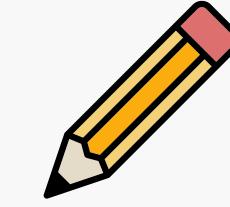

• • • •

TIME SERIES MODEL: VAR VS ARIMA

Nguyen Phuong Huyen 23229585
Nguyen The Hung 23229586
Nguyen Duc Thanh Vinh 23229582



Table of Contents



01

ARIMA
model

02

VAR model

03

Applications

04

Object

05

Orange
workflow

06

How we
improve
the result

ARIMA(a univariate model)

Auto-Regressive Integrated Moving Average

- **AR: Autoregression.** A regression model that uses dependence relationship between an observation and a number of lagged observations.
- **I: Integration.** Calculating the differences between observations at different time points, **aiming to make the time series stationary.**
- **MA: Moving Average.** This approach considers the dependence that may exist between observations and the error terms created when a moving average model is used on observations that have a time lag.

Stationarity

Stationary is essential for ARIMA to predict which includes:

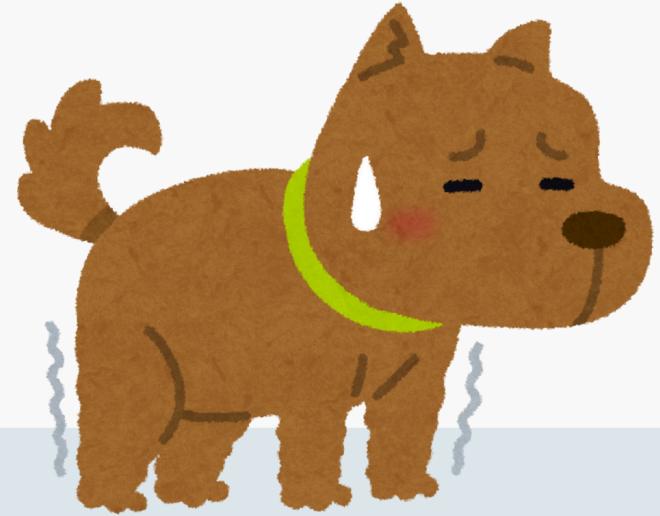
- Mean value is constant
- Variance is constant
- Autocovariance depends on lag

Strengths



- The models are very **flexible**, successfully modeling an astonishing variety of irregular or quasi-periodic, smooth or choppy, constant or variable mean light curves
- **The dimensionality of the model is relatively low** with a moderate computational burden of the numerical optimization.
- **Error analysis** on the parameters naturally emerges through the **likelihood regression analysis**.
- They are **extensible to situations involving multivariate time series**, combinations of stochastic and deterministic behaviors, change points, and (moderately) irregular observation spacing

Weaknesses



- It is hard to model the nonlinear relationships between variables
- There is a constant standard deviation in errors in ARIMA model, which in practice it may not be satisfied

