Realized semicovariances: Empirical applications to volatility forecasting and portfolio optimization

Rafael Ricco[†]
Flavio A. Ziegelmann[‡]

Abstract We propose a two-fold empirical study applying the concept of realized semicovariances as introduced by Bollerslev et al. (2020): in the first part of the paper we aim to estimate and forecast the realized volatility of an equally weighted portfolio formed by Brazilian B3 asset returns, whereas in the second part we search and find an optimum portfolio for these returns. In both parts we use high frequency data of ten assets from different segments and among the most negotiated in B3 financial market from July 2018 to January 2021. In addition, we investigate whether a Markov Switching strategy fits well to our volatility modeling approach considering that our observed data starts some time before the Covid-19 pandemic and spans well into the pandemic period. Machine Learning Regularization (LASSO) methods are employed to select covariates and potentially improve volatility estimation and forecasting. In the portfolio optimization analysis we see that under higher frequency rebalancing periods, minimum variance portfolios using the negative semicovariance matrices present better performances in terms of risk-adjusted returns compared to those that use the standard realized covariance matrices. In general we see that the realized semicovariances bring improvements to the solutions of our two problems.

Keywords: High-frequency data; Volatility forecasting; Realized semicovariances; Portfolio optimization; Markov switching; LASSO; Economic performance.
JEL Code: C32, C53, G11, C58.

1. Introduction

Realized variance, the most commonly used realized measure, is constructed by summing up squared intra-daily returns. There is an extensive statistical theory for this subject derived in papers by Barndorff-Nielsen and Shephard (2002), Meddahi (2002), Andersen et al. (2003), and Mykland and Zhang (2009), among others.

This paper aims to explore the empirical contribution of realized semicovariance measures (described in Bollerslev et al. (2020)) in the Brazilian

Submitted on April 7, 2022. Revised on January 22, 2023. Accepted on March 22, 2023. Published online in September 2023. Editor in charge: Juan Carlos Arismendi Zambrano.

[†]Graduate Program in Economics, Universidade Federal do Rio Grande, Brazil: rafaricco@gmail.com
[‡]Department of Statistics, Universidade Federal do Rio Grande do Sul, Brazil: flavioaz@mat.ufrgs.br

financial market for two problems: i) realized volatility (RV) estimation and forecasting, and ii) portfolio optimization (PO). Our contribution is on the empirical side, showing that these measures can help achieve better results in both statistical (for RV) and economic (for PO) terms. Regarding RV, our most complex approach uses a Markov-Switching Semi-Covariance Heterogenous Autorregressive (MS-SCHAR) model, where a LASSO regularization is applied to select relevant variables (in a pre/in-pandemic data split). We use the popular HAR model of Corsi (2009) as the benchmark for comparison. In terms of portfolio optimization, our benchmark is the minimum variance optimal portfolio obtained with the traditional realized covariance matrix as a risk measure.

In the context of realized measures, Hansen et al. (2012) introduce the Realized GARCH, a GARCH model that incorporates realized measures as covariates. Patton and Sheppard (2015), through an empirical framework, show that future volatility is more strongly related to the volatility of past negative returns than to that of positive returns and that the impact of a price jump on volatility depends on the sign of the jump, with negative (positive) jumps leading to higher (lower) future volatility.

According to Bollerslev et al. (2020), in the multivariate context, a rapidly-growing recent literature has forcefully advocated for the use of high-frequency intra-daily data to more reliably estimate lower-frequency return covariance matrices as in Andersen et al. (2003); Barndorff-Nielsen and Shephard (2004); Barndorff-Nielsen et al. (2011).

Noureldin et al. (2012) propose the HEAVY models, a new class of multivariate volatility models that utilizes high-frequency data. Their empirical results suggest that the HEAVY model outperforms the multivariate GARCH model in an out-of-sample analysis, with the gains being particularly significant at shorter forecast horizons. Borges et al. (2015) show that covariance matrices based on higher frequency data lead to better performance indicators for a Brazilian stock portfolio. Aït-Sahalia and Xiu (2016) analyze that the crisis period of 2007-2010 did indeed result in an increase in quadratic variation in all the assets considered by them; however, it did not cause a significant change in the breakdown between their respective Brownian and jump contributions, with both moving consistently with one another. Bollerslev et al. (2018) depict through an empirical framework the relationship between volume, volatility, and public news announcements.

Under the assumption that investors care more about loss than gain, Barndorff-Nielsen et al. (2010) introduce the realized semivariance, a measure of risk that takes the return sign into account, and motivate Bollerslev et al. (2020) to propose a decomposition of the realized covariance matrix into three realized semicovariance matrix components dictated by the signs of the underlying high-frequency returns. According to the authors, the realized semicovariance matrices may be seen as a high-frequency multivariate extension of the semivariances originally proposed in Markowitz (1952); Mao (1970); Hogan and Warren (1972); Fishburn (1977).

Using high-frequency data for a large cross-section of U.S. equities, Bollerslev et al. (2020) find that models that incorporate the realized semicovariance measures have superior forecast performance than models that employ the realized semivariance measures or just the realized covariance matrix.

Çelik and Ergin (2014), using ISE-30 index futures data, find the superiority of high frequency based volatility forecasting models over traditional GARCH models. Given the different nature of volatility in emerging markets, as outlined in Aggarwal et al. (1999), and the challenges imposed on such markets due to idiosyncratic factors, studied by Bekaert and Harvey (2003), such as segmentation, capital flows, political risk, among others, realized semicovariance measures might bring extra relevant information to the game, particularly to improve the capability of realized volatility predicting and portfolio optimization in the Brazilian stock market.

Set against this background and given the importance of the covariance matrix of asset returns for portfolio management, based on Bollerslev et al. (2020), we attack a two-fold problem via realized semicovariance measures: Realized Volatility Forecasting and Portfolio Optimization. Our dataset is composed of the most traded assets (spread a priori over 6 different sectors) in the B3 Brazilian stock market from July 2018 to January 2021. Furthermore, since our data includes the Covid-19 pandemic, we also consider a Markov Switching modeling approach [1]. Machine Learning Regularization (LASSO) methods are employed to select covariates and potentially improve volatility estimation and forecasting. Our main findings suggest that i) the realized semicovariances help to better explain and forecast the realized portfolio returns variance; ii) under different regimes, the relation among these realized measures can change; iii) using LASSO, we see that including all realized semicovariances within a HAR Model can lead to model "overfitting"; and iv) under higher frequency rebalancing periods, using the realized semicovariances brings improvements to the minimum variance portfolio performance.

We must stress that there are many other possible approaches to be taken for forecasting realized volatility and finding optimal portfolios. Under the popularity of various recent machine learning methods, Kristjanpoller et al.

[1] For alternative approaches under stress market periods, see, for instance, Emre Alper et al. (2012), which use MIDAS modeling.

(2014) and Kim and Won (2018), for instance, employ hybrid Neural Network GARCH models for volatility forecasting. Nevertheless, their datasets are at a lower daily frequency, whereas ours are high frequency data. We emphasize that our goal is limited to evaluating the contribution of semicovariance realized measures, hence we try to isolate this contribution by comparing our proposal to similar models that do not use these measures.

