AUTO-REGRESSIVE METHODS IN PREDICTIVE VOLUME MODELING

YANHAO DONG, BEN GOLDMAN, XIAOTIAN LI, FANDI MENG, SUSHAN ZHAO

Abstract. An overview of a suite of improvements made over the two non-live updating components of the Bloomberg Quintet Model. These autoregressive models include a prediction for daily volume and for intraday volume. A further discussion is given to the unsuccessful model variants. This culminates in an enumeration of multiple avenues for further exploration.

Contents

1. Introduction and Data Processing 2
2. Industry Standard: The Bloomberg Model 3
3. Methods for Daily Model 4
4. Intraday Volume Prediction with Historical Price Volatility 11
5. Other Tested Models 15
6. Areas for Further Exploration 19
7. Acknowledgements 24
References 25

Date: June, 2022.

1. Introduction and Data Processing

Volume predictions give traders the ability to anticipate and develop strategies around advantageous times to purchase or dump large quantities of stocks. The standard model for these predictions is used given by Bloomberg (described below) which is used to determine an upper bound on the necessary amount of stock needed to have available at any given time. In essence, by improving on the Bloomberg model, traders can anticipate the excess stock that will be available at a given time and use this to their advantage.

In approaching the problem, we adopted a purely statistical approach with an emphasis on autoregressive methods. While Machine Learning algoirthms can improve on performance, it is in our interest to develop an understanding of the precise factors and relationships that improve the accuracy of prediction.

In addition to the original Bloomberg paper, we used TAQ data from the Wharton Research Data Services for records of Bloomberg terminal data. For intraday data, we pulled the data at 1 minute BIN intervals – these are the same interval sizes used in the original Bloomberg paper, and tend to be the standard for autoregressive or functional regressive methods which require a smoother curve. For examination of the minute-to-minute intraday valuation (i.e. price) of stocks, we queried in one-minute intervals by the average (minutely) price of the stock. It should be noted that initial attempts to define the minute-to-minute price by the last completed BIN were unsuccessful as they required copious amounts of millisecond data.

The stocks used for testing were varied based on the desired liquidity. For the study of more liquid stocks, tests were done on samples from the S&P 500 and the Russell 1000 – note that a representative random sample of one third of each index was used as the precise testing set. Other indices that were examined were the QQQ, IVV, VTI, VOO, and SPY depending on the desired level of liquidity. Likewise, the VIX price volatility index was used to compare volume(s) against measurements of volatility. For less-liquid or illiquid stocks, tests were performed on the Russell 2000 stocks (more precisely, another representative sample of the Russell 2000).

For autoregressive models, a training and testing set was necessary to build the model and assess performance. We train the model on training set (usually around 2 years of prior data), select the model with best hyper-parameters on the validation set, and get model evaluation results by evaluating it on the test set. Specifically, start with certain $t \leq T$, given information $\mathcal{F}_t$ and train the model using information in $\mathcal{F}_t$ to estimate parameters (e.g., maximizing the likelihood for ARMA model, minimizing mean square error for linear regression). Then, record one-step-forward predictions, $\hat{y}_{t+1}$, and repeat this procedure until $t = T - 1$ (note that we have generated predictions from $\hat{y}_{t+1}$ to $\hat{y}_T$).

The evaluation metric we are using is mean absolute percentage error, which is given by (for daily models)

$$\frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{y_i}, \tag{1.0.1}$$

where N is the number of predicted days in daily volume prediction. This metric was chosen by our research team since it does not give higher weight to higher volume trading days. In our case, we have $N_{days} = T - t$. An analogous computation is made for intraday volume, where $N \equiv N_{\text{minutes}}$, instead.

2. Industry Standard: The Bloomberg Model

The Quintet Volume Projections from Bloomberg are predictions for Close Auction Volume, Daily Total Volume, and Intraday Volumes with separate methods for Liquid Stocks, and Illiquid Stock [M17]. For our project, we were primarily concerned with liquid stocks and liquid models (though tests were performed to confirm that these methods were most effective on liquid stocks).

2.1. **Daily Total Volume Model.** Let y_t be the logarithmic daily volume. The ARMA$(1,1)$ model is used as baseline for predicting y_t, which takes the form

$$y_t = \varphi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}, \tag{2.1.1}$$

where φ, θ are parameters. Throughout this paper, we will continue to use y_t to represent the log daily volume at day t, and $\hat{y}_t$ as the corresponding prediction.

2.2. **Close Auction Volume Model.** A 20-day geometric average is used as a base prediction of the close auction volume, which has the form of

$$\hat{y}_t = \exp\left(\frac{1}{20}\sum_{i=1}^{20}\log(y_{t-i})\right). \tag{2.2.1}$$

2.3. **Intraday Volume Curve Model.** Let $y_i(t)$ denote the percentage of daily volume of the stock traded at minute, t, on day, i. The baseline model employs a functional regression of $y_i(t)$ on $x_i^1(t), x_i^2(t)$ which correspond to the overnight price gap (recorded between days $i-1$ and i) and the average percentage of volume traded at minute t

$$y_i(t) = \beta_0(t) + \beta_1(t)x_i^1 + \beta_2(t)x_i^2 + \varepsilon_{i,t}$$

This model estimates each minute's trade volume as a percentage of the daily total volume, generating the u-shaped volume smile.

3. Methods for Daily Model

For basic linear models (e.g., ARMA, linear regression) of daily volume, the performance is largely affected by outliers. The data includes several spikes which exacerbate the predictive errors made by the baseline models. We first designed a mechanism to regularize the spikes, and then combined it with the original ARMA$(1, 1)$ model to generate predictions. The new model (named Robust ARMA), has shown potential in outperforming ARMA by regularizing outliers. We further utilize the mechanism to perform residual regression, which incorporates the predictive power of overnight price gap and VIX.

In this section, we will first introduce Robust ARMA model and residual regression seperately, and we will conclude with the combined model and the requisite evaluation results.

3.1. **Robust ARMA.**

Motivation. The gray curve in figure 1 is the log daily volume of UPS, and we can see that there are several spikes in it. The red curve is ARMA$(1, 1)$ model's prediction and also our baseline, shown us the spikes largely affect the predictions on the subsequent days. Therefore, we want to design a mechanism that can regularize the spikes, and thus improve prediction accuracy.

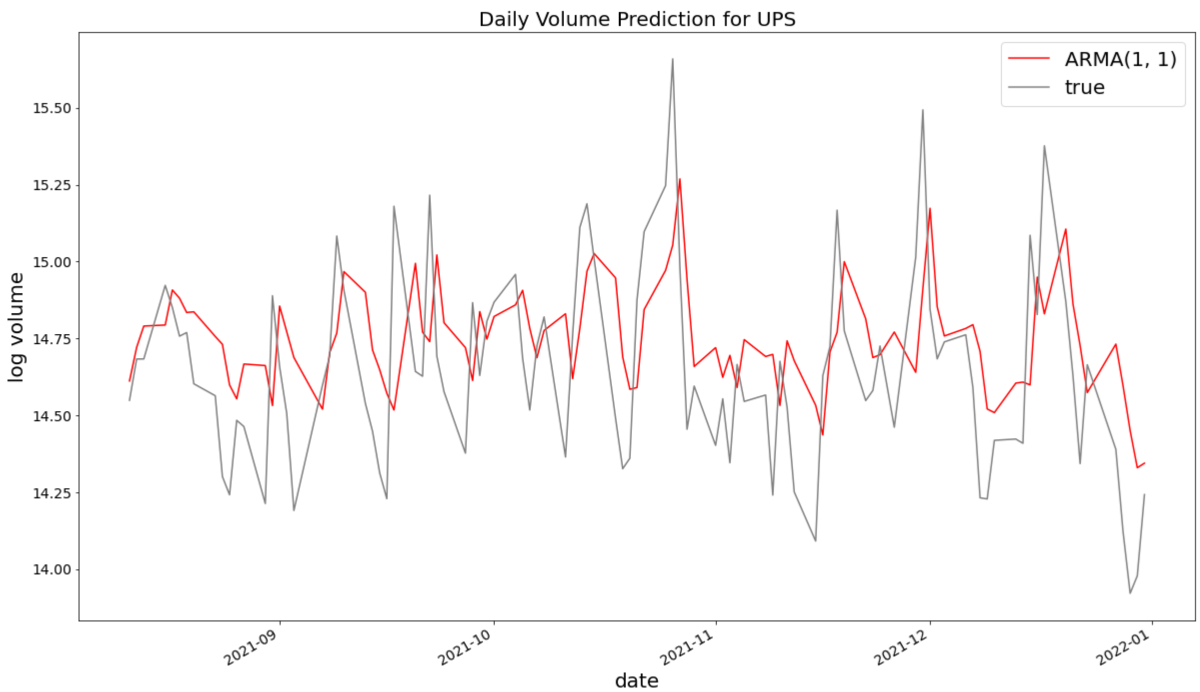


Figure 1. ARMA (1,1) Predictions for UPS Daily Volume in 2021

Generally, a spike is a point that are "largely" greater than the points before and after. Handling spikes should be careful, because the spikes might also contain information. Since we use ARMA model with order $(1, 1)$, we can safely define our spike to be a point that are "largely" greater than the former point and the latter point.

