CAPSTONE DESIGN

MODELLING THE SPREAD OF COVID-19 IN HO CHI MINH CITY: DATA-DRIVEN ANALYSIS

Supervisor: Tran Van Ly

Group 23 Presentors:
Vo Thi Thien My (team leader)
Nguyen Huynh Ngoc Que
Pham Ngoc Thu Uyen

TABLE OF CONTENTS

Presentation Outline

I. INTRODUCTION

1.1 Motivation

- Assist the city establishing effective resources planning and control measures.
- Inform the danger of the upcoming outbreak for better personal protection.

1.2 Problem statement

Previous research papers did not consider the external factors that contribute to the spread of COVID-19.

1. INTRODUCTION

1.3 Objectives

- Provide a suitable model to forecast the spread of COVID-19 in Ho Chi Minh city to increase forecasting accuracy and successfully include external elements that directly affect the epidemic spread.

1.4 Project requirements

- Provide a suitable model approach.
- Investigate the external factor.
- Consider medical standards on data collection and model forecast.

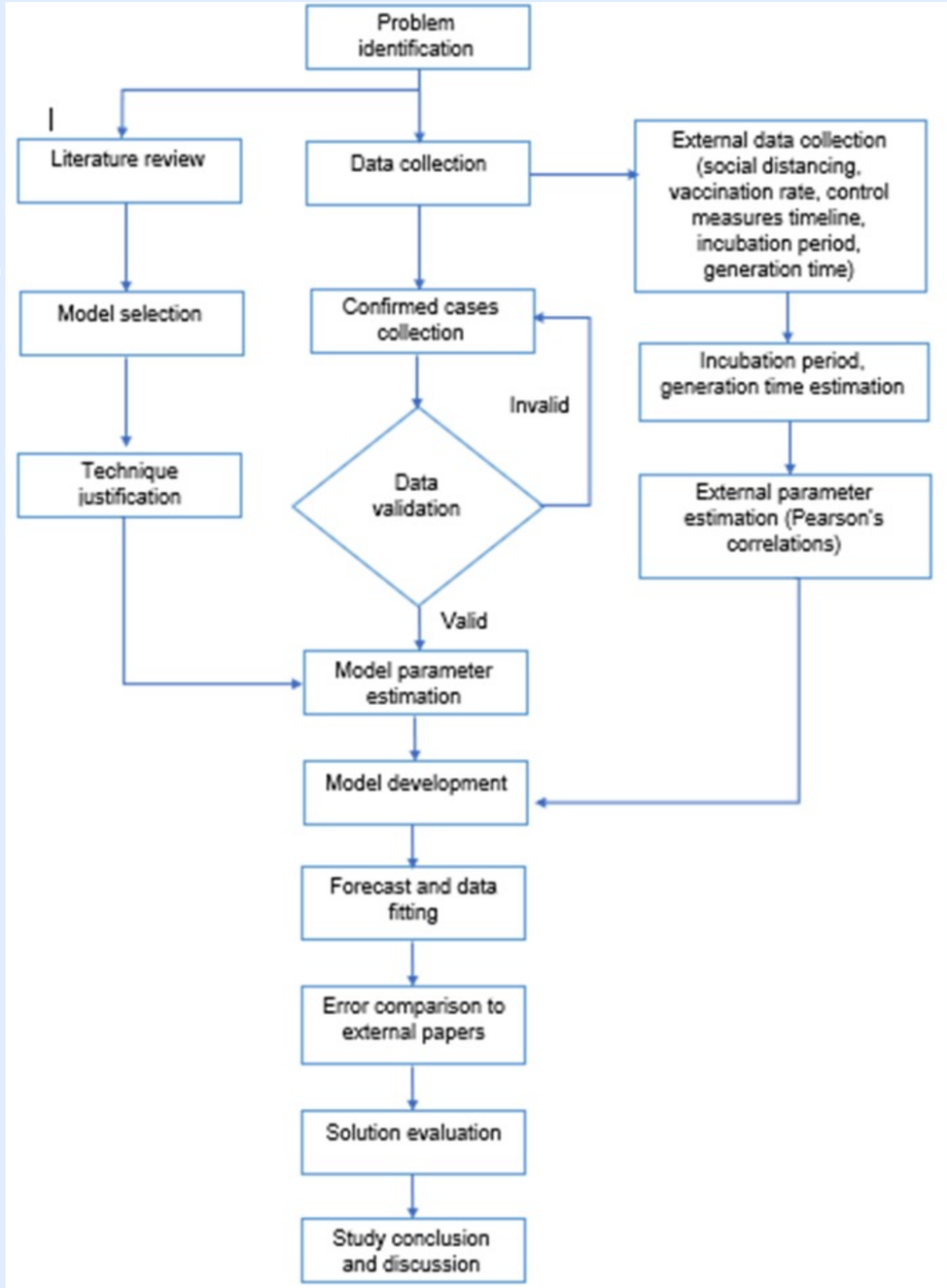
1. INTRODUCTION

1.5 Scope and limitation

- Forecasting the COVID-19 confirmed cases in HCMC from April 27, 2021 to November 24, 2021 with external factors included.
- Immigration and emigration rate, vaccination efficiency deviation excluded.
- Lack of knowledge and time for researching.
- Data may not be accurately updated.