Beyond this introduction, this work has six more sections. In section 2, we describe the variance components of asset returns in the univariate and multivariate contexts. In section 3, we describe the statistical models used in the present work to analyze the relationship between the portfolio realized variance and its components, whereas in section 4, we apply these models to evaluate how much the realized semicovariances help to explain and predict an equally weighted stock portfolio variance in the Brazilian financial market. We present the portfolio optimization theoretical framework in section 5. Section 6 brings the application of the realized measures presented in Bollerslev et al. (2020) to evaluate if their use brings improvements in the economic performance for stock portfolios. Finally, we make concluding remarks in section 7.

2. Realized Volatility Measures

Building upon the work of Barndorff-Nielsen et al. (2010), Bollerslev et al. (2020) expanded their research into the multivariate context. Let $\boldsymbol{X}_t = (X_{1,t},\dots,X_{d,t})^\top$ denote a d-dimensional log-price process, sampled on a regular time grid $0 = t_0 < t_1 < \cdots < t_n = T$ over a fixed time span $T > 0$. We denote the ith return as $\boldsymbol{\Delta}_i\boldsymbol{X} = \boldsymbol{X}_{t_i} - \boldsymbol{X}_{t_{i-1}}$. The realized covariance matrix is defined as:

$$\widehat{RC} = \sum_{i=1}^{n} (\boldsymbol{\Delta}_i\boldsymbol{X})(\boldsymbol{\Delta}_i\boldsymbol{X})^\top. \tag{1}$$

We use $[\cdot]^+ = \max\{x,0\}$ and $[\cdot]^- = \min\{x,0\}$ to denote the componentwise positive and negative elements of the real vector x. The corresponding "positive", "negative", and "mixed" realized semicovariance matrices are simply defined as:

$$\begin{aligned}
\widehat{RSC}_{positive} &= \sum_{i=1}^{n} [\boldsymbol{\Delta}_i\boldsymbol{X}]^+([\boldsymbol{\Delta}_i\boldsymbol{X}]^+)^\top, \\
\widehat{RSC}_{negative} &= \sum_{i=1}^{n} [\boldsymbol{\Delta}_i\boldsymbol{X}]^-([\boldsymbol{\Delta}_i\boldsymbol{X}]^-)^\top, \\
\widehat{RSC}_{mixed} &= \sum_{i=1}^{n} \left([\boldsymbol{\Delta}_i\boldsymbol{X}]^+([\boldsymbol{\Delta}_i\boldsymbol{X}]^-)^\top + [\boldsymbol{\Delta}_i\boldsymbol{X}]^-([\boldsymbol{\Delta}_i\boldsymbol{X}]^+)^\top\right).
\end{aligned} \tag{2}$$

Note that $\widehat{RC} = \widehat{RSC}_{positive} + \widehat{RSC}_{negative} + \widehat{RSC}_{mixed}$. Also, $\widehat{RSC}_{positive}$ and $\widehat{RSC}_{negative}$ are defined as sums of vector outer-products and are thus positive semidefinite, whereas $\widehat{RSC}_{mixed}$ is indefinite.

Motivated by empirical observations from the returns of each of the 30 Dow Jones Industrial Average (DJIA) stocks on two different days, Bollerslev et al. (2020) noted that estimates of $\widehat{RSC}_{positive}$ and $\widehat{RSC}_{negative}$ can diverge in response to the content of the news or event. The theoretical framework that illustrates the distinction between the information carried by the positive semicovariance and the negative semicovariance matrices can be found in Bollerslev et al. (2020).

3. Volatility forecasting

Various models have been developed to forecast the realized variance of a portfolio, and we discuss some of these below. The realized variance of a portfolio with weights $\boldsymbol{w}$ can be expressed as

$$\widehat{RV}^{port} = \boldsymbol{w}^{\top}\widehat{RC}\boldsymbol{w}, \tag{3}$$

where $\widehat{RC}$ is defined in (1). Since $\widehat{RC} = \widehat{RSC}_{positive} + \widehat{RSC}_{negative} + \widehat{RSC}_{mixed}$, we have

$$\begin{aligned} \widehat{RV}^{port} &= \boldsymbol{w}^{\top}\widehat{RC}\boldsymbol{w} \\ &= \boldsymbol{w}^{\top}\widehat{RSC}_{positive}\boldsymbol{w} + \boldsymbol{w}^{\top}\widehat{RSC}_{negative}\boldsymbol{w} + \boldsymbol{w}^{\top}\widehat{RSC}_{mixed}\boldsymbol{w} \\ &= \widehat{P}^{port} + \widehat{N}^{port} + \widehat{M}^{port}. \end{aligned} \tag{4}$$

A widely-used model for forecasting realized variance is the HAR model of Corsi (2009). It is defined by

$$\widehat{RV}^{port}_{t+1|t} = \alpha_0 + \alpha_d \widehat{RV}^{port}_{t} + \alpha_w \widehat{RV}^{port}_{t-1:t-4} + \alpha_m \widehat{RV}^{port}_{t-5:t-21}, \tag{5}$$

where $\widehat{RV}^{port}_{t-l:t-k} = \frac{1}{k-l+1}\sum_{s=l}^{k}\widehat{RV}^{port}_{t-s}$. As in Bollerslev et al. (2020), we consider it as our benchmark model for evaluating forecast performance.

Alongside the HAR model, we borrow the concept from Patton and Sheppard (2015) and consider a HAR extension, the Semivariance HAR (SHAR), which adds semivariances to the set of explanatory variables:

$$\begin{aligned} \widehat{PSV} &= \sum_{i=1}^{n}[\boldsymbol{w}^{\top}[\boldsymbol{\Delta}_i\boldsymbol{X}]^{+}]^2 \\ \widehat{NSV} &= \sum_{i=1}^{n}[\boldsymbol{w}^{\top}[\boldsymbol{\Delta}_i\boldsymbol{X}]^{-}]^2, \end{aligned}$$

where $\boldsymbol{w}$ represents the portfolio weights, $\boldsymbol{\Delta}_i\boldsymbol{X} = \boldsymbol{X}_{t_i} - \boldsymbol{X}_{t_{i-1}}$ is the i-th intradaily return of a d-dimensional log-price process $\boldsymbol{X}_t$, and $[\cdot]^+ = \max\{x,0\}$ and $[\cdot]^- = \min\{x,0\}$ represent the component-wise positive and negative elements of the real vector x. The forecasting scheme results in

$$\widehat{RV}^{port}_{t+1|t} = \alpha_0 + \alpha_{d,p}\widehat{PSV}_t + \alpha_{d,n}\widehat{NSV}_t + \alpha_w\widehat{RV}^{port}_{t-1:t-4} + \alpha_m\widehat{RV}^{port}_{t-5:t-21}. \quad (6)$$

This model allows us to examine whether semivariances provide additional information when forecasting the portfolio realized variance.

