Analyzing Factors and Predictive Modeling for ZRX to USD Exchange Rate Fluctuations

Chloe Yang[department of statistics], Jinghe Shen[department of statistics], & Ziqi Xu[department of statistics]

[1] UIUC

Author Note

The authors made the following contributions. Chloe Yang: Modelling - time + trading value + buying price, Writing - Original Draft Preparation; Jinghe Shen: Final Model Selection and Forecasting, Designing Slides - Original Draft Preparation; Ziqi Xu: Modelling - time + trading value + selling price, Designing Slides - Original Draft Preparation.

Abstract

This study aims to discover the factors behind ZRX to USD exchange rate fluctuations, utilizing regression analysis and market indicators beyond temporal variables, to enhance forecasting accuracy in the cryptocurrency market.

The main result of the study is the identification of using three significant predictors, selling price, stock volume, and trading date which contribute ZRX to USD exchange rate movements. This finding challenges previous assumptions by demonstrating that factors such as trading volumes and buying or selling activities play a more substantial role in exchange rate dynamics. It adds to existing knowledge by highlighting the need for considering multiple market indicators to achieve accurate exchange rate forecasting.

In the broader context of financial markets and digital assets, understanding the intricate interactions between market indicators and exchange rates is crucial for informed decision-making and risk management strategies. The prediction we made based on our selected model shows a trend of decreasing of ZRX to USD exchange rate in the next few trading days.

The approach of this study highlights the interdisciplinary nature of predictive modeling, emphasizing the importance of combining different data sources and analytical techniques to make robust predictions, which is a principle that applies across scientific disciplines seeking predictive insights from complex datasets.

Analyzing Factors and Predictive Modeling for ZRX to USD Exchange Rate Fluctuations

Introduction

Given that 0x Protocol (ZRX), a cryptocurrency, plays a protocol role in open-source infrastructure enabling the decentralized exchange of ERC20 tokens across various blockchains, the ZRX to USD exchange rate reflects the market dynamics and trading activities surrounding this protocol.

Launched in late 2017, ZRX has experienced an inflation emission with a 1,000,000,000 supply (Barbereau, 2023). Traders hold ZRX tokens as an investment, monitoring the ZRX to USD rate closely to make more informed decisions to pursue larger profits. Moreover, ZRX is used in various Decentralized Finance (DeFi) transactions(Nadler, 2020). Users may need to convert ZRX to USD or vice versa when interacting with these platforms, making the exchange rate important (Chu, 2023).

This analysis seeks to explore the determinants of the exchange rate from ZRX to USD using regression analysis, incorporating multiple independent variables beyond time alone. The data of the ZRX to USD ($) cryptocurrency pair from 2023-1-1 to 2024-02-27 obtained from the official BITFINEX website is used in this analysis. This data includes exchange rate metrics such as daily highs and lows, average prices, last traded rates, and buying/selling prices. The aim of the study is to identify significant predictors of change rate movements and forecast future rates with enhanced accuracy.