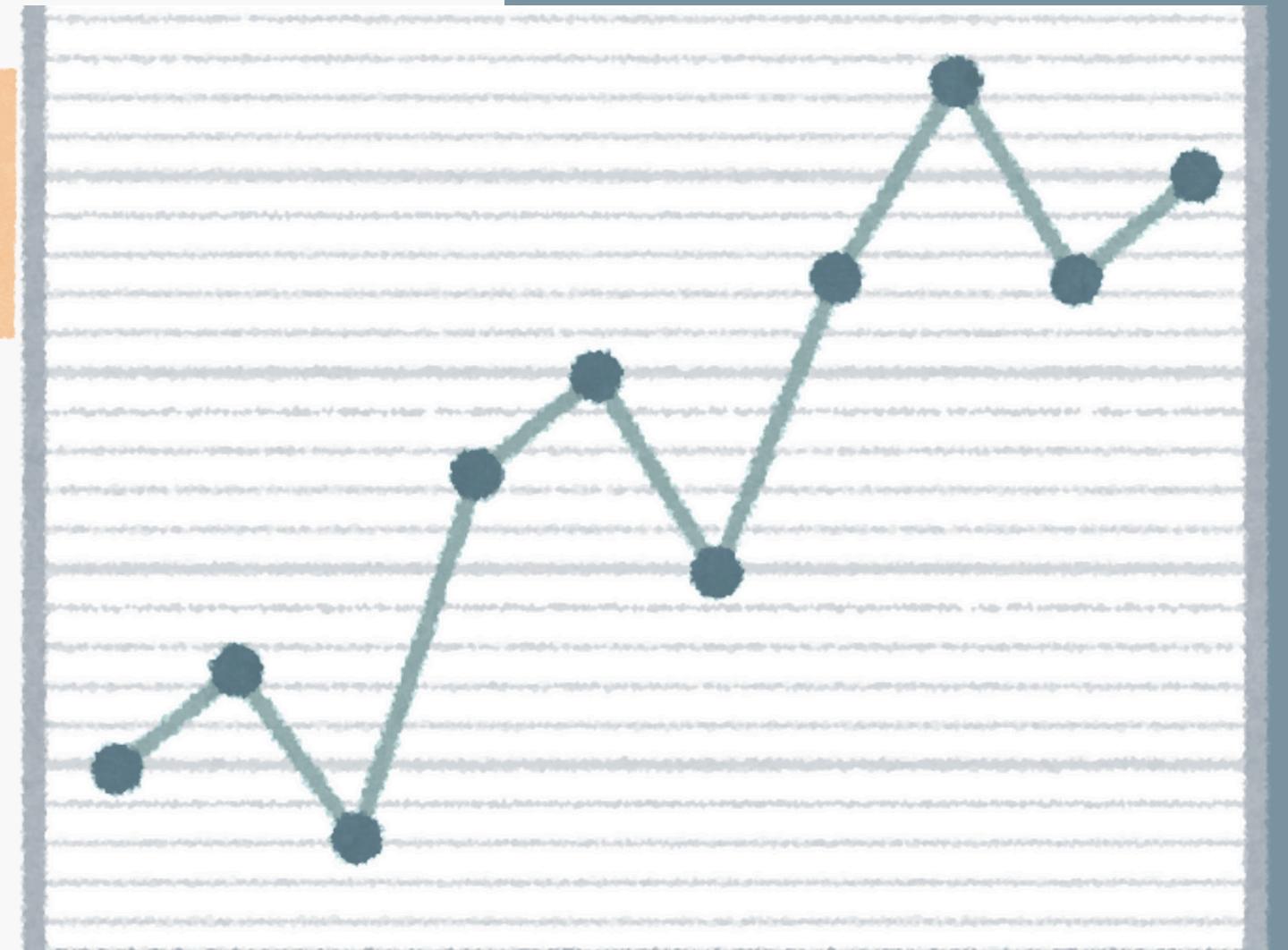
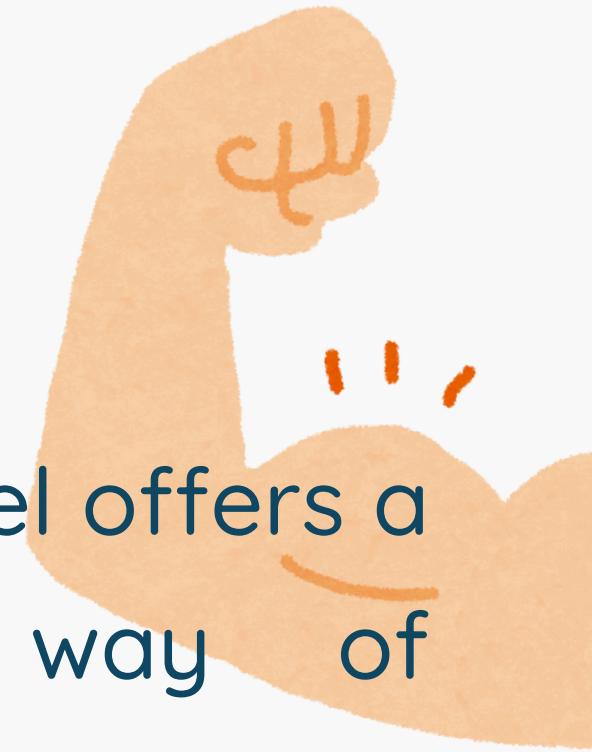
VAR (a multivariate model)

Vector Autoregression

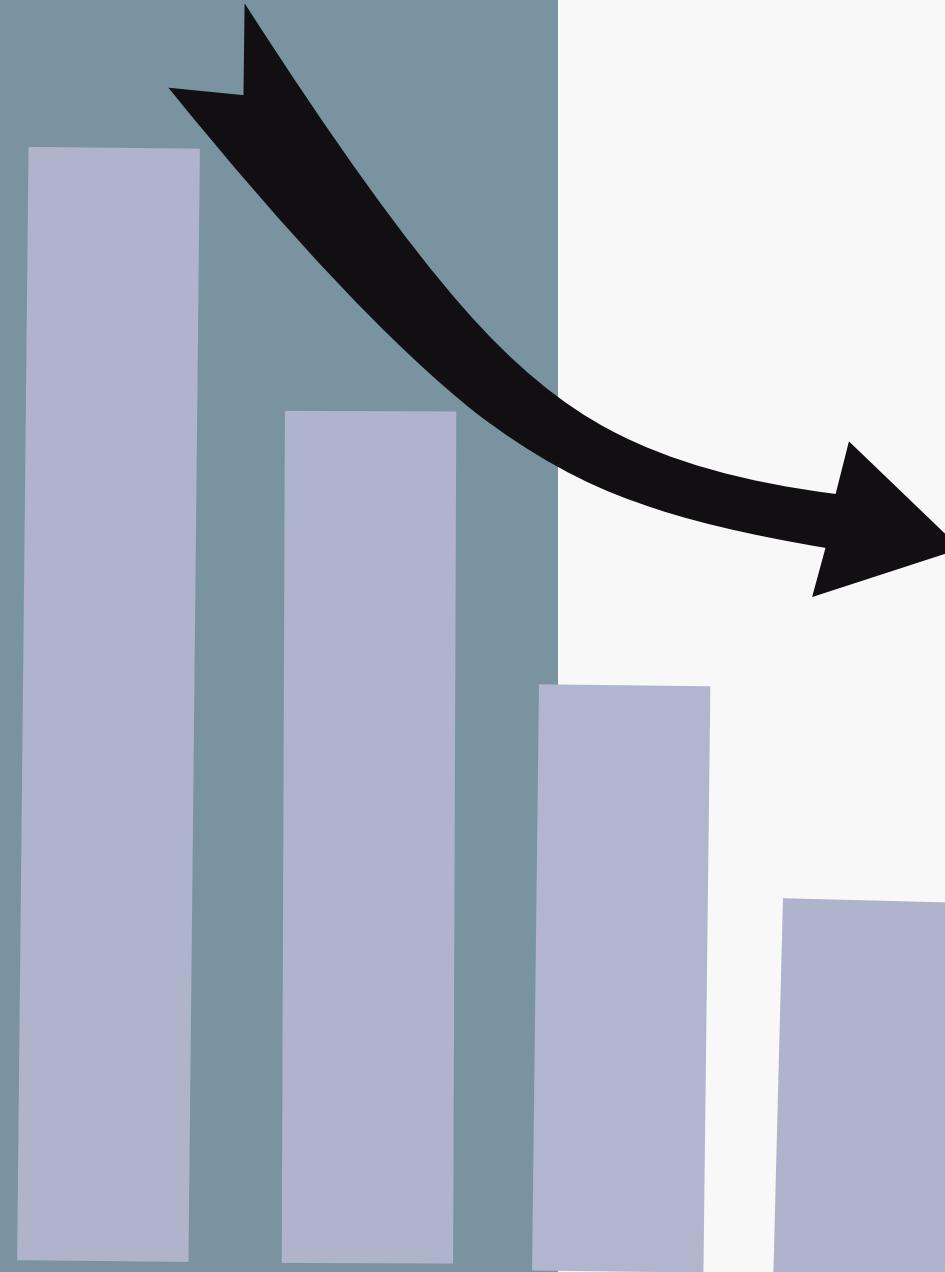
- **V: Vector.** The multivariate time series model involves two or more input variables, and leverages the interrelationship among the different time series variables.
- **AR: Autoregression.** A regression model that uses the dependence relationship between an observation and a number of lagged observations.