Among the HAR extensions, following Bollerslev et al. (2020), we consider another model, the SemiCovariance HAR (SCHAR). This model includes the semicovariance components shown in (4). The one-step-ahead forecast for the portfolio realized variance is

$$\begin{aligned}\widehat{RV}^{port}_{t+1|t} = \alpha_0 &+ \alpha_{d,p}\widehat{P}^{port}_t + \alpha_{w,p}\widehat{P}^{port}_{t-1:t-4} + \alpha_{m,p}\widehat{P}^{port}_{t-5:t-21} \\ &+ \alpha_{d,n}\widehat{N}^{port}_t + \alpha_{w,n}\widehat{N}^{port}_{t-1:t-4} + \alpha_{m,n}\widehat{N}^{port}_{t-5:t-21} \\ &+ \alpha_{d,m}\widehat{M}^{port}_t + \alpha_{w,m}\widehat{M}^{port}_{t-1:t-4} + \alpha_{m,m}\widehat{M}^{port}_{t-5:t-21}.\end{aligned} \quad (7)$$

In order to select the best predictors while avoiding "overfitting", we apply the Least Absolute Shrinkage and Selection Operator (LASSO) proposed by Tibshirani (1996) to (7), resulting in what we refer to as SCHAR-lasso-in.

Our LASSO estimator, considering α_0 as pre-estimated and $\widehat{RV}^{port}_t$ corrected by its mean, is given by the solution of

$$\widehat{\boldsymbol{\alpha}} = \arg\min\left[RSS(\alpha_1,\ldots,\alpha_9) + \lambda\sum_{j=1}^{9}|\alpha_j|\right], \quad (8)$$

where $\boldsymbol{\alpha} = (\alpha_1,...,\alpha_9) = (\alpha_{d,p},\ldots,\alpha_{m.m})$, $RSS(\alpha_1,...,\alpha_9) = \sum_{t=22}^{T}(\widehat{RV}^{port}_t - \alpha_1\widehat{P}^{port}_{t-1} - \cdots - \alpha_9\widehat{M}^{port}_{t-6:t-22})^2$.

The tuning parameter λ is typically chosen using data-driven techniques, such as cross-validation. In this work, we use a 10-fold cross-validation method. Alternatively, information criteria can also be used to select this parameter.

As the final model specification, we introduce a two-regime Markov-Switching (MS) SCHAR model. The state variable S_t in this model is an unobservable Markov chain. The estimation process is based on the Kim filter, as described in the work of Kim and Nelson (1999). Given that the probability distribution of the realized volatility deviates substantially from the normal distribution, we estimate the coefficients using the quasi-maximum like-

Table 1
Portfolio's stocks

Company's Name	Ticker Symbol	Sector
Ambev S.A.	ABEV3	Consumer Staples
B3	B3SA3	Financials
Bradesco S.A.	BBDC4	Financials
Intermedica S.A.	GNDI3	Health Care
Itaú S.A.	ITUB4	Financials
JBS S.A.	JBSS3	Consumer Staples
Magazine Luiza S.A.	MGLU3	Consumer Discretionary/IT
Petrobras S.A.	PETR4	Oil & Gas
Suzano S.A.	SUZB3	Industrials
Vale S.A.	VALE3	Industrial Materials

Stocks included in the portfolio. IT stands for Information Technology.

lihood estimation method, which was introduced by Lindsay (1988). Mathematically, this is expressed as:

$$\ell = \sum_{t=1}^{T} \log \left(\sum_{S_t=0}^{1} f\left(\widehat{RV}_t^{port} | S_t, \psi_{t-1}\right) Pr[S_t | \psi_{t-1}] \right). \tag{9}$$

where

$$f\left(\widehat{RV}_t^{port} | S_t, \psi_{t-1}\right) = \frac{1}{\sqrt{2\pi\sigma_{S_t}^2}} e^{-\frac{\left(\widehat{RV}_t^{port} - \alpha_{0,S_t} - \alpha_{d,p,S_t}\widehat{P}_{t-1}^{port} - \cdots - \alpha_{m,m,S_t}\widehat{M}_{t-6:t-22}^{port}\right)^2}{2\sigma_{S_t}^2}}.$$

Because the regimes S_t are unobservable, to evaluate the log-likelihood in (9), we need to calculate the weights $Pr[S_t|\psi_{t-1}]$ for $S_t = 0$ and $S_t = 1$ (the two states). We carry out this step using the Kim filter, as explained in Kim and Nelson (1999).

4. Empirical exercise 1

In this section, we forecast the realized volatility of an equally weighted stock portfolio using the models described in Section 3. Our portfolio consists of 10 stocks from the BOVESPA index, as shown in Table 1. These stocks were chosen for their high trading volumes and to ensure heterogeneity across different sectors of the economy.

Our sample period spans from July 2018 to January 2021, a total of 624 trading days.

Table 2
Descriptive statistics for intra-daily returns

Statistic	N	Mean	St. Dev.	Min	Max
ABEV3	51,486	0.00000	0.003	−0.105	0.098
B3SA3	51,486	0.00003	0.003	−0.195	0.171
BBDC4	51,486	0.00001	0.003	−0.118	0.151
GNDI3	51,486	0.00004	0.004	−0.226	0.179
ITUB4	51,486	0.00001	0.003	−0.107	0.124
JBSS3	51,486	0.00003	0.004	−0.235	0.185
MGLU3	51,486	0.00004	0.004	−0.262	0.197
PETR4	51,486	0.00002	0.004	−0.245	0.170
SUZB3	51,486	0.00001	0.003	−0.122	0.098
VALE3	51,486	0.00002	0.003	−0.181	0.233

Descriptive statistics for intra-daily returns.

Table 3
Descriptive statistics for portfolio realized measures

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
$\widehat{RV}^{port}$	624	0.0002	0.0005	0.00002	0.0001	0.0001	0.005
$\widehat{P}^{port}$	624	0.0001	0.0003	0.00002	0.0001	0.0001	0.003
$\widehat{N}^{port}$	624	0.0001	0.0003	0.00001	0.0001	0.0001	0.004
$\widehat{M}^{port}$	624	−0.0001	0.0001	−0.001	−0.0001	−0.00005	−0.00002

We construct the realized measures using 5-minute intra-daily returns[2], excluding the overnight returns. Table 2 presents basic summary statistics for the intra-daily returns of the stocks, while Table 3 provides descriptive statistics for $\widehat{RV}^{port}$,$\widehat{P}^{port}$,$\widehat{N}^{port}$,$\widehat{M}^{port}$.

4.1 In-sample analysis

To conduct the in-sample analysis, we first estimate three variations of HAR Models discussed previously: the HAR, SHAR, and SCHAR.

Table 4 presents the estimation results for each of these models. We observe that all coefficients are statistically significant at the 5% level in the first model. In the second model, we find that the daily positive semivariance ($\widehat{PSV}_{t-1}$) and both weekly and monthly realized covariances ($\widehat{RV}^{port}_{t-2:t-5}$, $\widehat{RV}^{port}_{t-6:t-22}$) are the main contributors to the realized portfolio variance ($\widehat{RV}_t$).

[2]The dataset can be obtained from https://github.com/rricco/realvol/tree/master/data

Lastly, in the third model, we note that all coefficients except the monthly negative semicovariance and monthly positive semicovariance ($\widehat{N}^{port}_{t-6:t-22}$ and $\widehat{P}^{port}_{t-6:t-22}$), are statistically significant at the 5% level. This contrasts with the analysis in Bollerslev et al. (2020), where all negative semicovariances are significant at the 5% level. Looking at the information criteria, both AIC and BIC measures suggest that the inclusion of realized semivariance and realized semicovariance components improves the explanation of the realized portfolio variance. In particular, SCHAR is identified as the best model according to these criteria.