Definition 3.1.1. Let T be the last day of our log volume series. Given thresholds T_f, T_b, then for every $t < T$, if $y_t - y_{t+1} \geq T_f$ and $y_t - y_{t-1} \geq T_b$, we call y_t a *spike*. We also use *forward differential* and *backward differential* to represent $y_t - y_{t+1}$ and $y_t - y_{t-1}$, respectively.

With the definition of spikes, we can regularize them accordingly. Before the regularization, we still need to find thresholds for spikes. One way to determine the thresholds is to use the quantile information of absolute backward differential series. In our experiment, we construct series $B_t = |y_t - y_{t-1}|, t \leq T$, then define threshold

$$T_f = T_b = B = 1.5 \times \left(B_t^{(0.75)} - B_t^{(0.25)}\right), \tag{3.1.2}$$

where the superscript $(\cdot)^{(0.75)}$ represents series $(\cdot)$'s 0.75-quantile. The multiplier 1.5 and $0.75, 0.25$ quantile in (3.1.2) are chosen subjectively, a grid search for parameter tuning does not give huge improvement over prediction in our case, thus not worth implementing it when balanced with increasing the model complexity. Note that the definition here only considers upward spikes, given that it's more risky to overestimate volume than underestimate, and in experience generally upward spikes are a lot more than downward spikes.

Following the convention in stochastic calculus, we use filtration $\mathcal{F}_t$ to represent the accessible information when performing model training. For each spike in model training data, we use the average of former point and latter point to replace the spike. Basically, we draw a straight line between the former point and the latter point. For the last day's volume y_t, we do not have the information of y_{t+1} and thus it's not possible to compute forward differential. We then only use one inequality condition in definition 3.1.1 to define spike, and regularize the spike such that we have the backward differential exactly equals to the threshold. The whole process of spike regularization is listed below.

Given $\mathcal{F}_t$, for $i \leq t$, the regularized series is given by

$$y_i' = \begin{cases} \dfrac{y_{i-1} + y_{i+1}}{2}, \text{ if } i < t, y_i - y_{i-1} \geq B, y_i - y_{i+1} \geq B; \\ y_{i-1} + B, \text{ if } i = t, y_i - y_{i-1} \geq B; \\ y_i, \text{ otherwise.} \end{cases}$$

Continuously, we use stock UPS as an example to give a sense of what the regularized series looks like. As we can see in figure 2, the regularized series (blue curve) has smoothed the series at several points, compared with the original series (red curve).

We call ARMA model incorporated with this regularization mechanism Robust ARMA model. This mechanism is used in the further modeling of daily volume prediction, thus we slightly abuse the notation y_t to represent log volume with regularization and abandon notation y_t'.

3.2. **Incorporating Price Gap and Volatility.** A natural step in the research was to search for other signals in the data that could improve the accuracy of the model. The baseline intraday model, for instance, makes use of the overnight price gap measurement (i.e. $OP_t - CP_{t-1}$ where OP is opening price, CP is closing price). However, the baseline daily model neglects to use any data beyond the previous day's volume. By normalizing the price gap accordingly, we observed signals for sufficiently high price gaps. In general, strong shifts in price (both positive and negative, alike)

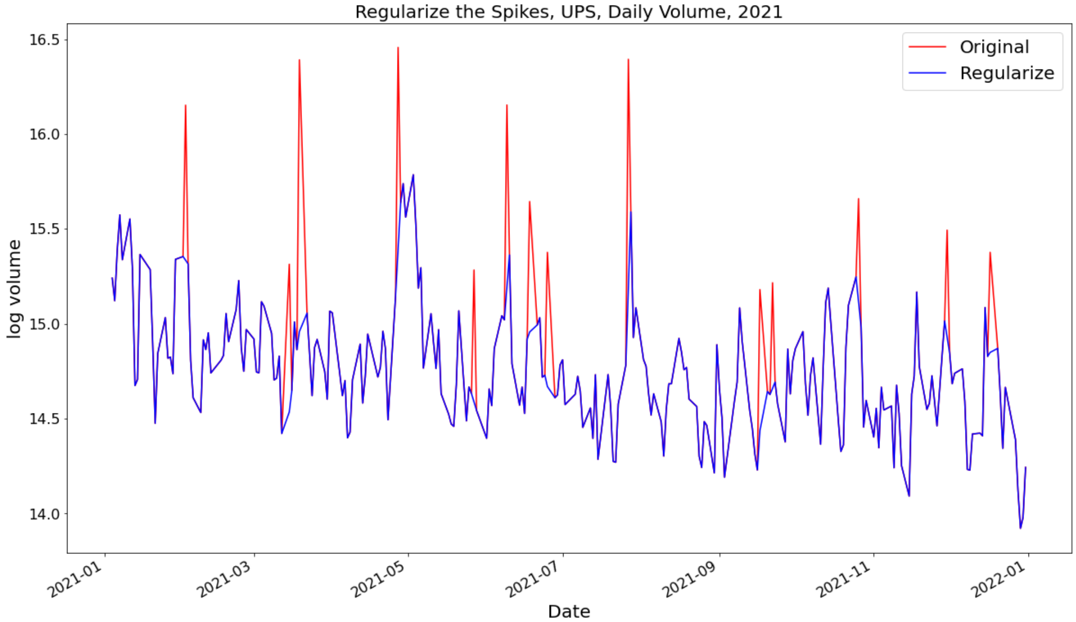


FIGURE 2. 2021 Daily volume for UPS before and after regularization

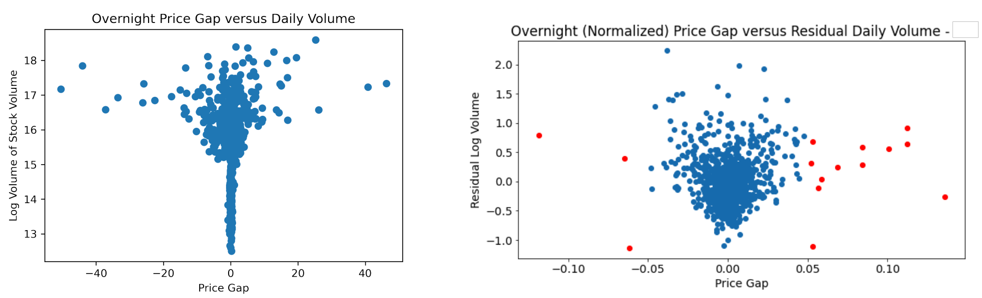


FIGURE 3. The relationship of residual log volume with respect to overnight price gap (right) and normalized price gap (left) on MRNA

would lead to an increase in the volume.

An initial test displayed that the residual logarithmic volume, $\varepsilon_t = y_t^{\text{actual}} - y_t^{\text{baseline}}$, roughly increased with the size of the price gap. However, this signal is clearest when we examined the residuals relative to a normalized price gap, $p_t \equiv \frac{OP_t - CP_{t-1}}{CP_{t-1}}$ (shown in Figure 3). Even with these substantial improvements, the noise in the data taken for insufficiently large price gaps was inconclusive. We therefore admitted only a certain threshold of price gaps when looking for a trend (shown in Figure 3 in red). The most simple approximation of this relationship is an absolute-value curve, achieved by taking a regression of y_t on $|p_t|$.

A similar, albeit less tedious, process was used to analyze the VIX volatility index and residuals. It did not require normalization nor a threshold.

As documented in subsection 5.1, a number of approaches to adding a predictive term into the ARMA were tested, but the most successful method was to regress further on residuals.

Definition 3.2.1. Let y_t be the logarithmic stock volume. A traditional ARMA Model gives

$$y_t = \varphi y_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t,$$

where θ, ε are computed by a maximum likelihood algorithm. Our model assumes $\varepsilon_t \sim |p_t| + v_{t-1}$ where $p_t = \frac{OP_t - CP_{t-1}}{CP_{t-1}}$ is normalized price gap and v_t is the VIX volatility index. To account for the threshold on price gap, we worked only with price gaps that were in the top 90% of measurements. Therefore,

$$\varepsilon_t = \phi_1|p_t| + \phi_2 v_{t-1} + \eta_t,$$

where η_t is the residual with price and volatility effects removed (i.e. the error given by this regression). Our final model is given by now incorporating the robust ARMA and these residual terms,

$$y_t = \varphi y_t^{\text{robust}} + \phi_1|p_t| + \phi_2 v_{t-1} + \theta\varepsilon_{t-1} + \eta_t.$$

3.3. **Performance of the Daily Model.** We first tested Robust ARMA model on QQQ index, S&P 500 and Russell. Table 3.3.1 is the statistics of prediction errors of ARMA and Robust ARMA. As it is shown, the error distribution of the Robust ARMA$(1,1)$ model's predictions has a smaller mean and a smaller standard deviation. The results of smaller error mean of Robust ARMA model are universal, here we mainly focus our attention on the testing results on Russell, since the Russell stocks set has more diversity over liquidity.

