Norwegian School of Economics
Bergen, spring semester 2020

ECN430: Term Paper

More is always better?

Comparison of variance forecast results using daily and intraday return series in a GARCH framework

NORWEGIAN SCHOOL OF ECONOMICS

Contents

Abstract

The declared aim of this paper is to forecast the conditional variance of a time series based on the current literature. We use intraday and daily data and compare the results according to their accuracy. For this purpose, we create two models, the first model being based on daily yields. The second model is based on intraday returns and thus uses a larger amount of data. Due to the high liquidity necessary (Banulescu et al. 2014) (Cont 2011) for intraday analysis we decided to use the time series of the S&P500. Both models make use of the plain vanilla ARCH-GARCH framework (Engle 1982) (Bollerslev 1986), which has become the most common in finance (Engle 2001). The empirical analysis should show that the additional information mentioned above, which flows into the model via the intraday returns, actually leads to a more accurate forecast. Because of the focus of this paper on the forecast, we are particularly interested in which of the two approaches to the test data has better accuracy. Consequently, we refrain from in-sample testingfor the most part. However, before we start with the empirical analysis, we would like to present the theoretical foundations in detail. For this, we heavily rely on the work of Andrea Bucci (Bucci 2017). This includes the theoretical derivation of the price process as well as the theoretical aspects of variance and the ARCH and GARCH framework.

Introduction

Financial economists and econometricians have long been studying the concept of variance in financial markets. Unlike prices, the variance (volatility) cannot be observed, not even ex-post. Nevertheless, it is an essential component of risk management, asset management and option pricing. The need to have a reliable measure of risk is much greater today than it was 20 years ago. Firstly, the trading volume of options and similar financial derivatives has risen sharply. The price or value of an option depends on the volatility of the underlying security. Second, the Basel Committee's agreements on banking supervision stipulate that banks and financial institutions must hold a multiple of their Value at Risk (VaR) capital, which is the minimum loss expected with a given probability over a certain period of time. Such a VaR value is therefore explicitly geared towards the future and requires volatility forecasts, which today are often based on GARCH models.

In this paper, we work with the intra-day returns measured every five minutes of the S&P 500. According to the efficient market hypothesis, the best prediction for tomorrow's return is today's value, because tomorrow's value is not correlated with today's value. According to the efficient market hypothesis, price changes are not predictable. But that does not mean that tomorrow's value must be stochastically independent of today's value. If the return is generated by an ARCH or GARCH process, the squared return tomorrow is correlated with the squared return today. So, although there is no correlation in the level of returns, the returns are still related over the squares. This is because tomorrow's conditional volatility depends on the square of today's return, and tomorrow's volatility in turn is responsible for tomorrow's return.

Theoretical foundation:

Price process

The theoretical foundations in this section as well as the theoretical foundation of the ARCH and GARCH models are strongly leaned on the work "forecasting realized volatility" by Andrea Bucci. We have followed the outline of this work both in the declination of the variables and in the logical structure (Bucci 2017).

With the introduction of algorithm-based trading of securities, the frequency with which prices are negotiated has become extremely shortened. So it can be assumed that price movements follow a continuous process. The univariate process of a logarithmic price $p(t)$ is defined in the probability space (Ω, I, P). The process takes place in a continuous time interval which is described as follows $[0, T]$. The total information available on the price during this period can be presented as follows and is part of or equivalent to the total amount of all information $(I_t)_{t\epsilon[0,T]} \subseteq I$.

Under the assumption of efficiency market theory, market prices reflect all available information and there is no possibility of making profit by arbitrage (Fama 1965). With the additional assumption of finite first moments the prize process can be assigned to the class of special semimartingale processes (Shiryaev 1999). The log price process $p(t)$ can be assumed to be a semimartingale process if there is decomposition into a drift component and a local martingale which for itself can be decomposed further into a realization of a continuous process and a jump component. (Andersen et al. 2001) So it should hold:

$$p(t) = p(t) + A(t) + M(t) = p(0) + A^C(t) + \Delta A(t) + M^C(t) + \Delta M(t)$$

The realizations of the continuous process are donated by $A^C(t)$ and $M^C(t)$ while $\Delta A(t)$ and $\Delta M(t)$ are the mentioned jump components.

If return between two observations in the interval $[t - h, t]$, for $0 \leq h \leq t \leq T$ is defined as

$$r(t, h) = p(t) - p(t - h);$$

And considering the return can be calculated from the interval $[0, t]$

$$r(t) = r(t, t) = p(t) - p(0)$$

It follows that

$$r(t, h) = p(t) - p(0) + p(0) - p(t - h)$$

$$= r(t) - r(t - h)$$

Another assumption is that the price development of the financial instrument is a finite process, i.e. that both $p(t)$ and $r(t)$ are well defined over the entire interval $[0, T]$.

It must apply that the squares of the price process as well as the squares of the return process are integrable.

$$r(t) \equiv p(t) - p(0) = \mu(t) + M(t) = \mu(t) + M^C(t) + M^J(t)$$

Thus, the return of an infinitesimal small time unit can be decomposed into a predicted process $\mu(t)$ and a local martingale process $M(t)$. The latter process can be decomposed again, into an infinite variation local martingale $M^C(t)$ and a compensated jump martingale $M^J(t)$.

Volatility

Volatility is a measure of risk and shows the fluctuation intensity of the price of an underlying asset within a certain period (French et al. 1987).

Let $X(t)$be a semimartingale process with a quadratic variation $[X,X]_t$. Further $X(t)$and $Y(t)$ donate a couple of semimartingale processes with the covariation $[X,Y]_t$. It follows that if $t \in [0,T]$

$$[X,X]_t = X(t)^2 - 2\int_0^t X(s-)dX(s)$$

$$[X,Y]_t = X(t)Y(t) - \int_0^t X(s-)dX(s) - \int_0^t Y(s-)dY(s)$$

$X(s-)$and Y$(s-)$ are cádlág processes with a well-defined integral. Thus, it follows that $[X,X]_t$the quadratic variation is a growing stochastic process with the following properties:

i. Assumed τ_m is a partition of [0,T], for $0 = \tau_{m,0} \leq \tau_{m,1} \leq \cdots \leq \tau_{m,m} = T$, such that $supj \geq 0(\tau_{m,j+1} - \tau_{m,j}) \to 0$ for $m \to \infty$, then (Bucci 2017)

$$\lim_{m\to\infty}\left\{\sum\nolimits_{j\geq 1}\left(X\left(t\wedge\tau_{m,j}\right) - X\left(t\wedge\tau_{m,j-1}\right)\right)^2\right\} \to [X,X]_t$$

the convergence is uniform and $t \wedge \tau \equiv min(t,\tau)$. The quadratic variation process represents the (cumulative) realized sample-path variability of $X(t)$ over the $[0,t]$ time interval. (Bucci 2017)

ii. The covariance between $X(t)$, and $Y(t)$ in the interval $[t-h,t]$ can be represented as follows:

$$Cov[X(t),Y(t)|I_{t-h}] = E([X|Y]_t|I_{t-h}) - [X,Y]_{t-h}$$

If $X(t)$, and $Y(t)$ are square integrable semimartingale.

iii. If the finite variation component $A(t)$ is also continuous, then

$$[X_i, X_j]_t = [M_i, M_j] = [M_i^C, M_j^C] + \sum_{0 \leq s \leq t} \Delta M_i(s) \Delta M_j(s)$$

It is therefore shown that the quadratic variation in continuous finite variation processes is zero. The component of the mean value can therefore be disregarded. It can therefore be assumed that the log price follows a diffusion process.