Additionally, Table 5 displays the coefficients selected by the Least Absolute Shrinkage and Selection Operator (LASSO) from the SCHAR Model. As shown, the shrinkage method identifies four covariates ($\widehat{P}^{port}_{t-1}$, $\widehat{P}^{port}_{t-2:t-5}$, $\widehat{M}^{port}_{t-1}$, $\widehat{M}^{port}_{t-6:t-22}$) as the best predictors of the portfolio realized variance.

4.2 In-sample analysis under different regimes

Our sample spans from 2 July 2018 to 8 January 2021, a period that includes the COVID-19 pandemic, which significantly increased financial market volatility. To account for this, we divide our dataset into a pre-pandemic period (from 2 July 2018 to 28 February 2020) with 388 observations, and a post-pandemic period (from 2 March 2020 to 8 January 2021) with 214 observations. We then apply the Least Absolute Shrinkage and Selection Operator (LASSO) to the SCHAR Model for each sub-period. The results for each period are presented in Table 6.

We select variables from Table 6 with non-zero coefficients for either of the two sub-periods ($\widehat{P}^{port}_{t-1}$,$\widehat{P}^{port}_{t-2:t-5}$, $\widehat{N}^{port}_{t-1}$, $\widehat{N}^{port}_{t-2:t-5}$, $\widehat{M}^{port}_{t-1}$, and $\widehat{M}^{port}_{t-6:t-22}$) and run a Markov Switching model for the full sample. Based on these results, we next adjust only those coefficients that were statistically significant for both regimes ($\widehat{P}^{port}_{t-1}$ and $\widehat{P}^{port}_{t-2:t-5}$). Table 7 shows the estimates for the MS-SCHAR Model. Notably, only the first two coefficients ($\widehat{P}^{port}_{t-1}$,$\widehat{P}^{port}_{t-2:t-5}$) vary according to the state variable (S_t), and all covariate coefficients are statistically significant.

Figures 1 and 2 illustrate a prominent regime change from February 2020 to July 2020, coinciding with the global COVID-19 pandemic outbreak. These figures suggest that the relationship between the portfolio's realized variance and its daily positive realized semicovariance component ($\widehat{P}^{port}_{t-1}$) strengthens in stressed scenarios. This is indicated in Table 7, where the coefficient of the daily positive semicovariance under regime 1 ($\widehat{P}^{port}_{t-1,1}$) is greater than the coefficient of the daily positive semicovariance under regime 0 ($\widehat{P}^{port}_{t-1,0}$). The transition probabilities from the Markov Switching Model are available in

Table 4
HAR estimates

	Dependent variable: $\widehat{RV}_t^{port}$		
	HAR	SHAR	SCHAR
$\widehat{RV}_{t-1}^{port}$	0.612*** (0.040)		
$\widehat{PSV}_{t-1}$		6.639*** (0.509)	
$\widehat{NSV}_{t-1}$		0.601*** (0.455)	
$\widehat{RV}_{t-2:t-5}^{port}$	0.306*** (0.047)	0.210*** (0.041)	
$\widehat{RV}_{t-6:t-22}^{port}$	−0.069** (0.034)	−0.096*** (0.030)	
$\widehat{P}_{t-1}^{port}$			1.232*** (0.093)
$\widehat{P}_{t-2:t-5}^{port}$			1.726*** (0.252)
$\widehat{P}_{t-6:t-22}^{port}$			−0.561 (0.860)
$\widehat{N}_{t-1}^{port}$			−0.509*** (0.074)
$\widehat{N}_{t-2:t-5}^{port}$			−1.408*** (0.212)
$\widehat{N}_{t-6:t-22}^{port}$			1.228 (0.772)
$\widehat{M}_{t-1}^{port}$			−0.937*** (0.222)
$\widehat{M}_{t-2:t-5}^{port}$			−1.249** (0.573)
$\widehat{M}_{t-6:t-22}^{port}$			2.655*** (0.796)
Observations	602	602	602
R^2	0.704	0.763	0.809
Adjusted R^2	0.703	0.761	0.807
Residual s.e.	0.0003 (df = 598)	0.0002 (df = 597)	0.0002 (df = 592)
F Statistic	474.398*** (df = 3; 598)	479.963*** (df = 4; 597)	279.366*** (df = 9; 592)
AIC	−9972.87	−10103.91	−10225.65
BIC	−8240.47	−8367.11	−8466.84

Parameter estimates and their respective standard errors in brackets for each model in (5), (6) and (7), respectively. The first column shows us the covariates used across models. $***$, $**$ and $*$ represent whether a coefficient is significant at 1%, 5% or 10% levels, respectively.

Table 5
SCHAR-lasso-in estimates

	Dependent variable: $\widehat{RV}_t^{port}$
$\widehat{P}_{t-1}^{port}$	0.9933
$\widehat{P}_{t-2:t-5}^{port}$	0.370
$\widehat{P}_{t-6:t-22}^{port}$	0
$\widehat{N}_{t-1}^{port}$	0
$\widehat{N}_{t-2:t-5}^{port}$	0
$\widehat{N}_{t-6:t-22}^{port}$	0
$\widehat{M}_{t-1}^{port}$	−0.5523
$\widehat{M}_{t-2:t-5}^{port}$	0
$\widehat{M}_{t-6:t-22}^{port}$	0.1529

Lasso parameter estimates for model (7). The covariates chosen by lasso are $\widehat{P}_{t-1}^{port}$, $\widehat{P}_{t-2:t-5}^{port}$, $\widehat{M}_{t-1}^{port}$, $\widehat{M}_{t-6:t-22}^{port}$.

Table 6
SCHAR-lasso-in estimates for each sub-period

	Dependent variable: $\widehat{RV}_t^{port}$	
	SCHAR-lasso-in-pre	SCHAR-lasso-in-post
$\widehat{P}_{t-1}^{port}$	0.3258	0.9780
$\widehat{P}_{t-2:t-5}^{port}$	0	0.3763
$\widehat{P}_{t-6:t-22}^{port}$	0	0
$\widehat{N}_{t-1}^{port}$	0.4056	0
$\widehat{N}_{t-2:t-5}^{port}$	0.1538	0
$\widehat{N}_{t-6:t-22}^{port}$	0	0
$\widehat{M}_{t-1}^{port}$	0	−0.6121
$\widehat{M}_{t-2:t-5}^{port}$	0	0
$\widehat{M}_{t-6:t-22}^{port}$	−0.3422	0.2565

Lasso parameter estimates for model (7) for each sub-period. The covariates chosen by lasso for the pre-pandemic period are $\widehat{P}_{t-1}^{port}$, $\widehat{N}_{t-1}^{port}$, $\widehat{N}_{t-2:t-5}^{port}$, $\widehat{M}_{t-6:t-22}^{port}$. On the other hand, the covariates chosen by lasso for the post-pandemic period are $\widehat{P}_{t-1}^{port}$, $\widehat{P}_{t-2:t-5}^{port}$, $\widehat{M}_{t-1}^{port}$, $\widehat{M}_{t-6:t-22}^{port}$

Table 7
Switching Markov SCHAR estimates

	Dependent variable: $\widehat{RV}_t^{port}$
$\widehat{P}_{t-1,0}^{port}$	0.5848*** (0.0974)
$\widehat{P}_{t-1,1}^{port}$	1.2234*** (0.1730)
$\widehat{P}_{t-2:t-5,0}^{port}$	1.4123*** (0.1213)
$\widehat{P}_{t-2:t-5,1}^{port}$	2.0681*** (0.1719)
$\widehat{N}_{t-1}^{port}$	−0.5286*** (0.0595)
$\widehat{N}_{t-2:t-5}^{port}$	−1.3821*** (0.1755)
$\widehat{M}_{t-1}^{port}$	−0.9829*** (0.2029)
$\widehat{M}_{t-6:t-22}^{port}$	0.6257*** (0.1230)

Parameter estimates and their respective standard errors in brackets. $***$, $**$ and $*$ represent whether a coefficient is significant at t 1%, 5% or 10% levels, respectively.

