1) model. The AR(1) model has an almost strictly negative daily R^2_{OOS}, and the VIX models seem to outperform the historical average in most part of the days across our sample.

Furthermore, Figure A.4 reports the average *out-of-sample* R^2 and the adjusted R^2 from equation 1-6 each month. The results obtained previously maintain in this analysis as well, confirming a better performance of volatility models. In the following section we replicate the same analysis by splitting the sample according to the occurrence of macroeconomic announcements.

1.3.1 Macroeconomic Announcements

To estimate the effects of information disclosure, we use selected macroeconomic announcements dates, available in the Bloomberg Economic Releases data to split the *out-of-sample* into announcements and non-announcements days. The events that compose the data set are detailed in Appendix A.1.2. As we are dealing with high-frequency data, and the estimations are computed within a day, we can simply break the sample of daily outputs presented in the last section in two groups of days. We follow Faust and Wright (2018) in order to select 9 major groups of events in which macroeconomic announcements happen.

The left columns in Table A.8 report the summary statistics for the subsample containing only non-announcement dates, while the columns in the right present the same statistics for the announcement dates subsample. It is noticeable that the announcement days SPY market returns are higher on average, as well as its kurtosis and the contemporary absolute correlation with the VIX index.

Table A.9, in turn, presents clear discrepancies in performance when comparing the average *out-of-sample* R^2 for announcement, in contrast with non-announcement dates. These compelling results are inflated when we exclude the crisis period, as shown in Table A.7. These results assist shedding a light on the question of whether and when market returns are predictable.

1.3.2 Minute by Minute Analysis

In this section, we confirm the previous phenomenons by another approach: Figure A.5 plots a minute by minute analysis, in which instead of calculating the metrics for each day, we compute it for each minute across all

days in the sample. In the figure, the plot in the top reports the *out-of-sample* $R^2_{OOS,m}$, calculated for each minute m as

$$R^2_{OOS,m} = 1 - \frac{\sum_{t=1}^{T}(r_{t,m-4:m} - \hat{r}_{t,m-4:m})^2}{\sum_{t=1}^{T}(r_{t,m-4:m} - \overline{r}_{t,m-4:m})^2}$$

in which we include the *in-sample* mean return at minute m ($\overline{r}_{t,m-4:m}$) as a naive prediction metric, among with the selected models. The plot in the bottom displays the Hit rate, or the percentage of correct direction assignment of the predicted returns when compared to the observed returns at each minute m, computed in the following manner

$$HitRate_m = \frac{1}{T} \cdot \sum_{t=1}^{T} 1\{\text{Sign}(r_{t,m-4:m}) = \text{Sign}(\hat{r}_{t,m-4:m})\}.$$

Once again, after calculating these minute-by-minute statistics, the outputs seem to confirm that our volatility factor models surpass the best performing benchmark in both *out-of-sample* $R^2_{OOS,m}$ and amount of correct sign attribution for the predicted values.

Additionally, we carry out the same analysis by excluding the crisis period in Figure A.6. The results obtained are quite close to the ones of the full sample. Interestingly, in both approaches the return predictability seem to differ from the intraday pattern in the minutes between 14:00 and 15:00pm.

1.4 Conclusions

Predicting high frequency returns can be a challenging task, but the advances in data processing capacity and the development of intraday models allowed a great evolution in this field of research. In this paper we generate intraday estimations through rolling windows, as a strategy to forecast equity market returns at minute frequency. We employ volatility factor models, in which we relate Cboe Volatility index (VIX) and the S&P 500 ETF (SPY) intraday data. The results obtained suggest that our method has a better forecasting performance than well-established autoregressive benchmarks, even when we exclude crisis periods. Additionally, the returns seem to be more predictable on dates in which macroeconomic announcements do not take place. Interestingly, this effect is amplified when we remove the crisis period.

2
Modeling and Forecasting Intraday Market Returns: a Machine Learning Approach

2.1
Introduction

The recent advances in high-frequency data estimation are associated not only to the technological development and growing processing capacity of *big data*, but also to the interest in understanding and predicting the behavior of variables in shorter time spaces. Machine Learning methods have been developed in parallel as increasingly accurate tools for estimating and predicting high-dimensional data. Both these fields of study can be easily accommodated into the economic and financial data environment. The adaptation of these methods to incorporate and update information over time allowed the development of robust predictive methods, progressively more relevant in time series analysis.

This paper examines the intrinsic relation between market returns and volatility measures, besides lagged returns themselves. Using minute-by-minute intraday data, we find that the CBOE Volatility Index (VIX) can be a strong predictor for the S&P 500 ETF (SPY) in high-frequency, especially through machine learning models. For each day in the sample, we implement a 30 minute rolling-window estimation procedure to forecast the subsequent minute, totaling 340 estimated market returns in a day. This framework is based on the estimation scheme of Chinco et al. (2019) for a cross-section of stock returns. We use different approaches to estimate these returns, from ordinary least squares (OLS) regression benchmarks to more sophisticated methods, focusing on nonlinear models. To build minute-by-minute machine learning models, we relied on the work of Masini et al. (2021).

Particularly, we apply Long-Short-Term Memory (LSTM) neural networks (Hochreiter and Schmidhuber, 1997) and Random Forest (RF) models (Breiman, 2001) to estimate multiple intraday rolling windows, using a range of predictive variables in different settings. LSTM models are a variant of Recurrent Neural Networks (RNNs), that differ from the standard neural networks on the ability to remember the previous states in time. This method is

broadly used in applications related to weather data, and speech and writing recognition. For economic and financial data in particular, we want this time dependence to be present, what makes the LSTM such an attractive model, specially due to its ability to remember what matters and forget what is irrelevant to the model. For a nice application of LSTMs in asset pricing see Chen et al. (2019).

Introduced by Breiman (2001), the Random Forest is an ensemble method, which means it combines several simpler models, producing an optimal improved aggregate version in the end. In this case, the base models are classification or regression trees, a nonparametric method based on the mechanism of recursive partitioning of the space of covariates. The average of the estimates generated by each tree are used to build the final forecast, known as Random Forest. RF models have been shown to be a very competitive forecasting tool. See, for example, Medeiros et al. (2021).

The choice to analyze volatility measures as potential predictors is an important issue in the economic literature. Among others, Corsi (2009) and Patton and Sheppard (2015) established a whole literature on realized volatility (RV), defined as the daily sum of the cumulative squared returns during business hours of a trading day. The forecasting methodology in this field of research uses predominantly autoregressive structures endowed with long memory to predict realized volatilities. McAleer and Medeiros (2008) provide a extensive review of theoretical developments and empirical applications concerning realized volatility.

Time-series properties of the VIX index are addressed by works like Fernandes et al. (2014), as well as its positive contemporaneous link with the volume of the S&P 500 index. Martin (2017) and Martin and Wagner (2019) explore the predictability of market and stock returns defining an associated volatility index named SVIX that provides a bound on the equity premium perceived by the investor. By associating implied variance (VIX^2) and realized variances measures, Bollerslev et al. (2009) introduce the *variance risk premium* (VRP) as the difference between those two variables, that can be used to explain the aggregate stock market returns. Bekaert and Hoerova (2014) extend these results by decomposing VIX^2 into a conditional variance of the stock market and the equity variance premium. The authors conclude that the variance premium is a significant predictor of stock returns.

This plethora of applications for volatility measures and machine learning methods raises our interest in combining these fields of study in a high frequency environment to predict future market returns. As will be discussed throughout the paper, our outputs indicate that machine learning models

as the LSTM may improve the performance of benchmark linear regression models in a minute-by-minute framework. Furthermore, if we choose the right regressors, particularly the VIX in our study, the predictive ability becomes even higher. On the other hand, Random Forests did not bring improvements to our estimations, when compared to the established benchmarks.

The paper is organized as follows. Section 2.2 reviews the methodology to be implemented in the machine learning framework. The high-frequency data details are presented in Section 2.3. Section 2.4 discusses the employment of different predictive models, as well as the benchmarks to our exercise. The results of intraday estimations and aggregate outputs are described in Section 2.5. Finally, we conclude the paper in Section 2.6.

2.2 Methodology

Define the following high-frequency forecasting model, where t is the day, and m the intraday minute:

$$Y_{t,m+h} = \mathcal{F}_h(\boldsymbol{X}_{t,m}) + U_{t,m+h}, \quad h = 1, \ldots, H, \quad t = 1, \ldots, T \tag{2-1}$$

where $\boldsymbol{X}_{t,m} := (Y_{t,m-1}, \ldots, Y_{t,m-p}, \boldsymbol{Z}'_{t,m}, \ldots, \boldsymbol{Z}'_{t,m-r})'$ is a n-dimensional vector of predictors, with $p \geq 1$ and $r \geq 0$. $\mathcal{F}_h : \mathbb{R}^n \to \mathbb{R}$ is an unknown measurable function and $U_{t,m+h} := Y_{t,m+h} - \mathcal{F}_h(\boldsymbol{X}_{t,m})$ has zero mean and finite variance.

For any chosen model framework and forecasting horizon $h = 1, \ldots, H$, we want to define the target function $\mathcal{F}_h$, to be estimated from the available data set. This function can be an ensemble of multiple models, and it can vary according to the horizon h.

2.2.1 Nonlinear Models

Masini et al. (2021) detail different machine learning methods, as well as their advances and applications in time series data environment. The authors discuss both linear and nonlinear models, of which we will focus on the last. The reason for this is that sometimes the linearity hypothesis may not encompass all the characteristics of variables such as volatility measures, specially in a high-frequency environment. In this case, we look for alternatives in the universe of models in statistical learning literature.

2.2.2 Long-Short-Term Memory Neural Networks

Recurrent Neural Networks (RNNs) are neural networks that allow for feedback among the hidden layers. RNNs can use their internal state (memory) to process sequences of inputs. A generic RNN can be written as

$$\begin{aligned} \boldsymbol{H}_{t,m} &= \boldsymbol{f}(\boldsymbol{H}_{t,m-1}, \boldsymbol{X}_{t,m}), \\ \widehat{Y}_{t,m+h|m} &= g(\boldsymbol{H}_{t,m}), \end{aligned}$$

where $\widehat{Y}_{t,m+h|m}$ is the prediction of $Y_{t,m+h}$ given observations only up to minute m at day t, $\boldsymbol{f}$ and g are functions to be defined and $\boldsymbol{H}_{t,m}$ is what we call the (hidden) state. From a time-series perspective, RNNs can be see as a kind of nonlinear state-space model.

RNNs can remember the order that the inputs appear through its hidden state (memory) and they can also model sequences of data so that each sample can be assumed to be dependent on previous ones. However, RNNs are hard to be estimated as they suffer from the vanishing/exploding gradient problem. Fortunately, there is a solution to the problem proposed by Hochreiter and Schmidhuber (1997): the Long-Short-Term Memory (LSTM) network . Figure A.7 shows the architecture of a typical LSTM layer. A LSTM network can be composed of several layers. In the figure, red circles indicate logistic activation functions, while blue circles represent hyperbolic tangent activation. The symbols "X" and "+" represent, respectively, the element-wise multiplication and sum operations. The RNN layer is composed of several blocks: the cell state and the forget, input, and ouput gates. The cell state introduces a bit of memory to the LSTM so it can "remember" the past. LSTM learns to keep only relevant information to make predictions, and forget non relevant data. The forget gate tells which information to throw away from the cell state. The output gate provides the activation to the final output of the LSTM block at day t and minute m. Usually, the dimension of the hidden state ($\boldsymbol{H}_{t,m}$) is associated with the number of hidden neurons.

Algorithm 2.1 describes how the LSTM cell works. $\boldsymbol{f}_{t,m}$ represents the output of the forget gate. It is a combination of the previous hidden-state ($\boldsymbol{H}_{t,m-1}$) with the new information ($\boldsymbol{X}_{t,m}$). Note that $\boldsymbol{f}_{t,m} \in [0,1]$ and it attenuates the signal coming com $\boldsymbol{c}_{t,m-1}$. The input and output gates have similar structure. Their goal is to filter the "relevant" information from the previous minute as well as from the new input. $\boldsymbol{p}_{t,m}$ scales the combination of inputs and previous information. This signal will be then combined with the output of the input gate ($\boldsymbol{i}_{t,m}$). The new hidden state will be an attenuation of

the signal coming from the output gate. The prediction is a linear combination of hidden states. Figure A.8 illustrates how the information flows in a LSTM cell.

Algorithm 2.1 *Mathematically, RNNs can be defined by the following algorithm:*

1. *Initiate with* $\boldsymbol{c}_{t,0} = 0$ *and* $\boldsymbol{H}_{t,0} = 0$.

2. *Given the input* $\boldsymbol{X}_{t,m}$, *for* $m \in \{1, \ldots, M\}$, *do:*

$$\begin{aligned}
\boldsymbol{f}_{t,m} &= \text{Logistic}(\boldsymbol{W}_f \boldsymbol{X}_{t,m} + \boldsymbol{U}_f \boldsymbol{H}_{t,m-1} + \boldsymbol{b}_f) \\
\boldsymbol{i}_{t,m} &= \text{Logistic}(\boldsymbol{W}_i \boldsymbol{X}_{t,m} + \boldsymbol{U}_i \boldsymbol{H}_{t,m-1} + \boldsymbol{b}_i) \\
\boldsymbol{o}_{t,m} &= \text{Logistic}(\boldsymbol{W}_o \boldsymbol{X}_{t,m} + \boldsymbol{U}_o \boldsymbol{H}_{t,m-1} + \boldsymbol{b}_o) \\
\boldsymbol{p}_{t,m} &= \text{Tanh}(\boldsymbol{W}_c \boldsymbol{X}_{t,m} + \boldsymbol{U}_c \boldsymbol{H}_{t,m-1} + \boldsymbol{b}_c) \\
\boldsymbol{c}_{t,m} &= (\boldsymbol{f}_{t,m} \odot \boldsymbol{c}_{t,m-1}) + (\boldsymbol{i}_{t,m} \odot \boldsymbol{p}_{t,m}) \\
\boldsymbol{h}_{t,m} &= \boldsymbol{o}_{t,m} \odot \text{Tanh}(\boldsymbol{c}_{t,m}) \\
\widehat{\boldsymbol{Y}}_{t,m+h|m} &= \boldsymbol{W}_y \boldsymbol{h}_{t,m} + \boldsymbol{b}_y
\end{aligned}$$

where $\boldsymbol{U}_f$, $\boldsymbol{U}_i$, $\boldsymbol{U}_o$,$\boldsymbol{U}_c$,$\boldsymbol{U}_f$, $\boldsymbol{W}_f$, $\boldsymbol{W}_i$, $\boldsymbol{W}_o$, $\boldsymbol{W}_c$, $\boldsymbol{b}_f$, $\boldsymbol{b}_i$, $\boldsymbol{b}_o$, *and* $\boldsymbol{b}_c$ *are parameters to be estimated.*

2.2.3 Regression Trees and Random Forests

A regression tree is a nonparametric model that approximates an unknown nonlinear function with local predictions using recursive partitioning of the space of the explanatory variables (predictors).