TABLE 3.3.1. Mean and variance of each model's prediction error, daily volume prediction on S&P500, 2021

	ARMA$(1,1)$	Robust ARMA$(1,1)$
mean	0.026844	0.026421
std	0.027668	0.027625

The Russell stocks have been inherently classified into two sets, which are Russell 1000 set and Russell 2000 set. The Russell 1000 stocks set is generally considered to be consisted of liquid stocks, whereas Russell 2000 is less liquid. For both sets of Russell stocks, the Robust ARMA model all has smaller average prediction errors across stocks, which in fact is 59.5% of stocks in total better performed on Robust ARMA. When we are looking at these two separated stocks sets, the percentage goes up to 62.5% on Russell 1000 and goes down to 57.8% on Russell 2000, which shows us our model performs better on liquid stocks than less liquid stocks. The statistics of prediction errors are also given by table 3.3.2 and 3.3.3 for completeness. We want to further identify whether the improvement is valid, and a natural way is to check whether the improvement is significant in the sense of statistics. We have two prediction error series, thus a non-parametric paired difference test is well suited for our purpose. The one-sided Wilcoxon signed-rank test can

TABLE 3.3.2. Mean and variance of each model's prediction error, daily volume prediction on Russell 1000, 2021

Russell 1000	ARMA$(1,1)$	Robust ARMA$(1,1)$
mean	1.832%	1.818%
std	0.00353	0.00356

TABLE 3.3.3. Mean and variance of each model's prediction error, daily volume prediction on Russell 2000, 2021

Russell 2000	ARMA(1, 1)	Robust ARMA(1, 1)
mean	3.054%	3.039%
std	0.0157	0.0161

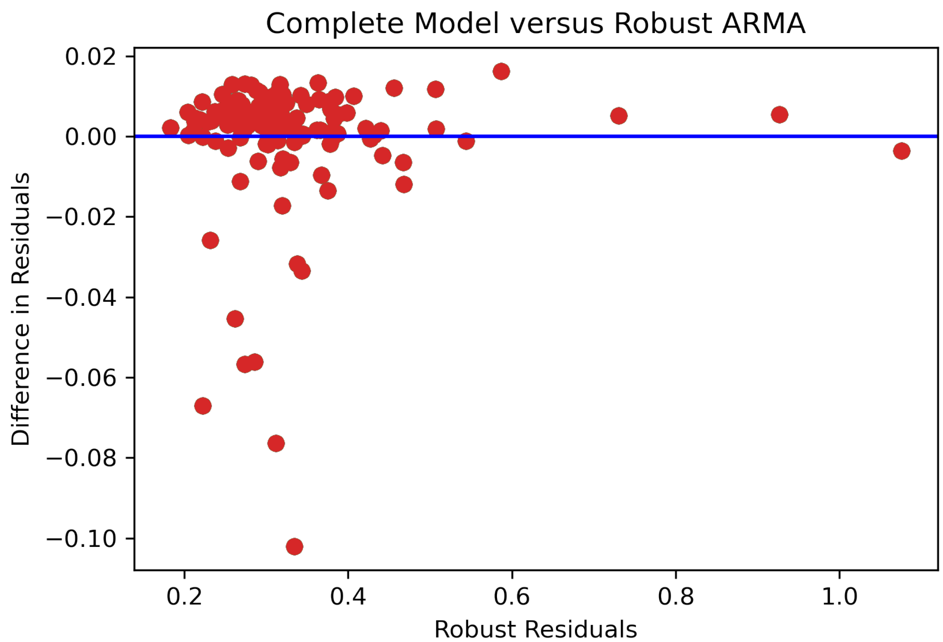


FIGURE 4. The difference between the Robust ARMA error and the complete model versus the Robust ARMA error – taken over the Russell 3000 – stocks above the blue line represent improvements made by the complete model over the Robust ARMA model

be used for testing whether the median of difference series is positive, and we performed it in this case. The Wilcoxon signed-rank test shows us the improvement of Robust ARMA on the Russell 1000 stocks set is significant; for the Russell 2000 stocks set, the p-value is 0.055 which is slightly larger than 0.05, showing us the improvement is not significant.

With the addition of the VIX and price gap terms, the model improves almost uniformly over our testing set. While these improvements are much smaller (the average improvement over the day and standard deviation are almost unchanged), they are consistent. In the plot above, stocks above the line indicate improvements when adding the VIX and Price Gap terms (and vice-versa for those below the line). Note that over 70% of the stocks tested had an improved performance with the addition of the VIX and Price Gap terms. For those that performed worse, the Baseline model tended to perform better than both the Robust ARMA and the complete model.
As indicated above, the original baseline performed best over a third of the Russell 3000 stocks. The complete model, however, performed the best over a plurality of the stocks. When comparing

Table 3.3.4. Best Performing Stocks in Russell 3000

Russell 3000	Baseline	Robust ARMA	Complete Model
Best Model Performance	33%	16 %	51 %

these results with the baseline, roughly two thirds of the measured stocks had improved predictions when using the complete model.

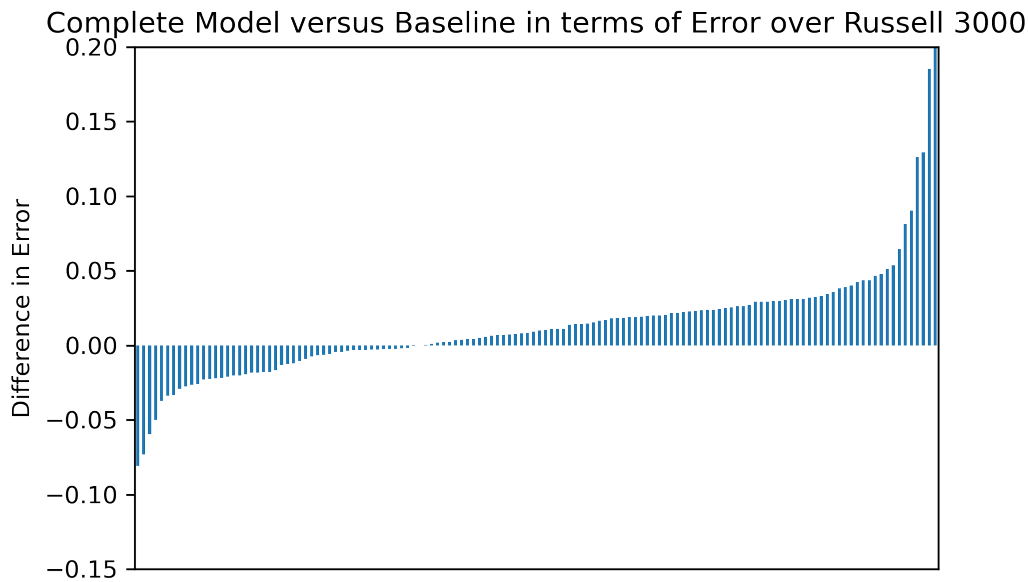


Figure 5. Difference Between the Baseline and Complete Model Errors across the Russell 3000

Our intuition for these improvements is twofold. The Robust ARMA model provides a much more stable base for the addition of the residual terms. It therefore avoids inconsistencies like the inaccurate spikes made by the baseline model in the GCBC example shown below. On the other hand, the volatility and price considerations allow for a more accurate predictor of whether the fluctuations from the smoother, "Robust" curve will be sharp. The complete model is therefore able to more closely emulate the periods of increased/decreased volume that occur over a shorter timespan (particularly where they may be less likely to be properly modeled by the standard ARMA models).

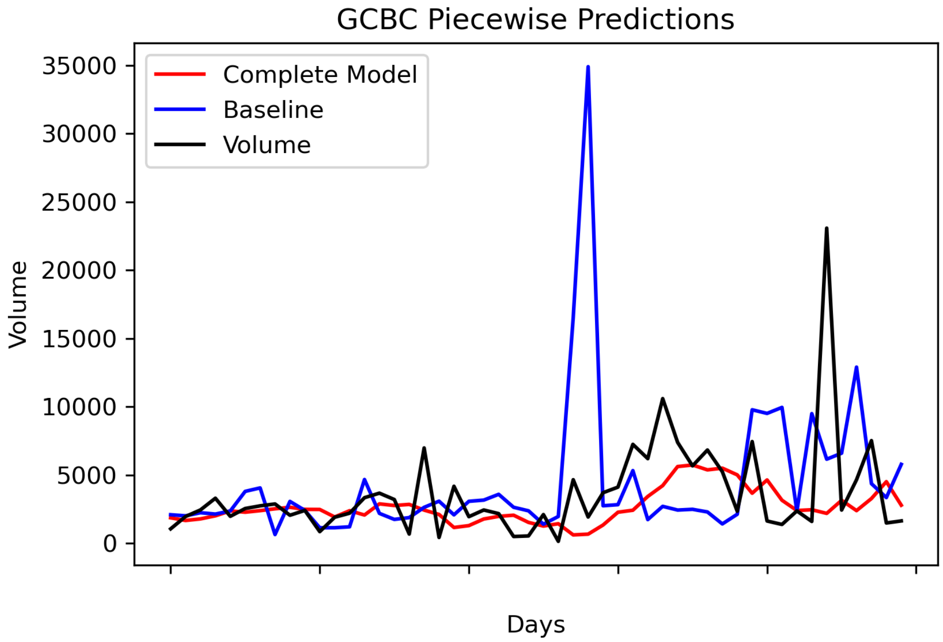


FIGURE 6. Volume Predictions for GCBC over 50 days for Complete Model Versus the Baseline

4. Intraday Volume Prediction with Historical Price Volatility

4.1. **Intraday Model.**

Motivation. Based on the impact of market volatility (VIX) in the daily volume predictions, we want to extend the idea of volatility to intraday volume prediction. We also wanted to intake recent historical information (ie. past few minutes) in the development of an online model.