Table 8
Estimated transition probabilities

		to:	
		0	1
from:	0	0.9908	0.0092
	1	0.0931	0.9069

Estimated transition probabilities for $S_t = 0,1$.

Table 8. In a first-order Markov chain with two possible states (regimes), the smoothed probabilities indicate a state ($S_t = 0$) characterized by low portfolio volatility and another state ($S_t = 1$) marked by high portfolio volatility (a stressed scenario) due to the COVID-19 pandemic.

4.3 Out-of-sample analysis

Our out-of-sample analysis employs the four models previously described: HAR, SHAR, SCHAR, and SCHAR-lasso-in. We use the same equally-weighted portfolio as in the in-sample analysis. We construct rolling out-of-sample one-step ahead forecasts based on each of these models, re-estimating the model parameters daily using the most recent 542 observations. [3]. Given that Bollerslev et al. (2020) suggests the SCHAR model may be "over-parameterized" in their out-of-sample analysis, we also include the SCHAR-lasso-out model,

[3]The code for running the rolling window analysis is available at https://github.com/rricco/realvol

Figure 1
Realized variance of the equally-weighted portfolio from Apr/2018 to Jan/2021

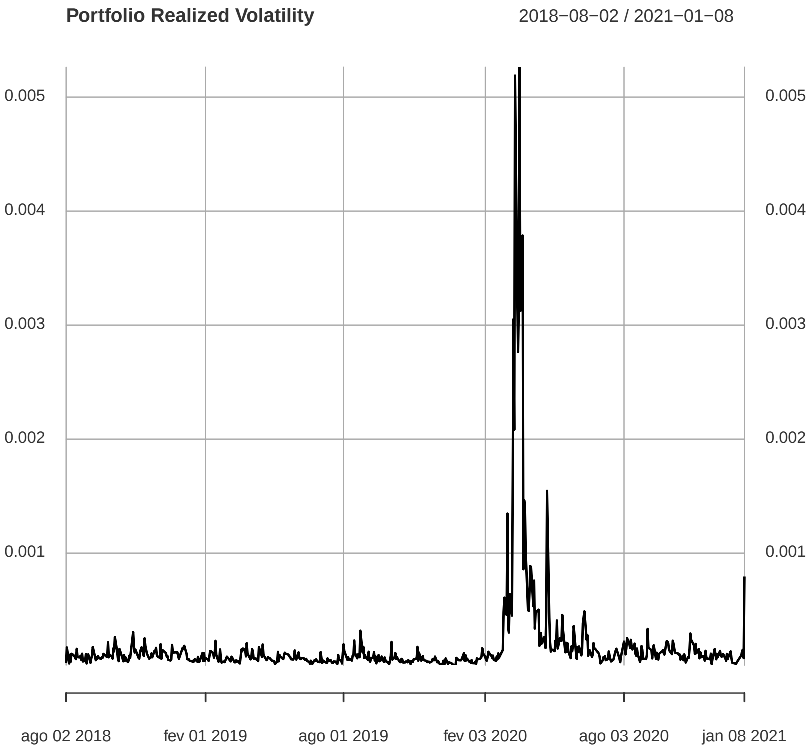


Figure 2
Smoothed probabilities for regime 0 ($S_t = 0$) from the Markov Switching SCHAR model

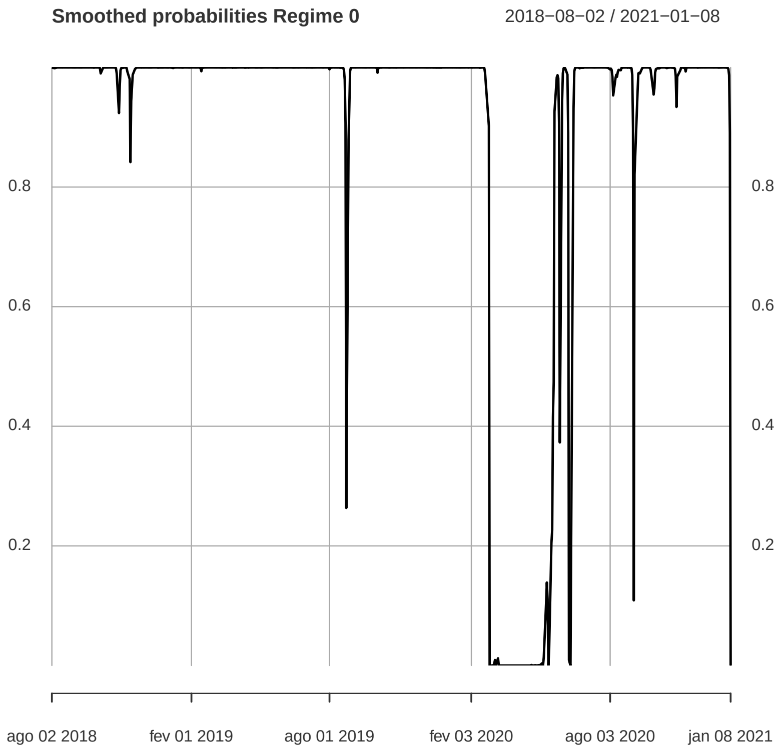


Table 9
Performance comparison for portfolio variance forecasts

Model	MSE	MAE
HAR	**1.00***	**1.00***
SHAR	1.02*	1.17*
SCHAR	1.38	1.59
SCHAR-lasso-in	1.01*	1.19*
SCHAR-lasso-out	1.02*	1.21*

Mean squared errors (MSE) and mean absolute errors (MAE) for the forecasts relative to the HAR Model. The values in bold indicate the method with lowest values of MSE and MAE. Cells with $*$ indicate that the method is included in the MCS constructed based on the T_{max} statistic using the squared/absolute errors with 5% of significance.

which applies LASSO to each rolling window estimation, in addition to the SCHAR-lasso-in model.