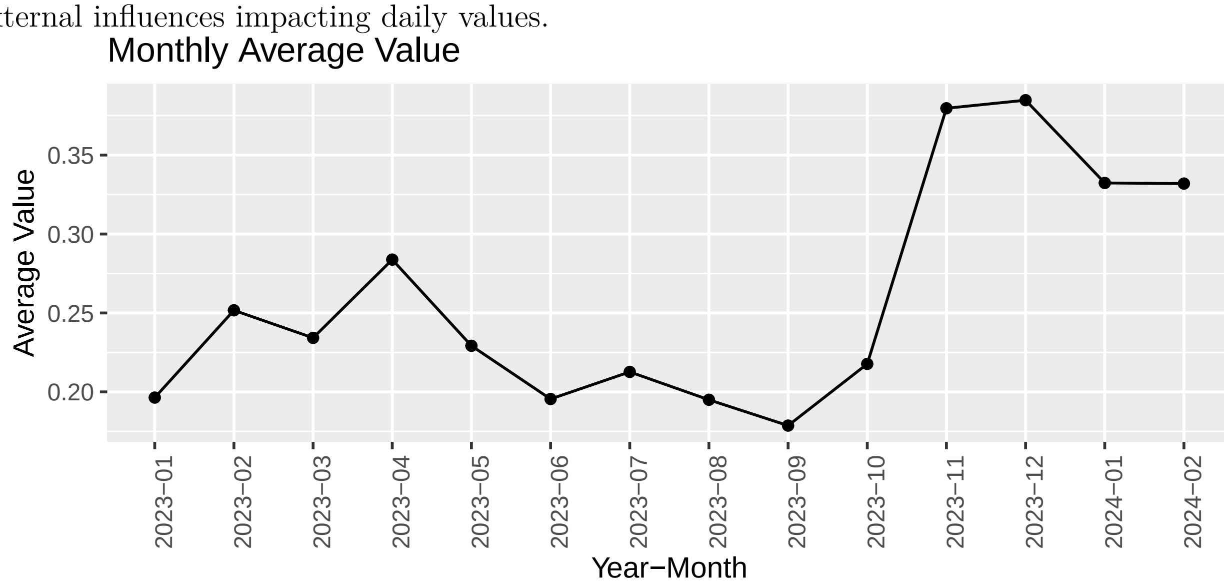
Methods

Data

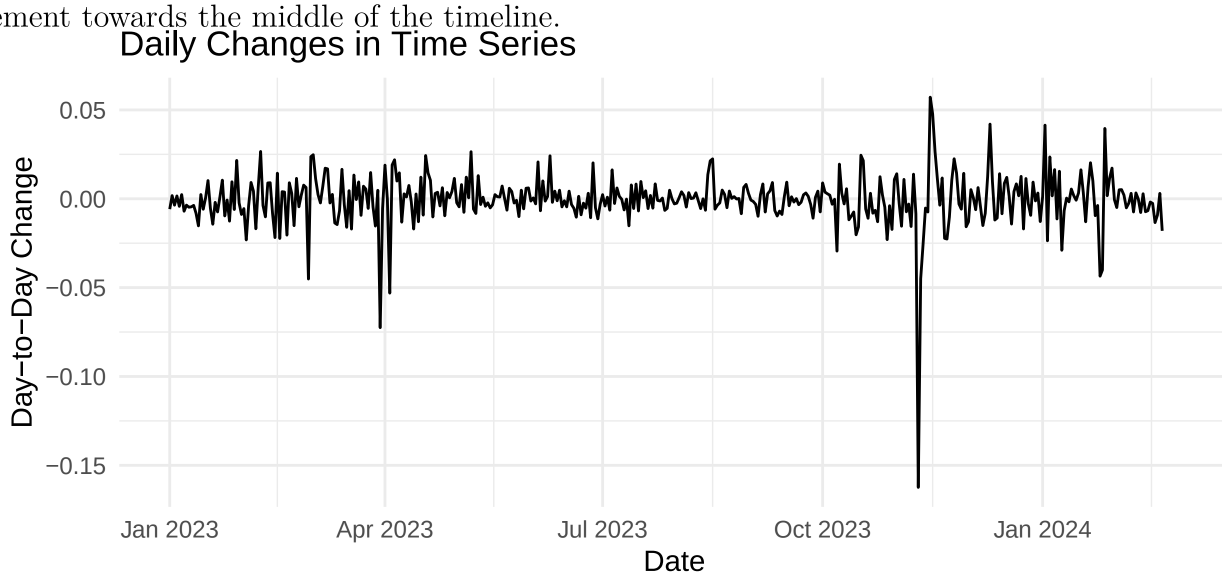
We are using `mid` (Average Daily Price) as response variable, and `time` (Date of record), `volume` (Trading volume at end of day), and `bid` / `ask` (Buying/Selling price at end of day) as predictor.

Preliminary Analysis

The daily time series plot displays the fluctuations in average value over time on a daily basis, beginning from 2023-1-1 to 2024-02-27. The pattern suggests non-stationarity due to the changes in mean and variance over time, highlighting periods of potential market events or external influences impacting daily values.



The plot shows the monthly average value plot traces the average value across several months, starting from January in the year 2023 to February of 2024. This representation smooths out the daily fluctuations seen in the first graph, providing a clearer view of the longer-term trend and potential seasonality. Initially, there is a dramatic decrease followed by a trough, after which the average values oscillate at a lower level with some upward movement towards the middle of the timeline.



Then we tried differencing method on the ZRX to USD data. By differencing, the the daily changes in the ZRX to USD exchange rate plot seems stationary as it shows constant expectation and variance over time. Additionally, the Augmented Dickey-Fuller test yielded a test statistic of -9.074 and a p-value of 0.01. The test's alternative hypothesis suggests stationarity, indicating that the differenced data is stationary.

Regression Analysis

Model 1: time + trading value + buying price

From the **Model 1**,we find the predictors' p-value are all lower than 0.05, indicating their significant contribution to the daily average price of ZRX to USD.