$$d\,p(t) = \mu(t)dt + \sigma(t)dW(t),$$

$W(t)$ is a Wiener process (Springer, Berlin, Heidelberg 2012), $\mu(t)$ is a finite variation predictable process and $\sigma(t)$ is a strictly positive and squarely integrable process

$$P\left[\int_{t-h}^{t} \sigma^2(s)ds < \infty\right] = 1$$

The realized return over the interval $[t-h, t]$ can therefore be represented as follows

$$r(t,h) = \mu(t,h) + M(t,h) = \int_{t-h}^{t} \mu(s)ds + \int_{t-h}^{t} \sigma(s)dW(s)$$

For the calculation of the quadratic variation (integrated variance) the following equation results:

$$QV_t = [p,p]_t - [p,p]_{t-h} = \int_{t-h}^{t} \sigma^2(s)ds$$

The concept of quadratic variation is the basis for the definition of notional volatility, quantified by the realized variance (Curci und Corsi 2004). Over the time interval $[t-h, t]$, notional variation equals the quadratic variation of a return series and can be expressed by the following equation

$$v^2 = [r,r]_t - [r,r]_{t-h} = \int_{t-h}^{t} \sigma^2(s)ds$$

The information parameter I_t contains, as already described, all information from r_t on $[t-h, t]$,, which results in the following expression for the calculation of the conditional volatility

$$Var(r_t|I_t) = E[\{r_t - E(r_t|I_t)\}^2|I_t]$$

By rearranging this equation, it can be shown that the conditional variance can also be written in the following measures.

$$Var(r_t|I_{t-h}) \cong E[v^2(t,h)|I_{t-h}] = E[QV(t,h)|I_{t-h}]$$

It can be seen that for $\mu(s) = 0$ or under the circumstance that $\mu(s)$ is measurable from the information I_{t-h}, the conditional variance is equal to the conditional expected value of the quadratic variation. One way, among others, to estimate volatility is to use parametric models. The most common model framework used by financial economists is the Autoregressive Conditional Heteroscedasticity model (ARCH). In this model, the information parameter I_{t-h}, contains all directly observable information about the returns in the period $[t-h, t]$ under consideration and is therefore dependent on them. This type of model requires a functional form of the stochastic process of local martingale $M(t)$ and for the finite variation process $\mu(t)$.

GARCH Framework

ARCH models have proven successful among financial economists due to their improved forecasting intervals in efficient markets; correctly specified regression residuals and valid statistical inference in the regression model; and risk measurement and forecasting, for example, for option pricing. However, not only option prices depend on the volatility of the underlying security, but also the price of a security itself can be seen as a function of its variance. This is because a riskier security is expected to yield a higher return on average. (Uwe Hassler 2003)

Models from the ARCH framework take advantage of this and work with the second conditional moments.

$$Var(r_t|I_{t-h}) = E[(r(t,h) - E[\mu(t,h)]|I_{t-h})^2|I_{t-h}]$$

ARCH models take advantage of the fact that the returns on financial instruments can be decomposed as described in the previous chapter.

$$y_t = x_t'b + \varepsilon_t$$

With validity of the Gauss-Markov acceptance to the expected value of the error under the information parameter $E[\varepsilon_t|I_{t-2}] = 0$ the innovation ε_t at time t can be further decomposed as follows

$$\varepsilon_t = u_t h_t^{1/2}$$

The stochastic part $u_t \sim i.i.d.D(0,1)$ is a standard process. h_t is the conditional variance of the error term. The following assumption must hold

$$Cov(\varepsilon_t \varepsilon_{t+k}) = 0$$

The process can be heteroskedastic, meaning that the conditional variance can change over time.

$$h_t = E[\varepsilon_t^2 | I_{t-1}] = Var(\varepsilon_t | I_{t-1})$$

This results in the following distribution of the innovations.

$$\varepsilon_t | I_{t-1} \sim N(0, h_t)$$

From the described assumptions a model for h_t can be derived, this is a linear function of the past squared realized error terms.

$$h_t = \omega + \sum\nolimits_{i=1}^{q} a_i \varepsilon_{t-j}^2$$

Both, the constant ω and the coefficient a_i need to be estimated in this ARCH model. Both must be positive in every lag.

$$\sum\nolimits_{i=1}^{q} a_i < 1$$

Is the necessary condition for a stationary process.

The GARCH model

The Generalized Autoregressive Conditional Heteroscedasticity Model adds an autoregressive component to the original ARCH model. The GARCH model for determining conditional variance under the same assumptions as before takes the following form

$$h_t = \omega + \sum\nolimits_{i=1}^{q} a_i \varepsilon_{t-j}^2 + \sum\nolimits_{j=1}^{p} \beta_i h_{i-j}$$

Thus, the conditional variance in the GARCH(p,q) model is dependent on the past weighted squared error terms as well as on the weighted past conditional variances themselves. Again, this model contains a constant.

To obtain a conditional variance stationary process, the following condition must be met. (Panorska et al. 1995)

$$\sum\nolimits_{i=1}^{q} a_i + \sum\nolimits_{j=1}^{p} \beta_i < 1$$

Realized volatility is computed by summing intraday squared returns. (Andersen et al. 2001)

Frequency:

Tick-by-tick prices are generally only available at unevenly spaced time points. For the calculation of evenly distributed returns it would be essential to calculate them by interpolation in even periods of time. As already shown in different literature, this interpolation can cause the returns to show a negative autocorrelation, which should be avoided. (Cont 2011) With the abandonment of data of a shorter frequency, information is inevitably lost. However, ultra-high frequency data contain market microstructure noise. Causes include bid-ask bounces, jumps or irregular data and omitted data. (Banulescu et al. 2014)

Taking into account the disadvantages of very short frequencies and long periods of time we decided to use a 5-minute sample. This is also possible because we are looking at a very liquid index. We can assume that our sample is free of measurement errors and yet slow enough to ensure that we do not have to fear market microstructures. (Duffie und Protter 1992)

Data Analysis and Forecast

This paper uses daily and intraday logarithmic returns of the S&P 500 from 2019/09/04 to 2020/03/17. We obtained the intraday prices with a five-minute frequency from Bloomberg and the daily prices S_t from Yahoo Finance. Both returns are calculated with the following formula:

$$r_{t,m} = \log S_{t,m} - \log S_{t-1,m}$$

The daily returns r_t are derived from daily closing prices. For the intraday returns, we used the next five-minute price, respectively. The m indicates the different frequencies.