Definition 4.1.1. The intraday price volatility model is comprised of a linear regression of 15-minute standard deviation in stock price on the residuals of the baseline intraday volume predictions. In doing so, we incorporate sensitivity to recent volatility into the u-curve. Formally, we define the model as follows:

$$\begin{aligned} y_i(t) &= \beta_0(t) + \beta_1(t)x_i^1 + \beta_2(t)x_i^2 + \varepsilon_{i,t}, \ \text{(Baseline Model)} \\ \varepsilon_{i,t} &= \alpha + \beta_3 x_i^3 + \eta_{i,t}, \\ \hat{y}_i(t) &= \alpha + \sum_j \beta_j x_i^j + \eta_{i,t}, \end{aligned} \tag{4.1.2}$$

where

- i is day and t is minute;
- $y_i(t)$ intraday volume size on day i at t minute;
- x_i^1 overnight price gap on day i;
- x_i^2 estimated daily volume percentile on day i;
- x_t^3 historical price volatility based on past 15 minutes rolling standard deviation.

To generate model predictions for a given day i, we first utilize the baseline intraday functional regression model to generate minute-level predictions for the prior day, $i-1$. Then we calculate the residual volume of the baseline model by subtracting the prediction values from the actual volume on day $i-1$. We then linearly regress the residuals on historical price volatility (defined as past 15 minutes rolling standard deviation) to get the price volatility coefficient and the intercept parameter. We then run the baseline intraday model for day i to generate baseline predictions for day i. Finally we add $\beta_3 \cdot x_t^3$, the price volatility coefficient multiplied by the price volatility at minute t, to the baseline predictions to generate our final predictions.

The result are predictions that vary more from the u-curve shape over the period of the day, as we see in Figure 7.

From Figure 8, we observe that the price volatility model has lower errors and thus performs better in the opening and closing hours of the day, where there is more volatility. Thus there is a stronger relationship between price volatility and volume when volatility is higher.

Unfortunately from Figure 9, we see that model performance is inconsistent across days, with the model performing better than the baseline model in just over half of the days we tested for NFLX.

When we test the model on the Russell 1000 sample set of stocks, we find that the price volatility model performs better than the baseline on average for 85% of stocks, as we see in Figure 10. However, as we see from Figure 11, we observe significantly worse results when we test the model

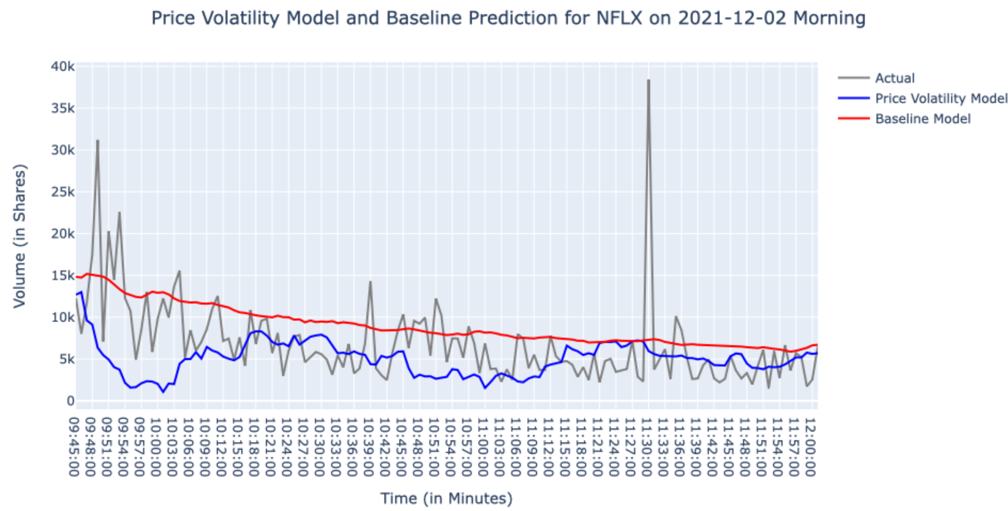


Figure 7. Price Volatility Model vs Baseline Model for NFLX on 2021-12-02)

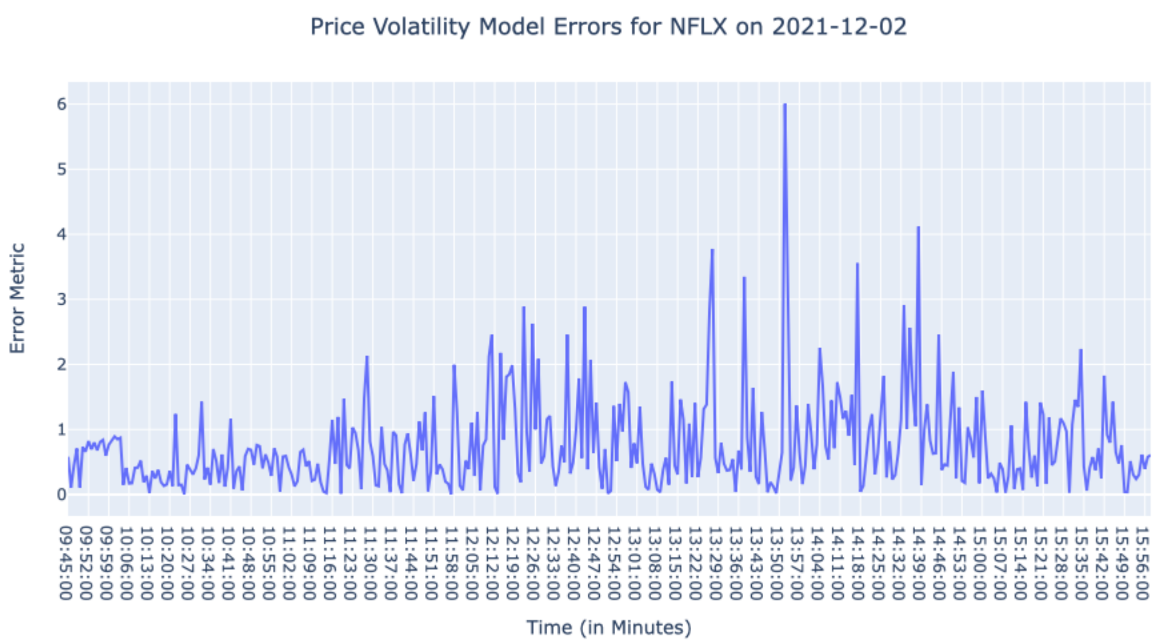


Figure 8. Price Volatility Model Errors for NFLX on 2021-12-02

on the Russell 2000 sample set of stocks, suggesting that liquidity is critical to the accuracy and performance of the model.