1. INTRODUCTION

FLOWCHART OF MODELLING THE SPREAD OF COVID-19 STUDY



II. DESIGN CONCEPTS CONSIDERATION

2. 2 Design concepts consideration

- use ARIMA time series model to forecast the spread of COVID-19

- include more external parameters into the model to improve forecast accuracy

III. MODEL DESIGN AND ANALYSIS

Approach Consideration and Selection

System Design Description

1. SYSTEM DESIGN DESCRIPTION

ACF and PACF

- ACF:
  - Determining how closely values connected to one another
  - Depicting the correlation coefficient versus the lag
- PACF: capturing the correlation between 2 variables
- Reason: Estimating the AR(p) and MA(q) parameters

ADF Test

- Giving results in hypothesis test with null and alternative hypotheses
- p-value from which would be needed to make inferences, whether it is stationary or not
- Hypothesis test:
  - H0: the time series has unit root (unstationary)
  - H1: the time series is stationary

1. SYSTEM DESIGN DESCRIPTION

Pearson's Correlation

- 4 direct external factors: daily vaccination coverage, social distancing by policy in accordance with control measures, the Covid variant incubation time and Covid variant generation time
- Pearson correlation coefficient methodology to seek the relationship between external factors and the main forecasting variables

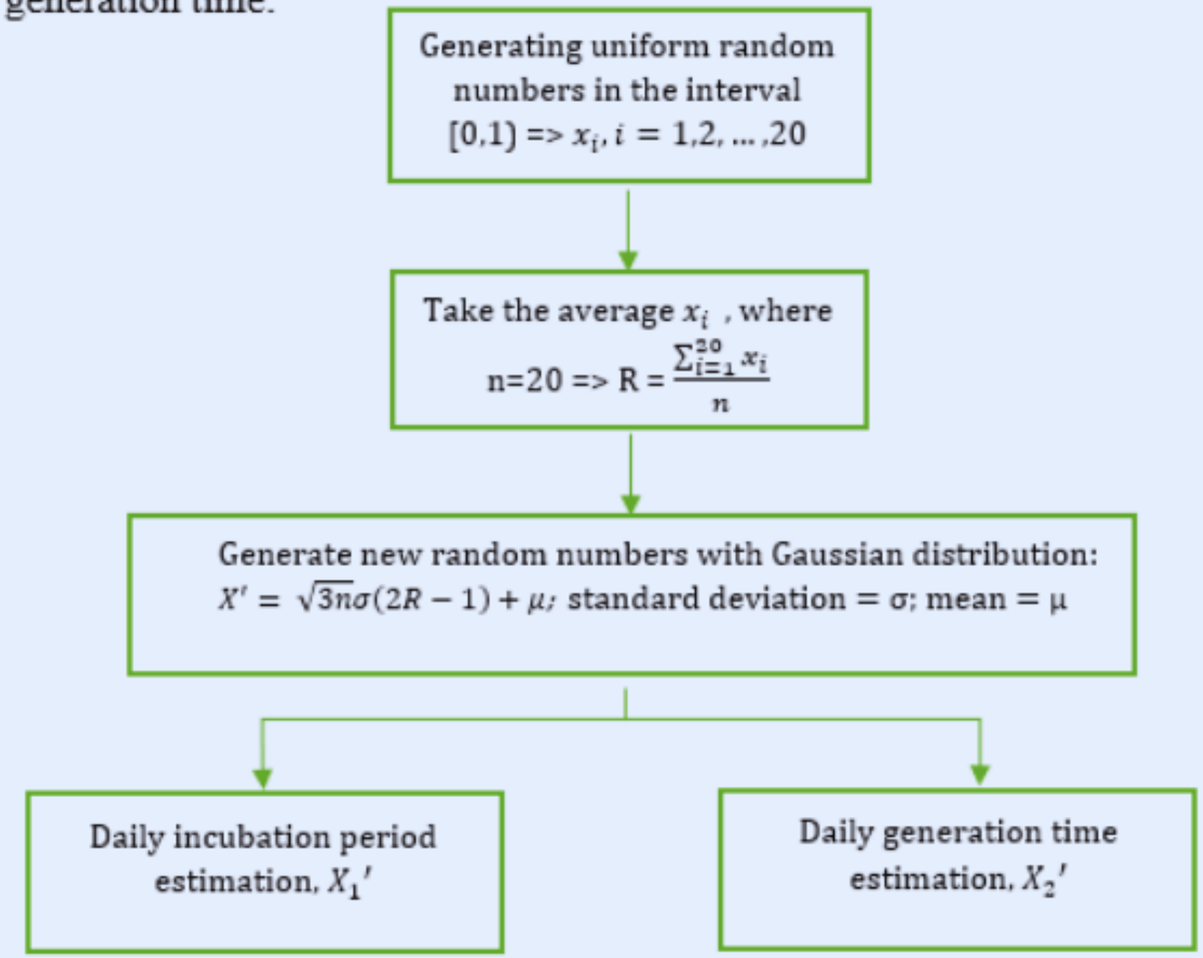
$$\rho\,(x,y) = \frac{cov\,(x,y)}{\sigma x^{*}\sigma y}$$