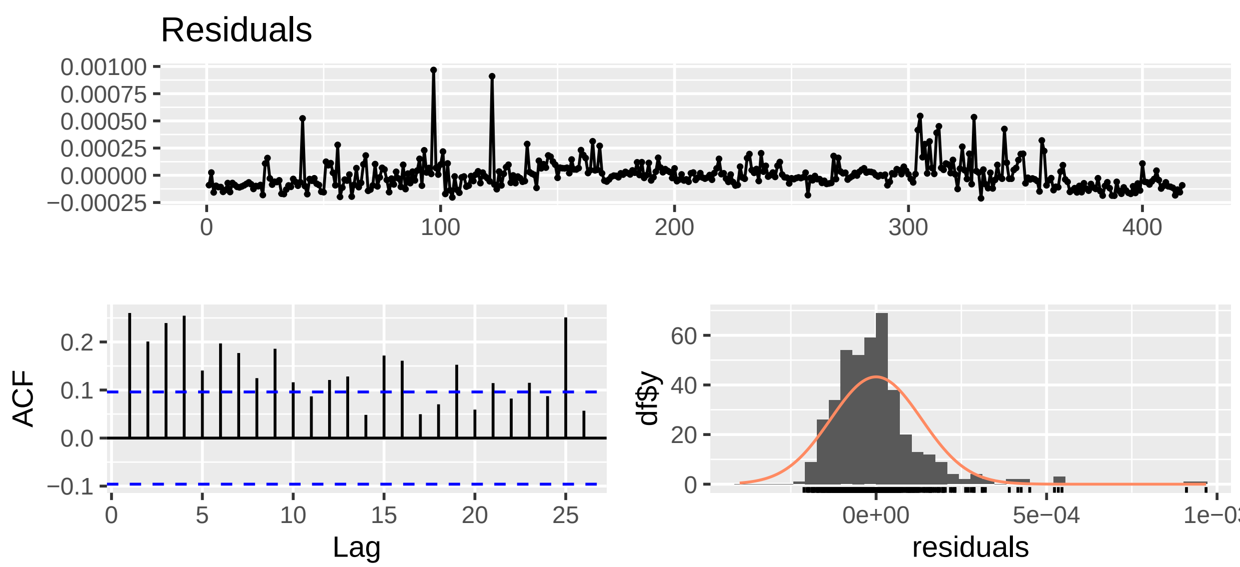
* The estimate parameter of date -3.540e-07 indicates a significant negative relationship between date and mid, suggesting that as time progresses, the average daily price decreases.

* The estimate parameter of bid 1.000e+00 indicates a strong positive relationship between bid (buying price at the end of the day) and mid, indicating that higher buying prices are associated with higher average daily prices.

* The estimate parameter of volume 6.276e-11 indicates a significant but very small

positive relationship between volume (trading volume at the end of the day) and mid, suggesting that higher trading volumes are associated with slightly higher average daily prices.

Additionally, the F-statistic of 4.32e+07 indicate that the model explains all the variability in the response variable and is highly significant. Then we run diagnostic on the **Model 1**:




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 162.5, df = 10, p-value < 2.2e-16
##
## Model df: 0.   Total lags used: 10
```

From the plot above, we can see for **Model 1**:

* The plot of residuals does not show a constant trend, and it fluctuates around 0.

* For the ACF plot, there are spikes out of the required limits, meaning that the residuals have some remaining autocorrelation. It suggests that there may be temporal

dependencies or patterns in the residuals, violating the assumption of independent residuals.

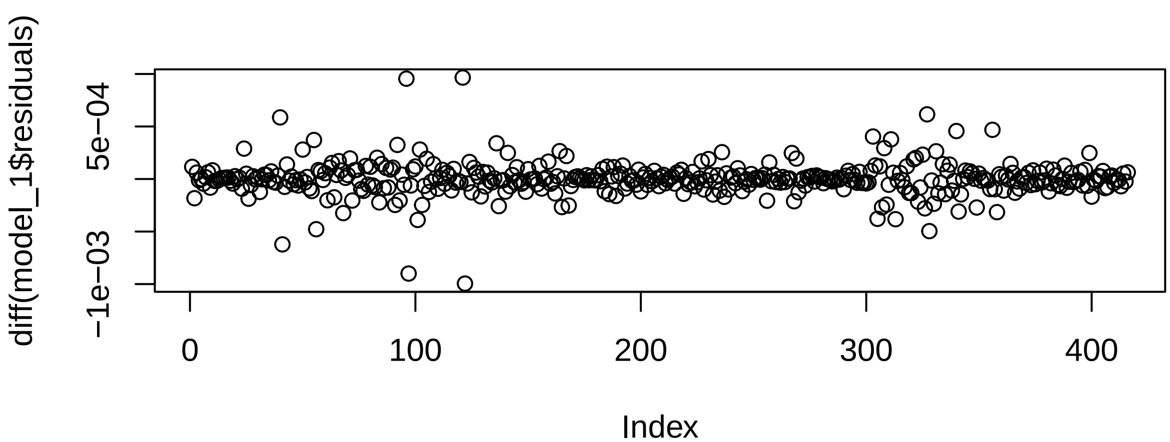
* The histogram shows that the residuals are not normally distributed with right skewed peak values, indicating that the residuals' distribution is not symmetric around zero, violating the assumption of normally distributed residuals.

* The p-value of Ljung-Box test is 2.2e-16 < 0.05, so we can reject H_0 and conclude we have enough evidence to show that the residuals are dependent.

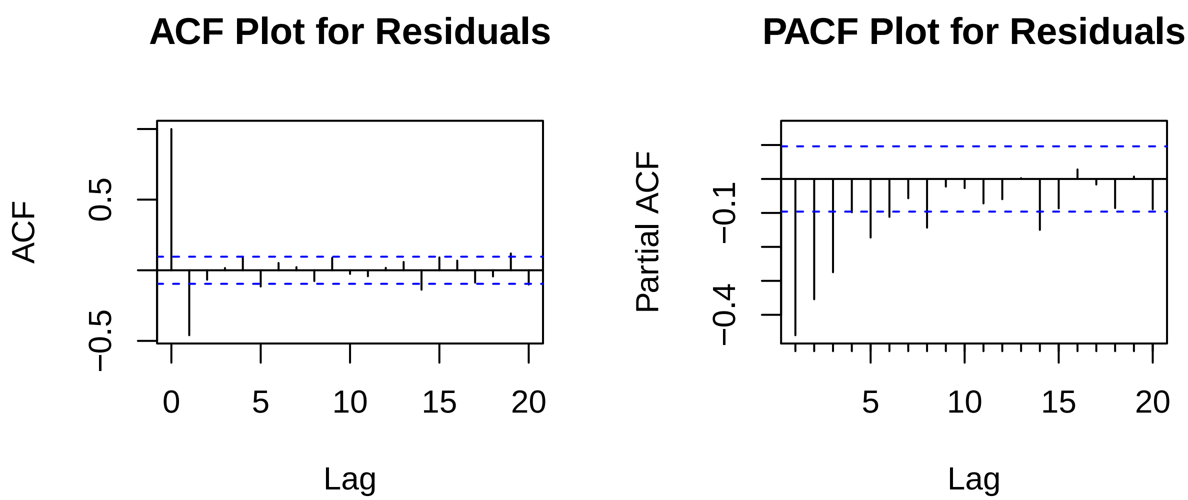
We consider using models that account for autocorrelated errors such as autoregressive integrated moving average (ARIMA) model.

By the residuals plot, it does not seems stationary with an increasing then decreasing trend.

It is not stationary as it shows an increasing then decreasing trend. So we tried differencing the residuals.



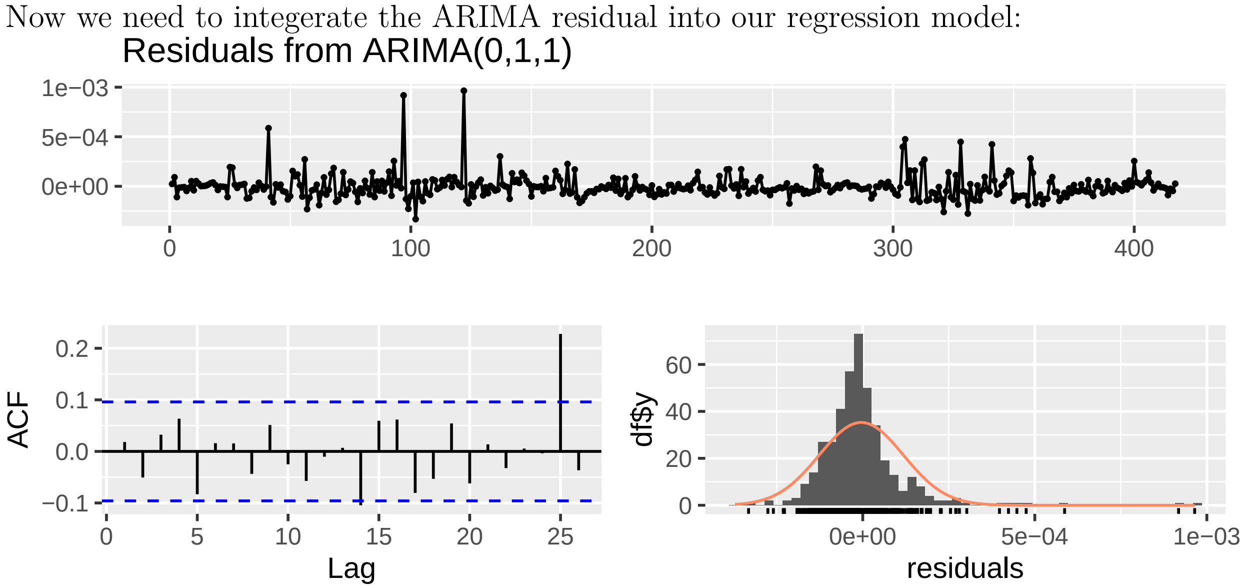
It now seems to have a constant mean and the variance is independent from t, so it is stationary now.



From the plot we can see that the PACF plot is tailing off and ACF plot cutts off at lag 1. This is an MA(1) process.

For parameter estimation, $p = 0, q = 1, d = 1$ for the residuals, so the ARIMA for residual would be `ARIMA(0,1,1)`

Now we need to integerate the ARIMA residual into our regression model:




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,1)
## Q* = 8.7127, df = 9, p-value = 0.4642
```

```
##
## Model df: 1.   Total lags used: 10
```

From the plot above, we can see:

The plot of residuals does not show a constant trend, and it fluctuates around 0. (though with some outliers)

For the ACF plot, there are very few spikes out of the required limits, meaning that the residuals almost have no autocorrelation.