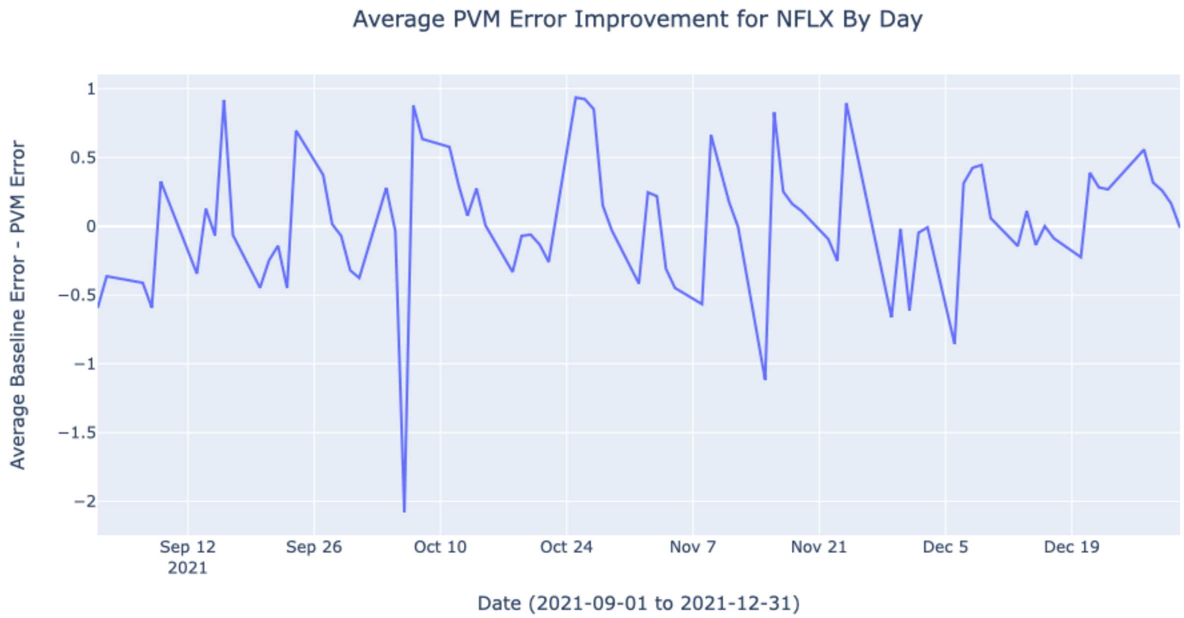


FIGURE 9. Average Improvement Over Baseline by Day for NFLX

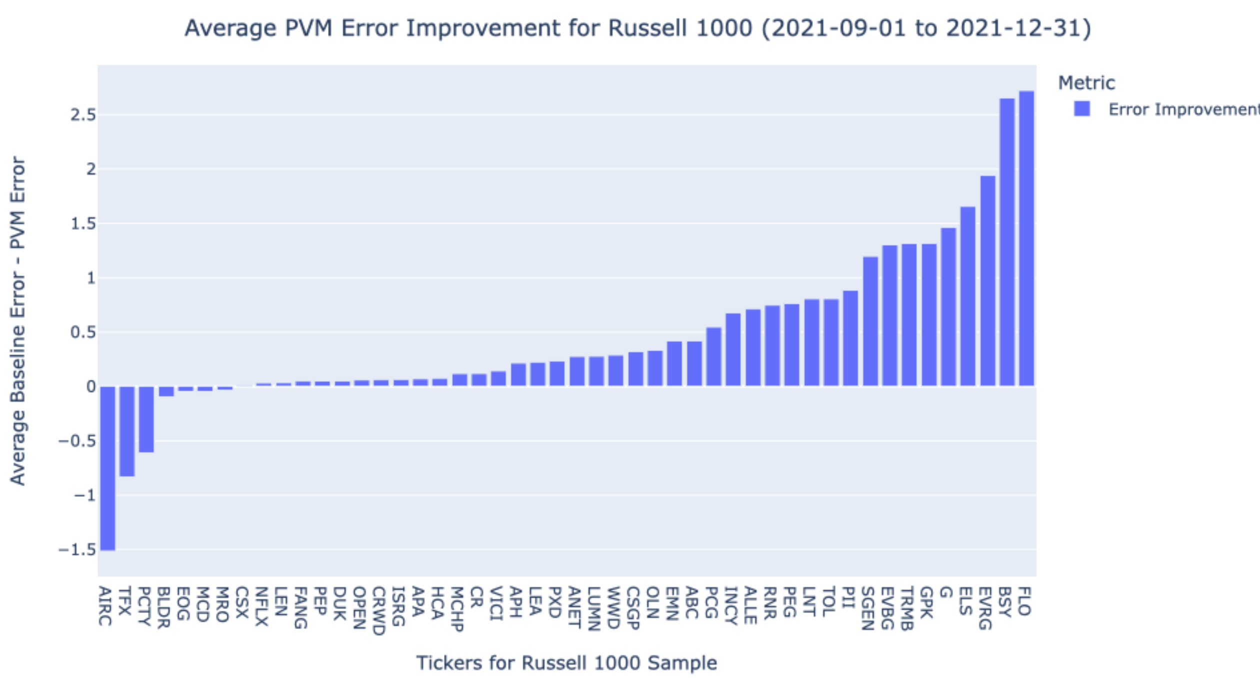


FIGURE 10. Price Volatility Model vs Baseline Model on Russell 1000

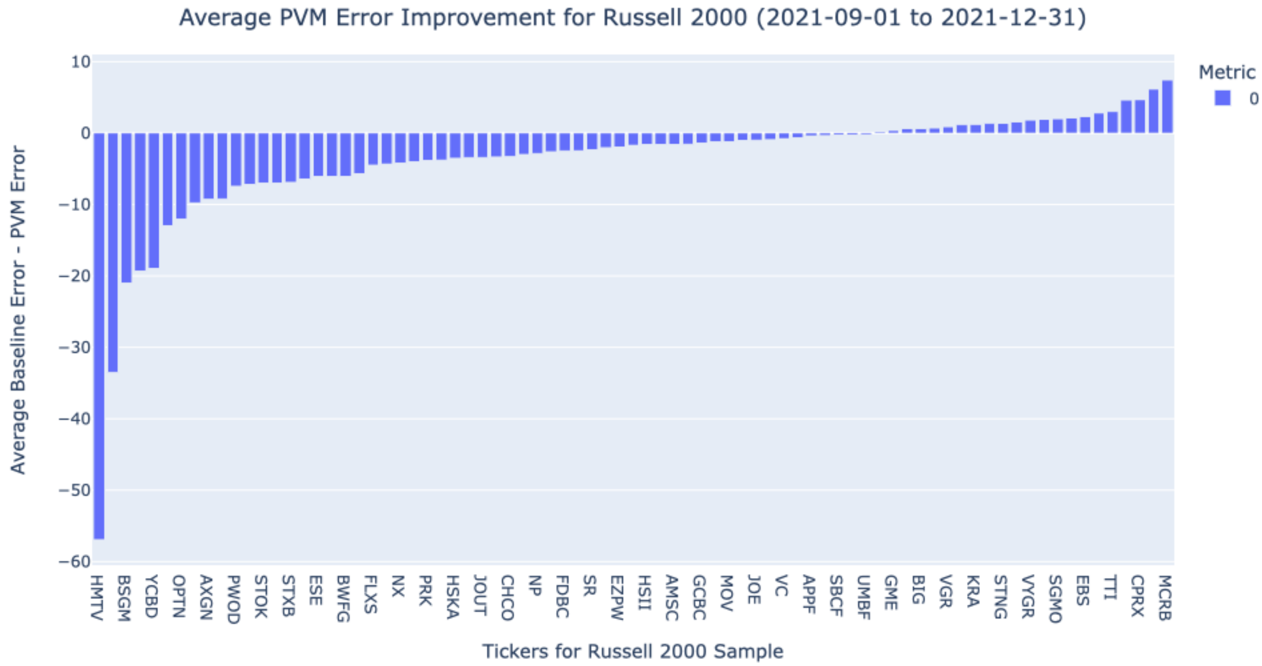


Figure 11. Price Volatility Model vs Baseline Model on Russell 2000

5. Other Tested Models

We also tried several other directions of modeling. However, they are not as effective as models we listed above. In this section, we give a brief description of other tested models for completeness.

5.1. **Tested Models for Daily Prediction.**

For daily volume prediction, mainly we tested three models: ARIMA-GARCH model, ARMA based on time series decomposition, and regression incorporated sector term.

Since our log volume data has many spikes, the conditional heteroskedasticity might exist. GARCH model is usually used for modeling the conditional heteroskedasticity, and it's usually combined with ARIMA model. By Ljung-Box test and Box-Pierce test, the log volume series is stationary and thus no need for difference. Then the ARIMA-GARCH model with order $(1,1)$ has the following form:

$$\begin{gathered} \hat{y}_t = \varphi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}, \\ \epsilon_t = \sigma_t z_t, \quad \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \end{gathered} \tag{5.1.1}$$

where z_t is distributed by standard normal, ω, α, β are parameters. As we can see, the GARCH model is based on the residual of ARIMA model. However, after several experiments, we found that ARIMA-GARCH model is not effective in our case.

Another natural modeling direction is to explore if there exists certain pattern of spikes, and periodicity is always the first thing to check. One way to check the periodicity (seasonality) is to use the basic time series (additive) decomposition model, which has the following form:

$$y_t = T_t + S_t + \varepsilon_t,$$

where T_t is trend term, S_t is seasonality term, and ε_t is residual. After separating the seasonality term, we may perform the ARMA model to predict the trend term, then combine the trend prediction with seasonality to generate final result. We have

$$\hat{y}_t = \hat{T}_t + S_t,$$

where

$$\hat{T}_t = \varphi T_{t-1} + \epsilon_t + \theta \epsilon_{t-1}.$$

This methodology is also not working, because the scale of seasonality term is much smaller than the residual term, thus is not able to improve the prediction accuracy. Also we should mention that the time series decomposition has an inherent period parameter needs to be specified, which is usually given by experience. In our case, we tried 2 to 30 and paid more attention to typical values: 5 (a week), 20 (a month), yet they all gave us not satisfactory results.

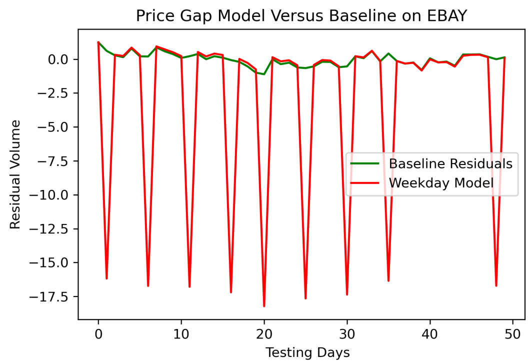
Weekday Model and the Choice of Residual Regressions

A simple predictor we examined for daily volume was the weekday. While this model unsuccessful, it confirmed that for our signals, a residual regression was generally the best measurement. We display the 3 models considered below, where D_i is an indicator variable for weekday.