- (x, y): pair of random variables used in correlation study
- Cov: covariance
- σx: standard deviation of variable x
- σy: standard deviation of variable y

External Parameter Estimation

- 2 parameters in transmission rate: incubation period and generation time
- Incubation period: time period in which the exposed individual is infected and starts to present illness
- Generation time: time interval between infection of primary case and its secondary cases

The following flowchart is the steps need to be taken so as to find incubation period and generation time:



Flow chart of generation time and incubation period estimation method

VI. PROTOTYPE DEVELOPMENT & IMPLEMENTATION

1. MODEL DEVELOPMENT

ARIMA Parameter Identification

- Autoregressive (AR): difference between observed value of the time series and its adjacent value in a defined time lag
- Integration (I): number of differencing time to transform non-stationary to stationary time series
- Moving Average (MA): difference between observed value and residual error in a defined time lag

3 notations:

- p: the order of AR, estimated by using statistical tools
- d: the order of differencing, estimated by counting
- q: the order of MA, estimated by using statistical tools

	ACF plot	PACF plot
White noise	All zero	All zero
MA (q)	Drop off after lag q	Die down
AR (p)	Die down	Drop off after lag p
ARMA (p, q)	Die down	Die down

Rules for estimation of p and q order in ARIMA model

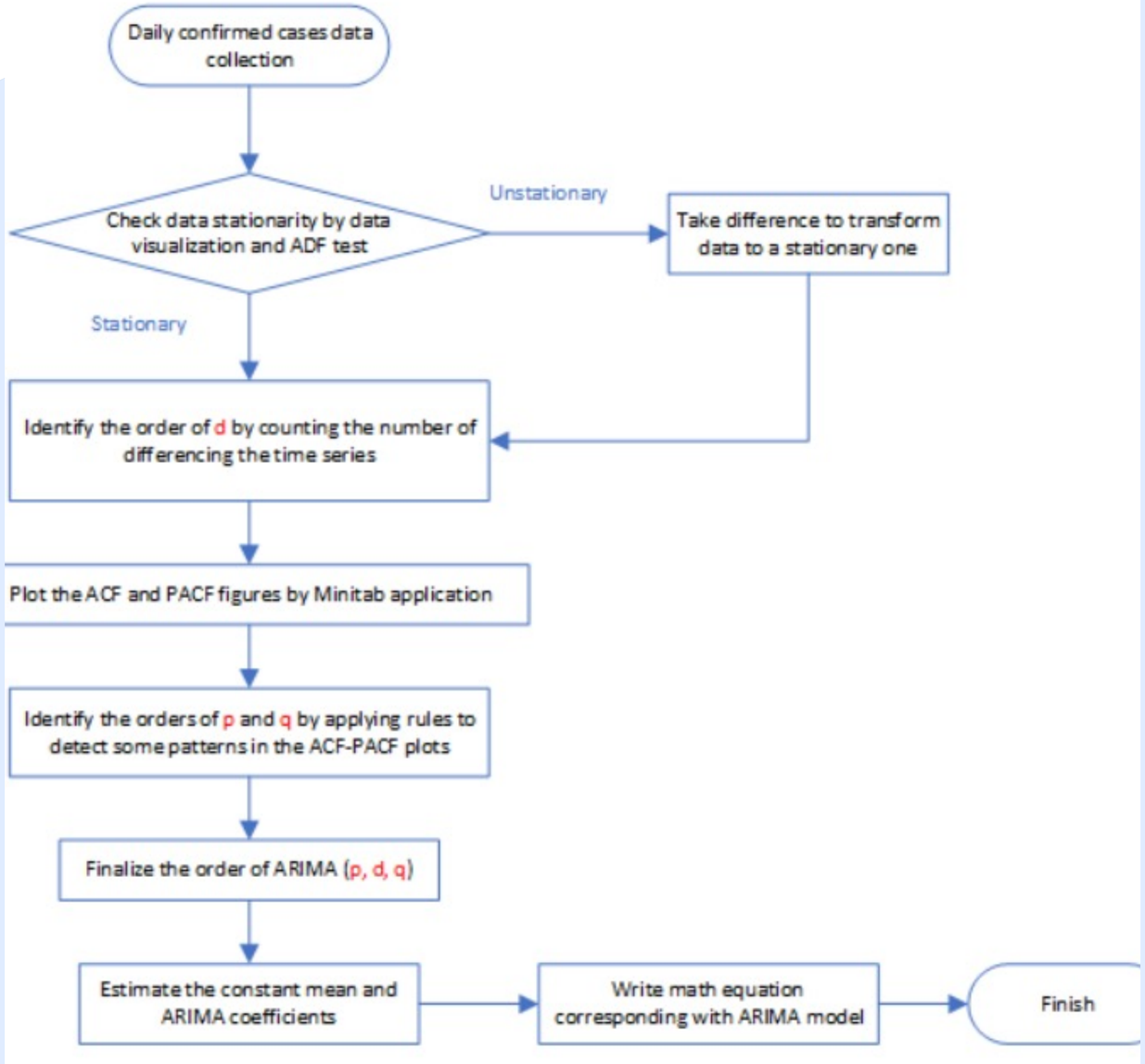
2. Model Parameters Estimation

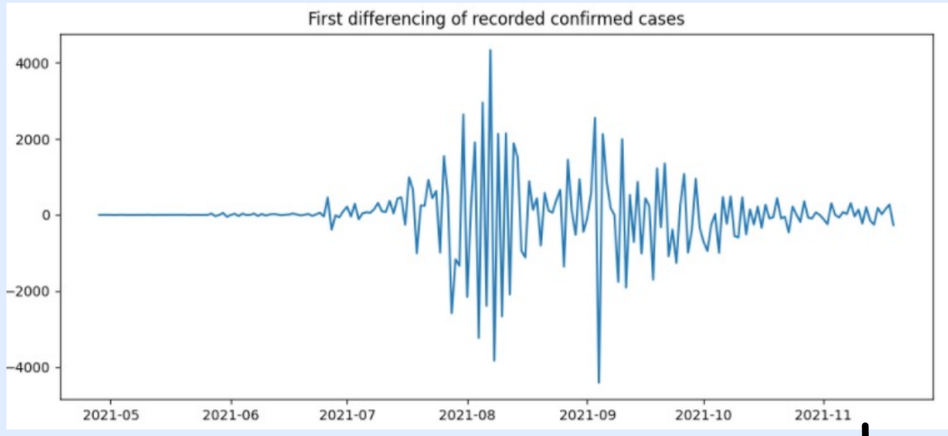
Part 1: Construction of ARIMA model

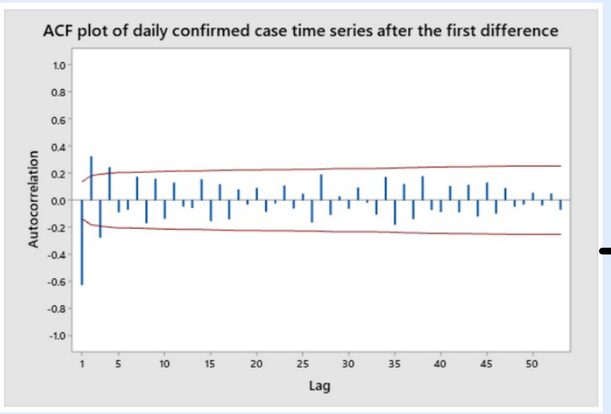
Part 2: Construction of External factor model