The histogram shows that the residuals are normally distributed. The p-value of Ljung-Box test is $0.4642 > 0.05$, so we fail reject H_0 and conclude that there are no obvious correlation between residuals.

The model with auto correlated errors (**Model 1.1**) satisfy the white noise assumption of following $N(0, \sigma_w^2)$, the residual diagnostics perform much better than the original (**Model 1**).

Model 2: time + trading volume + selling price

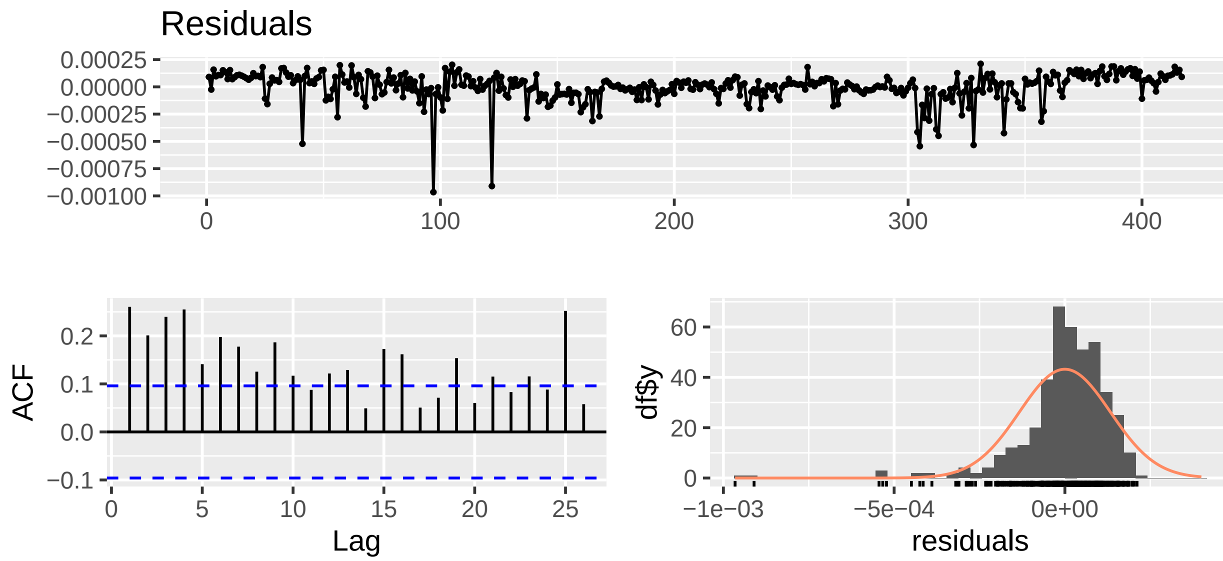
From the **Model 2**,we find the predictors' p-value are all lower than 0.05, indicating their significant contribution to the daily average price of ZRX to USD.

* The estimate parameter of date 3.571e-07 indicates significant positive relationship between date and mid, suggesting that as time progresses, the average daily price increases.

* The estimate parameter of ask 9.996e-01 indicates strong positive relationship between ask (selling price at the end of the day) and mid, indicating that higher selling prices are associated with higher average daily prices.

* The estimate parameter of volume -6.214e-11 indicates significant but very small negative relationship between volume (trading volume at the end of the day) and mid, suggesting that higher trading volumes are associated with slightly lower average daily prices.

Additionally, the F-statistic of 4.32e+07 indicate that the model explains all the variability in the response variable and is highly significant.