(1) One option is to use the built in ARMA package's additional term(s) feature. This produces a model for log daily volume of the form

$$y_t = \varphi y_{t-1} + \sum_i \phi_i D_i + \theta \varepsilon_{t-1} + \varepsilon_t,$$

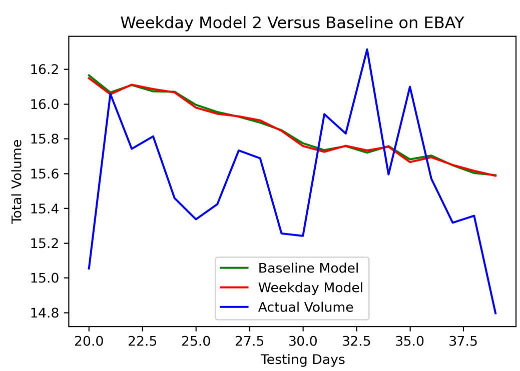
where the MLE estimator is used to determine each of the coefficients. The inherent issue with this model is that the smaller signals will throw off the optimization algorithm. This results in sharp spikes that return a high degree of error. Below we have graphed the error for the baseline against the error for this model on EBAY.



(2) Another option is to consider weekday as a multiplier on $\hat{y}_t$, the standard ARMA model (without the additional terms detailed above). In other words, we let

$$y_t = \big(1 + \sum_i \phi_i D_i\big)\hat{y}_t + \varepsilon_t,$$

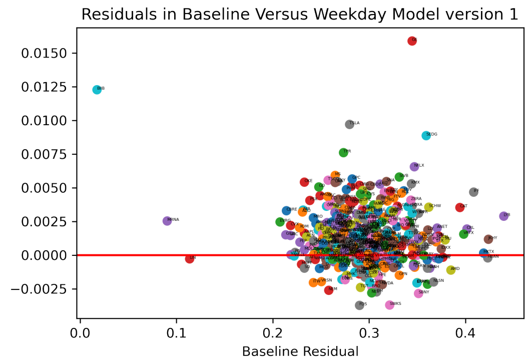
where ϕ_i is determined by OLS regression. The issue with this model is that the multiplier that is returned is simply too small, and it mirrored the baseline almost exactly (though it underperformed). Below we have plotted its performance against the baseline for EBAY.



(3) The third attempted method is the residual regression detailed in the description of the complete daily model. To recap, if ε_t represents the residuals due to the ARMA then one performs a regression,

$$\varepsilon_t = \sum_i \phi_i D_i + \eta_t,$$
$$\implies y_t = \varphi y_{t-1} + \sum_i \phi_i D_i + \theta \varepsilon_{t-1} + \eta_t.$$

While the improvement for this model was still subtle, the scatterplot of the difference in residual versus the baseline (taken against the baseline over a sample of the Russell 3000) shown below confirms that it improves on the majority of our tested stocks (stocks above the red line indicate improvement).



Sector Model

One of the first directions taken this quarter was a recategorization of stocks by sector. The motivation for this was also detailed in the mid-project report, though we provide a brief summary here as well. In the previous quarter, we examined relationships between the stock's volume and the (corresponding) aggregate sector volume.

Definition 5.1.2. Our hypothesis was that a shock may occur in a sector before it reaches some of its stocks. We therefore used our ARMA model to predict the aggregate sector volume (ARMA predictions of a quantity, x_t, are denoted by $\hat{x}_t$)

$$y_{\text{sector}} = \hat{y}_{\text{sector}} + \eta_t$$

where η_t is the residual/error produced from the ARMA model applied to the sector. Our model was therefore given by an assumption that $y_{\text{stock}} \sim \hat{y}_{\text{stock}} + \eta$ so that

$$y_{\text{stock},t} \approx \hat{y}_{\text{stock},t} + \phi\eta_{t-1}$$

Despite the tedium of re-aggregating the data in this manner, our models were largely unsuccessful. Below we have plotted residual comparison plots between the sector model and the baseline model. The model improved slightly on 1-2 of the sectors (such as Utilities), but greatly under-performed across most industries (like Health Care).

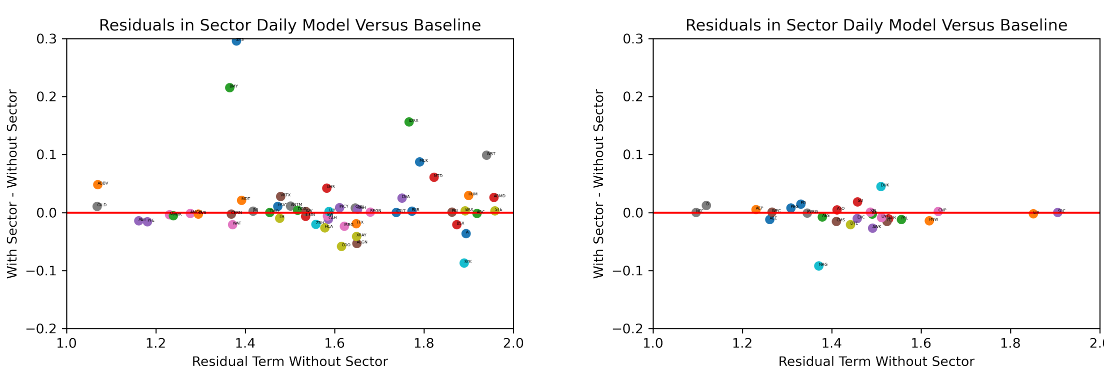


FIGURE 12. Residual Comparison Plot (improvements over baseline are above the red line(s)) for Health Care (Left) and Utilities (Right)

Regressions - Curve Fitting
We wanted to simplify the problem of predicting the intraday volume smile by parameterizing the volume smile and then develop models to predict the parameters. We started by looking for the best distributions to fit our data.

At first, we thought the beta distribution could be a representative u-shape curve, but the distribution was a poor fit due to the extended period of low trading volume in the middle of the day.

We then shifted to a piecewise approach, using different curves to fit different parts of the intraday u-curve. We separated the volume smile into three sections: the 30 minutes after market open, the middle of the day, and the 30 minutes before market close. Since the trade volume is quite flat during the middle of the day, we intend to continue to use a functional regression to model that period.

We focused on finding distributions to fit the opening and closing windows. Looking at more than 20 stocks from varying industries, we found out that a cubic polynomial fits the market opening period well and an exponential function fits the market close period well. The question then becomes how to predict the relevant parameters.

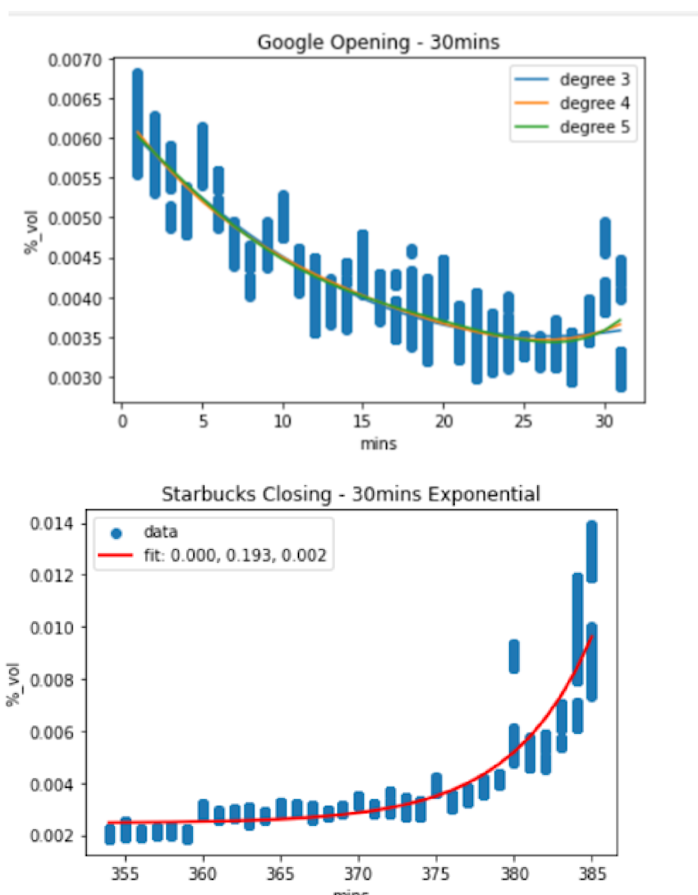


FIGURE 13. Polynomial regression fits well at the beginning of the day, exponential curve perform better at the time before the market close

6. Areas for Further Exploration

During our exploratory data analysis (EDA), we have noticed several phenomenons which are not universal. In particular, for a few stocks, we have observed that the jumps in intraday volume appear in every 5/10 minutes. In this section, we will give more details about such pattern.

6.1. **Jumps at Regular Intervals.**

Motivation. We observed spikes in volume that deviated significantly from the trend especially at every 5 minutes, when examining the intraday volume curve in Figure 14. Initially we wanted to simplify the problem of predicting the intraday volume smile by parameterizing the volume curve, and then develop models to predict the parameters. We started by looking for the best distributions to fit our data and used a piecewise approach to split the volume smile into three windows: the 30 minutes after market open, the middle of the day, and the 30 minutes before market close.