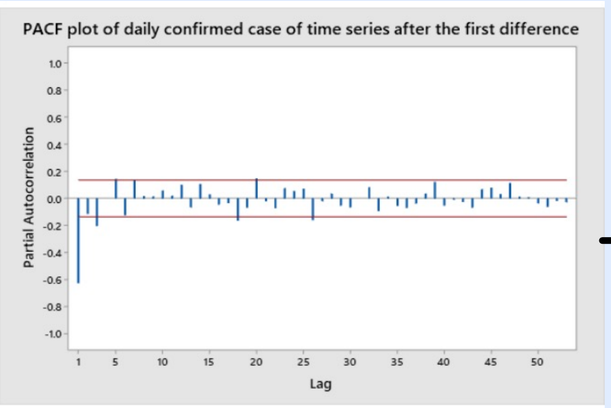
PART 1. CONSTRUCTION OF ARIMA MODEL

Flow chart of ARIMA modelling process







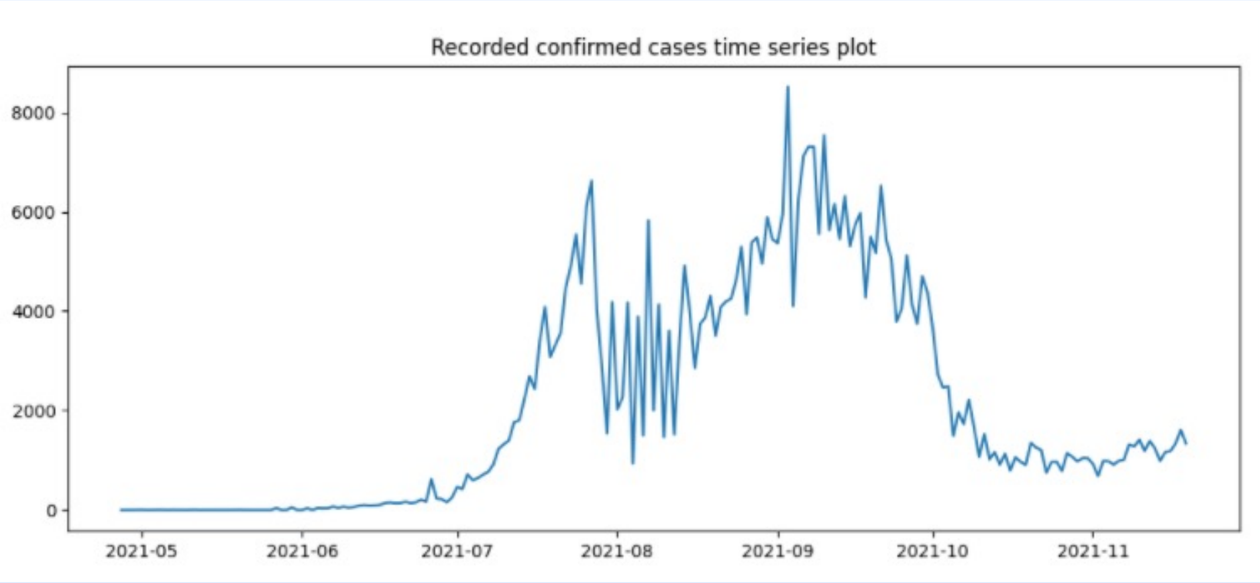


Model: ARIMA (1,1,0)

PART 1. CONSTRUCTION OF ARIMA MODEL

Data Stationary Checking: ADF test

A collection of confirmed case data in a daily time frame from 27 April, 2021 was performed and the research team continued to update the data daily into the data tracking file



Daily confirmed cases during the fourth wave of COVID-19 pandemic in Ho Chi Minh city

Null Hypothesis: SER01 has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 5 (Automatic - based on t-statistic, lagpval=0.05, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.229351	0.9012
Test critical values:	1% level	-4.004365	
	5% level	-3.432339	
	10% level	-3.139924	

*MacKinnon (1996) one-sided p-values.

ADF test results for stationarity of 0 differencing in confirmed cases

The test statistics were greater than the critical values of three different confidence levels. Hence, data was statistically non-stationary.

PART 1. CONSTRUCTION OF ARIMA MODEL

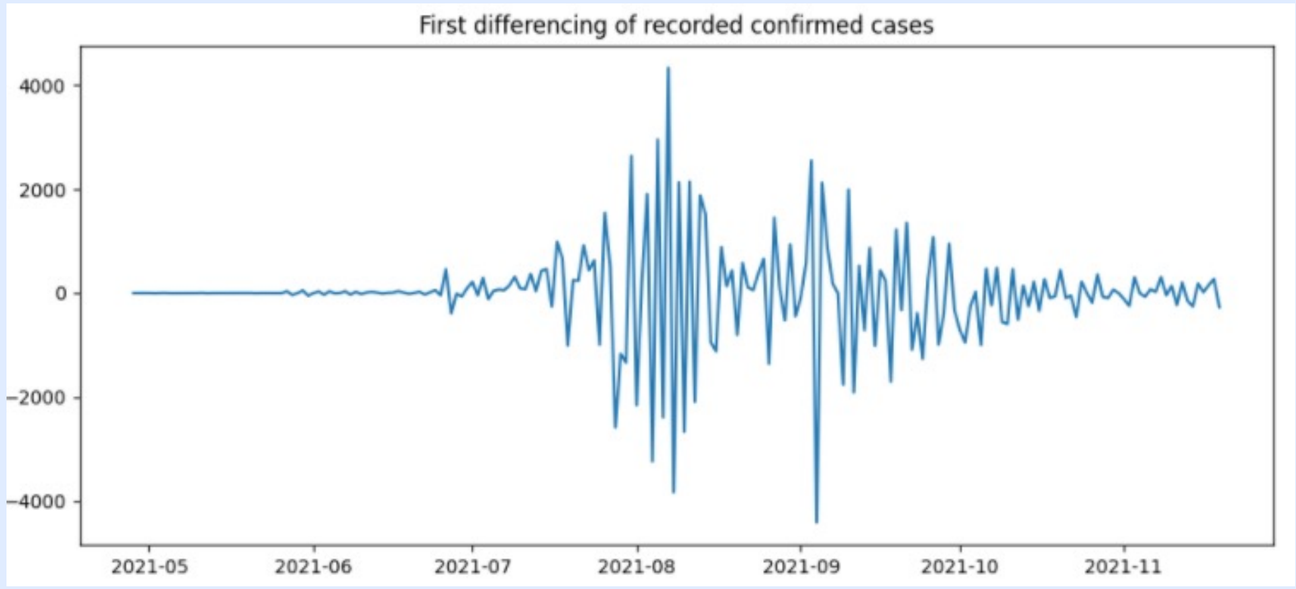
Data Stationary Checking: ADF test

Null Hypothesis: SER01 has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 5 (Automatic - based on t-statistic, lagpval=0.05, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.229351	0.9012
Test critical values:	1% level	-4.004365	
	5% level	-3.432339	
	10% level	-3.139924	

*MacKinnon (1996) one-sided p-values.