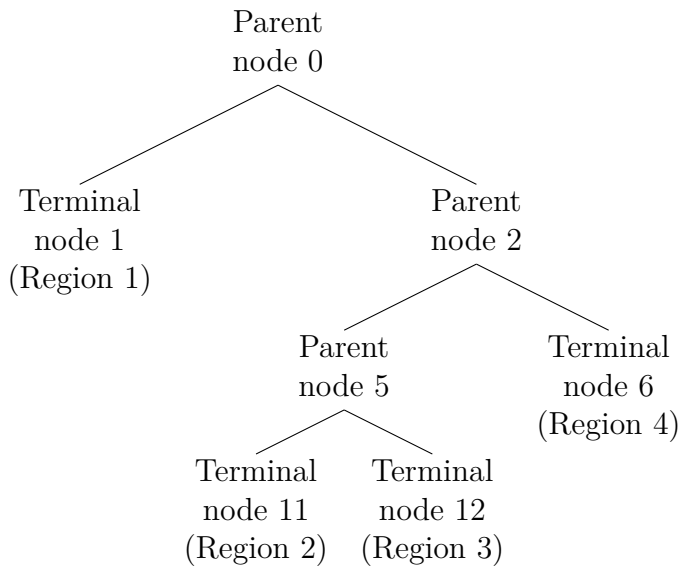
The idea of regression trees is to approximate $\mathcal{F}_h(\boldsymbol{X}_{t,m})$ in (2-1) by

$$h_D(\boldsymbol{X}_{t,m}) = \sum_{j=1}^{J} \beta_j I_j(\boldsymbol{X}_{t,m}), \quad \text{where} \quad I_k(\boldsymbol{X}_{t,m}) = \begin{cases} 1 & \text{if } \boldsymbol{X}_{t,m} \in \mathcal{R}_j, \\ 0 & \text{otherwise.} \end{cases}$$

From the above expression, it becomes clear that the approximation of $\mathcal{F}_h(\cdot)$ is equivalent to a linear regression on J dummy variables, where $I_j(\boldsymbol{X}_{t,m})$ is a product of indicator functions.

Let J and N be, respectively, the number of terminal nodes (regions, *leaves*) and parent nodes. Different regions are denoted as $\mathcal{R}_1, \ldots, \mathcal{R}_J$. The root node at position 0. The parent node at position j has two split (child) nodes at positions $2j+1$ and $2j+2$. Each parent node has a threshold (split) variable associated, $X_{s_j t}$, where $s_j \in \mathbb{S} = \{1, 2, \ldots, p\}$. Define $\mathbb{J}$ and $\mathbb{T}$ as the sets of parent and terminal nodes, respectively. Figure 2.1 gives an example. In the example, the parent nodes are $\mathbb{J} = \{0, 2, 5\}$ and the terminal nodes are $\mathbb{T} = \{1, 6, 11, 12\}$.

Figure 2.1: Example of tree with labels.



Therefore, we can write the approximating model as

$$h_D(\boldsymbol{X}_{t,m}) = \sum_{i \in \mathbb{T}} \beta_i B_{\mathbb{J}i}\left(\boldsymbol{X}_{t,m}; \boldsymbol{\theta}_i\right), \tag{2-2}$$

where

$$B_{\mathbb{J}i}\left(\boldsymbol{X}_{t,m}; \boldsymbol{\theta}_i\right) = \prod_{j \in \mathbb{J}} I(X_{s_j,t,m}; c_j)^{\frac{n_{i,j}(1+n_{i,j})}{2}} \times \left[1 - I(X_{s_j,t,m}; c_j)\right]^{(1-n_{i,j})(1+n_{i,j})}, \tag{2-3}$$

$$I(X_{s_j,t,m}; c_j) = \begin{cases} 1 & \text{if } X_{s_j,t,m} \leq c_j \\ 0 & \text{otherwise,} \end{cases}$$

$$n_{i,j} = \begin{cases} -1 & \text{if the path to leaf } i \text{ does not include parent node } j; \\ 0 & \text{if the path to leaf } i \text{ include the } \textbf{right-hand} \text{ child of parent node } j; \\ 1 & \text{if the path to leaf } i \text{ include the } \textbf{left-hand} \text{ child of parent node } j. \end{cases}$$

$\mathbb{J}_i$: indexes of parent nodes included in the path to leaf i. $\boldsymbol{\theta}_i = \{c_k\}$ such that $k \in \mathbb{J}_i$, $i \in \mathbb{T}$ and $\sum_{j \in \mathbb{J}} B_{\mathbb{J}i}\left(\boldsymbol{X}_{t,m}; \boldsymbol{\theta}_j\right) = 1$.

Random Forest (RF) is a collection of regression trees, each specified in a bootstrap sample of the original data. The method was originally proposed by Breiman (2001). Since we are dealing with time series, we use a block bootstrap. Suppose there are B bootstrap samples. For each sample b, $b = 1, \ldots, B$, a tree with K_b regions is estimated for a randomly selected subset of the original regressors. K_b is determined in order to leave a minimum number of observations in each region. The final forecast is the average of the forecasts of each tree applied to the original data:

$$\widehat{Y}_{t,m+h|m} = \frac{1}{B} \sum_{b=1}^{B} \left[\sum_{i=1}^{\mathbb{T}_b} \widehat{\beta}_{i,b} B_{\mathbb{J}i,b}(\boldsymbol{X}_{t,m}; \widehat{\boldsymbol{\theta}}_{i,b}) \right].$$

2.3 Data

Our dataset consists of 1 minute frequency data for both the S&P 500 ETF (SPY) and the CBOE Volatility Index (VIX), gathering a total of 3,000 business days between January 2005 and December 2016. We filter the data within each day to the observations between 09:40 and 15:50 (inclusive).

We calculate log-returns of SPY for each minute of day t, using a rolling five-minute return scheme, as follows:

$$r_{t,m-4:m} = \log(P_{t,m}) - \log(P_{t,m-4}), \tag{2-4}$$

where $r_{t,m_2:m_1}$ represents the log difference in prices computed between minutes m_1 and m_2, and $P_{t,m}$ is the price of SPY at day t and minute m. We explore the VIX index variable in level, and implement a transformation of the annualized VIX to an intraday minute variable, in order to facilitate the interpretation of our results[1]. Using this data set, we also compute the one-minute squared

[1] $VIX_{\text{Intraday}} = VIX_{\text{Annual}} \times (1/\sqrt{1440}) \times (1/\sqrt{252})$: 252 business days in a year and 1440 minutes in a day.

version VIX^2_m and first difference ΔVIX_m.

Inspired by Bollerslev et al. (2009), we generate a measure of what could be a high-frequency analogous to the *variance risk premium*, as the difference between the squared one-minute SPY return and the VIX^2 at minute m:

$$VRP_{t,m} = r^2_{t,m} - VIX^2_{t,m}. \tag{2-5}$$

The following estimations will be centered on models that use volatility measures as inputs, particularly the VIX. Despite presenting additional configurations and benchmark models, we focus on understanding how the minute-by-minute VIX relates to very high-frequency market returns, as well as its overall predictive capability across the models.

To illustrate the common behavior of market return and the volatility measure over the years, Figure A.9 displays the contemporary relation between the average minute-by-minute SPY log-return and VIX for each day of the sample. The darker the dots in the figure, the more recent the analyzed sample. It is noticeable that the relationship between both variables becomes less disperse over time, more concentrated in recent years, exhibiting lower returns for lower levels of volatility. Additionally, we present the summary statistics for minute-by-minute SPY log-returns and the annualized VIX in Table A.10, which shows that both variables are skewed right and leptokurtic. As expected, market returns and the VIX are negatively correlated.

Despite the clear intrinsic relationship between market returns and near-term volatility implied by stock index option prices, we want to investigate the forecasting potential of SPY returns in high-frequency using volatility measures. The subsequent section presents the fundamental estimation framework adopted in this paper, based on a minute-by-minute rolling window mechanism. This approach is similar to the one developed by Chinco et al. (2019), and allows the researcher to understand very short-term effects over intraday periodicity variables.

2.4 Empirical Analysis

To estimate our models, we implement a 30-minute rolling window estimation scheme for very high-frequency predictions. For each minute m and day t, we use the previous non-overlapping thirty minute window to forecast the one-minute-ahead five-minute log-return of SPY. At the end of each day

t, we run a total of 340 estimations between 09:40 and 15:50[2].

It is important to point out that our high frequency analysis uses *non-overlapping* input and output variables. This means that for each minute-by-minute log return $r_{t,m_2:m_1}$, we use lagged regressors that do not overlap minutes m_2 to m_1. For example, if we want to run a regression of the five-minute log return ($r_{t,m-4:m}$) on the lagged ΔVIX, we compute the *non-overlapping* variable as $\Delta\text{VIX}_{t,m-6:m-5} = \text{VIX}_{t,m-5} - \text{VIX}_{t,m-6}$. This strategy prevents from using future information to forecast minute $m + h$ when observations are available only up to minute m.

Machine learning algorithms perform better when numerical input variables are scaled. For each window, we perform a standard *MinMax* scaling to the data, by using the parameters from the *train* observations to scale both *train* and *test* data[3]. This transformation rescales variables into the range $[0, 1]$. We implement the standardization to the entire estimation data set, explored in both machine learning and benchmark models, as described in sequence.

2.4.1 Machine Learning Models

The architecture of the LSTM network allows the model to learn and forecast long sequences of data, a particularly useful attribute for time-series analysis. This approach can be put into action by running rolling window regressions in a one-shot multi-step framework for each day t. There are many options available to the researcher in terms of network structure and hyper-parameters to be chosen, including the number of layers, hidden states, loss function and optimizer. In our paper, we optimize the hyper-parameter by tuning it according the models performance on random subsamples.

As discussed in Section 2.2.1, Random Forest models (which we will also refer as RF from now on) fit classifying decision trees on data subsamples and employ an averaging approach to enhance the predictive power of the data. This ensemble learning method bootstraps the observations in blocks, using randomly sampled training sets (estimation intraday windows, in our case), controlling for potential overfit of the model. The choice of the number of trees to be used is a critical decision in a Random Forest, once it represents the main tuning engine in the model framework.

[2]We chose to avoid the first and last ten minutes of each business day in our estimations because these are the moments that usually contain a greater amount of missing values and/or repeated values.

[3]For each window ω rescale:

$$\tilde{x}^{train}_{\omega,m} = \frac{x^{train}_{\omega,m} - \min(x^{train}_{\omega})}{\max(x^{train}_{\omega}) - \min(x^{train}_{\omega})}; \quad \tilde{x}^{test}_{\omega,m} = \frac{x^{test}_{\omega,m} - \min(x^{train}_{\omega})}{\max(x^{train}_{\omega}) - \min(x^{train}_{\omega})}$$

For each machine learning model (LSTM and RF), we establish *three groups* of predictors $\boldsymbol{X}_{t,m}$. The first one is composed exclusively by the lagged volatility variable $\boldsymbol{X}_{t,m} = VIX_{t,m-5}$, while the second contains only the lagged five-minute return $\boldsymbol{X}_{t,m} = r_{t,m-9:m-5}$. We call these models VIX and AR(1), respectively. Finally, the last group gathers all the variables described in Section 2.3:

$$\boldsymbol{X}_{t,m} = \left(r_{t,m-9:m-5}, r^2_{t,m-9:m-5}, VIX_{t,m-5}, \Delta VIX_{t,m-6:m-5}, VRP_{t,m-5}\right)',$$

which we labeled as the aggregate model.

2.4.2 Benchmark Models

As benchmarks, we chose different setups for $\boldsymbol{X}_{t,m}$, implementing the same rolling-window approach. In addition to applying a traditional autoregressive model of order 1, we use lagged squared return, VIX, ΔVIX, and VRP as predictors. We estimate the coefficients in each window through standard ordinary least squares (OLS) regressions.

We define the benchmark models as follows:

$$\begin{aligned}
&\text{OLS-AR(1):} \quad r_{t,m-4:m} = \alpha_m + \beta_m \cdot r_{t,m-9:m-5} + \varepsilon_{t,m} \\
&\text{OLS-RV:} \quad r_{t,m-4:m} = \alpha_m + \beta_m \cdot r^2_{t,m-9:m-5} + \varepsilon_{t,m} \\
&\text{OLS-VIX:} \quad r_{t,m-4:m} = \alpha_m + \beta_m \cdot VIX_{t,m-5} + \varepsilon_{t,m} \\
&\text{OLS-}\Delta\text{VIX:} \quad r_{t,m-4:m} = \alpha_m + \beta_m \cdot \Delta VIX_{t,m-6:m-5} + \varepsilon_{t,m} \\
&\text{OLS-VRP:} \quad r_{t,m-4:m} = \alpha_m + \beta_m \cdot VRP_{t,m-5} + \varepsilon_{t,m}
\end{aligned} \tag{2-6}$$

We use the logic presented in the beginning of this section in order to regress the market returns on *non-overlapping* lagged variables. The benchmarks OLS-AR(1) and OLS-VIX are particularly important to our analysis, once it can be directly compared to the first two groups of predictors defined in the previous section. From both machine learning and benchmark models, we expect to understand the importance of the lagged log-return, as well as the VIX, in high-frequency forecasting.

This estimation practice was executed in a cloud environment, applying `scikit-learn`, `tensorflow`, and `keras` Python libraries to implement machine learning methods. We employ `Python 3.9.4 version` to perform

the estimations and `R 4.1.2 version` for output analysis. In the next section, we compare the performance metrics for both machine learning and benchmark models. We contrast specifically the regressions over different groups of predictors presented in Section 2.4.1 to the single input OLS models in the set of equations 2-6.

2.5 Estimation Results

Due to the high dimension of the estimated intraday outputs in each model (described in Section 2.4), we summarise our results through two performance metrics widely used in the literature: the daily Root-mean-square error (RMSE) and *out-of-sample* R^2. To obtain the mean and median RMSE, as well as the standard deviation, we calculate the daily measure

$$RMSE_t = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(r_{t,m-4:m} - \hat{r}_{t,m-4:m})^2} \tag{2-7}$$

and then compute its overall metric throughout the sample. Then, we calculate the same *out-of-sample* performance metric implemented by Welch and Goyal (2008) and Campbell and Thompson (2008) for each day t

$$R^2_{OOS,t} = 1 - \frac{\sum_{m=1}^{M}(r_{t,m-4:m} - \hat{r}_{t,m-4:m})^2}{\sum_{m=1}^{M}(r_{t,m-4:m} - \overline{r}_{t,m-4:m})^2} \tag{2-8}$$

to generate the average R^2_{OOS}, its median, and standard deviation across the days. This metric is relevant to introduce a relation between the mean squared error of our model estimate and the naive projection, represented by the *in-sample* historical mean within each estimation window. If the error of the model's prediction is lower than the error of the naive forecast, the R^2_{OOS} is necessarily positive. The same holds for the opposite case, in which a negative R^2_{OOS} represents a better performance of the historical average model.

2.5.1 Intraday Analysis

To exemplify how each predictive model works throughout a day, Figures A.10 and A.11 illustrate the intraday rolling prediction for a randomly selected date t (January 8, 2007). Each plot represents a machine learning model (LSTM and Random Forest) compared to the actual series, the historical mean based prediction, and the respective VIX or AR(1) benchmark. At each day t

calculate the daily performance metrics for each one of those series, and then compute the average, mean, and standard deviation for the entire sample.

The top plot in Figure A.10 presents the LSTM-VIX model (that uses VIX as predictive variable) minute-by-minute predictions, while the bottom one displays the same scenario for the RF-VIX model. Comparing both to the OLS-VIX benchmark, and the naive projection, one can see that the LSTM method tracks the observed series more smoothly, but also more precisely, while the Random Forest follows pretty much the same pattern established by the benchmark model. For this specific date, the R^2_{OOS} for the LSTM, Random Forest, and benchmark models were 21.15%, −37.10%, and 6.33%, respectively. We can see a clear superiority of the LSTM model, followed by the benchmark and, finally, the Random Forest model.