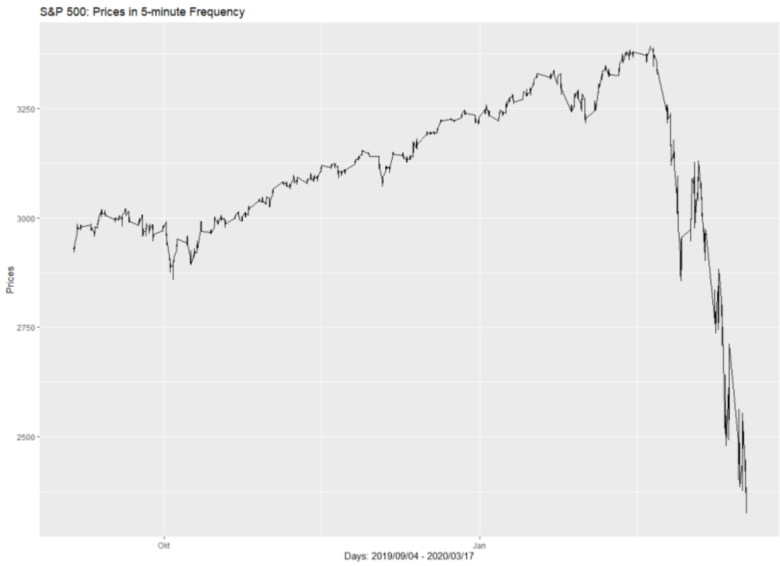


Figure 1: S&P 500: Prices in 5-minute frequency

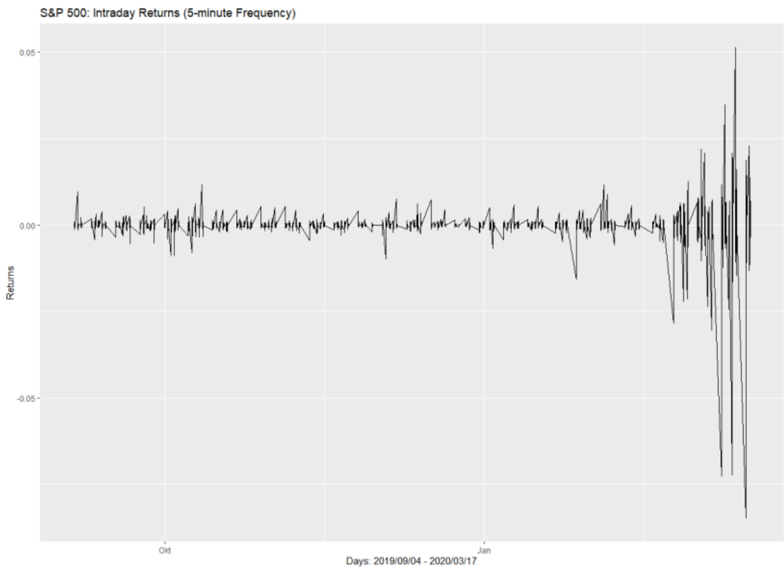


Figure 2: S&P 500: Intraday returns (5-minute frequency)

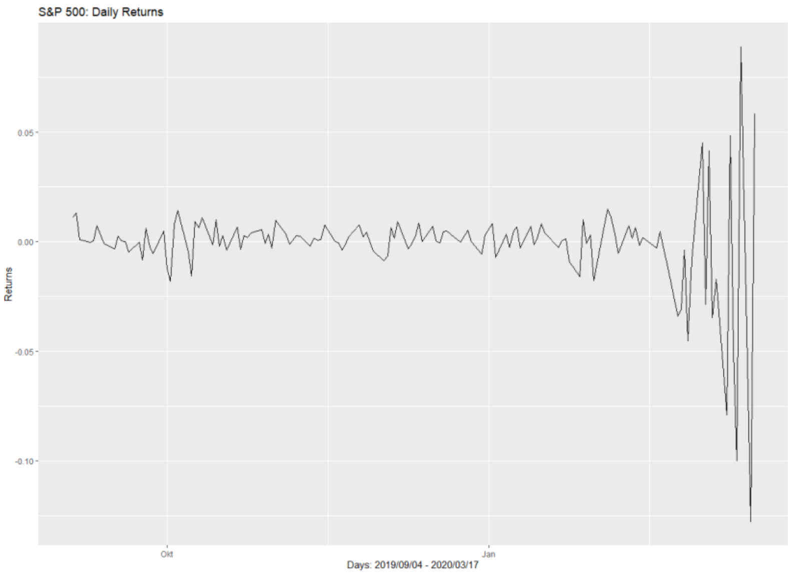


Figure 3: S&P 500: Daily returns

Stationarity

We then tested the series for stationarity using the Augmented Dickey-Fuller (ADF) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests. The ADF test tests the null hypothesis that there is a unit root, i.e. that the tested series is non-stationary. Hence, the alternative hypothesis states that the tested series is stationary. We received a p-value of 0.01888 and could therefore reject the null hypothesis at a significance level of five percent. The log return series consequently is stationary according to the ADF test.

The KPSS test, on the other hand, tests the null hypothesis that there is not a unit root. The alternative hypothesis therefore constitutes the lack of a unit root and suggests stationarity or trend-stationarity. The p-value for the return series is 0.01. We therefore have to reject the null hypothesis and assume the alternative hypothesis of non-stationarity.

Unfortunately, we receive opposite results from the ADF and KPSS tests as the existence of a unit roots constitutes H0 for the former and H1 for the latter and we were able to reject the null hypothesis both times.

Follow the result of the ADF test, we assume stationarity both for the intraday and daily return series.

Box-Jenkins methodology

Based on the assumption of stationarity for our series, we can employ the Box-Jenkins (BJ) methodology in order to identify an appropriate Autoregressive-moving-average (ARMA) model, estimate the parameters p and q for the AR and the MA part of the model and test its residuals (for white noise). Ideally, we would be able to forecast the series after the successful completion of those steps.

First, we generated plots of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the return series. We evaluated the plots according to the following criteria:

Table 3.1. Behavior of the ACF and PACF for ARMA Models

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Figure 4: Behavior of the ACF and PACF for ARMA Models (Stoffer und Shumway, 2017)

The plots look as follows:

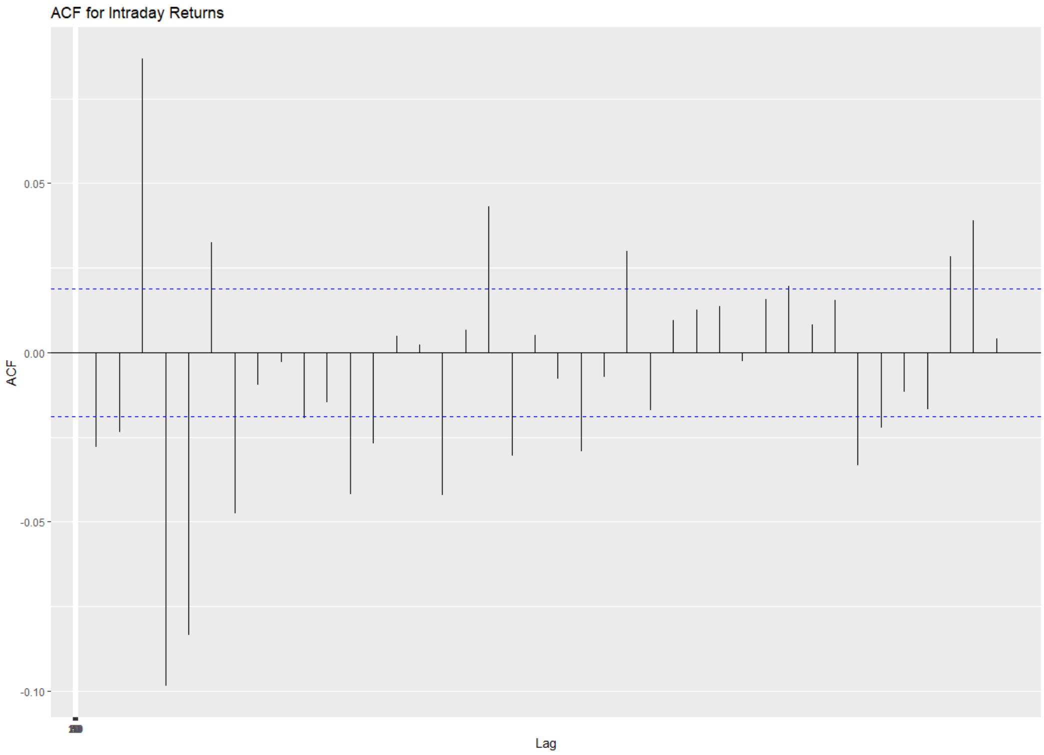


Figure 5: ACF for intraday returns

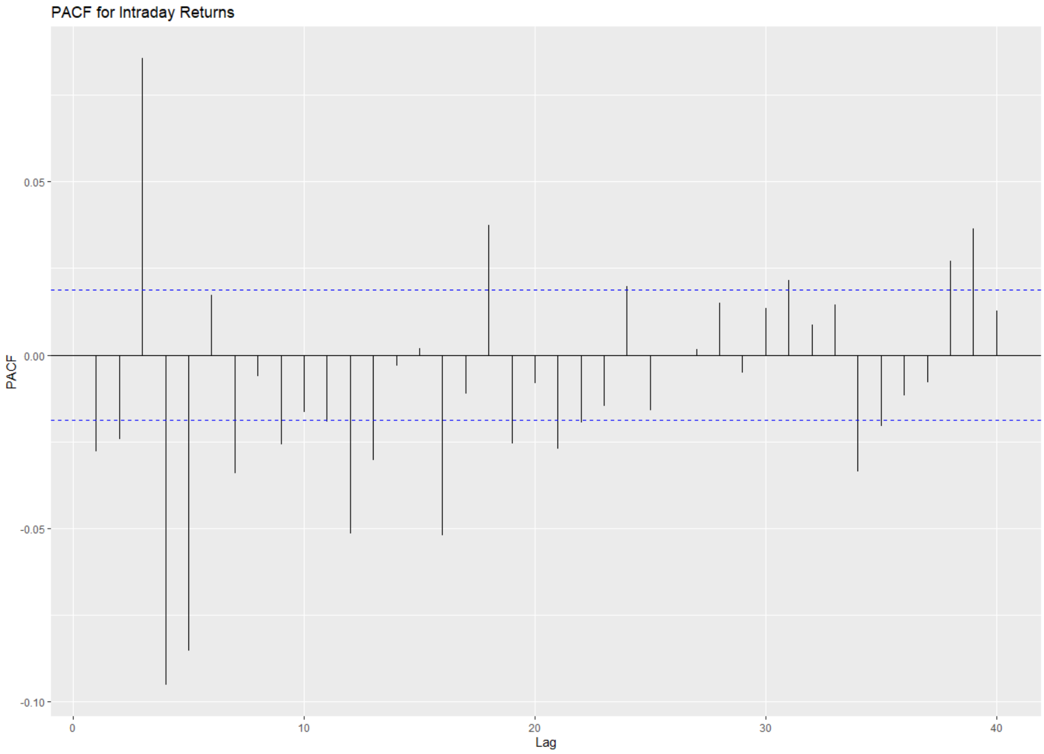


Figure 6: PACF for Intraday Returns

The ACF of the intraday return series very slowly tails off and does not have a clear cut-off. According to the table above, we therefore conclude an ARMA(p,q) model. The PACF tails off slightly more quickly as the lags seem to follow the same development but are less significant than in the ACF plot. This might suggest an ARMA(p,q) model with higher lag orders. Nonetheless, the lag order for an ARMA process cannot be reliably determined by visually examining the plots and therefore needs to be estimated in the next step.

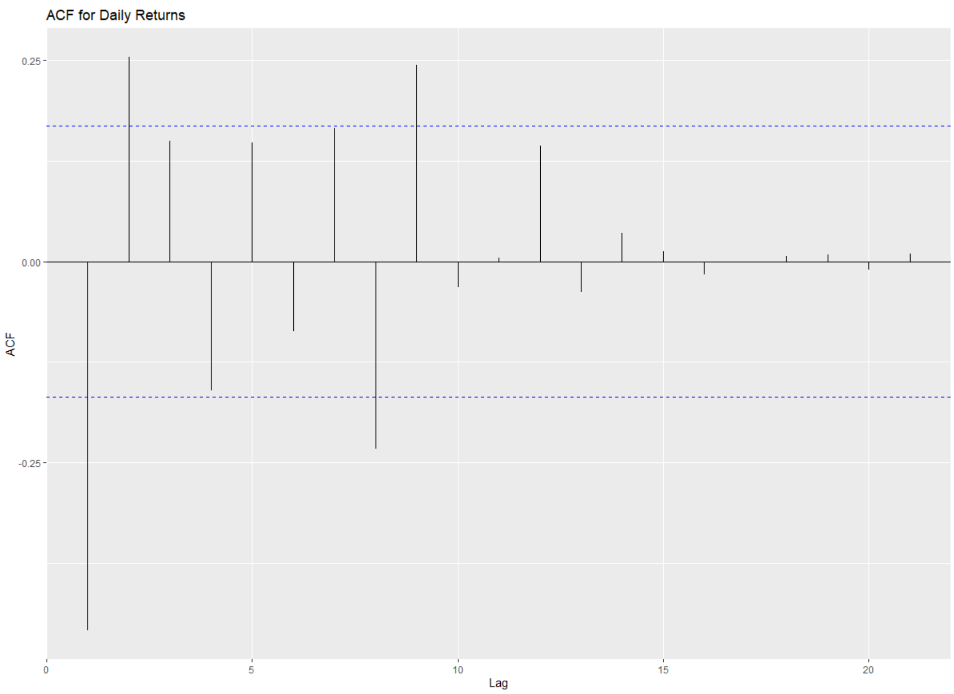


Figure 7: PACF for daily returns

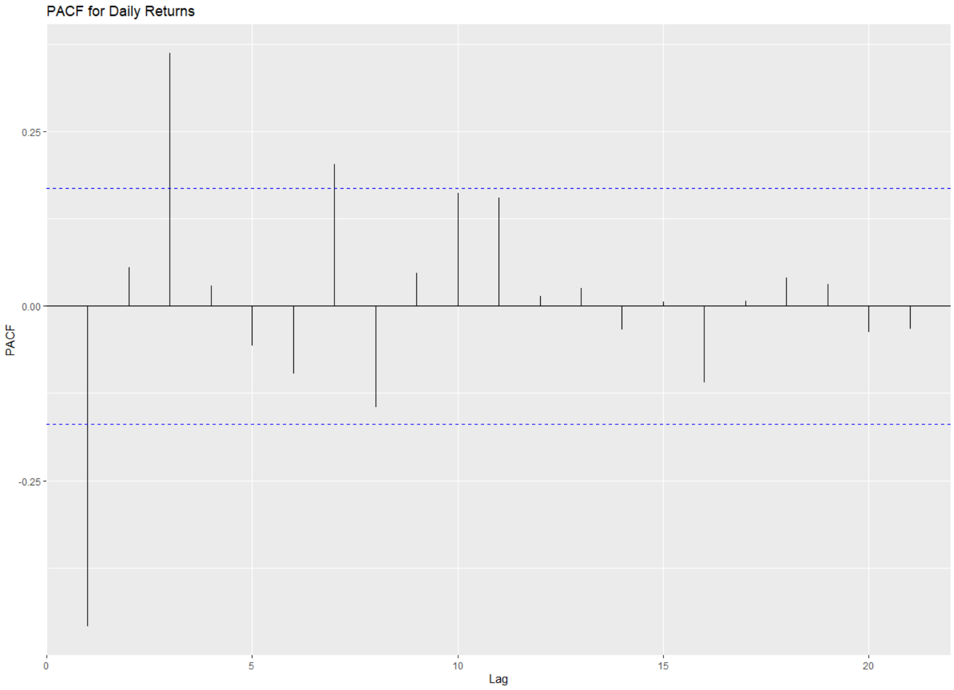


Figure 8: ACF for daily returns

Both the PACF and ACF tail off relatively quickly when examining the daily return series, suggesting an ARMA(p,q) process. Compared to the plots for the intraday returns, there are less significant lags. Thus, the ARMA model for the daily returns probably has a lower lag order than the model for the intraday returns.

In R, there is a simple function that automatically determines the best lag orders for the respective series. We apply said function and receive an ARMA(5,0), i.e. an AR(5) process with an $AIC = -100602.7$, for the intraday returns and an ARMA(3,1) for the daily returns with an $AIC = -693.26$.

$$y_t^{intra-day} = c + \varepsilon_t - 0.8242\, \alpha_{t-1} - 0.6507\, \alpha_{t-2} - 0.3983\, \alpha_{t-3} - 0.3220\, \alpha_{t-4} - 0.2180\, \alpha_{t-5}$$

$$y_t^{daily} = c + \varepsilon_t - 0.6007\alpha_{t-1} + 0.0816\alpha_{t-2} + 0.3360\, \alpha_{t-3} - 0.9099\, \varepsilon_{t-1}$$

As a last step in the Box-Jenkins methodology, we test the residuals to see if they resemble white noise. First, we simply take a look at the plotted squared residuals and its ACF and PACF plots for ARMA(5,0) and ARMA(3,1) as determined above in an effort to evaluate the remaining autocorrelation.