ADF test results for stationarity of 1st differencing in confirmed cases



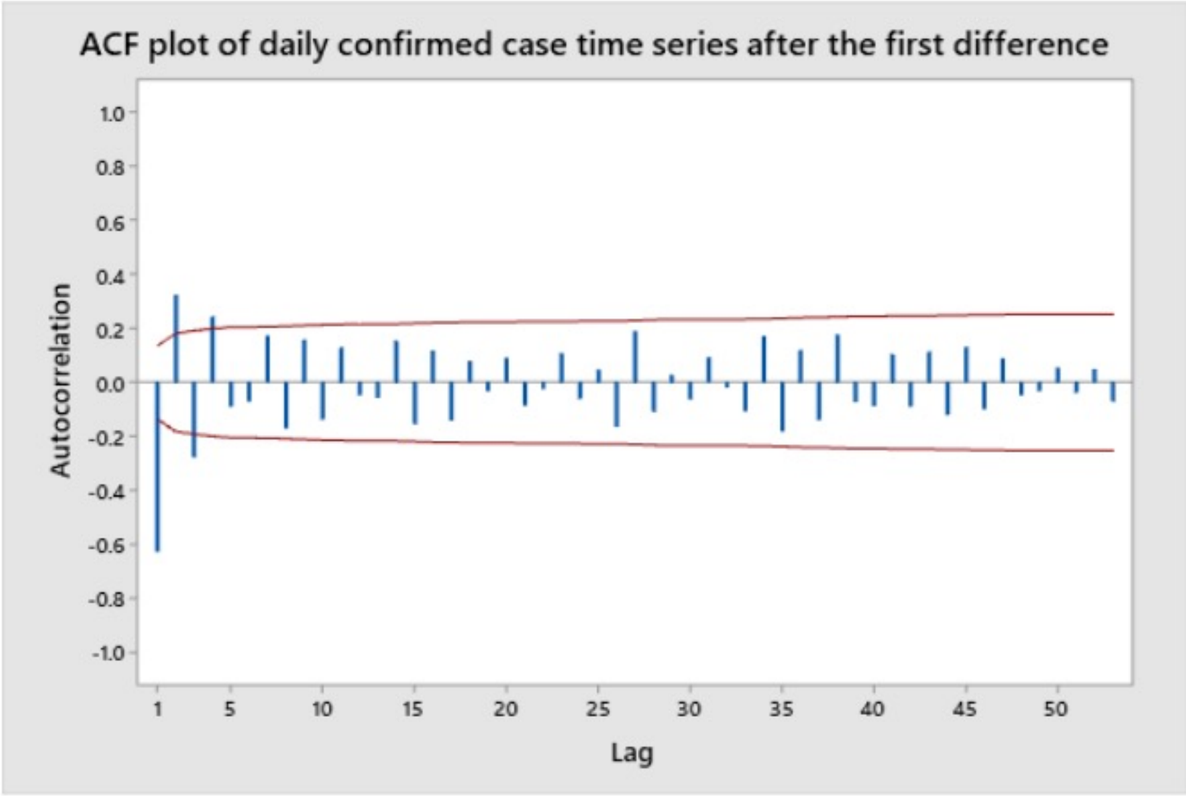
Daily confirmed cases distribution in the first differencing

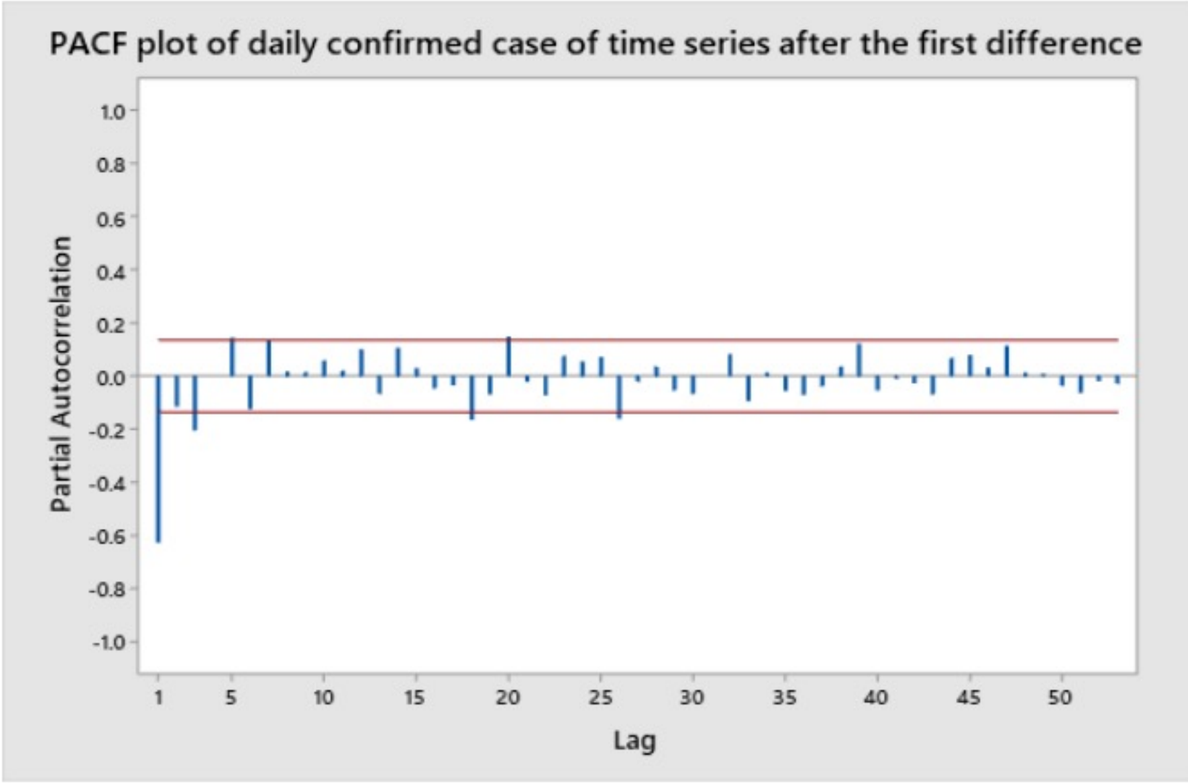
Taking the first difference to transform the data and did an ADF test again to check the stationarity of the first-difference data. The time series plot and ADF test both indicated the data was now stationary.