This pattern is repeated in Figure A.11, in which we perform the same analysis comparing autoregression based models. The minute-by-minute series forecasts follow an akin behavior to that observed in the previous figure. The curves are smoother, but also more precise for LSTM-AR(1) predictions. Even more interesting, the model now presents a *positive* R^2_{OOS}, compared to the negative ones obtained through the RF-AR(1) and OLS-AR(1) models on the same date: 6.93%, −21.87%, and −14.03%, respectively. Once again, the Random Forest output suggests a poor performance for this model. Obviously, these results are restricted to the analysis of a random day in the sample. The subsequent tables present and summarize the general outputs obtained by the estimation of the full sample, leading to similar conclusions.

2.5.2 Overall performance

Tables A.11 and A.12 present the average, median, and standard deviation of R^2_{OOS} and RMSE, respectively. These outputs are calculated over the performance metrics, using *out-of-sample* predictions for each day t, as described above[4]. Looking only at the bottom panel of each table, which contains the benchmarks presented in the set of equations 2-6, there is a clear hegemony of the OLS-VIX in terms of performance. The mean and median values of the *out-of-sample* R^2 outperform the naive prediction based on the historical mean ($R^2_{OOS} > 0$). As expected, this same conclusion can be drawn from the RMSE table, in which the VIX excels all the other variables as a predictor of market returns among the OLS equations.

[4]We trim the outputs in each model to restrict it to be between the 1^{st} and 99^{th} percentiles, excluding outliers in both directions.

When we include the LSTM and Random Forest machine learning models in the analysis, we can see critical improvements in terms of performance. The LSTM-VIX is the best performing model, followed by the OLS-VIX model. Even though the RF-VIX model performs well when compared to other models, it is not able to surpass the benchmark's performance. The LSTM approach for the autoregressive component also improves its performance when compared to the OLS-AR(1) model. Differently from the benchmark outputs, the mean and median R^2_{OOS} values for the LSTM-AR(1) are strictly positive, outperforming the historical mean projection. On the other hand, the RF-AR(1) model presents a drastic deterioration in performance when we use the lagged market return as predictor.

The LSTM and RF aggregate models (including all variables) do not present any improvement in its statistics when compared to the OLS-VIX benchmark, signaling that inserting other volatility measures and lagged market returns on the same model do not provide a better performance, compared to the models presented previously. To illustrate the different distributions of R^2_{OOS} metrics across the days in our sample, Figure A.12 presents the density for both machine learning (in the top row), and benchmark models (in the bottom). As one can see, the LSTM R^2_{OOS} distribution is predominantly left-skewed for every group of predictors, in accordance with the results presented in Tables A.11 and A.12. The RF models distributions present a reasonable performance in general, apart from the RF-AR(1). Still, these models are not able to overcome the OLS-VIX, which stands out as the best performing benchmark.

In a complementary way, Figure A.13 pursues to unravel the relationship between the results obtained by each machine learning model, compared to their respective benchmarks over the years. The vertical axis of each plot represent the LSTM and Random Forest models estimations using as regressor the AR(1) (in the first column of plots) and the VIX (in the second column). The horizontal axis displays the predicted values for the benchmark models. This visualization allows one to identify when machine learning models were able to overcome its benchmarks (the points below the 45° dotted line). The last years in our sample seem to present lower and less disperse estimation errors for all models combinations. Moreover, the VIX-based models got even better recently, when using both LSTM and RF models. These highlights synthesize the objective of this paper: contribute to the short-term forecasting literature, as well as the implementation of modern forecasting techniques through nonlinear rolling window models in a high-frequency setting.

2.6 Conclusion

There is growing interest in understanding how to predict high-frequency stock returns. Our paper introduces an innovative approach to estimate minute-by-minute market return (SPY) forecasts using volatility measures and lagged returns as predictors. In addition to standard benchmark models, we implement nonlinear machine learning methods as an attempt to capture idiosyncrasies of this kind of data, whose properties can differ substantially from variables at lower frequencies.

The outputs obtained in our estimations indicate a preliminary, but also encouraging path to better understand how to predict high-frequency market returns. We focus on neural networks (Long-Short-Term Memory) and tree-based (Random Forests) models to estimate multiple intraday rolling windows, using different regressors configurations. Models that use the Cboe Volatility Index (VIX) as predictor stand out, indicating that the VIX is a strong candidate predictor, when compared to other variables. The precision of the forecasts obtained using the OLS benchmark model gets even higher when we apply the LSTM to predict market returns using exclusively the VIX. Although the random forest model is not able to outperform the results obtained by the previous models (eventually even worsening it in some cases), the LSTM proved to be very promising in terms of predictive power, for both VIX and past market returns as regressors.

3
Factor Augmented High-Dimension Vector Autoregressive Models: Application to a Panel of Realized Volatilities

3.1
Introduction

In the presence of large data sets (e.g. financial) the number of parameters can be much larger than the observed sample size. In a high-dimensional environment one may need to impose some structure on the data as a way of dealing with that challenge, known as the *curse of dimensionality*. Modeling time-series in the context of *big data* generates additional demands to the researcher, especially regarding its asymptotic properties, as shown by Medeiros and Mendes (2012).

This paper carries out a combination of factor and shrinkage models, as a strategy to enhance standard *single-step* realized variance (RV) estimation procedures. Based on Fan et al. (2021)'s `FarmPredict` *three-step* estimation approach, we are able to boost a single factor model, initially with a poor performance for a single-step approach, but with a much higher predictive capacity when the multiple-step method is implemented. Moreover, this method improves the performance of well-known RV forecasting methods, as Corsi (2009)'s heterogeneous autoregressive (HAR) model. The most interesting output of this paper is the application of an easily replicable method, able to improve single factor models, delivering performances practically as good as those of consolidated methods, such as HAR and its variations. Finally, these outputs are ameliorated when using larger size windows for *one-step-ahead* daily estimations.

Applying factor models is a fairly convenient way of reducing the number of parameters to be estimated. This approach is not done in a completely *ad hoc* manner, since it aims to capture a few commonalities that properly explain the independent variable and concurrently reduce the dimension of the analyzed data. The main idea is to *"focus on what really matters"* to the model. Since the seminal work of Chamberlain and Rothschild (1983), the factor model structure has been widely explored by both practitioners and academics, as Stock and Watson (2002), Bai (2003), Bai and Ng (2003), Bernanke et al.

(2004), Bai and Ng (2006), and Fan et al. (2017).

In the macroeconomic forecasting literature, Stock and Watson (2002) develop a two stage dynamic factor model, in which the factors used to forecast the variable of interest are estimated by Principal Components Analysis (PCA), a practical approach to shrink large data sets. Bernanke et al. (2004) explore this framework to expand the vector autoregressive (VAR) model, by introducing factor-augmented vector autoregressive models (FAVARs). This method admits high dimensional data sets, assuming that the observable variables are directly related to the factors to be estimated, along with the variable to be predicted.

Statistical or implicit factor models as the Principal Components Regression (PCR) let the data tells us the relevant information, reducing the subjectivity of the process of choosing factors. On the other hand, analysis based on heuristically chosen observed factors are widely explored in financial econometrics and asset pricing literature. Welch and Goyal (2008) show that while some factors can be historically successful, they are also subject to different *out-of-sample* performances over time. A predictor may be useful in one period, but that does not guarantee that it will predict well in the future, specially in high-dimensional context. In a complementary way, Campbell and Thompson (2008) explore this fact in the context that most of the observable factors perform better when *out-of-sample* weak restrictions are imposed on the signs of coefficients and forecasts.

In the context of factor analysis, Feng et al. (2020) propose a model selection method to estimate and test the marginal importance of factors in pricing the cross section of returns, by extrapolating a large dimension set of potential factors (Cochrane (2011)'s *factor zoo*). The authors incorporate shrinkage methods to the factor approach in order to select the best control model, while taking into account potential selection errors. Brito et al. (2018) present a forecasting model for high-dimensional realized covariance matrices of returns, by combining firm-level factor decomposition, and sectorial restrictions. Their model also combines penalized methods and factor estimation.

In this paper we contribute to the literature on shrinkage and factor models, focusing on the FAVAR. We implement a generalization of the results in Fan et al. (2021). Their work assemble high-dimensional models in an unprecedented manner, by combining both factor and shrinkage models, and extracting its potential comparative advantages at different estimation steps. Moreover, the authors demonstrate that the combination of factors and a sparse regression strongly outperforms the traditional principal component regression. In an empirical application of this method, we assume that realized

variances are driven by a limited set of factors, which can be both observed and unobserved (latent). The goal of this exercise is to combine the stepwise method developed by Fan et al. (2021) to a multivariate panel prediction model. This strategy allows us to model the volatility panel as a factor model, while assuming that the idiosyncratic error term follows a (weakly) sparse VAR structure.

The paper is organized as follows: Section 3.2 introduces the model definition and some cases in which it can be applied. Section 3.3 discusses the panel data estimation, presenting each step of the three-stage method in detail. The empirical application of the realized variance panel data forecasting is addressed in Section 3.4. Finally, we conclude the paper in Section 3.5.

3.2 Model Definition

Consider the following model for a n_T-dimensional vector of time series:

$$\begin{aligned} \boldsymbol{Y}_t &= \boldsymbol{\Lambda}\boldsymbol{F}_t + \boldsymbol{V}_t, \\ &= \boldsymbol{C}_t + \boldsymbol{V}_t, \quad t = 1, 2, \ldots, T, \end{aligned} \tag{3-1}$$

where $\boldsymbol{F}_t \in \mathbb{R}^m$, $m < n_T$, is a vector of, possibly unobserved, common factors, $\boldsymbol{\Lambda}$ is a $(n_T \times m)$ matrix of unknown factor loadings, and $\boldsymbol{V}_t \in \mathbb{R}^n$ is the vector of idiosyncratic components. We assume that the number of factor m is fixed.

The vector of idiosyncratic components, $\boldsymbol{V}_t$, follows a n_T-dimensional, (weakly) sparse, vector autoregressive (VAR) model

$$\boldsymbol{V}_t = \boldsymbol{A}_0 + \boldsymbol{A}_1\boldsymbol{V}_{t-1} + \ldots + \boldsymbol{A}_{t-p}\boldsymbol{V}_{t-p} + \boldsymbol{U}_t \tag{3-2}$$

where $p := p_T$, such that both the dimension and the lag order of the VAR are allowed to grow with the sample size. Furthermore, $\boldsymbol{U}_t$ a zero-mean martingale difference error term ,$\boldsymbol{A}_0$ is a $(n_T \times 1)$ vector of parameters and $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p$ are $(n_T \times n_T)$ matrices of parameters.

To complete the model, the factors follows a zero-mean VAR model:

$$\boldsymbol{F}_t = \boldsymbol{B}_1\boldsymbol{F}_{t-1} + \ldots + \boldsymbol{B}_q\boldsymbol{F}_{t-p} + \boldsymbol{W}_t \tag{3-3}$$

where $\boldsymbol{W}_t$ is also a zero-mean martingale difference error term. $\boldsymbol{B}_1, \ldots \boldsymbol{B}_q$ are $(m \times m)$ matrices of parameters. For simplicity we assume that both VARs have the same order.

Substituting equations (3-2) and (3-3) into (3-1) the model can be written as:

$$\begin{aligned}\boldsymbol{Y}_t &= \boldsymbol{A}_0 + \boldsymbol{A}_1\boldsymbol{Y}_{t-1} + \cdots + \boldsymbol{A}_p\boldsymbol{Y}_{t-p} + \boldsymbol{\Lambda}\boldsymbol{F}_t - \boldsymbol{A}_1\boldsymbol{\Lambda}\boldsymbol{F}_{t-1} - \cdots - \boldsymbol{A}_p\boldsymbol{\Lambda}\boldsymbol{F}_{t-p} + \boldsymbol{U}_t \\ &= \boldsymbol{A}_0 + \boldsymbol{A}_1\boldsymbol{Y}_{t-1} + \cdots + \boldsymbol{A}_p\boldsymbol{Y}_{t-p} + (\boldsymbol{\Lambda}\boldsymbol{B}_1 + \boldsymbol{A}_1\boldsymbol{\Lambda})\boldsymbol{F}_{t-1} + \cdots + (\boldsymbol{\Lambda}\boldsymbol{B}_p + \boldsymbol{A}_p\boldsymbol{\Lambda})\boldsymbol{F}_{t-p} + \boldsymbol{U}_t + \boldsymbol{\Lambda}\boldsymbol{W}_t\end{aligned}$$

Therefore, the final specification is a Factor-Augmented VAR model (FAVAR):

$$\begin{aligned}\begin{pmatrix}\boldsymbol{Y}_t \\ \boldsymbol{F}_t\end{pmatrix} = \begin{pmatrix}\boldsymbol{D}_0 \\ \boldsymbol{0}\end{pmatrix} + \begin{pmatrix}\boldsymbol{A}_1 & \boldsymbol{D}_1 \\ \boldsymbol{0} & \boldsymbol{B}_1\end{pmatrix}\begin{pmatrix}\boldsymbol{Y}_{t-1} \\ \boldsymbol{F}_{t-1}\end{pmatrix} + \cdots \\ \cdots + \begin{pmatrix}\boldsymbol{A}_p & \boldsymbol{D}_p \\ \boldsymbol{0} & \boldsymbol{B}_p\end{pmatrix}\begin{pmatrix}\boldsymbol{Y}_{t-p} \\ \boldsymbol{F}_{t-p}\end{pmatrix} + \begin{pmatrix}\boldsymbol{U}_t + \boldsymbol{\Lambda}\boldsymbol{W}_t \\ \boldsymbol{W}_t\end{pmatrix}.\end{aligned} \tag{3-4}$$

Defining $\boldsymbol{Z}_t = (\boldsymbol{Y}_t', \boldsymbol{F}_t')'$, equation (3-4) can be written in compact form as

$$\boldsymbol{Z}_t = \boldsymbol{\Pi}_0 + \boldsymbol{\Pi}_1\boldsymbol{Z}_{t-1} + \cdots + \boldsymbol{\Pi}_p\boldsymbol{Z}_{t-p} + \boldsymbol{E}_t. \tag{3-5}$$

This general specification nests several interesting special cases as discussed in the following examples.

Example 1 (Diagonal Model) *If all matrices $\boldsymbol{A}_i$, $i = 1, \ldots, p$ are diagonal, the idiosyncratic terms follow autonomous AR models augmented by a factor structure.*

Example 2 (Block-Diagonal Model) *One generalization of the above example is to assume that each $\boldsymbol{A}_i$, $i = 1, \ldots, p$, is block-diagonal, representing, for instance, the cross-section dependence among the idiosyncratic shocks.*

3.3 Estimation Methodology

The panel data model is described by

$$Y_{it} = \boldsymbol{\beta}_i' \underbrace{\boldsymbol{X}_{it}}_{\substack{\text{Observed}\\ \text{factors}}} + \overbrace{\boldsymbol{\lambda}_i' \underbrace{\boldsymbol{F}_t}_{\substack{\text{Latent}\\ \text{factors}}} + \underbrace{U_{it}}_{\substack{\text{Idiosyncratic}\\ \text{term}}}}^{=:R_{it}} = \boldsymbol{\beta}_i'\boldsymbol{X}_{it} + R_{it} \tag{3-6}$$

where $\boldsymbol{X}_{it}$ is a $k-$dimensional observable random vector[1], $\boldsymbol{F}_t$ is a r-dimensional vector of common latent factors, and U_{it} is a zero mean idiosyncratic shock. We also can write the equation (3-6) in the matrix aggregate form. For $\boldsymbol{Y}_t := (Y_{1t}, \ldots, Y_{nt})'$, and $\boldsymbol{U}_t := (U_{1t}, \ldots, U_{nt})'$, define:

$$\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{X} + \boldsymbol{\Lambda}\boldsymbol{F}' + \boldsymbol{U} = \boldsymbol{B}\boldsymbol{X} + \boldsymbol{R} \tag{3-7}$$

composed by $(n \times T)$ dependent variable $\boldsymbol{Y} := (\boldsymbol{Y}_1, \ldots \boldsymbol{Y}_T)'$ and idiosyncratic shock $\boldsymbol{U} := (\boldsymbol{U}_1, \ldots \boldsymbol{U}_T)'$ matrices; a $(nk \times T)$ covariates matrix $\boldsymbol{X}_i := (\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{iT})$; and a $(r \times T)$ matrix of latent factors $\boldsymbol{F}' = (\boldsymbol{F}_1, \ldots, \boldsymbol{F}_T)$. $\boldsymbol{B}$ is a $(n \times nk)$ matrix of observable factor parameters $\beta_{i,k(i)}$ and $\boldsymbol{\Lambda}$ a $(n \times r)$ matrix of factor loadings λ_{ir}. We will predict $Y_{i,t+h}$ through an extension of the three-stage estimation procedure proposed by Fan et al. (2021) known as `FarmPredict`, explained subsequently.