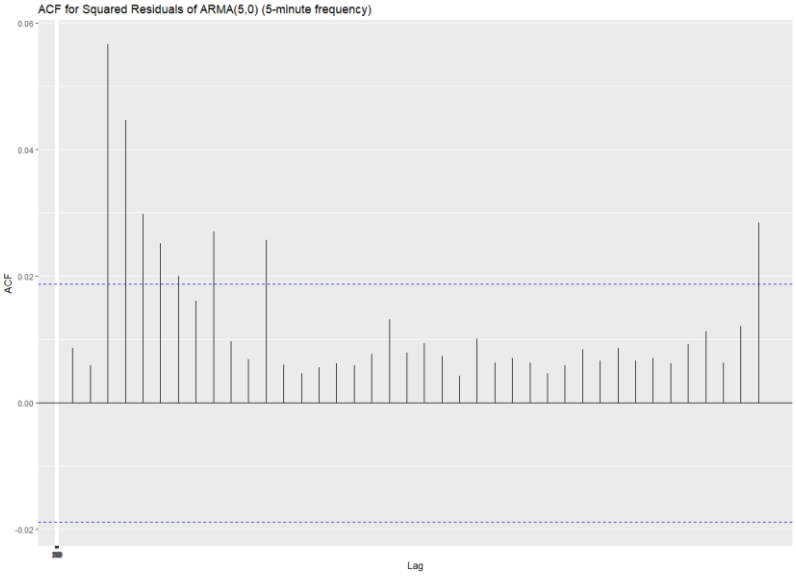


Figure 9: ACF for squared residuals of ARMA(5,0) (5-minute frequency)

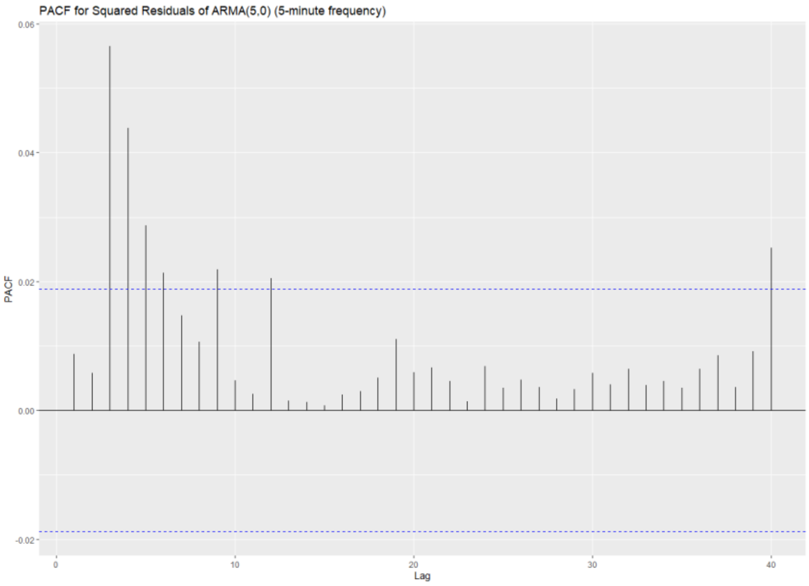


Figure 10: PACF for squared residuals of ARMA(5,0) (5-minute frequency)

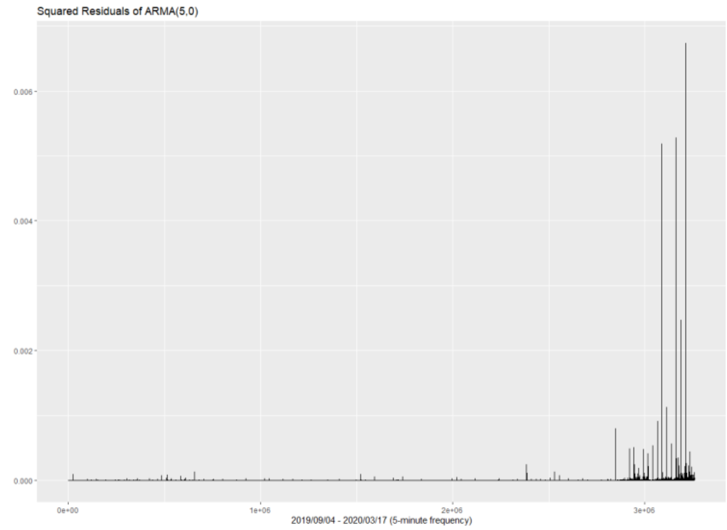


Figure 11: Squared residuals of ARMA(5,0)

From the ACF and PACF of the intraday squared residuals of an ARMA(0,5) model, it is obvious that a considerable amount of autocorrelation still remains even after fitting the first model as there are seven and eight significant lags, respectively. The plot of the squared residuals shows volatility clustering in the residuals towards the end of the series. If the model were sufficient, we would expect the residuals to resemble white noise. As this is clearly not the case, we have to work on improving our model.

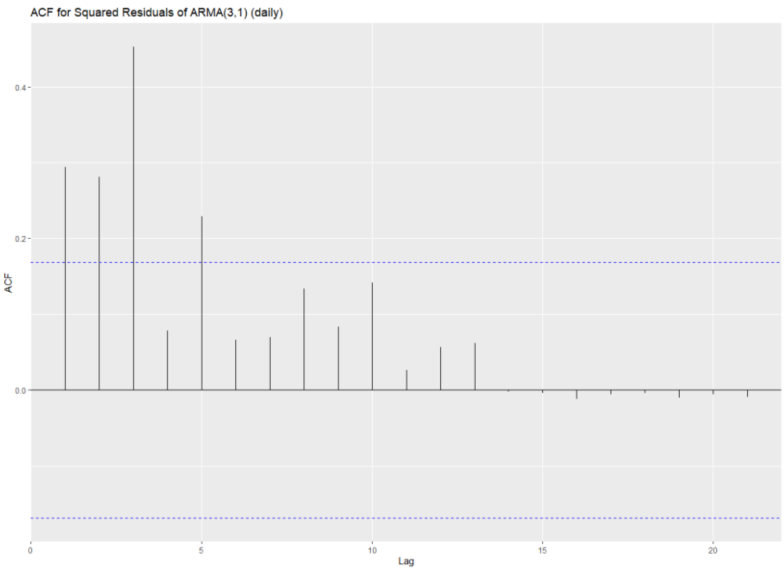


Figure 12: ACF for squared residuals of ARMA(3,1) (daily)

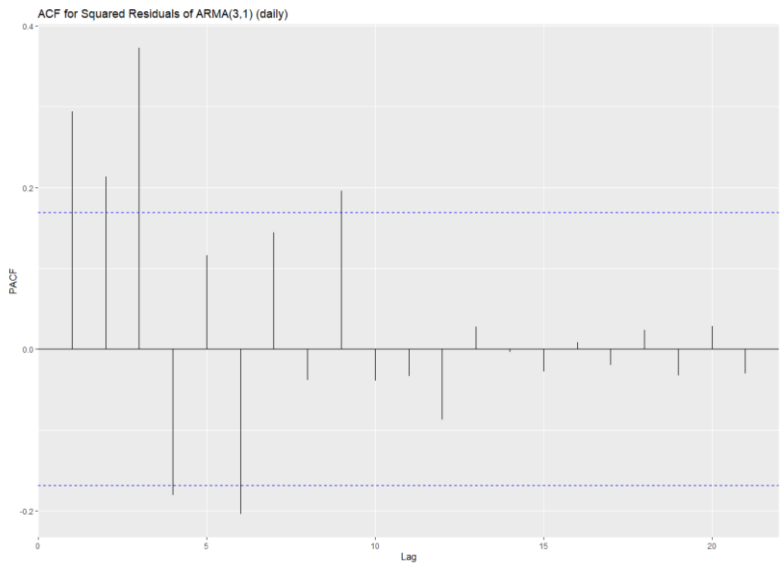


Figure 13: ACF for squared residuals of ARMA(3,1) (daily)