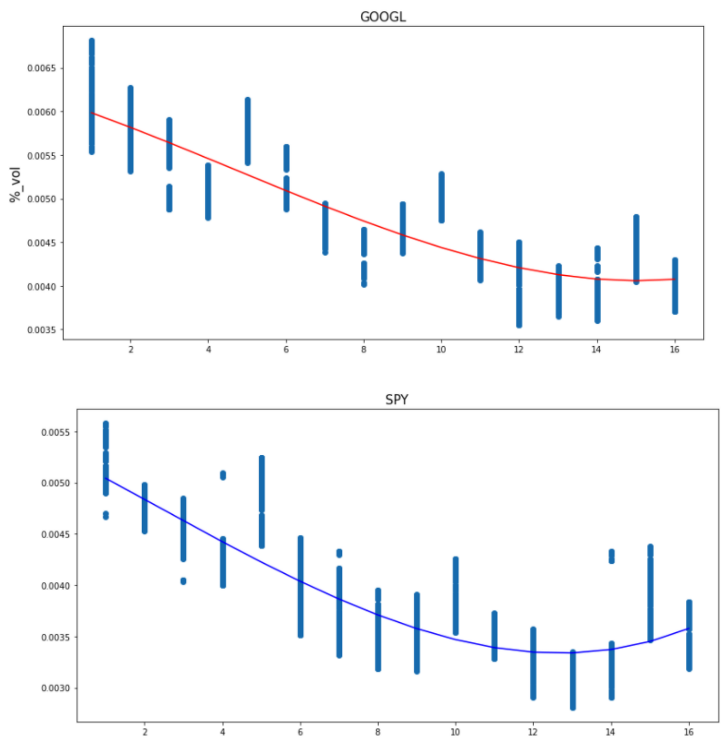


Figure 14. Opening 30 minutes Pattern in 3 years - High Liquid Stocks

First of all, we found these jumps by fitting curves on stock volumes in the past 3 years, from the beginning of the 2019 to the end of the 2021. After a survey of several hundred stocks across the Russell 3000, we found that consistent jumps appear more regularly on highly liquid stocks such as Google and the S&P 500 index (SPY), but are not observable on illiquid stocks – for instance, CRM (salesforce) and EL (Estée Lauder) at Figure 15.

In order to testify our hypothesis that liquidity matters in the jumps observations, we further investigated the top five liquid ETFs: IVV, QQQ, VTI, VOO, and SPY. We highlighted

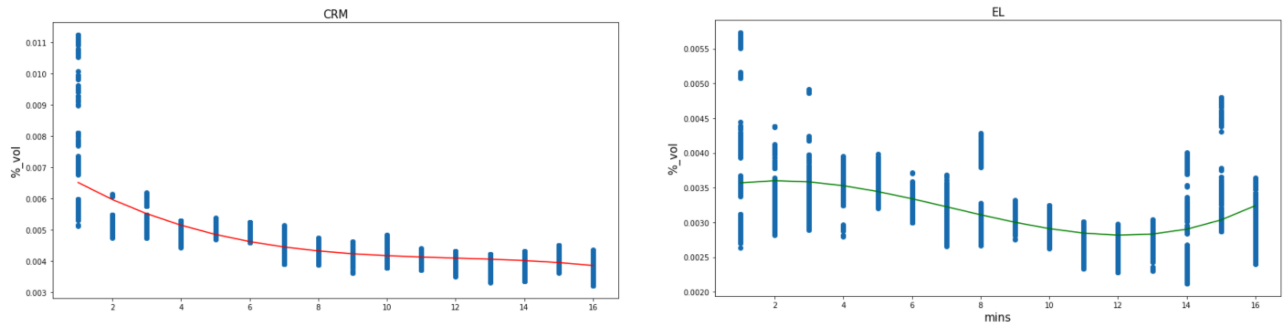


FIGURE 15. Opening 30 minutes Pattern in 3 years - CRM and EL

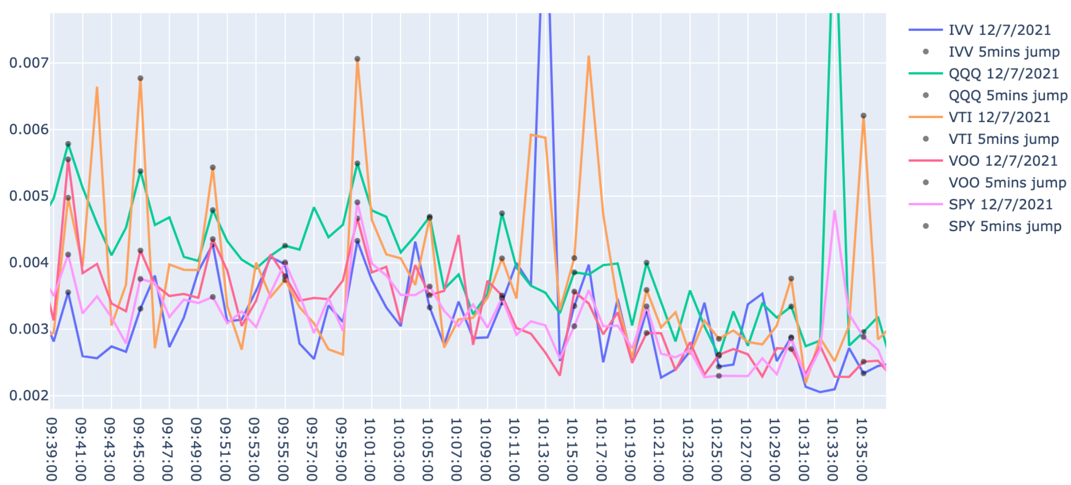


FIGURE 16. Opening: Trades at Every 5 minutes Intervals and Spikes

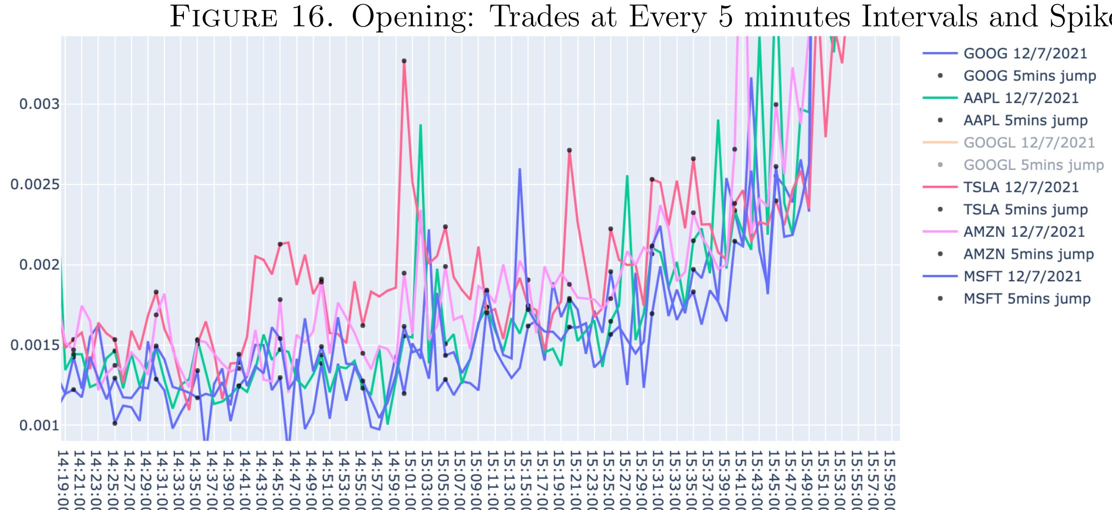


FIGURE 17. Closing: Jump Patterns are Nosier but Occur Mostly at Regular Interval.

trades at every 5 minutes in a sample day. As we can see from Figure 16 below, the trading volume at every 5 minutes is located at the spikes with persistent and clear patterns at the first hour after the market open and the last hour before the market close. We also selected the top five liquid stocks in 2021 as examples e.g. Google, Apple, Tesla, Amazon and Microsoft in Figure

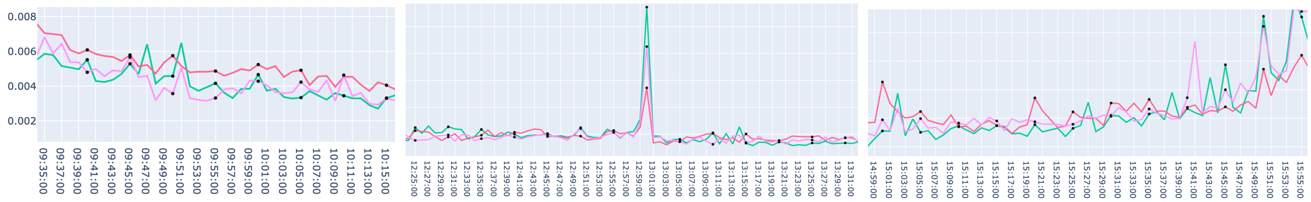


Figure 18. Green:AAPL, Red: TSLA, Pink: AMZN

17. These stocks' jump patterns are relatively nosier in comparison with those in the top five ETFs, but the majority of trades at every 5 minutes is still located at the spikes.