3.3.1 Three-Stage Method

The methodology in this paper involves a sequence of estimation steps. For each i regress Y_{it} on $\boldsymbol{X}_{it}$ to estimate the first-stage residuals $\hat{R}_{it}$. Compute the Principal Components Analysis (PCA) of the aggregated set of residuals $\hat{\boldsymbol{R}}_t$ to estimate the latent factors $\hat{\boldsymbol{F}}_t$ and its respective loadings. Gather the second-stage residuals in a vector $\hat{\boldsymbol{U}}_t$. Then, estimate equation-wise a (weakly) sparse Vector Autoregression (VAR) model of order p for $\hat{\boldsymbol{U}}_t$, through a penalized estimation method. Finally, combine the models and construct the forecasting equation. This procedure is presented in detail below, as well as the models adopted in each stage.

3.3.1.1 First Stage

For each $i \in \{1, \ldots, n\}$ run the following regression on the observable factors

$$Y_{it} = \boldsymbol{\beta}_i' \boldsymbol{X}_{it} + R_{it}, \quad t = 1, \ldots, T \tag{3-8}$$

using linear regression models. In this stage, we choose to estimate the coefficients and residuals through linear methods. In accordance with Masini et al. (2021), the estimator $\hat{\boldsymbol{\beta}}_i$ for the unknown desired parameter minimizes

[1] k may also be a function of i: $\boldsymbol{X}_{it} = (\boldsymbol{X}_{it}^{(1)}; \boldsymbol{X}_t^{(2)})$.

$$\mathcal{L}(\boldsymbol{\beta}_i) = \sum_{t=1}^{T}(Y_{it} - \boldsymbol{\beta}_i'\boldsymbol{X}_{it})^2 - \rho(\boldsymbol{\beta}_i)$$

where $\rho(\boldsymbol{\beta}_i)$ is the penalization term that depends on a tuning parameter[2] $\lambda \geq 0$. After choosing the method that will be implemented, compute $\hat{R}_{it} := Y_{it} - \hat{\boldsymbol{\beta}}_i'\boldsymbol{X}_{it}$ and write the first-stage residuals as a n-dimensional vector

$$\hat{\boldsymbol{R}}_t := (\hat{R}_{1t}, \ldots, \hat{R}_{nt})' \tag{3-9}$$

As pointed out in equation 3-7, this vector corresponds to $\boldsymbol{R}_t = \boldsymbol{\Lambda}\boldsymbol{F}_t + \boldsymbol{U}_t$. We assume the implicit premise that the observable factors $\boldsymbol{X}_{it}$ are independent of the latent factors $\boldsymbol{F}_t$.

3.3.1.2 Second Stage

Compute the PCA of $\hat{\boldsymbol{R}}_t$ to estimate the latent factors $\boldsymbol{F}_t$ and run a regression on these principal components for each i

$$\hat{R}_{it} = \boldsymbol{\Lambda}\boldsymbol{F}_t + \hat{U}_{it} \tag{3-10}$$

using models with potential for selection of latent factors. To determine the factors to be used among the principal components, one can adopt shrinkage models as the (adaptive) LASSO, or apply selection methods as the Bai and Ng (2002) information criteria, and Ahn and Horenstein (2013) eigenvalue ratio test, and simply estimate it through OLS. For each i, calculate the second-stage residuals $\hat{U}_{it} = \hat{R}_{it} - \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{F}}_t$ (equivalent to the idiosyncratic term in equation 3-7). Collect the residuals in a n-dimensional vector

$$\hat{\boldsymbol{U}}_t := (\hat{U}_{1t}, \ldots, \hat{U}_{nt})' \tag{3-11}$$

This aggregate object will be used to compute the final stage outputs.

3.3.1.3 Third Stage

Define $\hat{\boldsymbol{U}}_t$ as a sparse Vector Autoregressive model VAR(p) with n variables and lag p

[2]Notice that when $\lambda = 0$, we have the standard Ordinary Least Squares (OLS) estimation.

$$\begin{pmatrix} U_{1t} \\ \vdots \\ U_{nt} \end{pmatrix} = \mathbf{\Gamma}_1 \begin{pmatrix} U_{1,t-1} & \dots & U_{n,t-1} \\ \vdots & \ddots & \vdots \\ U_{1,t-1} & \dots & U_{n,t-1} \end{pmatrix} + \cdots + \mathbf{\Gamma}_p \begin{pmatrix} U_{1,t-p} & \dots & U_{n,t-p} \\ \vdots & \ddots & \vdots \\ U_{1,t-p} & \dots & U_{n,t-p} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ \vdots \\ e_{nt} \end{pmatrix}$$

$$\boldsymbol{U}_t = \mathbf{\Gamma}_1 \boldsymbol{U}_{t-1} + \cdots + \mathbf{\Gamma}_p \boldsymbol{U}_{t-p} + \boldsymbol{e}_t \tag{3-12}$$

where the matrices $\Gamma_{l=1,\dots,p}$ are autoregressive parameters and the vector $\boldsymbol{e}_t = (e_{1t}, \dots, e_{nt})'$ is the error term. The standard VAR approach is to estimate OLS equation by equation. However, due to the large dimension of the data, it may be preferable to use an adaptive LASSO approach for the estimation of a sparse VAR model.

Define $\boldsymbol{U}_i = (\boldsymbol{U}_{i,p+1}, \dots, \boldsymbol{U}_{i,T})'$; $\boldsymbol{U}_L = (\mathbf{\Upsilon}_{p+1}, \dots, \mathbf{\Upsilon}_T)'$, with $\mathbf{\Upsilon}_t = (\boldsymbol{U}'_{t-1}, \dots, \boldsymbol{U}'_{t-p})'$; $\mathbf{\Gamma}_i = (\mathbf{\Gamma}_{i,1}, \dots, \mathbf{\Gamma}_{i,p})$; and $\boldsymbol{E}_i = (\boldsymbol{e}_{i,p+1}, \dots, \boldsymbol{e}_{i,T})'$. Rewrite the VAR in matrix notation

$$\boldsymbol{U}_i = \boldsymbol{U}_L \mathbf{\Gamma}_i + \boldsymbol{E}_i \tag{3-13}$$

To estimate $\mathbf{\Gamma}_i^*$ through LASSO-VAR(p), we minimize

$$\mathcal{L}(\mathbf{\Gamma}_i) = \frac{1}{T} \|\boldsymbol{U}_i - \boldsymbol{U}_L \mathbf{\Gamma}_i\|^2 + 2\boldsymbol{\lambda}_T \|\mathbf{\Gamma}_i\|_{\ell_1}$$

Defining a weighting scheme as the inverse of the LASSO estimator, estimate the adaptive-LASSO-VAR(p) through the minimization

$$\tilde{\mathcal{L}}(\mathbf{\Gamma}_i) = \frac{1}{T} \|\boldsymbol{U}_i - \boldsymbol{U}_{L,\tilde{J}(\mathbf{\Gamma}_i)} \hat{\mathbf{\Gamma}}_{i,\tilde{J}(\hat{\mathbf{\Gamma}}_i)}\|^2 + 2\boldsymbol{\lambda}_T \sum_{j \in J(\hat{\mathbf{\Gamma}}_i)} \frac{|\mathbf{\Gamma}_{i,j}|}{|\hat{\mathbf{\Gamma}}_{i,j}|}$$

where $J(\hat{\mathbf{\Gamma}}_i) = \{j \in \mathbb{R}^{kp} : \hat{\mathbf{\Gamma}}_{i,j} \neq 0\}$ represent the indices of the coefficients in the i'th equation deemed zero by the LASSO and $\tilde{J}(\hat{\mathbf{\Gamma}}_i) = \{1\} \cup \{J(\hat{\mathbf{\Gamma}}_i) + 1\}$, following the framework of Callot et al. (2017). The penalization works in the following manner: if $\hat{\mathbf{\Gamma}}_{i,j}$ is small, and hence $1/\hat{\mathbf{\Gamma}}_{i,j}$ is large, the penalty on $\mathbf{\Gamma}_{i,j}$ will be large. Alternatively, if the estimated $\hat{\mathbf{\Gamma}}_{i,j}$ is large, the penalty will be small.

3.4 Application to a Panel of Realized Volatilities

3.4.1 Data

We use Intraday data for constituents of the Dow Jones Industrial Average Index comprising all business days between January 2006 and February 2017, totaling 2802 days. We filter the data to include observations during the regular market trade hours, between 09:30 and 16:00 (inclusive). We consider firms that remained in the index for the full sample period, culminating in a total of 28 stocks.

Out of this high-dimensional data set, we are able to construct the daily realized variance measures (RV_{it}) using 5-minutes returns as an estimator of the *ex-post* volatility on day t for a given stock i. In addition to the standard autoregressive framework described in the following section, we chose the first-difference lagged squared CBOE Volatility Index (VIX) as observed volatility factor $\boldsymbol{X}_{it}$. This fundamental relation between RV and the VIX has been already explored in literature by Bollerslev et al. (2009), Bekaert and Hoerova (2014), and, Martin (2017), among others. Ultimately, *train-test split* establishes different estimation windows to determine the forecast in a one-step-ahead framework.

3.4.2 Realized Variance Forecasting

The daily *realized variance* (RV) is given by the sum of all available intraday high frequency squared returns

$$RV_{it} = \sum_{m=1}^{M} r_{itm}^2 \tag{3-14}$$

Where M represents the number of intradaily intervals m for a given asset i on day t. In a similar fashion to what is done by Liu et al. (2015) and Bollerslev et al. (2018) we use subsampled *5-minute* RV, which seems to work well in the majority of volatility forecasting exercises.

For each stock i and day t define $Y_{it} := log(RV_{it})$ as the observed daily logarithmic realized variance. This approach is also implemented by Bekaert and Hoerova (2014), who argue that due to right-skewed distributions of variances, it is preferable to predict instead its logarithmic transformation,

which tends to have near Gaussian distributions. Using the panel data model represented by equation 3-6, we want to predict for each given i

$$\hat{Y}_{i,t+1|t} := \hat{\boldsymbol{\beta}}_i' \boldsymbol{X}_{it} + \hat{\boldsymbol{\lambda}}_i' \hat{\boldsymbol{F}}_{t+1|t} + \hat{U}_{i,t+1|t} = \hat{\boldsymbol{\beta}}_i' \boldsymbol{X}_{it} + \hat{R}_{i,t+1|t} \tag{3-15}$$

through the stepwise methodology presented in Section 3.3.1. As pointed out by McAleer and Medeiros (2008) the volatility of returns seems to be relatively easier to forecast, when compared to daily returns of financial assets *per se*: a multitude of realized volatility forecasting methods have been explored in literature, including Corsi (2009), Patton and Sheppard (2015), and Bollerslev et al. (2016). That motivates us to implement the multiple-step procedure over some of the main realized volatility estimation methods, detailed in the next section.

3.4.3 Setup

The empirical application in this paper employs a multivariate extension of Fan et al. (2021)'s `FarmPredict`. This procedure allows the researcher to explore a number of distinct combinations of factor and shrinkage models. In the first two stages, we decided to use combinations of non-penalized (OLS) and shrinkage (PCA) models, while in the third stage we restrict our analysis to the sparse VAR estimation, focusing on the adaptive LASSO method.

The intuition underlying the first stage relates to the ability of determining the adequate observable variables, and remove its contribution in order to properly estimate the second stage, both described in Section 3.3.1. In the standard OLS approach, which we call the Volatility Factor Model, define the observable variable as: $\boldsymbol{X}_{it} := \Delta VIX_t^2 = VIX_t^2 - VIX_{t-1}^2$.

Additionally, we explore the heterogeneous autoregressive (HAR) structure, proposed by Corsi (2009), as well as its extensions. We select conventional models utilized to predict realized volatility: the HAR-J (jump augmented), the HAR-RS (realized semi-variance augmented), and the HAR-Q (introducing the *"realized quarticity"* to the model), as well as its fully time-varying specification, the HARQ-F. The fundamental HAR equation of Corsi (2009) is characterized as

$$\text{(HAR):} \quad RV_{i,t+1} = \mu + \phi_d RV_{it}^d + \phi_w RV_{it}^w + \phi_m RV_{it}^m + R_{it}$$

In each model, $RV_{it}^{d} = RV_{it}$ denotes the lagged *daily* realized variance, $RV_{it}^{w} = \frac{1}{4}\sum_{d=1}^{4} RV_{i,t-d}$ the lagged *weekly* realized variance, and $RV_{it}^{m} = \frac{1}{17}\sum_{d=5}^{21} RV_{i,t-d}$ the lagged *monthly* realized variance. This alternative autoregressive format aims to capture long memory in financial volatility, by including the mean realized variances computed for the preceding week and month, besides the daily lagged component. McAleer and Medeiros (2008) discuss this additive hierarchical structure, specified as a sum of components over different horizons.