PART 1. CONSTRUCTION OF ARIMA MODEL

ARIMA Model Selection: ACF & PACF

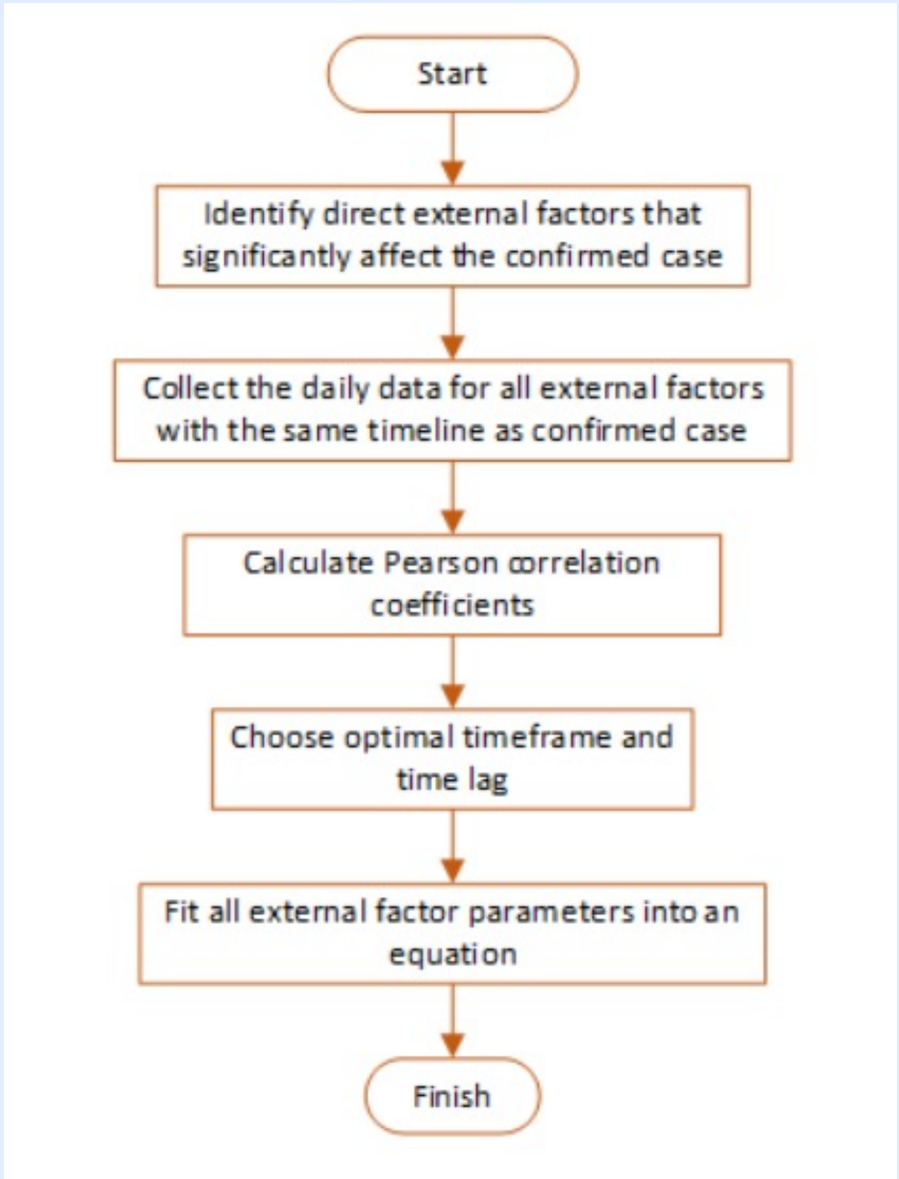
The order of the ARIMA model was identified based on the ACF and PACF of the stationary time series, which was the first-difference confirmed case data.