Strengths

- A vector autoregression model offers a **systematic** yet **flexible** way of capturing **complicated** real-world **behavior**.
- VAR models can provide **better forecasting performance** compared to simpler models

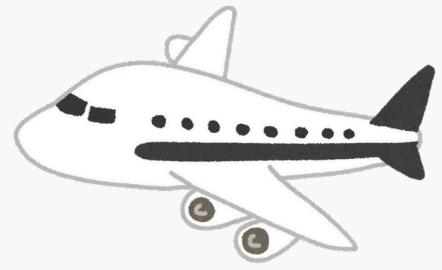


Weaknesses



- VAR models require **a large amount of data** and it's important to carefully choose the number of lags for accurate forecasting.
- Because the **error terms across equations are related**, it's difficult to isolate the effect of a single shock or variable on the whole system
- VAR models are relatively flexible but still **require assumptions about stationarity and variable relationships**.

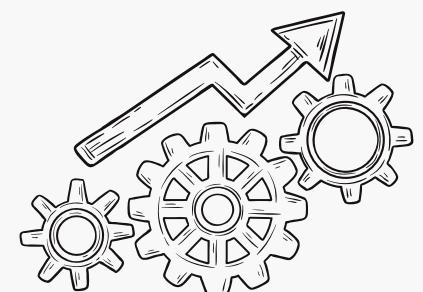
Application



Airplanes Forecast: In real-life airlines, ARIMA models can help airlines to make better decisions about their operations, such as scheduling flights, allocating resources, and adjusting prices



Stock Prices Prediction: VAR models estimate the relationships between different stocks, and ARIMA models can be used to predict the future values of a single stock



Engineering: These models can be used in a variety of real-life applications, such as predicting equipment failures, optimizing production processes, and forecasting demand for products

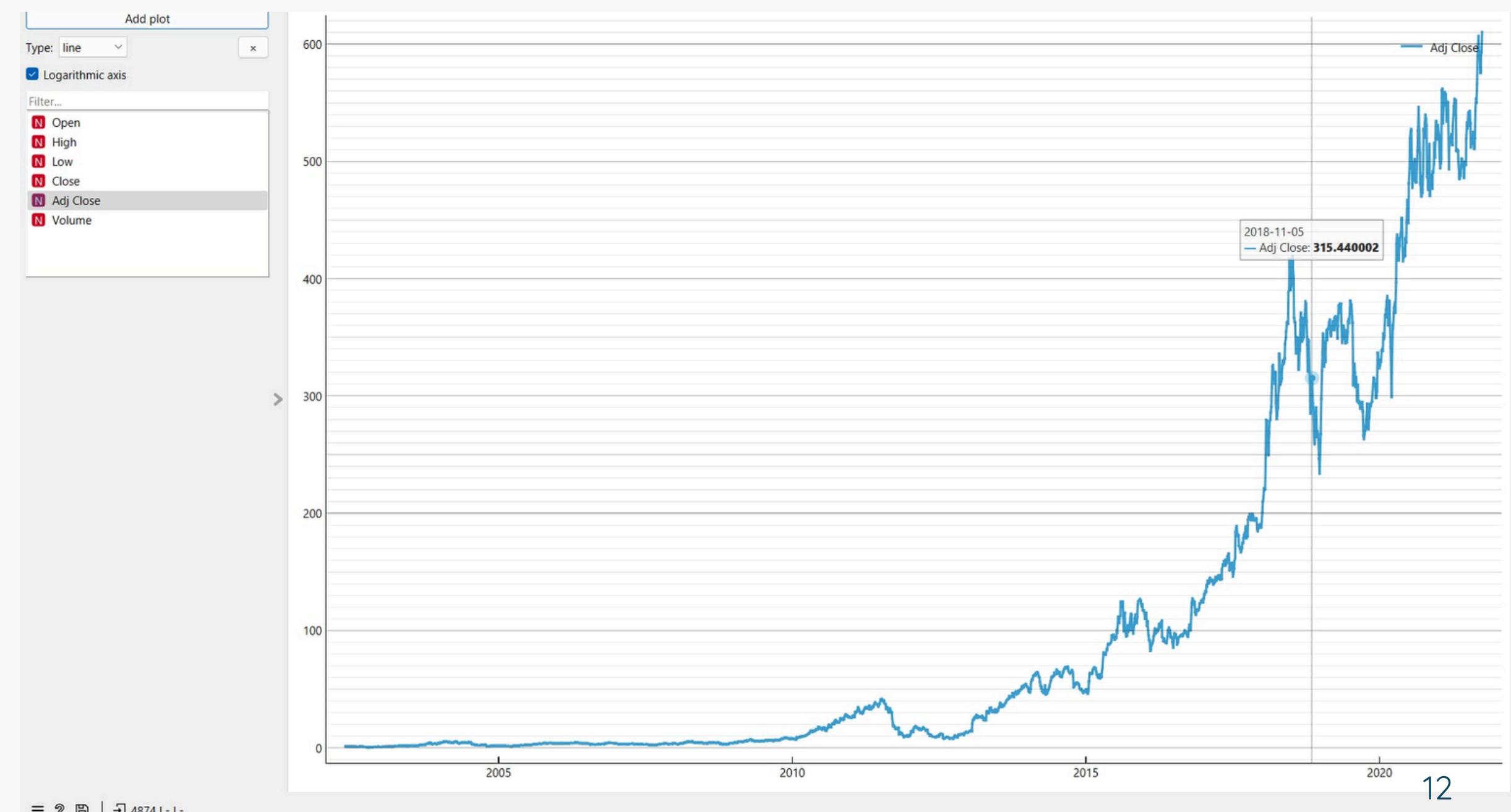
We are using Netflix's stock market dataset