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 163.23, df = 10, p-value < 2.2e-16
##
## Model df: 0.   Total lags used: 10
```

From the plot above, we can see for **Model 2**:

* The plot of residuals show a constant trend over time, and it fluctuates around 0.

* For the ACF plot, there are spikes out of the required limits, meaning that the residuals have some remaining autocorrelation. It suggests that there may be temporal dependencies or patterns in the residuals, violating the assumption of independent residuals.

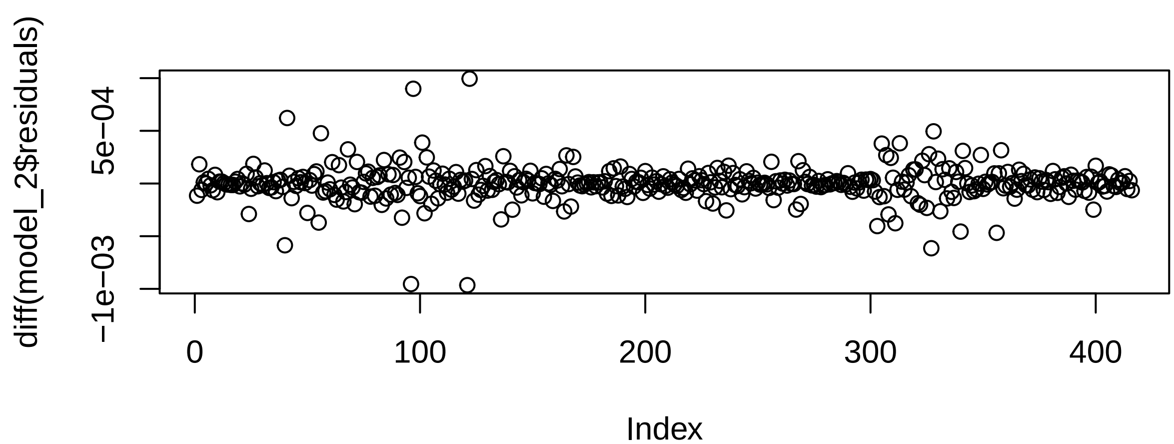
* The histogram shows that the residuals are not normally distributed with left skewed peak values, indicating that the residuals' distribution is not symmetric around zero, violating the assumption of normally distributed residuals.

* The p-value of Ljung-Box test is 2.2e-16 < 0.05, so we can reject H_0 and conclude we have enough evidence to show that the residuals are dependent.

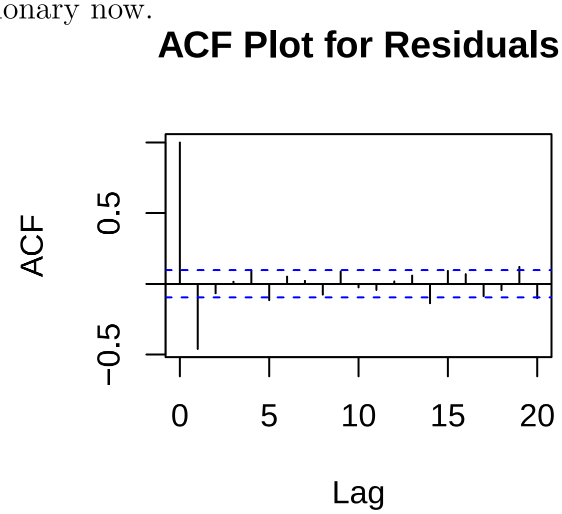
We consider using models that account for autocorrelated errors such as autoregressive integrated moving average (ARIMA) model.

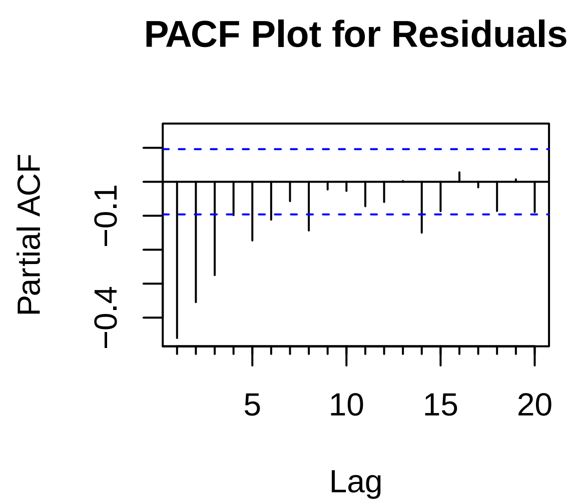
By the residuals plot, it does not seems stationary with an increasing then decreasing trend.

It is not stationary as it shows an decreasing then increasing trend. So we tried differencing the residuals.



It now seems to have a constant mean and the variance is independent from t, so it is stationary now.





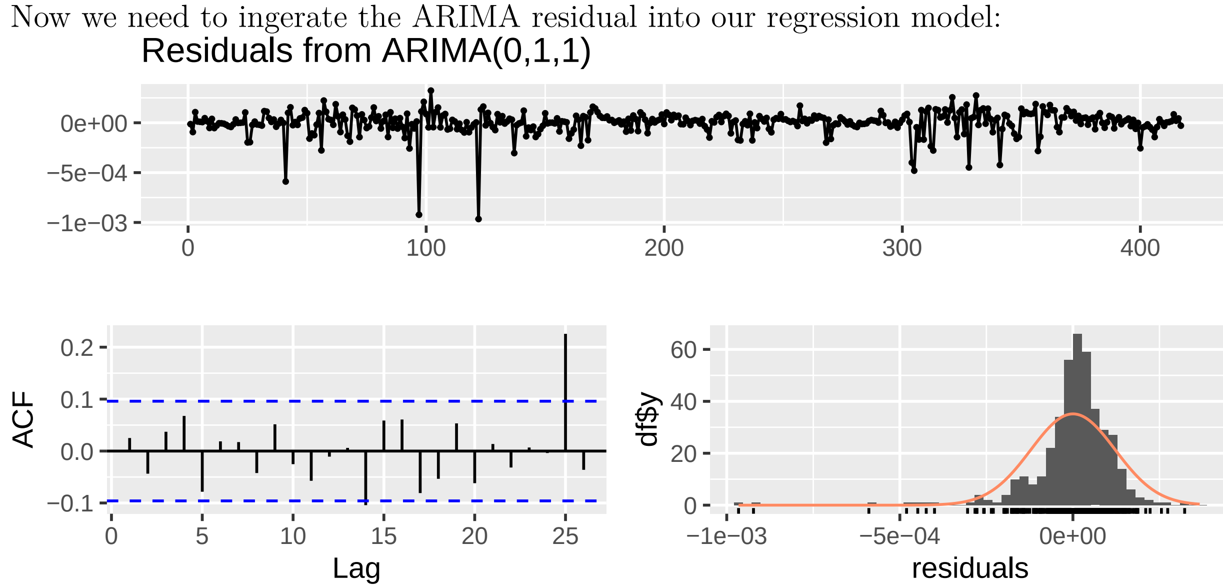
From the plot we can see that the PACF plot is tailing off and ACF plot cutts off at lag 1. This is an MA(1) process.

For parameter estimation, $p = 0, q = 1, d = 1$ for the residuals.

So the ARIMA for residual would be `ARIMA(0,1,1)`.