As done in Andersen et al. (2007) and Patton and Sheppard (2015), we extend the HAR model by including jumps (J_{it}), positive (RS_{it}^{+}), and negative (RS_{it}^{-}) realized semi-variances. The jump augmented model is given by

$$\text{(HAR-J):} \quad RV_{i,t+1} = \mu + \phi_d RV_{it}^{d} + \phi_w RV_{it}^{w} + \phi_m RV_{it}^{m} + \phi_J J_{it} + R_{it}$$

The jumps are defined as $J_{it} = \max\{(RV_{it} - BV_{it}), 0\}$, where $BV_{it} = (2/\pi)^{-1}\sum_{m=2}^{M} |r_{it,m}| \cdot |r_{it,m-1}|$ is the *bipower variation*, proposed by Barndorff-Nielsen (2004). The realized semi-variance augmented model is

$$\text{(HAR-RS):} \quad RV_{i,t+1} = \mu + \phi_d^{+} RS_{it}^{+} + \phi_d^{-} RS_{it}^{-} + \phi_w RV_{it}^{w} + \phi_m RV_{it}^{m} + R_{it}$$

in which the positive and negative realized semi-variances are constructed as follows

$$RS_{it}^{+} = \sum_{m=1}^{M} r_{itm}^{2} \cdot 1\{r_{it} > 0\}, \qquad RS_{it}^{-} = \sum_{m=1}^{M} r_{itm}^{2} \cdot 1\{r_{it} < 0\}$$

This approach bounds the realized variance measure according to the sign of the return in each intradaily interval m. The concept of setting apart realized semi-variances is widely discussed in the realized volatility forecasting literature, as in Barndorff-Nielsen et al. (2008) and Bollerslev et al. (2020). Finally the realized quarticity (RQ) augmented model of Bollerslev et al. (2016) is specified as

$$\text{(HAR-Q):} \quad RV_{i,t+1} = \mu + \phi_{d,t+1} RV_{it}^{d} + \phi_w RV_{it}^{w} + \phi_m RV_{it}^{m} + R_{it}$$

Define $\phi_{d,t+1} = \phi_d + \phi_{dQ} RQ_{it}^{1/2}$, where $RQ_{it} = \frac{M}{3} \sum_{m=1}^{M} r_{itm}^2$ is the realized quarticity, a consistent estimation of the integrated quarticity presented in Barndorff-Nielsen and Shephard (2002). To compute the daily RQ for each i, we use the same high-frequency specification as in equation 3-14. The fully time-varying specification includes the same transformations for the weekly and monthly realized variances

$$\text{(HARQ-F):} \quad RV_{i,t+1} = \mu + \phi_{d,t+1} RV_{it}^{d} + \phi_{w,t+1} RV_{it}^{w} + \phi_{m,t+1} RV_{it}^{m} + R_{it}$$

where $\phi_{w,t+1} = \phi_w + \phi_{wQ}(\frac{1}{4} \sum_{d=1}^{4} RQ_{i,t-d})^{1/2}$ and $\phi_{m,t+1} = \phi_m + \phi_{mQ}(\frac{1}{17} \sum_{d=5}^{21} RQ_{i,t-d})^{1/2}$.

As detailed in Section 3.3.1, after estimating the first-stage residuals, the second-stage involves an objective process to define the non-observed factors. While using a non-penalized method, we select the number of latent factors from the set of estimated principal components, by applying the information criteria designed by Bai and Ng (2002), or Ahn and Horenstein (2013)'s eigenvalue ratio test for the number of factors. If we choose to implement a penalized method, the principal components selection is automatically executed by the shrinkage mechanism inherent to the model.

At all estimation steps we are able to combine shrinkage (penalized) models: we can obtain the first-stage residuals, select latent factors, calculate the second-stage residuals, and impose a sparse structure at the third stage to estimate the VAR of order p. We follow Garcia et al. (2017) and include up to four lags ($p = 4$) for each candidate variable in the third step of our model. The (adaptive) LASSO framework and its implementation are described in Section 3.3.1.

Through this setup, we are able to estimate all the steps of the `FarmPredict` method for a panel of realized volatilites. Gathering the outputs in each stage, one can predict $\hat{Y}_{i,t+1|t}$ for stock i as in equation 3-15. The aggregate results of this exercise are presented in the following section.

3.4.4 Forecasting Results

Using the model combinations described in Section 3.4.3, we report the accuracy of our model in Tables A.13 and A.14. The panels report the *out-of-sample* performance metrics of the *one-step-ahead* rolling estimation scheme, using different sized windows. For each one, we present the average measure

across DJIA stocks, for the first, second, and third stages of the `FarmPredict` method described in Section 3.3.1.

The panels on Table A.13 evaluate the performance of the forecasts through an overall *out-of-sample* (OOS) R^2 statistic, that can be directly compared to the in-sample R^2. Analogously to Welch and Goyal (2008) and Campbell and Thompson (2008), we calculate

$$R^2_{OOS} = 1 - \frac{\sum_{t=1}^{T}\sum_{i=1}^{n}(Y_{i,t} - \hat{Y}_{i,t})^2}{\sum_{t=1}^{T}\sum_{i=1}^{n}(Y_{i,t} - \overline{Y}_i)^2} \tag{3-16}$$

where $\hat{Y}_{i,t}$ represents the fitted value estimated through the forecasting equation, as a combination of each stage of the stepwise model, and $\overline{Y}_i$ is the historical average of $Y_{i,t}$, computed for each i. This measure of predictive power of the forecasted models works in the following manner: if the R^2_{OOS} is positive, the predictive regression has lower average mean squared error prediction error than the historical average approach, outperforming it. Nonetheless, a negative *out-of-sample* R^2 indicates that the model was not able to overcome the naive prediction accuracy. In a complementary way, Table A.14 reports for each panel the mean absolute error (MAE), using different estimation windows.

The outputs indicate that, in general, the `FarmPredict` procedure applied to the stocks' realized volatilities outperforms its single step estimation benchmarks, located in the first column of each panel. Moreover, it seems to improve not only traditional forecasting models performance when the second and third stages are implemented over it, but enhance dramatically the performance of the single volatility factor model, specially for larger estimation windows. These compelling results are displayed in Tables A.13 and A.14.

We run rolling estimations using different window sizes (1, 2, 4, and 5 years) for all models. In general, the performance of HAR family models seem to improve slightly for longer windows, having the best absolute performance for 5-year windows, followed by 1-year estimation windows. On the other hand, the volatility factor model presents a dramatic improvement in all scenarios, showing a monotonic progression on larger size estimation windows. These results show an extremely poor performance when estimating solely the first-step, but a massive improvement when applying the following steps over it.

As described in equation 3-16, the R^2_{OOS} summarises a direct comparison between the method and the naive historic approach. When comparing these statistics in Table A.13, the volatility factor model initially appears not to contain enough information to explain future RV. However, the application of the following steps seems to put in evidence the existence of potentially

unobserved (latent) factors, derived in the second and third steps from the residual terms. The second step delivers a better result for all windows, while the third step enhances exclusively the 1 year length estimation window. The results indicate the possibility of improvements of third step outputs for larger estimation windows, by modeling the second stage residuals more efficiently. Finally, for some scenarios, the volatility factor model gets quite close to the traditional HAR in terms of performance, as well as some of its family constituents, a challenging task in the RV literature.

3.5 Conclusions

In this paper we implement an extension the `FarmPredict` method to a multivariate approach, allowing to combine both factor-augmented autoregressive (FAVAR) and shrinkage models for high-dimensional data sets. In order to predict some dependent variable, we assume that its time-series dynamic is driven by a limited set of factors, that can be observed or unobserved. Using a three step procedure, this method allows one to estimate the unobserved (latent) factors, while modeling the idyosincratic term. We document an application to a panel of realized volatilities, revealing an increase in *out-of-sample* performance driven by our method.

We are able to boost a single factor model, that initially shows poor performance for single-step estimations, but a much higher forecasting accuracy when estimated through multiple-steps. The method improves the performance of well-known RV forecasting methods, as the heterogeneous autoregressive (HAR) model and its variations, and specially a factor model based solely on the CBOE VIX. The application of this method improves substantially single factor models, delivering performances practically as good as those of consolidated methods in the literature. Finally, the method seems to work better when using larger size windows for *one-step-ahead* daily estimations.

A
Appendix

A.1
Forecasting intraday returns using VIX

A.1.1
Data Pre-Processing

We pre-process the historical prices for each variable before running the predictive models. For both SPY and VIX, we fill the minute by minute daily sample by using the last available observation. For example, for a day t, if we have a missing values between the minutes 09:40 and 09:44, we can adjust the prices by doing the following

Time	*Close Price*	*Adjusted Close Price*
09:40	19.22	$= 19.22$
09:41	NA	$\rightarrow 19.22$
09:42	19.42	$= 19.42$
09:43	19.45	$= 19.45$
09:44	NA	$\rightarrow 19.45$
$\vdots$	$\vdots$	$\vdots$

By doing this, we assume that all the information disclosed in a missing price at minute m comes from the last observation. After filling all missing observations, we generate the variables of interest that will be used in each regression. We consider only the days when both indices have been reported, totaling a period consisting of 3032 trading days between February 2005 and February 2017.

Although the market's regular business hours are between 09:30 and 16:00, we chose limit our analysis to returns calculated between 09:40 and 15:50, since the trimmed minutes may suffer from distortions or a greater occurrence of missing data. On the other hand, once we are working with five-minute log-return data, the first returns are effectively calculated only after 09:35, which means that our decision leads to an even smaller loss of data in practice.

A.1.2 Macroeconomic Announcements

We use the Bloomberg Economic Releases from the Economic Calendar[1] to extract the announcements dates data set. These data contain a wide variety of economic announcements data, including National Accounts (GDP), Prices, Labor Market, Retail and Wholesail Sector, Industrial Sector, Services Sector, Surveys and Cyclical Indicators, Purchasing Managers' Index, Housing and Real Estate, Personal/Household Sector, International Trade and Balance of Payments, Governance Finance and Debt, Monetary Sector, and Financial Indicators.

In order to organize the data, we select major macroeconomic announcements as done by Faust and Wright (2018), and arrange them into 9 groups: FOMC, GDP, Household, Housing, Industrial, Labor, Prices, Retail and Wholesale, and Trade Balance. We extract both the announcements (1580 days) and the non-announcements subsets (1452 days) from the original data set, composed by 3032 days between February 2005 to February 2017. The events enclosed in each group are listed as follows

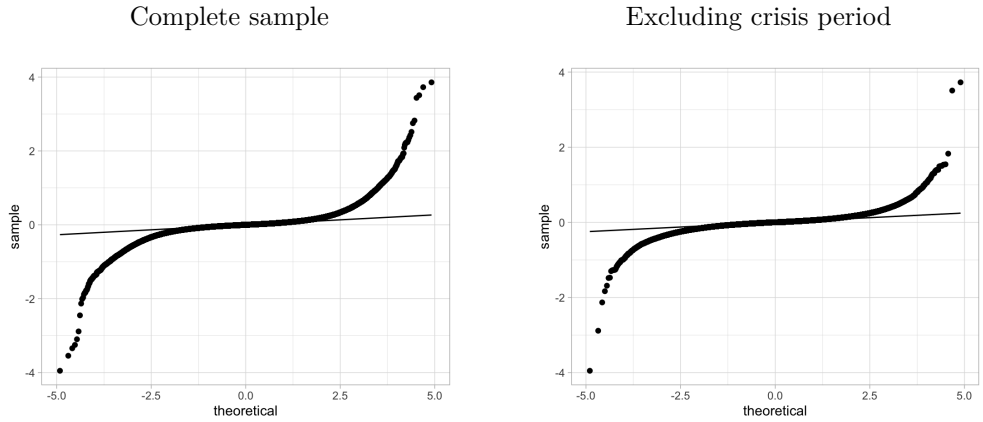
Group	Macroeconomic Announcement	Number of Events
FOMC	*U.S. Federal Reserve Releases Beige Book*	198
FOMC	*Fed Releases Minutes from FOMC Meeting*	198
FOMC	*FOMC Rate Decision (Upper Bound)*	198
FOMC	*FOMC Rate Decision (Lower Bound)*	198
GDP	*Personal Consumption*	363
GDP	*GDP Annualized*	363
GDP	*GDP Price Deflator*	363
Household	*Personal Income*	181
Housing	*Housing Starts*	303
Industrial	*Industrial Production*	413
Industrial	*Durable Goods Orders*	413
Labor	*Initial Jobless Claims*	1149
Labor	*Unemployment Rate*	1149
Labor	*Change in Nonfarm Payrolls*	1149
Prices	*PPI*	578
Prices	*CPI*	578
Retail and Wholesale	*Retail Sales Advance*	181
Trade Balance	*Trade Balance*	181

Several of the announcements above take place on the same day. All events are recurring at different frequencies: while some events as the Initial Jobless Claims happen on a weekly basis, the FOMC announcements only occur a few times a year.

[1] See `https://www.bloomberg.com/markets/economic-calendar` (accessed in July 2021)

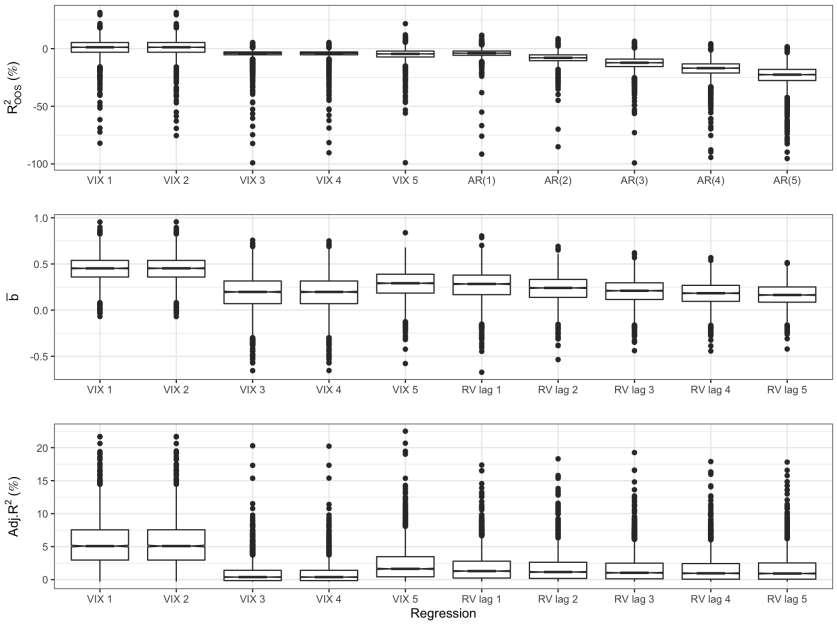
A.1.3 Figures

Figure A.1: Intraday SPY returns Q–Q plot



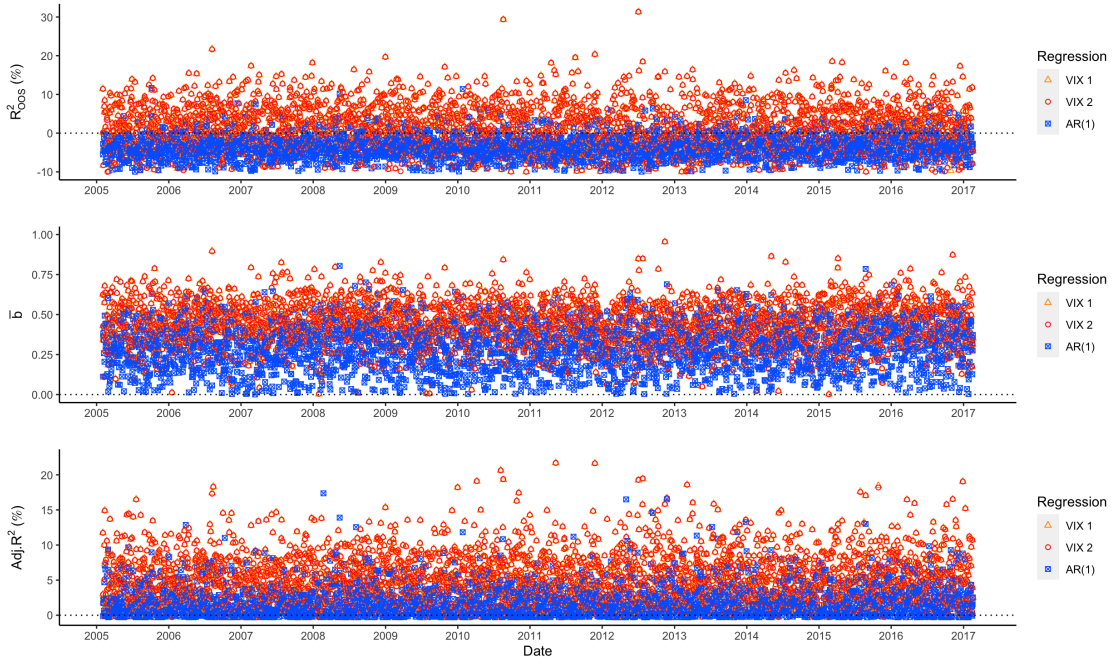
The Q-Q plots above represent the intraday SPY distribution for the entire sample and for the same dates, excluding crisis period between August 1, 2008 and July 31, 2009, as done in Martin (2017). Each figure plots sample data quantiles in sorted order versus quantiles from a standard Normal distribution. Both plots present heavy tails when compared to a normal distribution. This format is often present in log-returns empirical distributions.