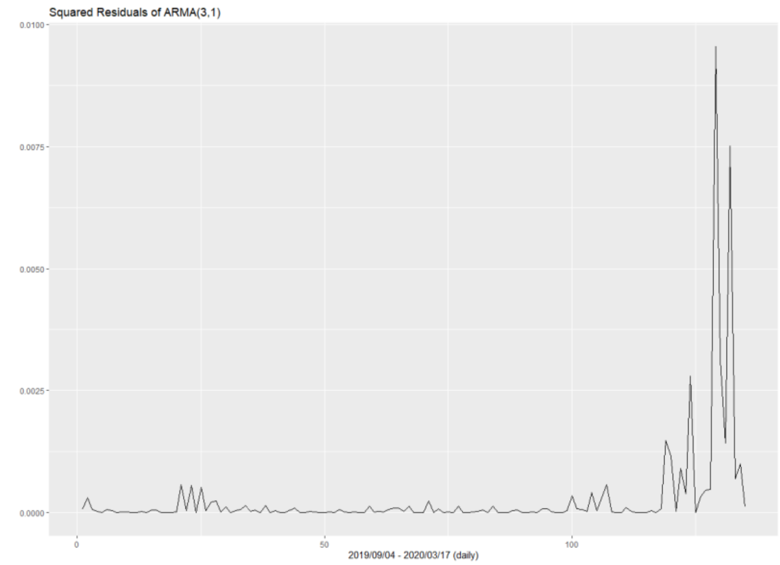


Figure 14: Squared residuals of ARMA(3,1)

Unsurprisingly, the squared residuals of the ARMA(3,1) for the daily return series show a similar pattern. Lags 1 – 3 of the ACF and PACF are clearly significant and the plotted squared residuals show considerable volatility clustering.

In addition to the visual examination of the plotted squared residuals and the ACF and PACF plots, we can conduct two tests in order to test for autocorrelation in the residuals.

For this, we apply the Box-Pierce test to the residuals of both our models. The Box-Pierce test tests the null hypothesis of white noise against the alternative hypothesis of a remaining underlying pattern. The p-value for our intraday return series amounts to 0.8772 and leads us to keep the null hypothesis of random values for the residuals. The p-value for our daily return series is 0.4632 and suggests the same.

Another option to test the residuals is the so-called Autoregressive Conditional Heteroscedasticity Lagrange multiplier (ARCH LM) test that evaluates significance of Autoregressive Conditional

Heteroscedasticity (ARCH) effects . The null hypothesis states that all the autocorrelation is captured by the employed model. The alternative hypothesis is that there is still remaining autocorrelation in the squared residuals of the examined series. After employing the test, we receive a p-value < 0.01 for the intraday returns and a p-value of < 0.01 for the daily returns. This means that we have to reject H0 and assume that both the ARMA(5,1) and the ARMA(3,1) model do not account for all autocorrelation in the daily log return series.

In addition to the plots of the squared residuals and the Box-Pierce test, the ARCH LM test (Lee 1991) leads to the conclusion that our models can still be improved. For this, we fit a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model to capture all the autocorrelation in the series.

Estimating an ARMA-GARCH model

It therefore makes sense to apply a GARCH(p,q) model to remove the remaining autocorrelation in both return series in order to be able to forecast the volatility of the series. In accordance with a wide range of econometrics literature and a variety of empirical proofs of the model, we choose a GARCH(1,1) model to estimate the parameters of the model.

In summary, we now have an ARMA(5,0)-GARCH(1,1) model for the intraday return series and an ARMA(3,1)-GARCH(1,1) model for the daily return series. In order to be able to forecast the conditional variance, we additionally need to pass each model the conditional variance for each day that we have intraday data for. For this purpose, we use the daily specification of ARMA(3,1)-GARCH(1,1) and perform a rolling forecast over the whole data set of daily returns. The resulting fitted conditional variance will later be passed to fit the models for the training set.

Then, we also calculate the realized variance RV_t for each day as it is an unbiased estimator of the conditional variance. It is calculated as the sum of squared intraday returns $r_{t,m}$:

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

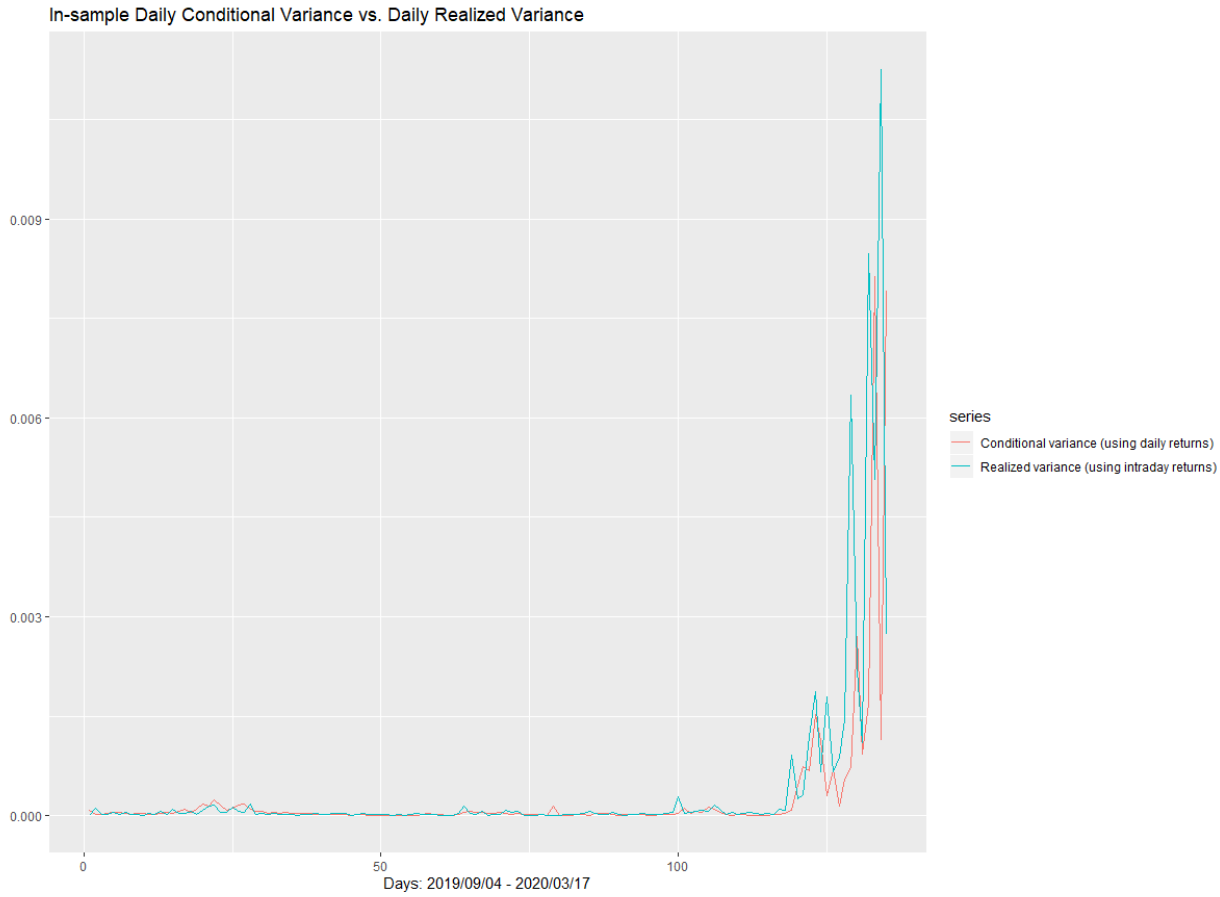


Figure 15: In-sample daily conditional variance vs. daily realized variance

As can be seen in the plot above, the daily realized covariance and the daily conditional variance follow approximately the same pattern. RV_t slightly lags behind.