2002-2021

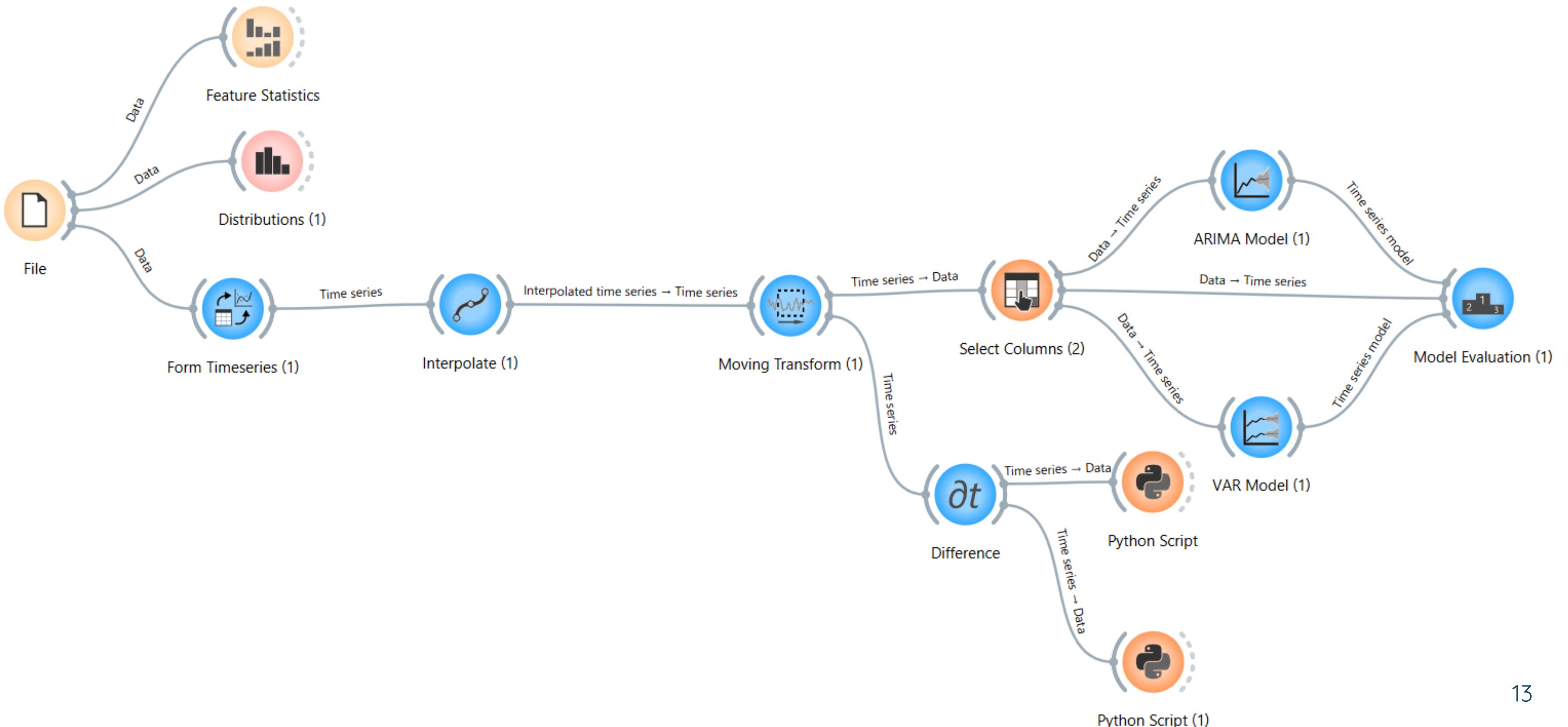


Research Object:

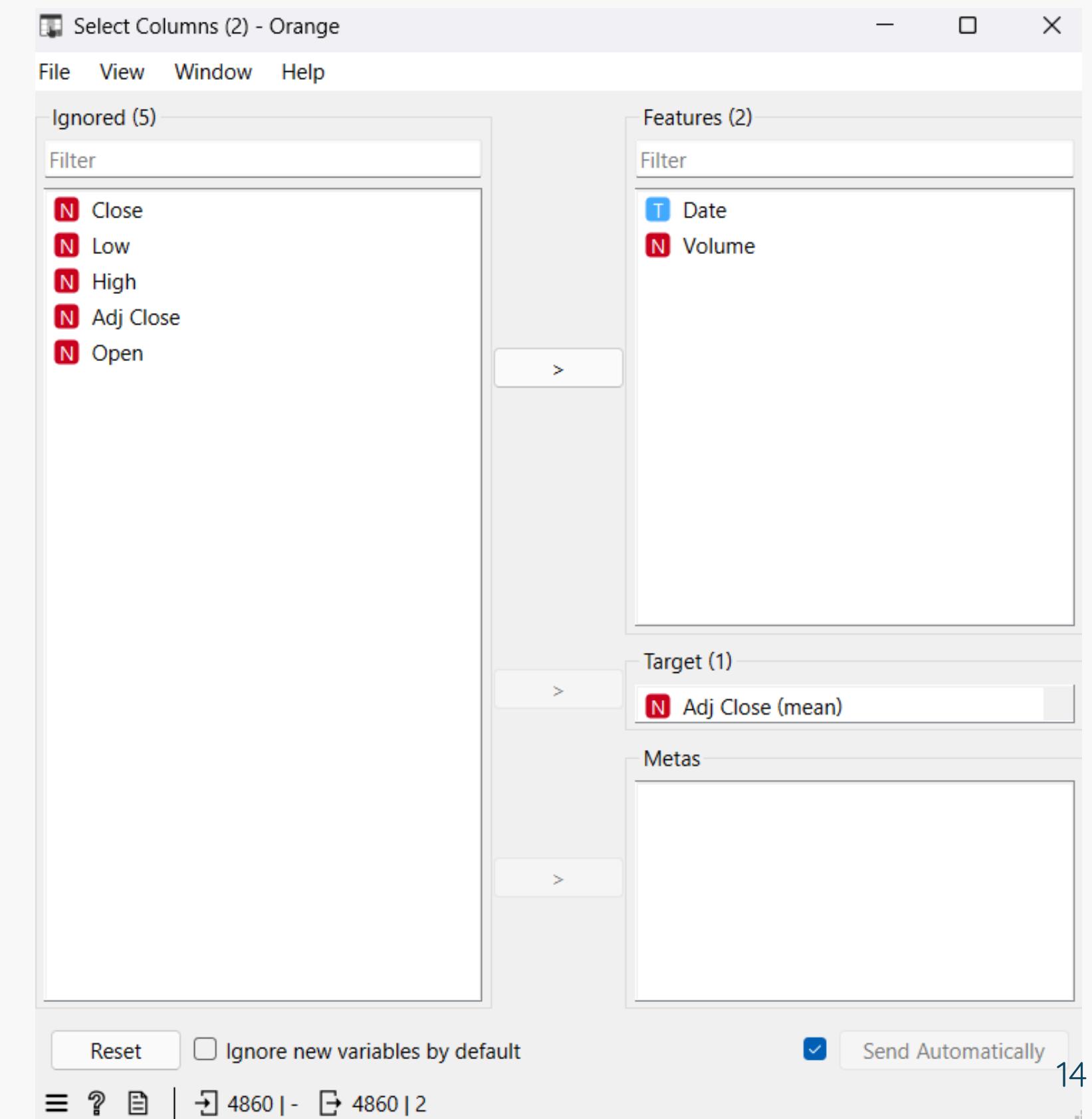
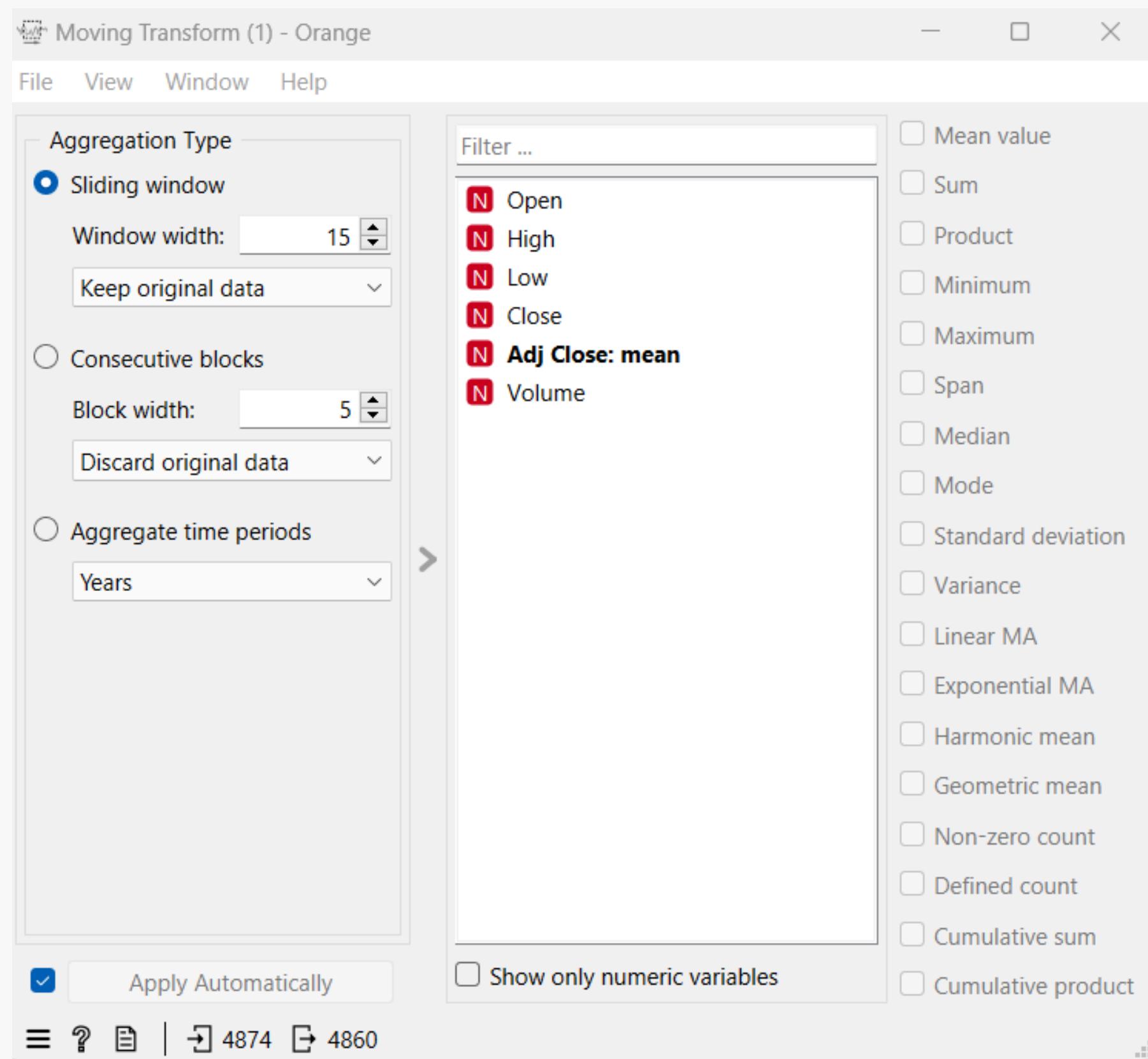
- + *To model and forecast Netflix stock price using an ARIMA and Var model on adjusted closing prices.*
- + *Optimize Arima Model.*