Figure A.2: Metrics boxplot for each model



The boxplot displays the distribution of the daily *out-of-sample* R^2_{OOS}, *out-of-sample* fit $\bar{b}$, and adjusted R^2 for each model presented in Section 1.2.3, reporting the median, 25% and 75% percentiles, and potential outliers. The sample period extends from February 2005 to February 2017.

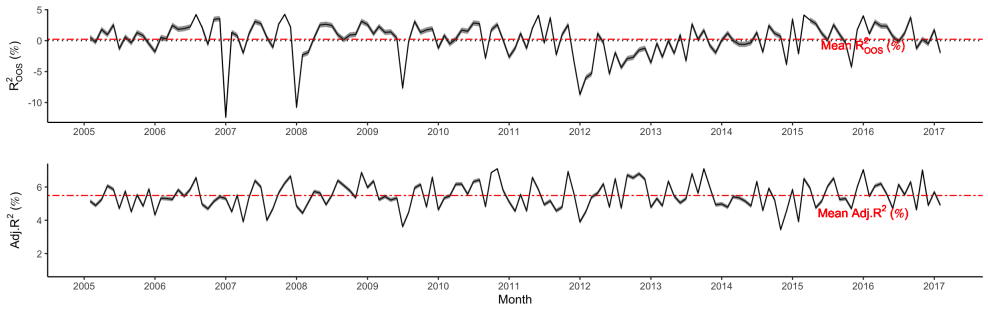
Figure A.3: Daily performance metrics - VIX models *vs* AR(1) benchmark



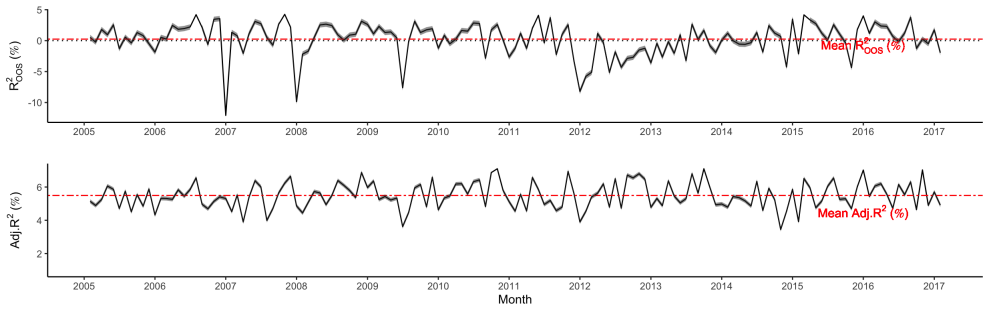
For each day in the sample, the set of figures above report the computed *out-of-sample* R^2_{OOS} (%), and *out-of-sample* fit coefficient $\bar{b}$ and adjusted R^2 (%) for selected models presented in Section 1.2.3. Each figure displays the performance metrics for the VIX 1, VIX 2 models, and AR(1) benchmark. The sample period extends from February 2005 to February 2017.

Figure A.4: Average monthly statistics

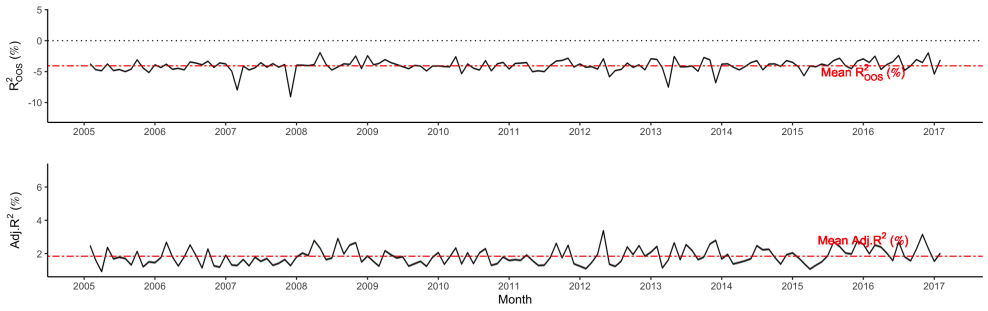
VIX 1 Model



VIX 2 Model

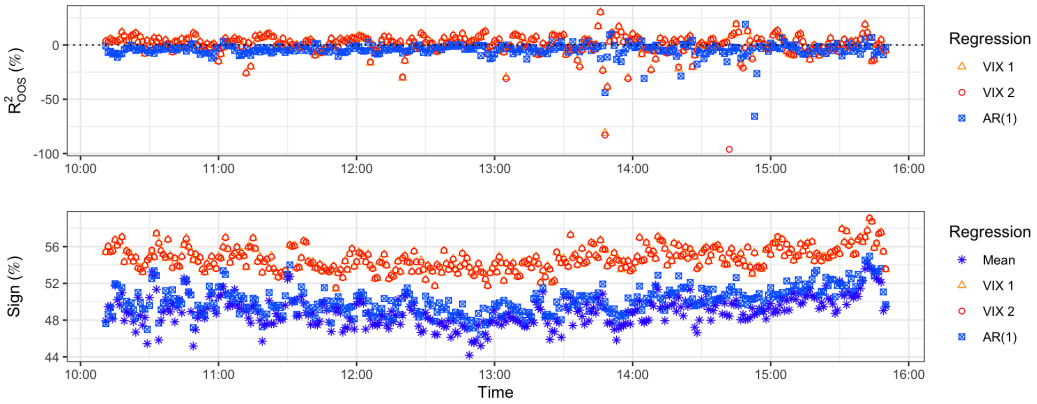


AR(1) Model



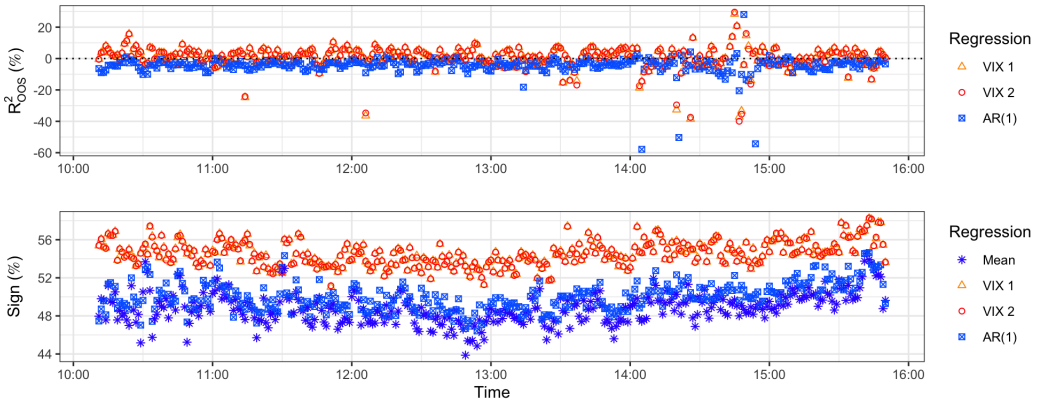
The figures above plot for each model the average *out-of-sample* R^2 and adjusted R^2 each month. These statistics are described in section 1.3. The dashed red line indicate the time-series average, and the grey bands represent the 99.9% confidence interval. The sample period extends from February 2005 to February 2017.

Figure A.5: Minute by minute analysis



The figure on the top reports for each model the *out-of-sample* R^2 for each minute m across all days (in %). The figure on the bottom represents the mean hit rate output, calculated as the percentage of equal signs between observed and predicted return at each minute m. The sample period extends from February 2005 to February 2017.

Figure A.6: Minute by minute analysis - excluding crisis period (Aug 1, 2008 – Jul 31, 2009)



The figure on the top reports for each model the *out-of-sample* R^2 for each minute m across all days (in %). The figure on the bottom represents the mean sign output, calculated as the percentage of equal signs between observed and predicted return at each minute m. The sample period extends from February 2005 to February 2017, excluding days between August 1, 2008 and July 31, 2009, as done in Martin (2017).

A.1.4 Tables

Table A.1: Summary Statistics

Variable	SPY	VIX
Descriptive Statistics (%)		
Mean	0.0002	19.410
Std. Deviation	0.097	9.381
Skewness	0.292	2.485
Kurtosis	41.716	8.153
Correlations (%)		
SPY	100	-0.391
VIX	-0.391	100

The summary statistics and correlations are reported for intraday variables in percentage form for both the minute by minute SPDR S&P 500 ETF Trust (SPY) and the annualized Cboe Volatility Index (VIX) at minute frequency. The sample period extends from February 2005 to February 2017.

Table A.2: Model performance metrics - *Out-of-sample* R^2

Model	R^2_{OOS} (%)	
	Mean	95% CI
VIX 1	0.25	[-0.15 , 0.65]
VIX 2	0.27	[-0.12 , 0.67]
VIX 3	-5.12	[-5.44 , -4.79]
VIX 4	-5.11	[-5.43 , -4.79]
VIX 5	-5.21	[-5.50 , -4.93]
AR(1)	-4.07	[-4.23 , -3.91]
AR(2)	-8.47	[-8.79 , -8.15]
AR(3)	-13.15	[-13.58 , -12.73]
AR(4)	-18.36	[-18.91 , -17.80]
AR(5)	-24.67	[-25.64 , -23.71]

The table reports the average *out-of-sample* R^2_{OOS} (%) described in equation 1-5 for each model across all days. 95 % CI are the 95% confidence intervals for population averages. The top panel of rows represents models presented in Section 1.2.3, and the bottom panel represents the benchmarks. The sample period extends from February 2005 to February 2017.

Table A.3: Model performance metrics - *Out-of-sample* fit

Model	$\bar{b}$		Adj. R^2 (%)		$\tilde{b}$		Adj. $\tilde{R}^2$ (%)	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
VIX 1	0.45	[0.442 , 0.451]	5.50	[5.38 , 5.62]	0.45	[0.448 , 0.457]	5.75	[5.62 , 5.87]
VIX 2	0.45	[0.442 , 0.452]	5.50	[5.38 , 5.62]	0.45	[0.448 , 0.457]	5.75	[5.63 , 5.87]
VIX 3	0.19	[0.179 , 0.193]	0.92	[0.87 , 0.98]	0.23	[0.227 , 0.240]	1.33	[1.27 , 1.40]
VIX 4	0.19	[0.179 , 0.193]	0.92	[0.87 , 0.97]	0.23	[0.227 , 0.240]	1.33	[1.27 , 1.40]
VIX 5	0.28	[0.276 , 0.287]	2.27	[2.18 , 2.35]	0.31	[0.305 , 0.315]	2.77	[2.68 , 2.87]
AR(1)	0.27	[0.262 , 0.274]	1.84	[1.76 , 1.91]	0.30	[0.297 , 0.308]	2.34	[2.25 , 2.42]
AR(2)	0.23	[0.226 , 0.236]	1.75	[1.67 , 1.82]	0.26	[0.255 , 0.265]	2.19	[2.10 , 2.27]
AR(3)	0.20	[0.196 , 0.206]	1.66	[1.59 , 1.74]	0.23	[0.222 , 0.231]	2.05	[1.97 , 2.13]
AR(4)	0.18	[0.173 , 0.183]	1.6	[1.53 , 1.67]	0.20	[0.195 , 0.205]	1.95	[1.87 , 2.03]
AR(5)	0.16	[0.159 , 0.168]	1.63	[1.56 , 1.71]	0.18	[0.178 , 0.186]	1.95	[1.87 , 2.03]

The table reports the average *out-of-sample* coefficients and adjusted R^2 (%), for each model across all days. The first two columns represent the performance metrics for the regression including the intercept (equation 1-6), while the last two columns represent the same outputs in which the intercept is forced to be zero (equation 1-7). 95 % CI are the 95% confidence intervals for population averages, reported in brackets. The top panel of rows represents models presented in Section 1.2.3, and the bottom panel represents the benchmarks. The sample period extends from February 2005 to February 2017.

Table A.4: Summary Statistics - excluding crisis period (Aug 1, 2008 – Jul 31, 2009)

Variable	SPY	VIX
Descriptive Statistics (%)		
Mean	0.0002	17.489
Std. Deviation	0.076	5.770
Skewness	0.082	1.386
Kurtosis	32.565	2.190
Correlations (%)		
SPY	100	-0.728
VIX	-0.728	100

The summary statistics and correlations are reported for intraday variables in percentage form for both the minute by minute SPDR S&P 500 ETF Trust (SPY) and the annualized Cboe Volatility Index (VIX) at minute frequency. The sample period extends from February 2005 to February 2017, excluding days between August 1, 2008 and July 31, 2009, as done in Martin (2017).

Table A.5: Model performance metrics - R^2_{OOS} (%) - excluding crisis period (Aug 1, 2008 – Jul 31, 2009)

Model	R^2_{OOS} (%)	
	Mean	95% CI
VIX 1	0.22	[-0.20 , 0.64]
VIX 2	0.24	[-0.17 , 0.65]
VIX 3	-5.12	[-5.46 , -4.78]
VIX 4	-5.11	[-5.44 , -4.77]
VIX 5	-5.28	[-5.58 , -4.97]
AR(1)	-4.11	[-4.28 , -3.94]
AR(2)	-8.53	[-8.87 , -8.18]
AR(3)	-13.25	[-13.7 , -12.79]
AR(4)	-18.49	[-19.09 , -17.89]
AR(5)	-24.86	[-25.91 , -23.81]

The table reports the average *out-of-sample* R^2_{OOS} (%) described in equation 1-5 for each model across all days. 95 % CI are the 95% confidence intervals for population averages. The first panel of rows represents models presented in Section 1.2.3, while the last panel represents the benchmarks. The sample period extends from February 2005 to February 2017, excluding days between August 1, 2008 and July 31, 2009, as done in Martin (2017).

Table A.6: Model performance metrics - *Out-of-sample* fit - excluding crisis period (Aug 1, 2008 – Jul 31, 2009)

Model	$\overline{b}$		Adj. R^2 (%)		$\tilde{b}$		Adj. $\tilde{R}^2$ (%)	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
VIX 1	0.44	[0.440 , 0.450]	5.49	[5.36 , 5.61]	0.45	[0.446 , 0.456]	5.73	[5.61 , 5.86]
VIX 2	0.44	[0.440 , 0.450]	5.49	[5.36 , 5.61]	0.45	[0.446 , 0.456]	5.74	[5.61 , 5.87]
VIX 3	0.18	[0.176 , 0.190]	0.91	[0.86 , 0.97]	0.23	[0.224 , 0.238]	1.32	[1.25 , 1.39]
VIX 4	0.18	[0.176 , 0.190]	0.91	[0.86 , 0.97]	0.23	[0.224 , 0.238]	1.32	[1.25 , 1.39]
VIX 5	0.28	[0.273 , 0.284]	2.24	[2.15 , 2.33]	0.31	[0.302 , 0.313]	2.74	[2.64 , 2.84]
AR(1)	0.27	[0.260 , 0.272]	1.83	[1.75 , 1.90]	0.30	[0.295 , 0.306]	2.32	[2.23 , 2.41]
AR(2)	0.23	[0.224 , 0.235]	1.74	[1.66 , 1.82]	0.26	[0.253 , 0.264]	2.17	[2.09 , 2.26]
AR(3)	0.20	[0.195 , 0.205]	1.65	[1.57 , 1.73]	0.23	[0.220 , 0.230]	2.04	[1.95 , 2.12]
AR(4)	0.18	[0.172 , 0.182]	1.59	[1.51 , 1.67]	0.20	[0.194 , 0.204]	1.94	[1.85 , 2.02]
AR(5)	0.16	[0.158 , 0.167]	1.62	[1.54 , 1.70]	0.18	[0.177 , 0.186]	1.94	[1.85 , 2.02]

The table reports the average *out-of-sample* coefficients and adjusted R^2 (%), for each model across all days. The first two columns represent the performance metrics for the regression including the intercept (equation 1-6), while the last two columns represent the same outputs in which the intercept is forced to be zero (equation 1-7). 95 % CI are the 95% confidence intervals for population averages, reported in brackets. The first panel of rows describe the VIX models presented in Section 1.2.3, while the last panel represents the benchmarks. The sample period extends from February 2005 to February 2017, excluding days between August 1, 2008 and July 31, 2009, as done in Martin (2017).