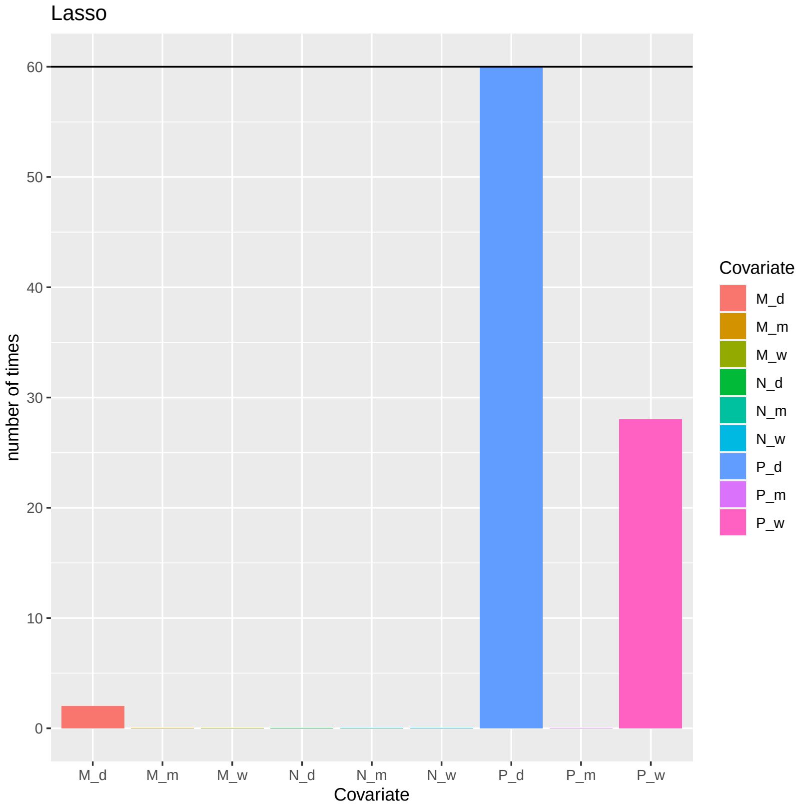
We evaluate the forecast performance of the models using the model confidence set (MCS) procedure introduced by Hansen et al. (2011). The MCS aims to determine the set of models containing the best performers with a given probability, based on a loss function such as mean squared error (MSE). This procedure employs bootstrap implementation to compute the p-values for all models.

Table 9 presents the forecast accuracy of each model, measured by both MSE and mean absolute error (MAE). The smallest MSE and MAE are indicated in bold. The cells with $*$ denote the models chosen by the MCS procedure. Interestingly, the simple HAR Model is the most accurate, even though it explained the dependent variable variance the least in the previous section. It is worth noting that the SHAR, SCHAR-lasso-in, and SCHAR-lasso-out models were included in the MCS procedure. Moreover, considering the sample size and the portfolio's dimension are not as large in this work as in Bollerslev et al. (2020), our results align with theirs because:

- In Bollerslev et al. (2020), the discrepancy in forecast accuracy among the models is greater for larger dimensional portfolios.
- The authors propose that the SCHAR Model might be "over-parameterized" and, as such, it could underperform in out-of-sample analysis. However, when we mitigate the "overfitting" in the SCHAR Model through the LASSO method, we achieve better results as shown in Table 9.

We also investigate which covariates are chosen in each rolling window estimation in the SCHAR-lasso-out Model. Figure 3 indicates that the covariate chosen most frequently by LASSO is $\widehat{P}_t^{port}$, selected 60 times out of 60,

Figure 3
Number of times each covariate is chosen by lasso from SCHAR-lasso-out model

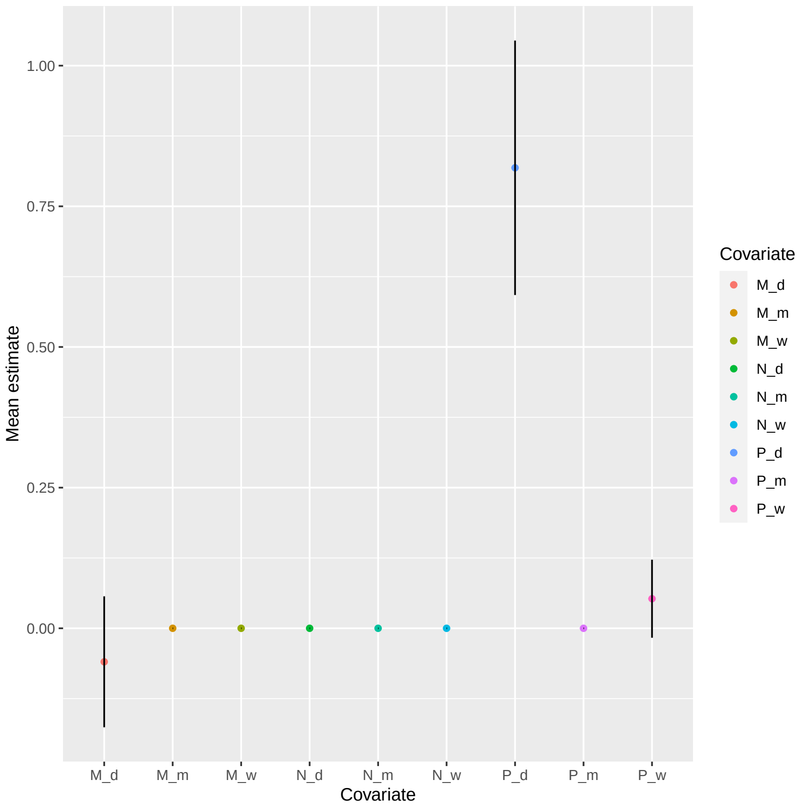


Legend: $P_d = \widehat{P}_t^{port}$, $P_m = \widehat{P}_{t-5:t-21}^{port}$, $P_w = \widehat{P}_{t-1:t-4}^{port}$, $N_d = \widehat{N}_t^{port}$, $N_m = \widehat{N}_{t-5:t-21}^{port}$, $N_w = \widehat{N}_{t-1:t-4}^{port}$, $M_d = \widehat{M}_t^{port}$, $M_m = \widehat{M}_{t-5:t-21}^{port}$, $M_w = \widehat{M}_{t-1:t-4}^{port}$.

followed by $\widehat{P}_{t-1:t-4}^{port}$, chosen 27 times out of 60, and finally $\widehat{M}_t^{port}$, selected 3 times out of 60.

Figure 4 displays the sample mean estimate for each covariate and its sample confidence interval using two sample standard deviations in the rolling window analysis. The sample mean estimate for $\widehat{P}_t^{port}$ is 0.8185, for $\widehat{P}_{t-1:t-4}^{port}$ it is 0.0525, and for $\widehat{M}_t^{port}$ it is –0.0595.

Figure 4
Mean estimate for each covariate (dots on the plot) and its sample confidence interval using two sample standard deviations (vertical black line)



Legend: $P_d = \widehat{P}_t^{port}$, $P_m = \widehat{P}_{t-5:t-21}^{port}$, $P_w = \widehat{P}_{t-1:t-4}^{port}$, $N_d = \widehat{N}_t^{port}$, $N_m = \widehat{N}_{t-5:t-21}^{port}$, $N_w = \widehat{N}_{t-1:t-4}^{port}$, $M_d = \widehat{M}_t^{port}$, $M_m = \widehat{M}_{t-5:t-21}^{port}$, $M_w = \widehat{M}_{t-1:t-4}^{port}$.

