



# Generalization of ARIMA(2,2) Stock Forecasting Across Various Industries

Orhan Koc  
orhankoc@uw.edu

Jyunghyun Noh  
jyungn@uw.edu

Siew Kit Liew  
siewkit@uw.edu

February 10, 2023

## Abstract

This study aims to assess how well ARIMA(2,2) price forecasting generalizes for stocks across across different indsutries. The assessment is conducted using Analysis of Variance (ANOVA) to assess if there is significant difference in forecasting error among different stocks. We will be employing Randomized Complete Block Design, acquiring an observation for each stock at the end of every day. Data will be blocked with respect to observations ( $t_n$ ) in the granularity of days. The results of the study will provide insights into the robustness and generalization of the ARIMA(2,2) model in stock price forecasting across different stocks and markets. The findings of this study will be valuable for investors, financial analysts, and academics interested in stock price forecasting and time series modeling.

# 1 Introduction

Price forecasting is a widely studied subject usually implemented with a variety of well-tested machine learning algorithms. A price forecasting algorithm is said to be successful if the logic can be generalized to an extent. **How is ARIMA's forecasting error margin affected by different stock prices ?**

**Hypothesis H<sub>0</sub>:** Mean forecasting error of ARIMA(2,2) is statistically the same across all stocks,  $\mu_1 = \mu_2 = \mu_3 \dots \mu_n$

**Hypothesis H<sub>1</sub>:** Mean forecasting error of ARIMA(2,2) is statistically different in at least one stock,  $\mu_a \neq \mu_b$  where  $\mu_a, \mu_b \in \{\mu_1, \mu_2, \mu_3, \dots, \mu_n\}$

# 2 Experiment Variables

There is one dependent variable in our experiment, and one independent variable with multiple levels.

**Dependent Variable** of this experiment is the error margin of the price forecasted from ARIMA(2,2). The calculation of error margin for each cell is as follows:

$$y_{i,j} = |PA_{i,j} - PF_{i,j}|$$

where  $PA_{i,j}$  denotes actual price of a stock and  $PF_{i,j}$  denotes the forecasted price in  $i$ th asset of  $j$ th day. It's important to note we are not concerned with the direction of error, and only concerned with magnitude.

**Independent Variable** of this experiment are stock prices from different industries, with multiple levels representing different stocks with the following tickers:

- GOOG
- AMZN
- APPL
- AXP
- ADSK
- ...

# 3 Design of Experiment

The experiment will use Analysis of Variance with a Randomized Complete Block Design, where observations (timestamps) will be used as blocks.

Stock — Observations					
Stock Ticker	$t_1$	$t_2$	$t_3$	...	$t_n$
GOOG	$y_{1,1}$	$y_{1,2}$	$y_{1,3}$	...	$y_{1,n}$
AMZN	$y_{2,1}$	$y_{2,2}$	$y_{2,3}$	...	$y_{2,n}$
APPL	$y_{3,1}$	$y_{3,2}$	$y_{3,3}$	...	$y_{3,n}$
...	...	...	...	...	...
level <sub>a</sub>	$y_{a,1}$	$y_{a,2}$	$y_{a,3}$	...	$y_{a,n}$