Table A.7: Model performance metrics - *Out-of-sample* R^2 - excluding crisis period (Aug 1, 2008 – Jul 31, 2009)

Model	Non Announcement R^2_{OOS} (%)		Announcement R^2_{OOS} (%)	
	Mean	95% CI	Mean	95% CI
VIX 1	0.54	[0.13 , 0.94]	-0.07	[-0.79 , 0.65]
VIX 2	0.54	[0.14 , 0.94]	-0.03	[-0.73 , 0.67]
VIX 3	-4.58	[-4.90 , -4.26]	-5.61	[-6.20 , -5.02]
VIX 4	-4.59	[-4.93 , -4.25]	-5.58	[-6.15 , -5.02]
VIX 5	-5.16	[-5.55 , -4.78]	-5.38	[-5.84 , -4.92]
AR(1)	-3.99	[-4.18 , -3.80]	-4.22	[-4.50 , -3.94]
AR(2)	-8.11	[-8.38 , -7.83]	-8.91	[-9.52 , -8.30]
AR(3)	-12.69	[-13.14 , -12.25]	-13.76	[-14.53 , -12.98]
AR(4)	-17.73	[-18.31 , -17.14]	-19.20	[-20.22 , -18.18]
AR(5)	-23.49	[-24.15 , -22.83]	-26.11	[-28.03 , -24.20]

The table reports the average *out-of-sample* R^2_{OOS} (%) described in equation 1-5 for each model across all days. 95 % CI are the 95% confidence intervals for population averages. The top panel of rows represents models presented in Section 1.2.3, the bottom panel represents the benchmarks. The first two columns represent the statistics for the subsample containing non-announcement dates, and the last two columns report the subsample for macroeconomic announcements days. The sample period extends from February 2005 to February 2017, excluding days between August 1, 2008 and July 31, 2009, as done in Martin (2017).

Table A.8: Summary Statistics

Variable	Non Announcement		Announcement	
	SPY	VIX	SPY	VIX
Descriptive Statistics (%)				
Mean	0.00002	19.474	0.0003	19.351
Std. Deviation	0.091	9.383	0.103	9.378
Skewness	0.298	2.421	0.284	2.543
Kurtosis	18.463	7.659	53.453	8.612
Correlations (%)				
SPY	100	−0.029	100	−0.684
VIX	−0.029	100	−0.684	100

The summary statistics and correlations are reported for intraday variables in percentage form for both the minute by minute SPDR S&P 500 ETF Trust (SPY) and the annualized Cboe Volatility Index (VIX) at minute frequency. The first two columns represent the statistics for the subsample containing non-announcement dates, while the last two columns report the subsample for macroeconomic announcements days. The sample period extends from February 2005 to February 2017.

Table A.9: Model performance metrics - *Out-of-sample* R^2

Model	Non Announcement R^2_{OOS} (%)		Announcement R^2_{OOS} (%)	
	Mean	95% CI	Mean	95% CI
VIX 1	0.49	[0.07 , 0.90]	0.04	[-0.63 , 0.71]
VIX 2	0.49	[0.08 , 0.90]	0.07	[-0.58 , 0.72]
VIX 3	-4.65	[-4.98 , -4.32]	-5.54	[-6.09 , -5.00]
VIX 4	-4.66	[-5.00 , -4.32]	-5.52	[-6.05 , -5.00]
VIX 5	-5.14	[-5.51 , -4.77]	-5.29	[-5.71 , -4.86]
AR(1)	-3.97	[-4.15 , -3.79]	-4.17	[-4.43 , -3.90]
AR(2)	-8.09	[-8.35 , -7.83]	-8.81	[-9.37 , -8.25]
AR(3)	-12.66	[-13.08 , -12.24]	-13.61	[-14.32 , -12.89]
AR(4)	-17.66	[-18.20 , -17.11]	-19.00	[-19.94 , -18.05]
AR(5)	-23.40	[-24.02 , -22.79]	-25.84	[-27.61 , -24.07]

The table reports the average *out-of-sample* R^2_{OOS} (%) described in equation 1-5 for each model across all days. 95 % CI are the 95% confidence intervals for population averages. The top panel of rows represents models presented in Section 1.2.3, the bottom panel represents the benchmarks. The first two columns represent the statistics for the subsample containing non-announcement dates, and the last two columns report the subsample for macroeconomic announcements days. The sample period extends from February 2005 to February 2017.

A.2
Modeling and Forecasting Intraday Market Returns: a Machine Learning Approach

A.2.1
Figures

Figure A.7: Architecture of the Long-Short-Term Memory Cell (LSTM)

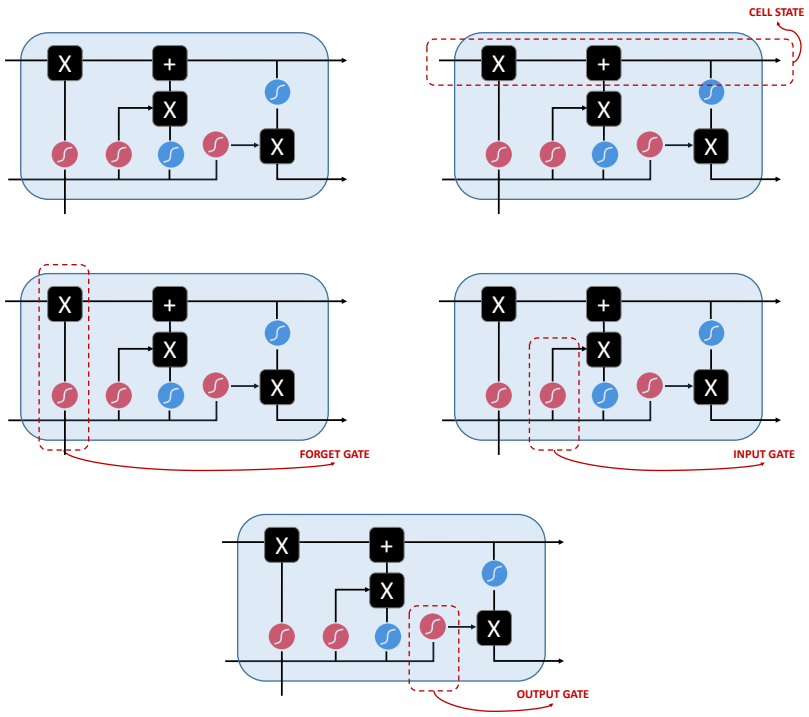


Figure A.8: Information flow in a LTSM Cell

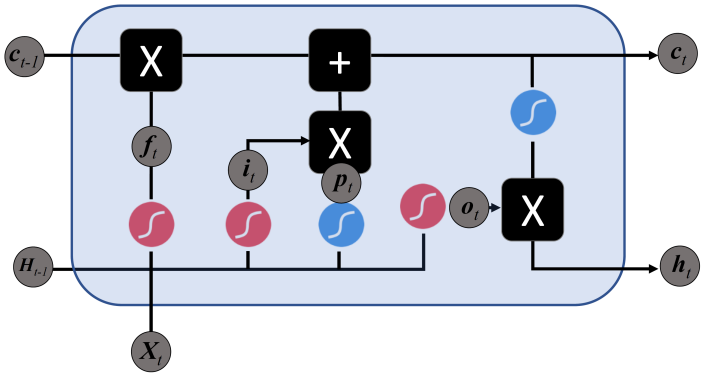
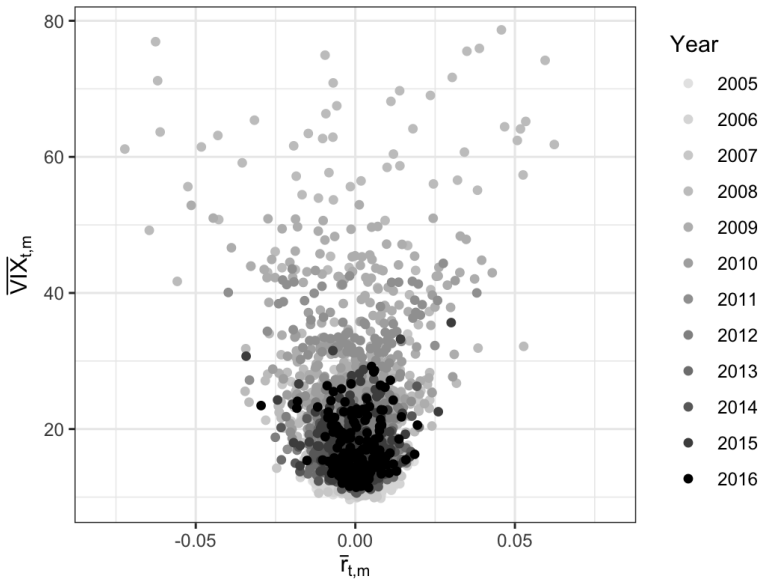


Figure A.9: Average $r_{t,m}$ and $VIX_{t,m}$ at each day t

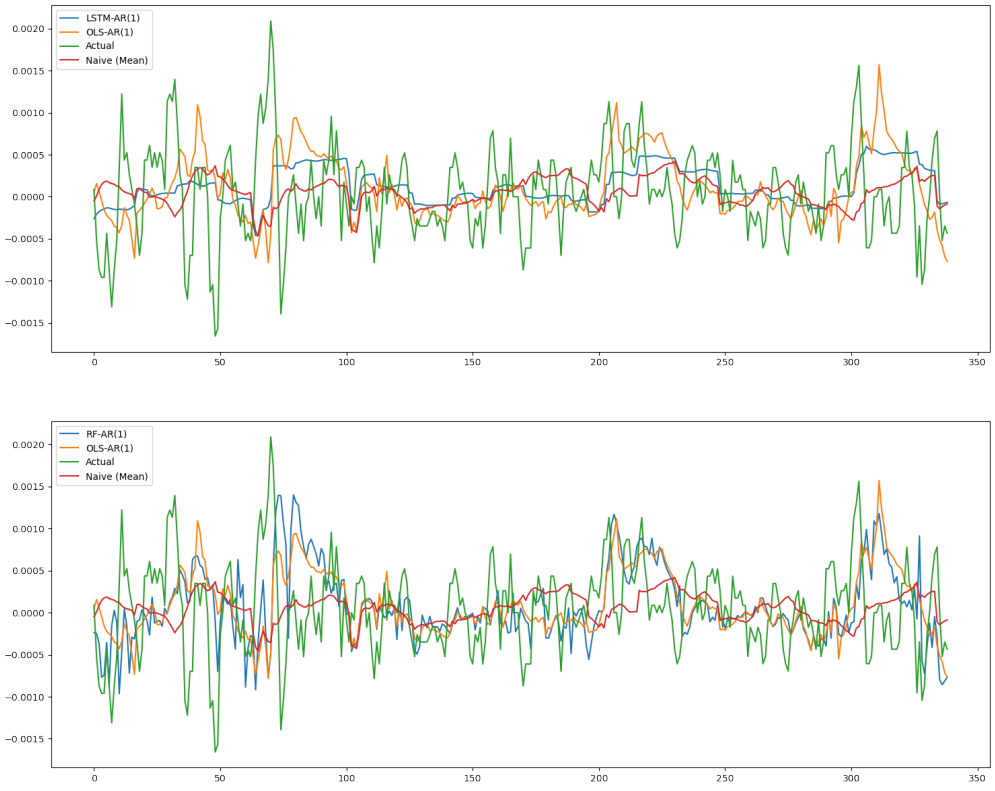


The figure presents the daily average $r_{t,m}$ and $VIX_{t,m}$ computed minute-by-minute within each day. The color scale represents the different years analyzed in the sample. The sample period extends from January 2005 to December 2016.

Figure A.10: Intraday prediction: VIX models (January 8, 2007)

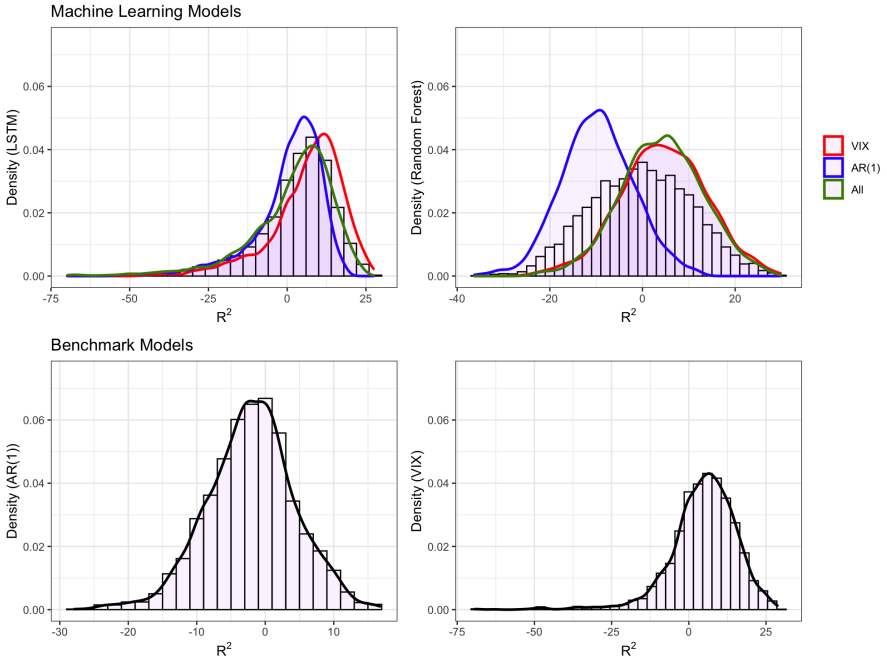
The plots above represent the intraday rolling prediction of minute-by-minute log-returns on the market. As detailed in Section 2.4, we run 340 estimations between 09:40 and 15:50, where the first prediction takes place at 10:11 (after the first 30 minute window). The plots represent the LSTM and Random Forest VIX-based models compared to its respective OLS-VIX benchmark, as well as the naive projection. The sample period extends from January 2005 to December 2016.