5. Portfolio Optimization

The covariance matrix plays a key role in portfolio optimization theory. In the seminal work by Markowitz (1952), the task becomes either minimizing the portfolio's risk (variance) for a fixed mean return or maximizing the mean return for a fixed risk. Formally, we can derive the optimum portfolio as the solution to the following problem:

$$\min_{\boldsymbol{w}} \ \boldsymbol{w}^\intercal C_{t|t-1} \boldsymbol{w} - \frac{1}{\lambda} \boldsymbol{w}^\intercal \mu_{t|t-1},$$

where $\boldsymbol{w}$ is the vector of portfolio weights, $C_{t|t-1}$ is the conditional covariance matrix, $\mu_{t|t-1}$ is the vector of the conditional mean, and λ is the investor's risk aversion coefficient.

To avoid issues related to estimation errors associated with estimating conditional return means, we focus on minimum variance portfolios as in Engle and Sheppard (2008); Borges et al. (2015), among many others. Mathematically, our problem becomes the following:

$$\min_{\boldsymbol{w}} \ \boldsymbol{w}^\intercal C_{t|t-1} \boldsymbol{w} \qquad \text{subject to} \qquad \boldsymbol{w}^\intercal \mathbf{1} = 1, \tag{10}$$

where $\mathbf{1}$ is the vector of ones.

One of the pioneering works in empirical applications of portfolio allocation using intra-daily data is the paper by Fleming et al. (2003). The authors suggest that covariance matrices estimated using intra-daily returns can lead to gains in portfolio performance compared to those estimated using daily returns.

Another significant contribution to this topic is the paper by Liu (2009). The author concludes that the gains generated by intra-daily returns depend on the rebalancing frequency and the prediction horizon. According to the author, if an investor rebalances his portfolio monthly with at least the previous 12 months of data, daily and intra-daily returns yield similar performance results. However, it is beneficial to use intra-daily data when the portfolio is rebalanced daily. Hautsch et al. (2013) demonstrate that large-scale portfolio covariance matrices based on high-frequency data result in significantly lower portfolio volatility than methods employing daily returns.

In this context, Borges et al. (2015) compare different covariance matrix estimators based on intra-daily or daily data for the Brazilian market. Their analysis suggests that conditional covariance estimates outperform unconditional estimators, and that covariance matrix forecasts based on high-frequency data yield lower portfolio volatility compared to using daily returns.

Given this backdrop, we investigate the performance of a Brazilian stock portfolio consisting of 10 stocks from the Ibovespa index, using the realized semicovariance measures of Bollerslev et al. (2020) and comparing them to those obtained using the standard realized covariance matrix.

Our portfolio analyses rely on the realized covariance matrix, the positive semicovariance matrix, and the negative semicovariance matrix measures defined in (1) and (2) respectively. We estimate the conditional realized measure (RM) for a single period as

$$\widehat{RM}_{t|t-1} = \frac{1}{\ell}\sum_{i=1}^{\ell} RM_{i,t-1}, \tag{11}$$

where ℓ is the length (in days) of the rebalancing period $t-1$ and $RM_{i,t-1}$ is the realized measure ($\widehat{RC}$, $\widehat{RSC}_{positive}$, or $\widehat{RSC}_{negative}$ as defined in section 2) for day i within the rebalancing period $t-1$.

The data used to run the analyses are the same as those used in section 4. For each period, we compute $\widehat{RM}_{t|t-1}$ in (10) to generate the minimum variance portfolio. Subsequently, we evaluate the portfolio performance on a daily basis in terms of the Sharpe ratio (S_a), Sortino ratio (S_o), and turnover (T_o), for each realized measure and rebalancing period (daily, weekly, or monthly). Assuming the risk-free rate as zero, these statistics are calculated as follows:

$$S_a = \frac{\hat{\mu}}{\hat{\sigma}}, \qquad S_o = \frac{\hat{\mu}}{\hat{\sigma}_-}, \qquad T_o = \frac{1}{T}\sum_{t=1}^{T} |(\boldsymbol{w}_{t+1} - \boldsymbol{w}_t)|^\intercal \mathbf{1},$$

where T is the length of the out-of-sample period, N is the number of stocks, $\boldsymbol{w}_t$ is the $(1\times N)$ vector of the portfolio weights at day t, $\boldsymbol{r}_t$ is the $(1\times N)$ vector of the assets' returns at day t, $\mathbf{1}$ is the $(1\times N)$ vector of ones, and

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{w}_t^\intercal \boldsymbol{r}_t$$

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} (\boldsymbol{w}_t^\intercal \boldsymbol{r}_t - \hat{\mu})^2$$

$$\hat{\sigma}_-^2 = \frac{1}{T}\sum_{t=1}^{T} [(\boldsymbol{w}_t^\intercal \boldsymbol{r}_t - \hat{\mu})\mathbb{I}_{\boldsymbol{w}_t^\intercal \boldsymbol{r}_t < \hat{\mu}}]^2 \quad \text{(Downside Semivariance).}$$

6. Empirical Exercise 2

In this section, we compare the out-of-sample portfolio performance using different realized measures ($\widehat{RC}$, $\widehat{RSC}_{positive}$, or $\widehat{RSC}_{negative}$) for the covariance matrix, based on high-frequency data. The portfolios, rebalanced daily,

Table 10
Minimum variance optimum portfolio performance (daily rebalancing)

	Average return	Standard deviation	Sharpe ratio	Sortino ratio	Turnover
$\widehat{RSC}_{negative}$	0.001	0.020	0.045*	0.060*	1.093
$\widehat{RSC}_{positive}$	0.0005	0.020	0.024	0.033	1.109
$\widehat{RC}$	0.001	0.019	0.040	0.053	0.829*

Out-of-sample performance of the minimum variance portfolio using 10 assets traded at B3 stock exchange based on daily returns. The best results for Sharpe, Sortino and Turnover are with $*$.

Table 11
Minimum variance optimum portfolio performance (weekly rebalancing)

	Average return	Standard deviation	Sharpe ratio	Sortino ratio	Turnover
$\widehat{RSC}_{negative}$	0.001	0.019	0.051	0.069	0.164
$\widehat{RSC}_{positive}$	0.001	0.020	0.063	0.086*	0.166
$\widehat{RC}$	0.001	0.018	0.064*	0.085	0.112*

Out-of-sample performance of the minimum variance portfolio using 10 assets traded at B3 stock exchange based on daily returns. The best results for Sharpe, Sortino and Turnover are with $*$.

weekly, and monthly, are evaluated according to their performance in terms of the Sharpe ratio, Sortino ratio, and turnover.

Tables 10 to 12 show the performances of the above-mentioned indicators for daily, weekly, and monthly rebalancing, respectively.

The results in Table 10 indicate that when the investor rebalances the portfolio daily, the $\widehat{RC}$ measure results in a portfolio with a lower standard

Table 12
Minimum variance optimum portfolio performance (monthly rebalancing)

	Average return	Standard deviation	Sharpe ratio	Sortino ratio	Turnover
$\widehat{RSC}_{negative}$	0.001	0.019	0.037	0.051	0.030
$\widehat{RSC}_{positive}$	0.0005	0.020	0.024	0.032	0.029
$\widehat{RC}$	0.001	0.019	0.045*	0.060*	0.019*

Out-of-sample performance of the minimum variance portfolio using 10 assets traded at B3 stock exchange based on daily returns. The best results for Sharpe, Sortino and Turnover are with $*$.

deviation. However, in terms of risk-adjusted returns, the $\widehat{RSC}_{negative}$ measure presents the highest Sharpe and Sortino ratios.