Orange workflow



Orange workflow



Model Evaluation (1) - Orange

— □ ×

View Window Help

Evaluation Parameters

Number of folds: 20

Forecast steps: 5

	RMSE	MAE	MAPE	POCID	\hat{R}^2	AIC	BIC
ARMA(1,0,0)	6.637	3.037	0.009	47.5	0.952	13382	13401
VAR(1,n)	6.188	3.277	0.009	65.7	0.958	55.4	55.4
ARMA(1,0,0) (in-sample)	1.823	0.094	0.005	89.7	1.000	13998	14018

View Window Help

Evaluation Parameters

Number of folds: 20

Forecast steps: 10

 Apply

	RMSE	MAE	MAPE	POCID	\hat{R}^2	AIC	BIC
ARMA(1,0,0)	12.0	6.597	0.017	44.7	0.739	12682	12701
VAR(1,n)	11.6	6.147	0.016	58.8	0.755	55.3	55.3
ARMA(1,0,0) (in-sample)	1.823	0.094	0.005	89.7	1.000	13998	14018
VAR(1,n) (in-sample)	1.002	0.092	0.005	88.3	1.000	55.4	55.4

View Window Help

Evaluation Parameters

Number of folds: 20

Forecast steps: 15

 Apply

	RMSE	MAE	MAPE	POCID	\hat{R}^2	AIC	BIC
VAR(1,n)	14.2	6.612	0.020	53.5	0.658	55.1	55.2
ARMA(1,0,0)	13.8	8.566	0.020	47.8	0.676	11807	11826
ARMA(1,0,0) (in-sample)	1.823	0.094	0.005	89.7	1.000	13998	14018
VAR(1,n) (in-sample)	1.002	0.092	0.005	88.3	1.000	55.4	55.4

We try to improve the result

- We set $d = 1$
-

$$\Delta X_t = X_t - X_{t-1}$$

where:

- ΔX_t is the differenced series,
- X_t is the value of the series at time t ,
- X_{t-1} is the value of the series at time $t - 1$.

Arima(1,0,0)