The daily conditional variance and the intraday ARMA(5,0)-GARCH(1,1) model *can then be put together for the intraday model specification* to fit the intraday and daily models for the training set.

Forecasting

We use the period from 2019/09/04 up to 2020/03/09 as a training set and the week from 2020/03/09 up to and including 2020/03/13 as our test set. The two fitted models to estimate the conditional variance consist of the intraday model specification, the daily conditional variances, the time frame of the training set and the intraday or daily returns, respectively.

$$h_t = 0.012425 + 0.037442\ \varepsilon_{t-1}^2 + 0.962559\ h_{i-j}$$

We apply the models to the test set and plot the results:

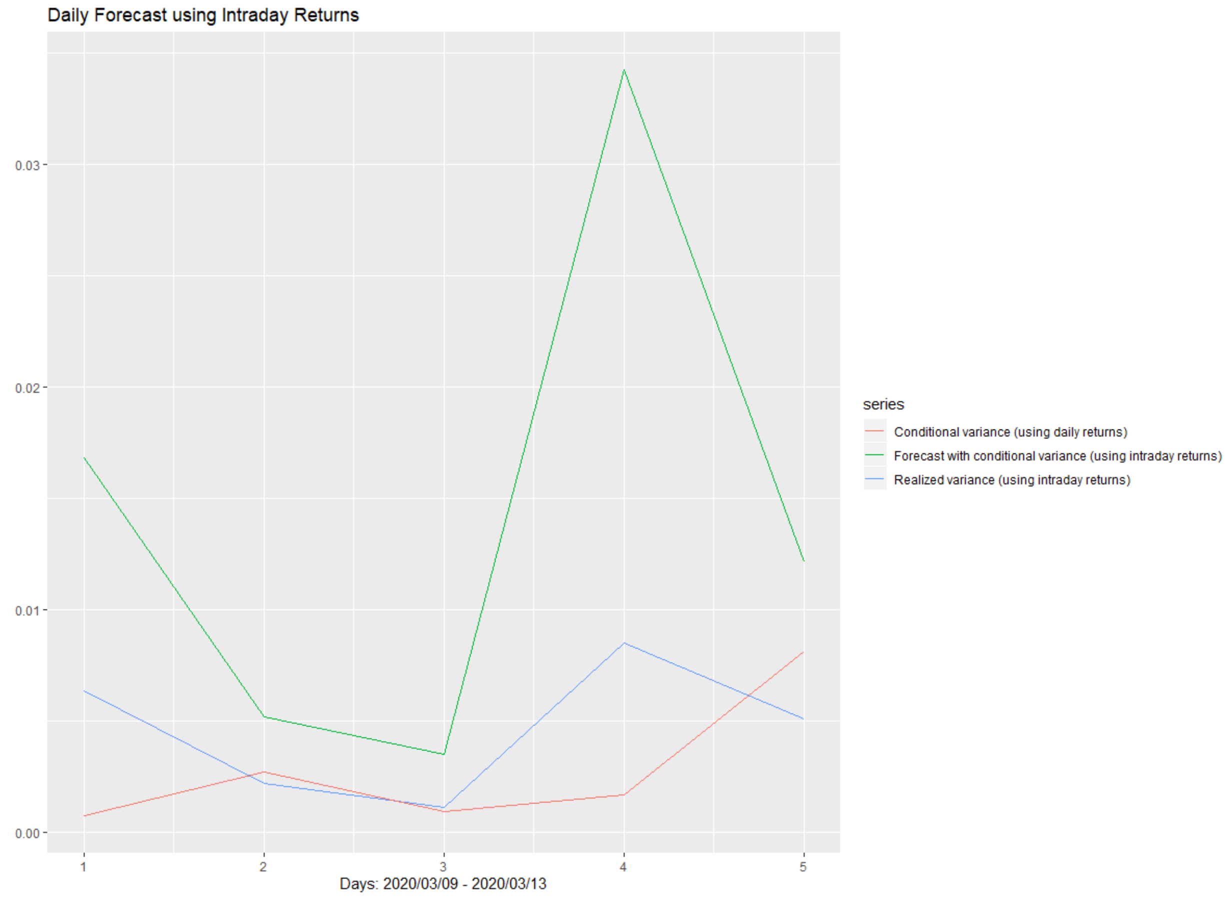
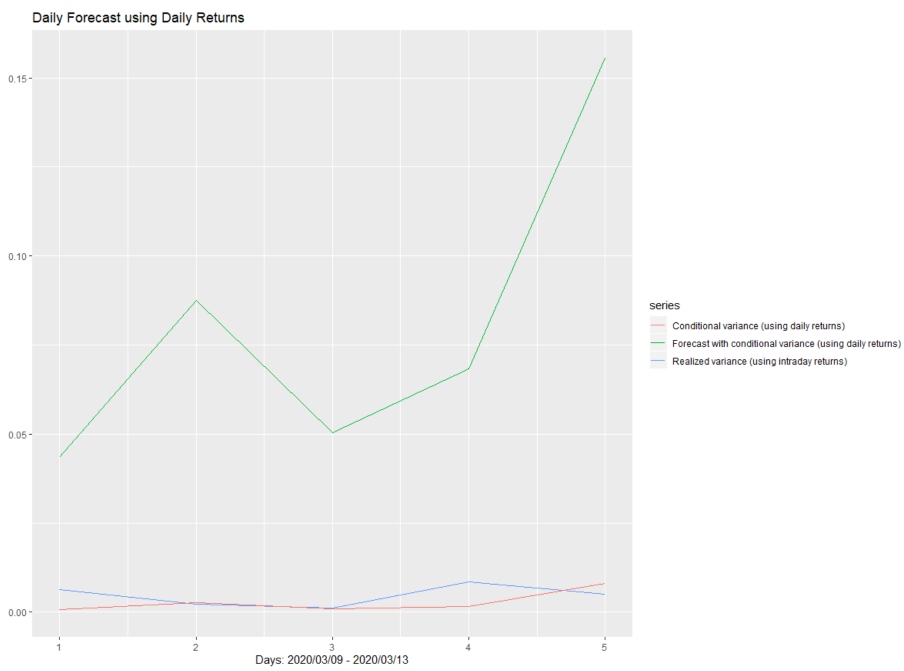


Figure 16: Daily forecast using intraday returns

The forecast of the conditional variance using the model with intraday return data nicely follows the movement of the realized variance. This makes sense because the realized variance is calculated exclusively using intraday returns. Interestingly, the forecast does not follow the movement of the conditional variance as well.



The forecast using daily return data shows a stark difference to both the fitted conditional variance and the daily realized variance. In this case, it follows the development of the conditional variance more closely than that of the realized variance.