Considering transaction costs, the $\widehat{RC}$ measure presents the best performance (lowest turnover). Table 11 shows the results when the investor rebalances the portfolio weekly. Once again, the $\widehat{RC}$ measure presents the lowest standard deviation. In terms of risk-adjusted returns, the $\widehat{RC}$ measure presents the highest Sharpe ratio, whereas the $\widehat{RSC}_{positive}$ has the highest Sortino ratio. With respect to transaction costs, the $\widehat{RC}$ measure, again, shows the lowest turnover.

Finally, Table 12 displays the results when the investor rebalances the portfolio monthly. Similar to the two previous rebalancing periods, the $\widehat{RC}$ measure presents the lowest standard deviation, and in terms of risk-adjusted returns, the $\widehat{RC}$ measure presents the highest Sharpe and Sortino ratios and the lowest turnover.

These results suggest that the realized components of the covariance matrix align better with higher frequency rebalancing periods in terms of economic performance. Following the literature, these results are consistent with Bollerslev et al. (2020), who find that realized semicovariance matrices ($\widehat{RSC}_{positive}$, $\widehat{RSC}_{negative}$) generally respond to new information faster than the realized covariance matrix ($\widehat{RC}$). Moreover, this feature helps us justify the higher turnover presented by the semicovariance measures ($\widehat{RSC}_{positive}$, $\widehat{RSC}_{negative}$).

7. Conclusions

The primary aim of this study is to demonstrate how realized semicovariances, as developed by Bollerslev et al. (2020), can contribute to two key problems in quantitative finance: volatility forecasting and portfolio optimization. We have a particular interest in addressing these matters within the context of the Brazilian stock market.

Our volatility forecasting yields some compelling conclusions. In our in-sample analysis, it is evident that incorporating semicovariance components into the model improves the goodness of fit for the realized portfolio variance model. In addition, we illustrate that a Markov Switching Model is applicable for our designated period of analysis. This finding suggests that the onset of the Covid-19 pandemic corresponds to a period of elevated volatility. It also implies that the relationship between the realized portfolio variance and its semicovariance components can fluctuate under different regimes. Our out-of-sample analysis reveals that the SCHAR-lasso-in and SCHAR-lasso-out are included in the Model Confidence Set (MCS), whereas the SCHAR is not. This finding indicates that the SCHAR Model could potentially suffer from "overfitting".

Lastly, our portfolio optimization analysis reveals that, under higher frequency rebalancing periods, minimum variance portfolios utilizing negative semicovariance matrices perform better in terms of risk-adjusted returns compared to those that use standard realized covariance matrices. This observation supports the analysis in Bollerslev et al. (2020), which suggests that semicovariance components react more quickly to new information.

References

Aggarwal, R., Inclan, C. and Leal, R. (1999). Volatility in emerging stock markets, *Journal of Financial and Quantitative Analysis* **34**(1): 33–55.

Aït-Sahalia, Y. and Xiu, D. (2016). Increased correlation among asset classes: Are volatility or jumps to blame, or both?, *Journal of Econometrics* **194**(2): 205–219.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility, *Econometrica* **71**(2): 579–625.

Barndorff-Nielsen, O., Hansen, P. R., Lunde, A. and Shephard, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading, *Journal of Econometrics* **162**(2): 149–169.

Barndorff-Nielsen, O. 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.

Barndorff-Nielsen, O. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics, *Econometrica* **72**(3): 885–925.

Barndorff-Nielsen, O., Kinnebrock, S. and Shephard, N. (2010). Measuring downside risk: Realised semivariance, *in* T. Bollerslev, J. Russell and M. Watson (eds), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, Oxford University Press.

Bekaert, G. and Harvey, C. R. (2003). Emerging markets finance, *Journal of Empirical Finance* **10**(1-2): 3–55.

Bollerslev, T., Li, J. and Xue, Y. (2018). Volume, volatility, and public news announcements, *The Review of Economic Studies* **85**(4): 2005–2041.

Bollerslev, T., Li, J., Patton, A. J. and Quaedvlieg, R. (2020). Realized semicovariances, *Econometrica* **88**(4): 1515–1551.

Borges, B., Caldeira, J. F. and Ziegelmann, F. A. (2015). Selection of minimum variance portfolio using intraday data: An empirical comparison among different realized measures for bm&fbovespa data, *Brazilian Review of Econometrics* **35**(1): 23–46.

Çelik, S. and Ergin, H. (2014). Volatility forecasting using high frequency data: Evidence from stock markets, *Economic Modelling* **36**: 176–190.

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

Emre Alper, C., Fendoglu, S. and Saltoglu, B. (2012). MIDAS volatility forecast performance under market stress: Evidence from emerging stock markets, *Economics Letters* **117**(2): 528–532.

Engle, R. and Sheppard, K. (2008). Evaluating the specification of covariance models for large portfolios, *Working paper*, New York University.

Fishburn, P. C. (1977). Mean-risk analysis with risk associated with below-target returns, *The American Economic Review* **67**(2): 116–126.

Fleming, J., Kirby, C. and Ostdiek, B. (2003). The economic value of volatility timing using "realized" volatility, *Journal of Financial Economics* **67**(3): 473–509.

Hansen, P. R., Huang, Z. and Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility, *Journal of Applied Econometrics* **27**(6): 877–906.

Hansen, P. R., Lunde, A. and Nason, J. M. (2011). The model confidence set, *Econometrica* **79**(2): 453–497.

Hautsch, N., Kyj, L. M. and Malec, P. (2013). Do high-frequency data improve high-dimensional portfolio allocations?, *Technical report*, SSRN.

Hogan, W. W. and Warren, J. M. (1972). Computation of the efficient boundary in the ES portfolio selection model, *Journal of Financial and Quantitative Analysis* pp. 1881–1896.

Kim, C. J. and Nelson, C. R. (1999). *State-space models with regime switching: Classical and Gibbs-sampling approaches with applications*, MIT Press.

Kim, H. Y. and Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models, *Expert Systems with Applications* **103**: 25–37.

Kristjanpoller, W., Fadic, A. and Minutolo, M. C. (2014). Volatility forecast using hybrid Neural Network models, *Expert Systems with Applications* **41**(5): 2437–2442.

Lindsay, B. G. (1988). Composite likelihood methods, *Contemporary Mathematics* **80**(1): 221–239.

Liu, Q. (2009). On portfolio optimization: How and when do we benefit from high-frequency data?, *Journal of Applied Econometrics* **24**(4): 560–582.

Mao, J. C. T. (1970). Survey of capital budgeting: Theory and practice, *Journal of finance* pp. 349–360.

Markowitz, H. (1952). Portfolio selection, *Journal of Finance* pp. 77–91.

Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility, *Journal of Applied Econometrics* **17**(5): 479–508.

Mykland, P. A. and Zhang, L. (2009). Inference for continuous semimartingales observed at high frequency, *Econometrica* **77**(5): 1403–1445.

Noureldin, D., Shephard, N. and Sheppard, K. (2012). Multivariate high-frequency-based volatility (HEAVY) models, *Journal of Applied Econometrics* **27**(6): 907–933.

Patton, A. J. and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility, *Review of Economics and Statistics* **97**(3): 683–697.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society Series B* **58**(1): 267–288.