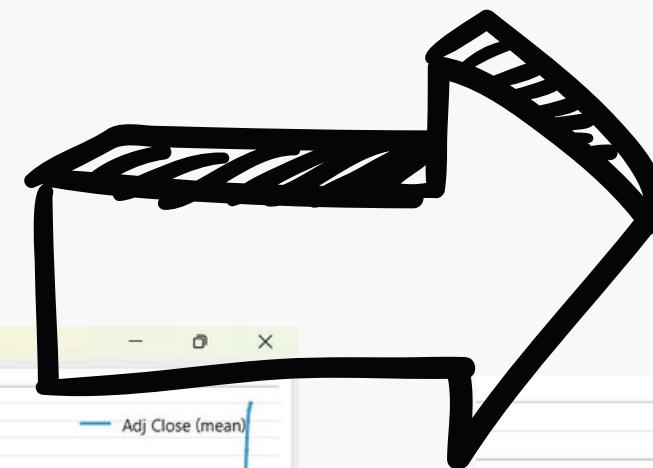
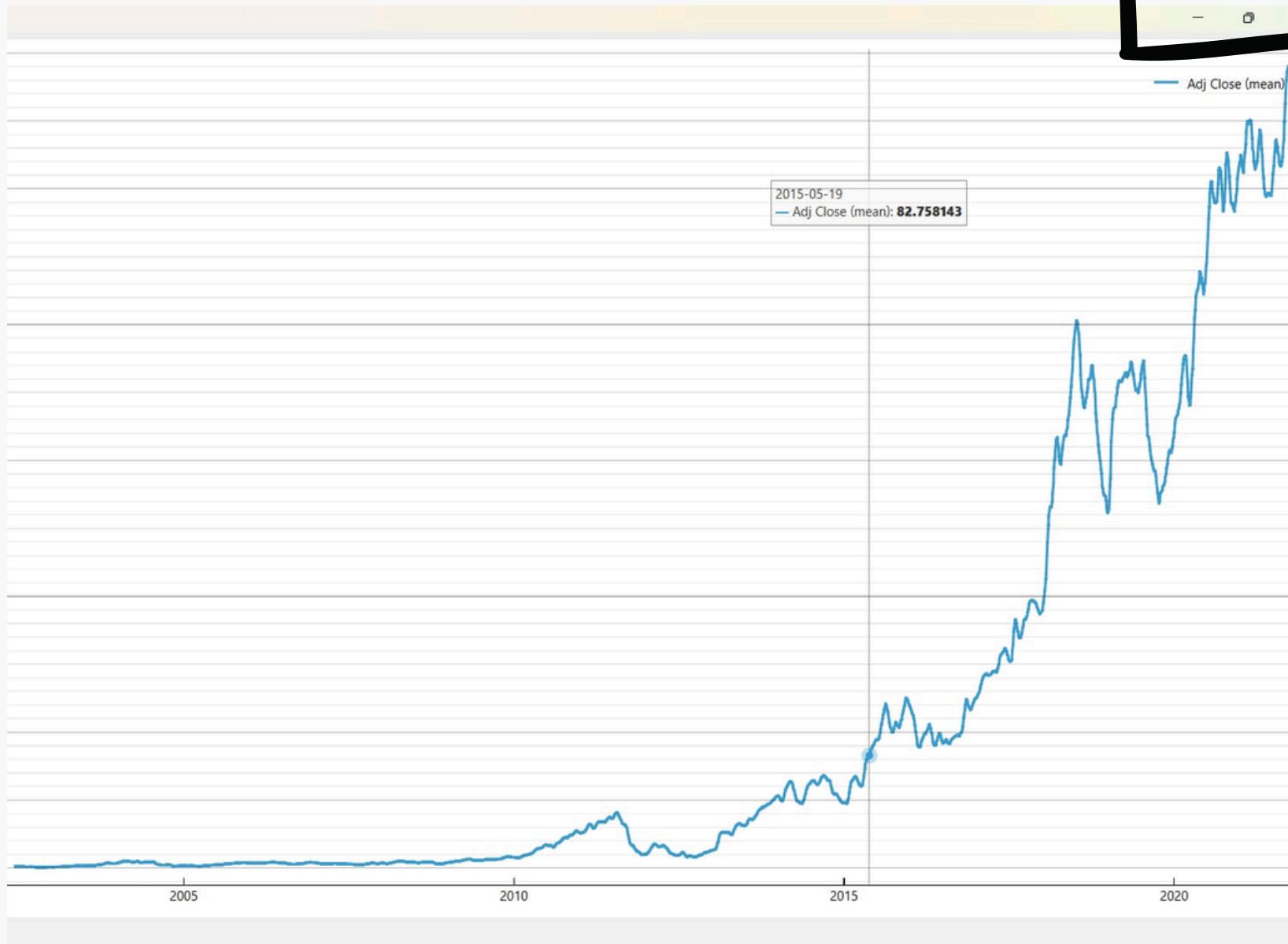


Arima(1,1,0)