```
##
## Call:
## arima(x = model_2$residuals, order = c(p = 0, q = 1, d = 1))
##
## Coefficients:
##           ma1
##       -0.8907
## s.e.   0.0288
##
## sigma^2 estimated as 1.537e-08:  log likelihood = 3151.04,  aic = -6298.07
```

Now we need to ingerate the ARIMA residual into our regression model:




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,1)
## Q* = 8.63, df = 9, p-value = 0.4721
##
```

```
## Model df: 1.   Total lags used: 10
```

From the plot above, we can see:

The plot of residuals does not show a constant trend, and it fluctuates around 0. (though with some outliers)

For the ACF plot, there are very few spikes out of the required limits, meaning that the residuals almost have no autocorrelation.

The histogram shows that the residuals are normally distributed.

The p-value of Ljung-Box test is $0.4721 > 0.05$, so we fail reject H_0 and conclude that there are no obvious correlation between residuals.

The model with auto correlated errors (**Model 2.1**) satisfy the white noise assumption of following $N(0, \sigma_w^2)$, the residual diagnostics perform much better than the original (**Model 2**).

Results

Comparison and evaluation

- AIC and BIC

  Extract the AIC and BIC values from each model, which are automatically calculated when you fit an ARIMA model using the arima or auto.arima function from the forecast package. Lower values of AIC and BIC generally indicate a better model fit with a good balance of model complexity and goodness of fit.

```
## Model 1.1 - AIC: -6303.15, BIC: -6283.00
```

```
## Model 2.1 - AIC: -6305.72, BIC: -6285.56
```

- AIC Comparison:

  Model 2.1 has a lower AIC (-6305.717) compared to Model 1.1 (-6303.15). This suggests that Model 2.1 has a slightly better fit to the data considering the trade-off between goodness of fit and model complexity.

- BIC Comparison:

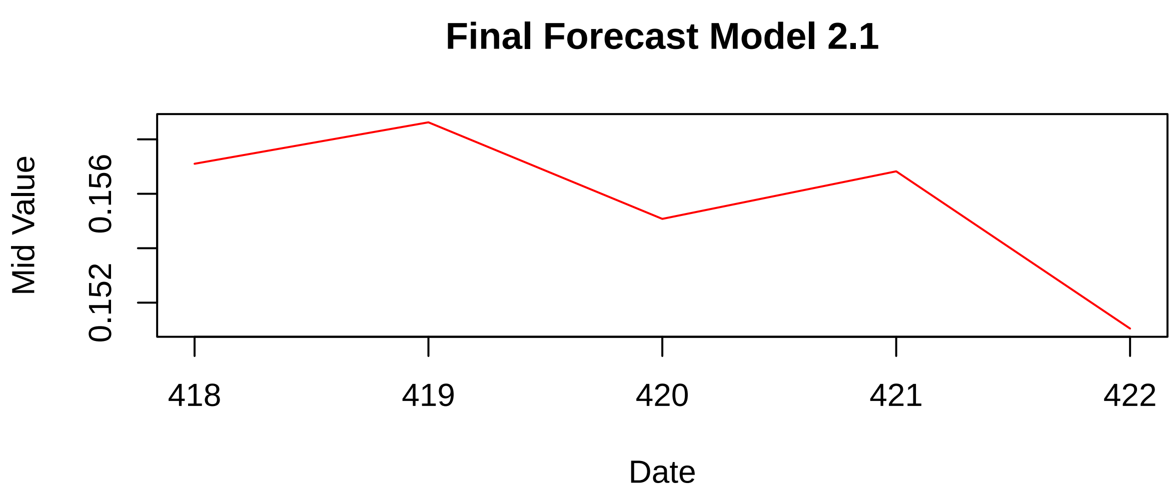
  Similarly, Model 2.1 has a lower BIC (-6285.564) compared to Model 1.1 (-6282.996). This indicates that when the penalty for the number of parameters is considered more stringently, Model 2.1 still performs better, suggesting it might be the more appropriate model among the two.

  Then we look at the residuals of each model to check for any patterns or autocorrelation that might suggest model inadequacies.

- Residual Analysis: Both models seem to perform well in terms of residual analysis including Residual plot, ACF plot, histogram and Ljung-Box test.

  Based on the overall statistical metrics, including the Box-Ljung test results, AIC, and BIC values, **Model 2.1** might be preferred slightly over Model 1.1, although both models exhibit robust statistical properties.

Pediction




```
## # A tibble: 5 x 2
##   Date       Predicted_Mid
##   <date>             <dbl>
## 1 2024-02-28         0.157
```

```
## 2 2024-02-29          0.159
## 3 2024-03-01          0.155
## 4 2024-03-02          0.157
## 5 2024-03-03          0.151
```

Discussion

Cryptocurrencies are volatile and unpredictable, so it's challenging for our model to fully capture the values when the market or policy changes a lot. This high volatility is from their immature market structure, speculative trading, and frequent regulatory changes. Although our models incorporate predictive factors like trading volume and buying/selling prices, their predictive power is limited due to external market forces. Thus, sudden, large-scale events such as regulatory announcements can skew predictions and distort the exchange rate.

Moreover, the short data period confines our analysis to short-term fluctuations. Identifying and quantifying seasonal or annual patterns is challenging. Cryptocurrency markets often exhibit patterns aligned with broader economic and technological cycles. While our models effectively analyze immediate movements, predicting cyclical changes that may emerge over longer periods is difficult.

Despite these challenges, this analysis offers valuable insights into which factors influence the ZRX to USD exchange rate, providing a foundation for predicting future trends and making more informed investment decisions.It is crucial to remain cautious and continuously refine models with more data and variables to improve predictive accuracy.

References

Barbereau, S., T. (2023). Decentralised finance's timocratic governance: The distribution and exercise of tokenised voting rights. *Technology in Society*, *102251*(73). Retrieved from https://doi.org/10.1016/j.techsoc.2023.102251

Chu, C., J. (2023). An analysis of the return–volume relationship in decentralised finance (DEFI). *International Review of Economics &Amp; Finance*, *85*, 236–254.

Nadler, &. S., M. (2020). Decentralized finance, centralized ownership? An iterative mapping process to measure protocol token distribution. *arXiv.org.*

Appendix (Optional)

Any R codes or less important R outputs that you wanted to keep- can go in here.

- To clarify the **Regression with autocorrelated residual** part, we provide code for how we obtain **Model 1.1** and **Model 2.1**

```
model_1_1 = arima(data$mid, c(0, 1, 1), xreg = cbind(data$date, data$bid, data$volume))
```

```
model_2_1 = arima(data$mid, c(0, 1, 1), xreg = cbind(data$date, data$ask, data$volume))
```

- To clarify **Prediction** part, we provide following code:

```
future_dates <- seq(max(data$date), by="day", length.out=6)[-1]
future_volume <- tail(data$volume, 5)
future_ask <- tail(data$ask, 5)


future_data_2 <- data.frame(date=future_dates, ask=future_ask, volume=future_volume)
predicted_mid_2 <- predict(model_2, newdata=future_data_2)
arima_resid_2 <- auto.arima(resid(model_2))
resid_forecast_2 <- forecast(arima_resid_2, h=5)
final_forecast_2 <- predicted_mid_2 + resid_forecast_2$mean
```

- Full description of data

Column	Description	Type
code	Unique cryptocurrency exchange pair identifier	String
date	Date of record	Date
high	Highest daily price	Double
low	Lowest daily price	Double
mid	Average daily price	Double
last	Last traded price	Double
bid	Buying price at end of day	Double
ask	Selling price at end of day	Double
volume	Trading volume at end of day	Double