Additionally, we discovered from these observable stocks that jumps are steady throughout the day but become less significant in fewer transaction period. In the middle of the day, jumps become harder to capture and unpredictable at the 5 minute level but a spike appears on every day at 13:00. Furthermore, we investigated the occurrences of jumps per day and by inspection, noted that these jump patterns occur in 3 cycles as in Figure 18: 1) the first one hour after market open, 2) the last one hour before the market close and 3) at 13:00 and 15:54 each day.

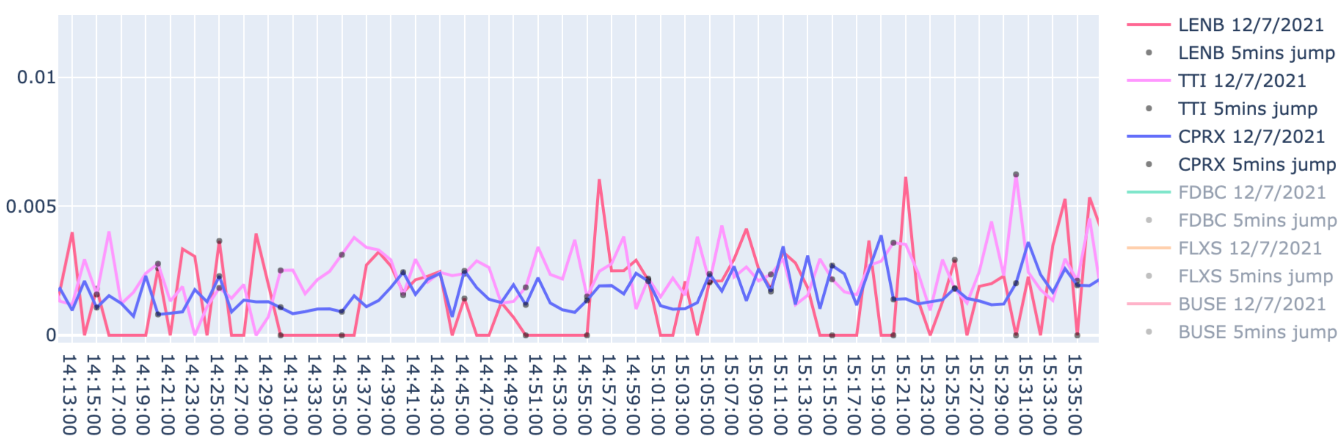


Figure 19. Jumps in Russell 1000 and 2000 Scaled in % change of volumes

These two findings provide evidence for our hypothesis that there may be institutions hedging ETF's or larger liquid stocks systematically along the day. Plus, there is a positive correlation between liquidity of a stock and the size and consistency of jumps. According to Figure 19, the 5 minute jumps are not significant across the Russell 1000 or 2000. Less liquid stocks lack a lot of the fluctuations and spikes that were characteristic of the stocks that presented clear jumps. Only highly liquid stocks have the most pronounced and steady jumps at 5-minute intervals. Moreover, jump patterns can be traced in 3 cycles per day, becoming less significant during extended periods of low trading volume.

It was important to verify that the five minute intervals, in particular, were the most consistent markers of volume jumps. One simple metric was to define a relative size of each jump by diving the "jump volume" by a symmetric average of the surrounding volume measurements. More specifically, the percent jump size, J_t, at time, t, where t for an interval of size N is given by the quotient of volume divided by a symmetric average

$$J_t \equiv \frac{y_t}{\frac{1}{N}\sum_{i=-N/2}^{N/2} y_{t+i}} - 1$$

Observe that in Table 6.1, the 5 minute measurements display much clearer signals than the other tested interval sizes (the 3/4 minute intervals even showed a slight average decrease for AMZN and MSFT). Here, the average for 5 minute jumps is taken over times at regular five minute intervals (and analogously for 3/4 minute jumps). Note that GOOG reflects a comparatively larger average volumetric jump size, comfortably surpassing the other liquid stocks presented above.

Table 6.1.1. Average Percentage Jump Size at varying intervals over GOOG, AMZN, MSFT on 5/12/2021

Avg. Percent Jump Size	Avg. 5 Minute Jump	Avg. 4 Minute Jump	Avg. 3 Minute Jump
GOOG	23.25%	1.18 %	1.08 %
AMZN	6.45%	2.19%	–5.26 %
MSFT	5.77%	–0.86 %	–1.35%
AAPL	5.88%	–2.69 %	–1.54%
TSLA	10.87%	–0.54 %	–2.41%

For completeness sake, we validated the existence of these jumps by examining the stocks' millisecond trading volumes. Here, in the example of the SPY– one of the largest ETF's– we first divided volumes at the millisecond level by the stock's daily trading volume size. At the beginning of the market open time, the percentage of volume change at the millisecond level was smaller then those toward the market close.

In conclusion, for stocks like Google, Amazon or most liquid ETFs such as QQQ, SPY, most of their spikes are systematic. We recommend researchers to implement additional parameters in regression to indicate these jumps (especially for highly liquid stocks and ETF indexes). The purpose of our Robust ARMA is to detect these unexpected jumps and regularize them so that it will not impact on our curve fitting substantially. Our definition of those regular jumps are also part of these spikes evaluations. In terms of curve fitting, without those outliers, this Robust ARMA can at least guarantee the model performance. To further quantify these jumps or spikes, thresholds of detecting jumps or regularizing spikes for each stock need to be considered independently. We expect to see more evaluations and justifications to conclude the best threshold for stocks in large, middle and small market cap for the best outcome in the following research.

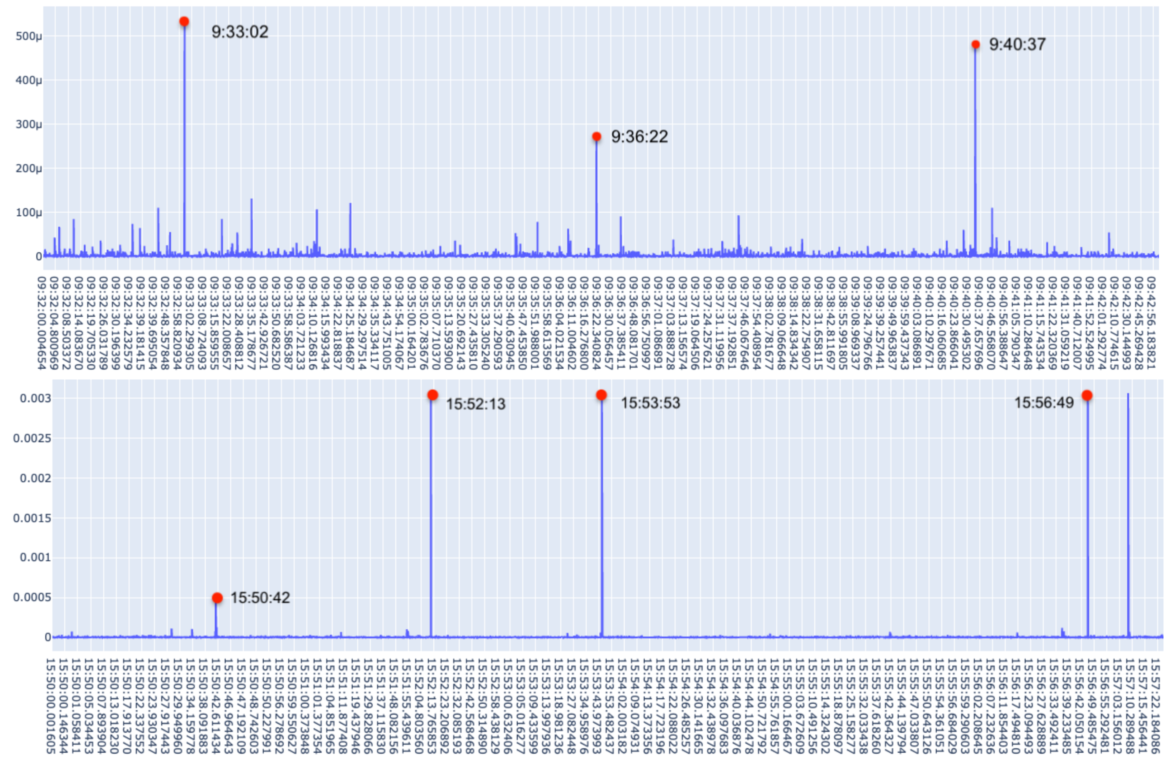


FIGURE 20. Jumps pattern in SPY on 12-07-2021 mostly appear every 3 minute or 4 minute.

7. Acknowledgements

We wish to thank DRW Holdings and the University of Chicago's Data Science Institute for providing us the opportunity to work on this project.

We extend our sincerest gratitude to our mentors from DRW and the University of Chicago: Phillip Lynch, David Uminsky, Dan Nicolae, Cole Bryant and Nick Gondek for their guidance and support.

References

[M17] Markov, Vladimir, Olga Vilenskaia, and Vlad Rashkovich. "Quintet Volume Projection." The Journal of Trading 12.2 (2017): 28-43.