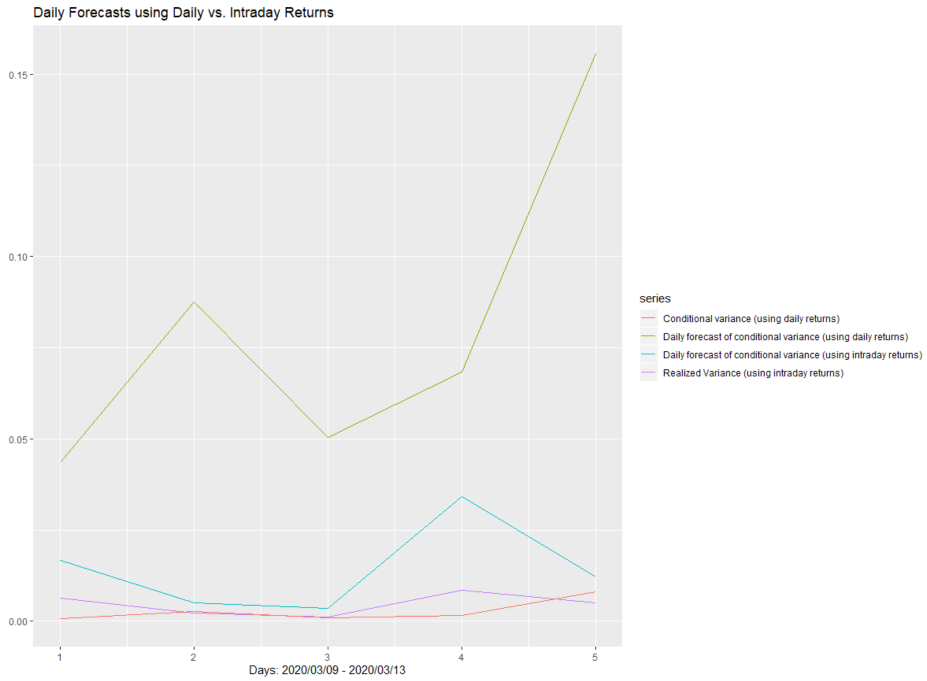


Figure 17: Daily forecast using daily returns

When plotting both forecasts and the realized and conditional variance together, it is obvious that the forecast improves with intraday data as compared to daily data. To substantiate those results, we calculated four measurements to determine the quality of our forecast. The mean squared error (MSE), the root-mean-square-error (RMSE) the mean absolute error (MAE) and the mean absolute percentage error (MAPE) all provide a measure of accuracy by comparing forecasting errors of different models. Hence, the aim is to have as low values as possible as that means that the errors are low.

Mean Absolute Error (MAE)

$$MAE_t = \frac{1}{|T|}\sum_{t=1}^{T} |\hat{\sigma}_t^2 - h_t|$$

Mean squared error (MSE):

$$MSE_t = \frac{1}{|T|}\sum_{t=1}^{T} (\hat{\sigma}_t^2 - h_t)^2$$

Root mean squared error (RMSE):

$$RMSE_t = \sqrt{\frac{1}{|T|}\sum_{t=1}^{T} (\hat{\sigma}_t^2 - h_t)^2}$$

Mean absolute percent error (MAPE):

$$MAPE_t = \left(\frac{1}{|T|}\sum_{t=1}^{T}\frac{|\hat{\sigma}_t^2 - h_t|}{|h_t|}\right) * 100$$

	Forecast with intraday returns	Forecast with daily returns
MSE	0.00016753	0.00746243
RMSE	0.01294334	0.08638536
MAE	0.00973520	0.07642995
MAPE	0.12929428	1.30754206

As can be seen in the table above, the forecast using intraday returns beats the forecast with daily returns in every single measurement. In combination with the plotted forecasts, it is clear that using intraday data to forecast conditional variance and therefore volatility constitutes a considerable improvement to the forecast.

Conclusion

In this paper, we discussed the theoretical foundations for the GARCH model, volatility and frequency and used them to forecast the variance of the S&P 500. We did so by calculating the returns, testing for stationarity in the time series and following the Box-Jenkins methodology. Then, we estimated a GARCH(p,q) model and the daily conditional variance and applied them to the time series. We produced two different forecasts for the conditional variance of the S&P 500 using an intraday and a daily time series, respectively, and compared them according to quality criteria, mainly RMSE.

Our goal was to show how the sample frequency of a time series improves or deteriorates the variance forecast of the S&P 500. The plotted forecasts show that intraday data greatly improves the variance forecast as compared to a daily price process. The quality criteria MSE, RMSE, MAE and MAPE confirm this result.As already mentioned in the abstract, this paper confirms that without a doubt, forecast accuracy increases with a higher sampling frequency.

Our results can be applied in risk management and to improve portfolio theory. Due to its volatile structure, our methodology lends itself to an application in the crypto market. Over the course of our research, we discovered the Bloomberg Crypto Galaxy Index that covers approximately the same percentage of the market as the S&P 500 as chosen by us. Research in this area might be fruitful.

References

Andersen, Torben G.; Bollerslev, Tim; Francis, Diebold X.; Labys, Paul (2001): The Distribution of Realized Exchange Rate Volatility. In: *Journal of the American Statistical Association* (96), S. 42–55.

Banulescu, Georgiana-Denisa; Candelon, Bertrand; Hurlin, Christophe; Laurent, Sébastien (2014): Do We Need Ultra-High Frequency Data to Forecast Variances? Online verfügbar unter https://halshs.archives-ouvertes.fr/halshs-01078158, zuletzt geprüft am 02.04.2020.

Bollerslev, Tim (1986): Generalized Autoregressiv Conditional Heteroskedasticity. In: *Journal of Econometrics* (31), S. 307–327.

Bucci, Andrea (2017): Forecasting realized volatility: a review. Online verfügbar unter https://mpra.ub.uni-muenchen.de/83232/.

Cont, Rama (2011): Statistical Modeling of High-Frequency Financial Data. In: *IEEE Signal Process. Mag.* 28 (5), S. 16–25. DOI: 10.1109/MSP.2011.941548.

Curci, Giuseppe; Corsi, Fulvio (2004): A Discrete Sine Transform Approach for Realized Volatility Measurement.

Duffie, Darrell; Protter, Philip (1992): From Discrete- to Continuous-Time Finance: Weak Convergence of the Financial Gain Process. In: *Mathematical Finance* (2).

Engle, Robert (2001): GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics. In: *Journal of Economic Perspectives* (4), S. 157–168.

Engle, Robert F. (1982): Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. In: *Econometrica* (50), S. 987–1007. Online verfügbar unter https://www.jstor.org/stable/1912773.

Fama, Eugene F. (1965): The Behavior of Stock-Market Prices. In: *The Journal of Business* (38).

French, Kenneth R.; Schwert, G.William; Stambaugh, Robert F. (1987): Expected stock returns and volatility. In: *Journal of Financial Economics* 19 (1), S. 3–29. DOI: 10.1016/0304-405X(87)90026-2.

Lee, John H.H. (1991): A Lagrange multiplier test for GARCH models. In: *Economics Letters* (37), S. 265–271.

Panorska, Anna K.; Mittnik, Stefan; Rachev, Svetlozar T. (1995): Stable GARCH Models for Financial Time Series. In: *Applied Mathematics Letters* (8), S. 33–37.

Shiryaev, Albert Nikolayevich (1999): Essentials of Stochastic Finance: Facts, Models, Theory: facts, models, theory. Singapore: World Scietific Publishing Co. Pte. Ltd.

Springer, Berlin, Heidelberg (Hg.) (2012): Handbook of Computational Finance. Numerical Solution of Stochastic Differential Equations in Finance. Unter Mitarbeit von In: Duan JC., Härdle W., Gentle J. Berlin, Heidelberg: Springer (Springer Handbooks of Computational Statistics).

Stoffer, David S.; Shumway,, Robert H. (2017): Time Series Analysis and Its Applications: With R Examples: Springer Texts in Statistics.

Uwe Hassler (2003): Zeitabhängige Volatilität und instationäre Zeitreihen. Zum Nobelpreis an Robert F. Engle und Clive W.J. Granger. In: *Wissenschaft für die Praxis*, S. 811–816.