Figure A.11: Intraday prediction: AR(1) models (January 8, 2007)



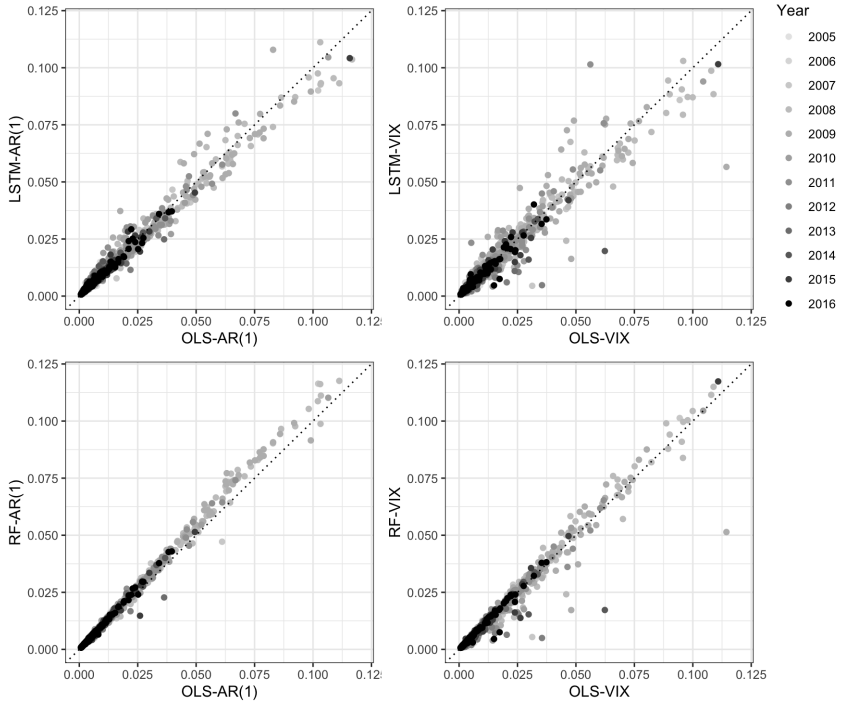
The plots above represent the intraday rolling prediction of minute-by-minute log-returns on the market. As detailed in Section 2.4, we run 340 estimations between 09:40 and 15:50, where the first prediction takes place at 10:11 (after the first 30 minute window). The plots represent the LSTM and Random Forest AR(1)-based models compared to its respective OLS-AR(1) benchmark, as well as the naive projection. The sample period extends from January 2005 to December 2016.

Figure A.12: R^2_{OOS} distribution and density plots



The figure presents the density plots for the daily R^2_{OOS} distribution generated by each model. The first row of plots display machine learning models, divided by the groups of regressors described in Section 2.4.1, while the second row exhibits benchmark OLS models detailed in equations 2-6. The sample period extends from January 2005 to December 2016.

Figure A.13: Daily RMSE: Machine Learning models *vs* Benchmarks



The plots above represent each machine learning model (vertical axis) against its respective benchmark (horizontal axis). The first column of plots displays the OLS-AR(1) model, while the second reports the OLS-VIX model. The rows of plots represent the LSTM and Random Forest models using AR(1) and VIX as single regressors, respectively. The sample period extends from January 2005 to December 2016.

A.2.2 Tables

Table A.10: Summary Statistics

Variable	SPY	VIX
Descriptive Statistics (%)		
Mean	0.0001	19.519
Std. Deviation	0.099	9.404
Skewness	0.168	2.483
Kurtosis	42.886	8.140
Correlations (%)		
SPY	100	−0.432
VIX	−0.432	100

The summary statistics and correlations are reported for intraday variables in percentage form for both the minute by minute SPDR S&P 500 ETF Trust (SPY) and the annualized Cboe Volatility Index (VIX) at minute frequency. The sample period extends from January 2005 to December 2016.

Table A.11: Out-of-sample R^2

R^2_{OOS} (%)	Mean			Median			Std. Deviation		
	VIX	AR(1)	Agg.	VIX	AR(1)	Agg.	VIX	AR(1)	Agg.
LSTM	6.52	0.41	1.68	8.76	2.89	4.51	11.46	11.25	13.45
RF	4.45	-9.73	4.17	4.43	-9.75	4.32	9.27	7.77	8.86
OLS-AR(1)	-	-2.11	-2.11	-	-1.99	-1.99	-	6.46	6.46
OLS-RV	-	-	-14.07	-	-	-7.20	-	-	28.22
OLS-VIX	4.76	-	4.76	5.80	-	5.80	10.81	-	10.81
OLS-ΔVIX	-	-	-3.42	-	-	-2.98	-	-	7.52
OLS-VRP	-	-	-14.11	-	-	-7.20	-	-	28.44

The table reports mean, median, and standard deviation of the *out-of-sample* R^2 (R^2_{OOS}) throughout the sample. The top panel reports the estimation outputs of machine learning models, while the bottom panel displays the ordinary least squares regression benchmark models, detailed in Section 2.4. For each metric, the columns represent the groups of predictors described in Section 2.4.1: VIX, AR(1), and Aggregate. The sample period extends from January 2005 to December 2016.

Table A.12: Root-mean-square error (RMSE)

$RMSE$ ($\times 10^4$)	Mean			Median			Std. Deviation		
	VIX	AR(1)	Agg.	VIX	AR(1)	Agg.	VIX	AR(1)	Agg.
Naive	0.0085	0.0086	0.0086	0.0043	0.0043	0.0043	0.0125	0.0129	0.0127
LSTM	0.0080	0.0086	0.0084	0.0041	0.0044	0.0043	0.0117	0.0127	0.0123
RF	0.0081	0.0095	0.0082	0.0041	0.0048	0.0041	0.0121	0.0143	0.0123
OLS-AR(1)	-	0.0087	0.0087	-	0.0044	0.0044	-	0.0128	0.0128
OLS-RV	-	-	0.0098	-	-	0.0051	-	-	0.0145
OLS-VIX	0.0081	-	0.0081	0.0041	-	0.0041	0.0121	-	0.0121
OLS-ΔVIX	-	-	0.0089	-	-	0.0046	-	-	0.0131
OLS-VRP	-	-	0.0098	-	-	0.0051	-	-	0.0145

The table reports mean, median, and standard deviation of the Root-mean-square error (RMSE) throughout the sample. The top panel reports the estimation outputs of machine learning models, while the bottom panel displays the ordinary least squares regression benchmark models, detailed in Section 2.4. For each metric, the columns represent the groups of predictors described in Section 2.4.1: VIX, AR(1), and Aggregate. The sample period extends from January 2005 to December 2016.

A.3
Factor Augmented High-Dimension Vector Autoregressive Models: Application to a Panel of Realized Volatilities

A.3.1
Tables

Table A.13: Model performance metrics (R^2_{OOS})

1-year estimation window

R^2_{OOS}	First Step	Second Step	Third Step
HAR	23.3%	23.6%	23.7%
HARJ	24.2%	24.2%	24.2%
HARQ	25.0%	25.0%	25.0%
HARQF	24.8%	24.6%	24.7%
HARRS	23.6%	23.9%	23.9%
Vol. Factor OLS	-1316.5%	12.7%	13.3%

2-year estimation window

R^2_{OOS}	First Step	Second Step	Third Step
HAR	18.7%	19.5%	19.5%
HARJ	19.7%	20.1%	20.0%
HARQ	20.6%	21.0%	20.9%
HARQF	20.5%	20.7%	20.7%
HARRS	19.1%	19.8%	19.7%
Vol. Factor OLS	-1410.2%	14.9%	14.8%

4-year estimation window

R^2_{OOS}	First Step	Second Step	Third Step
HAR	22.0%	22.9%	22.8%
HARJ	22.9%	23.5%	23.5%
HARQ	23.5%	24.1%	24.1%
HARQF	23.7%	24.1%	24.1%
HARRS	22.4%	23.2%	23.2%
Vol. Factor OLS	-1352.8%	21.0%	19.3%

5-year estimation window

R^2_{OOS}	First Step	Second Step	Third Step
HAR	24.7%	25.5%	25.5%
HARJ	25.6%	26.1%	26.1%
HARQ	26.0%	26.6%	26.5%
HARQF	26.2%	26.6%	26.6%
HARRS	25.1%	25.8%	25.7%
Vol. Factor OLS	-1303.9%	24.4%	22.9%

The tables above report the *out-of-sample* R^2 (R^2_{OOS}) for each model using different estimation windows. In each panel, the columns represent the chosen model outputs computed in the first, second, and third estimation stages, including the heterogeneous autoregressive (HAR) model, and its extensions HAR-J, HAR-RS, HAR-Q and HARQ-F, as well as the standard OLS volatility factors approach. The sample period extends from January 2006 to February 2017. We use estimation windows of 252 ($\approx$ 1 year), 504 ($\approx$ 2 years), 1008 ($\approx$ 4 years), and 1260 business days ($\approx$ 5 years). For comparison purposes, all metrics are computed using the same *out-of-sample* range as the of 5-year estimates.

Table A.14: Model performance metrics (MAE)

1-year estimation window

MAE	First Step	Second Step	Third Step
HAR	0.279	0.2766	0.2765
HARJ	0.2728	0.2727	0.2726
HARQ	0.2669	0.267	0.2667
HARQF	0.2685	0.2696	0.2692
HARRS	0.277	0.275	0.2749
Vol. Factor OLS	95.2189	0.359	0.3551

2-year estimation window

MAE	First Step	Second Step	Third Step
HAR	0.2759	0.2707	0.2709
HARJ	0.2695	0.2669	0.2671
HARQ	0.2636	0.2611	0.2612
HARQF	0.2642	0.2626	0.2625
HARRS	0.2732	0.269	0.2692
Vol. Factor OLS	94.7516	0.3026	0.3034

4-year estimation window

MAE	First Step	Second Step	Third Step
HAR	0.2751	0.2686	0.269
HARJ	0.2683	0.2644	0.2646
HARQ	0.2646	0.2601	0.2604
HARQF	0.2633	0.2602	0.2603
HARRS	0.272	0.2665	0.2668
Vol. Factor OLS	94.6134	0.2829	0.295

5-year estimation window

MAE	First Step	Second Step	Third Step
HAR	0.2746	0.2687	0.2691
HARJ	0.268	0.2644	0.2647
HARQ	0.2655	0.2611	0.2615
HARQF	0.2636	0.2607	0.261
HARRS	0.2716	0.2667	0.267
Vol. Factor OLS	94.5559	0.2771	0.2886

The tables above report the mean absolute error (MAE) in the right panel for each model using different estimation windows. In each panel, the columns represent the chosen model outputs computed in the first, second, and third estimation stages, including the heterogeneous autoregressive (HAR) model, and its extensions HAR-J, HAR-RS, HAR-Q and HARQ-F, as well as the standard OLS volatility factors approach. The sample period extends from January 2006 to February 2017. We use estimation windows of 252 ($\approx$ 1 year), 504 ($\approx$ 2 years), 1008 ($\approx$ 4 years), and 1260 business days ($\approx$ 5 years). For comparison purposes, all metrics are computed using the same *out-of-sample* range as the of 5-year estimates.

Bibliography

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue Ratio Test for the Number of Factors. *Econometrica*, 81(3):1203–1227.

Ait-Sahalia, Y. and Jacod, J. (2014). *High-Frequency Financial Econometrics*. Princeton University Press, 1 edition.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *The Review of Economics and Statistics*, 89(4):701–720.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71:2135–171.

Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71:135–171.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor augmented regressions. *Econometrica*, 74:1133–1155.

Barndorff-Nielsen, O. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37.

Barndorff-Nielsen, O. E., Kinnebrock, S., and Shephard, N. (2008). Measuring downside risk-realised semivariance. Economics Papers 2008-W02, Economics Group, Nuffield College, University of Oxford.

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

Bekaert, G. and Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, 183(2):181–192.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2004). Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. NBER Working Papers 10220, National Bureau of Economic Research, Inc.

Bollerslev, T., Hood, B., Huss, J., and Pedersen, L. H. (2018). Risk Everywhere: Modeling and Managing Volatility. *Review of Financial Studies*, 31(7):2729–2773.

Bollerslev, T., Li, S. Z., and Zhao, B. (2020). Good Volatility, Bad Volatility, and the Cross Section of Stock Returns. *Journal of Financial and Quantitative Analysis*, 55(3):751–781.

Bollerslev, T., Patton, A., and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1):1–18.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected Stock Returns and Variance Risk Premia. *Review of Financial Studies*, 22(11):4463–4492.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Brito, D., Medeiros, M. C., and Ribeiro, R. (2018). Forecasting large realized covariance matrices: The benefits of factor models and shrinkage. Working Paper 3163668, SSRN.

Callot, L. A. F., Kock, A. B., and Medeiros, M. C. (2017). Modeling and Forecasting Large Realized Covariance Matrices and Portfolio Choice. *Journal of Applied Econometrics*, 32(1):140–158.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies*, 21(4):1509–1531.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–304.

Chen, L., Pelger, M., and Zhu, J. (2019). Deep learning in asset pricing. Technical Report 3350138, SSRN.

Chinco, A., Clark-Joseph, A. D., and Ye, M. (2019). Sparse Signals in the Cross-Section of Returns. *Journal of Finance*, 74(1):449–492.

Cochrane, J. H. (2011). Presidential Address: Discount Rates. *Journal of Finance*, 66(4):1047–1108.

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

Fan, J., Masini, R., and C. Medeiros, M. (2021). Bridging factor and sparse models. Working Paper 3789141, SSRN.

Fan, J., Xue, L., and Yao, J. (2017). Sufficient forecasting using factor models. *Journal of Econometrics*, 201(2):292–306.

Faust, J. and Wright, J. H. (2018). Risk Premia in the 8:30 Economy. *Quarterly Journal of Finance (QJF)*, 8(03):1–19.

Feng, G., Giglio, S., and Xiu, D. (2020). Taming the Factor Zoo: A Test of New Factors. *Journal of Finance*, 75(3):1327–1370.

Fernandes, M., Medeiros, M., and Scharth, M. (2014). Modeling and predicting the CBOE market volatility index. *Journal of Banking & Finance*, 40(C):1–10.

Garcia, M. G., Medeiros, M., and Vasconcelos, G. F. (2017). Real-time inflation forecasting with high-dimensional models: The case of brazil. *International Journal of Forecasting*, 33(3):679–693.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9:1735–1780.

Liu, L., Patton, A., and Sheppard, K. (2015). Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187(1):293–311.

Lucca, D. and Moench, E. (2015). The pre-fomc announcement drift. *Journal of Finance*, 70(1):329–371.

Martin, I. (2011). Simple Variance Swaps. NBER Working Papers 16884, National Bureau of Economic Research, Inc.

Martin, I. (2017). What is the expected return on the market? *The Quarterly Journal of Economics*, 132(1):367–433.

Martin, I. W. R. and Wagner, C. (2019). What Is the Expected Return on a Stock? *Journal of Finance*, 74(4):1887–1929.

Masini, R. P., Medeiros, M. C., and Mendes, E. F. (2021). Machine learning advances for time series forecasting.

McAleer, M. and Medeiros, M. C. (2008). Realized volatility: A review. *Econometric Reviews*, 27(1-3):10–45.

Medeiros, M. C. and Mendes, E. F. (2012). Estimating high-dimensional time series models. Textos para discussão 602, Department of Economics PUC-Rio (Brazil).

Medeiros, M. C., Vasconcelos, G., Veiga, A., and Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business and Economic Statistics*, 39:98–119.

Moran, M. T. and Liu, B. (2020). The VIX Index and Volatility-Based Global Indexes and Trading Instruments - A Guide to Investment and Trading Features. CFA Institute Research Foundation Briefs 978-1-944960-95-7, CFA Institute.

Patton, A. J. and Sheppard, K. (2015). Good Volatility, Bad Volatility: Signed Jumps and The Persistence of Volatility. *The Review of Economics and Statistics*, 97(3):683–697.

Stock, J. and Watson, M. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20:147–162.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.