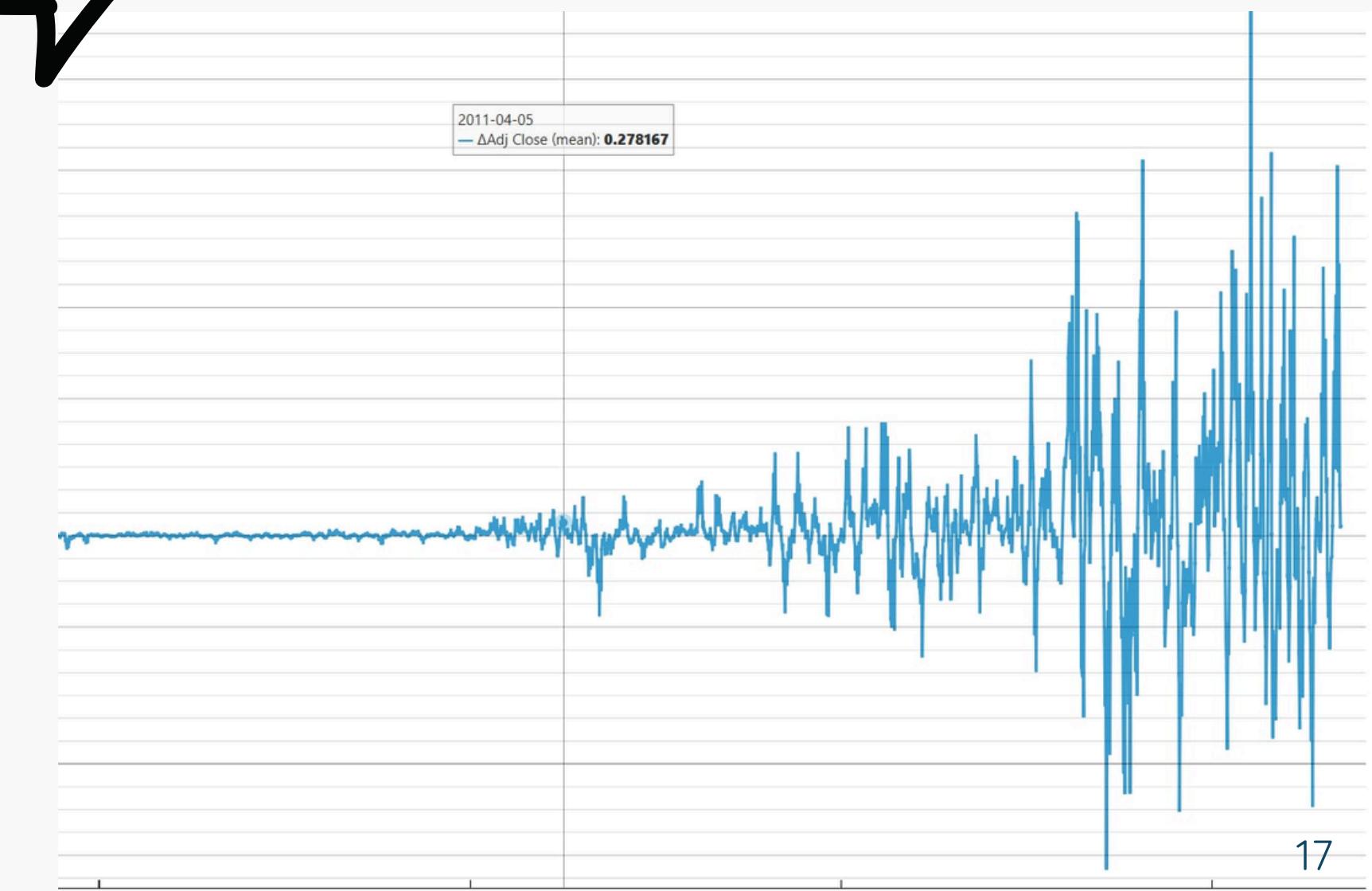
Stationary

We try to improve the result

Arima(1,0,0)



Arima(1,1,0)



Augmented Dickey-Fuller test (ADF test)

Result	Interpretation
p-value < 0.05	Reject $H_0 \Rightarrow$ Series is stationary
p-value ≥ 0.05	Fail to reject $H_0 \Rightarrow$ Series is non-stationary

```
Console
File "C:\Users\lemin\AppData\Local\Programs\Orange\Lib\site-pac
domain = tab.domain
^^^^^^^^^
AttributeError: 'NoneType' object has no attribute 'domain'
>>>
Running script:
ADF Statistic: -9.589463872315827
p-value: 2.0545699889893552e-16
>>>
Running script:
>>>
```

The ADF test helps you decide:

Null hypothesis (H_0): The series has a unit root (i.e., non-stationary)

Alternative hypothesis (H_1): The series is stationary

Conclusion from new model

After performing model optimization and comparing evaluation metrics, ARIMA(1,1,0) outperformed the VAR(1,n) model. Specifically, ARIMA demonstrated a higher R^2 (indicating better explanatory power) and a lower RMSE (indicating better prediction accuracy).

Parameters	ds:	20	RMSE	MAE	MAPE	POCID	\hat{R}^2
			6.188	3.277	0.009	65.7	0.958
		5	3.377	1.542	0.004	83.8	0.988
	>	ARIMA(1,1,0)					

Parameters	ds:	20	RMSE	MAE	MAPE	POCID	\hat{R}^2
			11.6	6.147	0.016	58.8	0.755
		10	7.849	4.192	0.011	68.3	0.888
	>	ARIMA(1,1,0)					

Conclusion

- **Strengths & Weaknesses:**

- **ARIMA:** Effective for stationary data with strong autocorrelations but struggles with complex relationships among multiple variables.
- **VAR:** Handles multiple time series with interdependencies but requires a larger dataset and assumes variable stationarity.

- **Choosing ARIMA vs. VAR:**

- ARIMA when analyzing a single time series with clear patterns.
- VAR when multiple time series influence one another and relationships need to be modeled.

References

Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. *Future Internet*, 15(8), 255. <https://doi.org/10.3390/fi15080255>

Feigelson, E. D., Babu, G. J., & Caceres, G. A. (2018). Autoregressive Times Series Methods for Time domain Astronomy. *Frontiers in Physics*, 6. <https://doi.org/10.3389/fphy.2018.00080>

Siami-Namini, S., & Namin, A. S. (2018). Forecasting Economics and Financial Time Series: ARIMA vs. LSTM. *arXiv* (Cornell University). <https://doi.org/10.48550/arxiv.1803.06386>

Bhattacharyya, R. (2025b, January 9). Vector autoregression. What Is It, Examples, Assumptions, Types. <https://www.wallstreetmojo.com/vector-autoregression/>



.....

Thank you

.....