ACF and PACF plot of daily confirmed cases during the first differencing

PART 2. CONSTRUCTION OF EXTERNAL FACTOR MODEL



Flow chart of estimation of external factors and implementation into forecasting model

PART 2. CONSTRUCTION OF EXTERNAL FACTOR MODEL

In this case, the Pearson correlation coefficients were 0.52, 0.832, -0.043, 0.131 for daily vaccination coverage (x1), social distancing policy (x2), incubation time (x3), generation time (x4), respectively. Subsequently, the best β parameters for our model are:

$$\beta_1 = \frac{\rho_2}{\rho_1+\rho_2+\rho_3} = \frac{0.52}{0.52+0.832-0.043+0.131} \approx 0.361$$

$$\beta_2 = \frac{\rho_2}{\rho_1+\rho_2+\rho_3} = \frac{0.832}{0.52+0.832-0.043+0.131} \approx 0.577$$

$$\beta_3 = \frac{\rho_2}{\rho_1+\rho_2+\rho_3} = \frac{-0.034}{0.52+0.832-0.043+0.131} \approx -0.03$$

$$\beta_4 = 1 - \beta_1 - \beta_2 - \beta_3 = 1 - 0.361 - 0.577 + 0.03 \approx 0.092$$

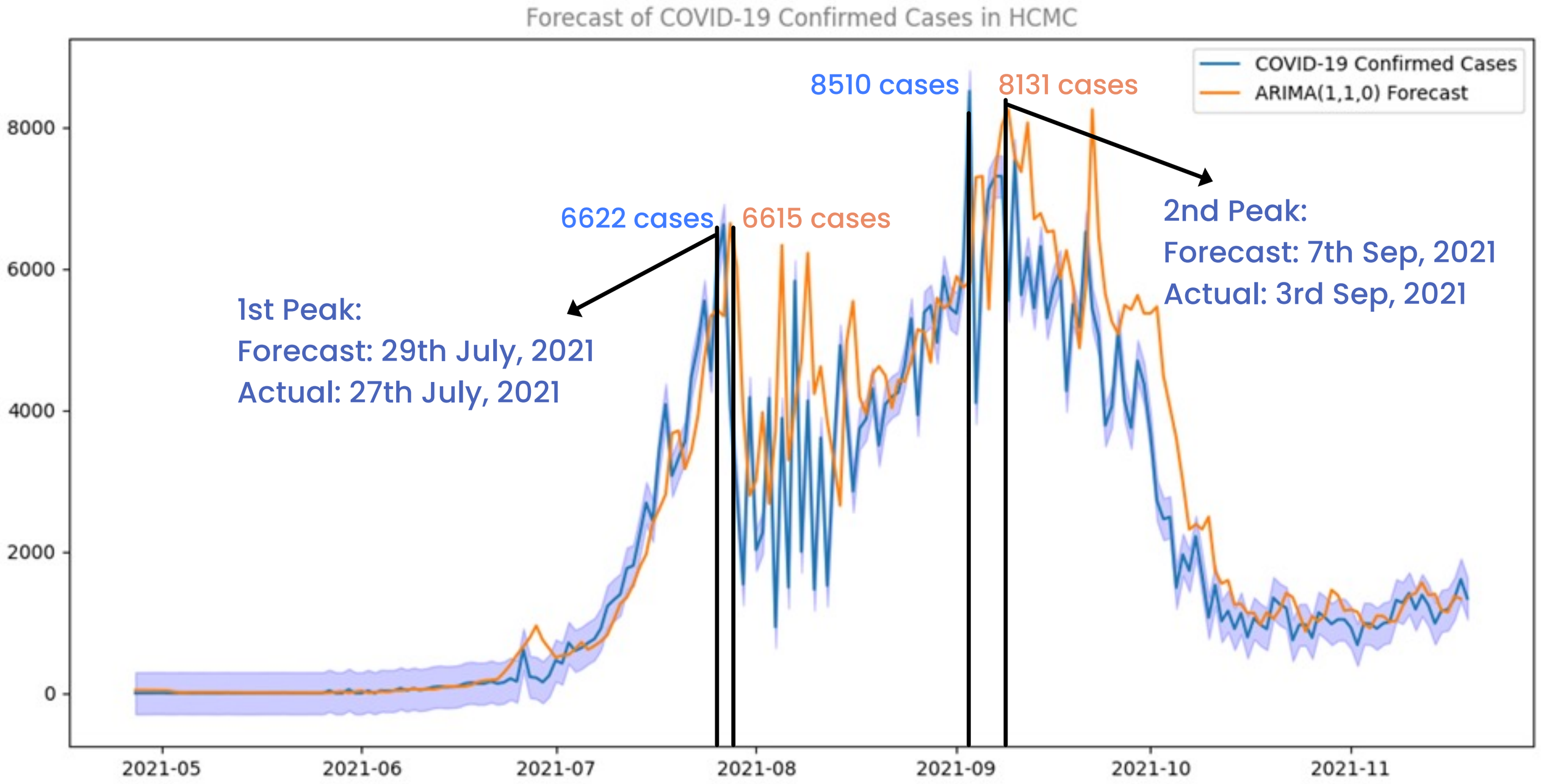
With the calculated β, the final forecasting model established in this study was as follows:

$$y(i) = \text{ARIMA}(1, 1, 0)(i) + 0.361\, x_1(i-2) + 0.577\, x_2(i-2) - 0.03\, x_3(i-2) + 0.092\, x_4(i-2)$$

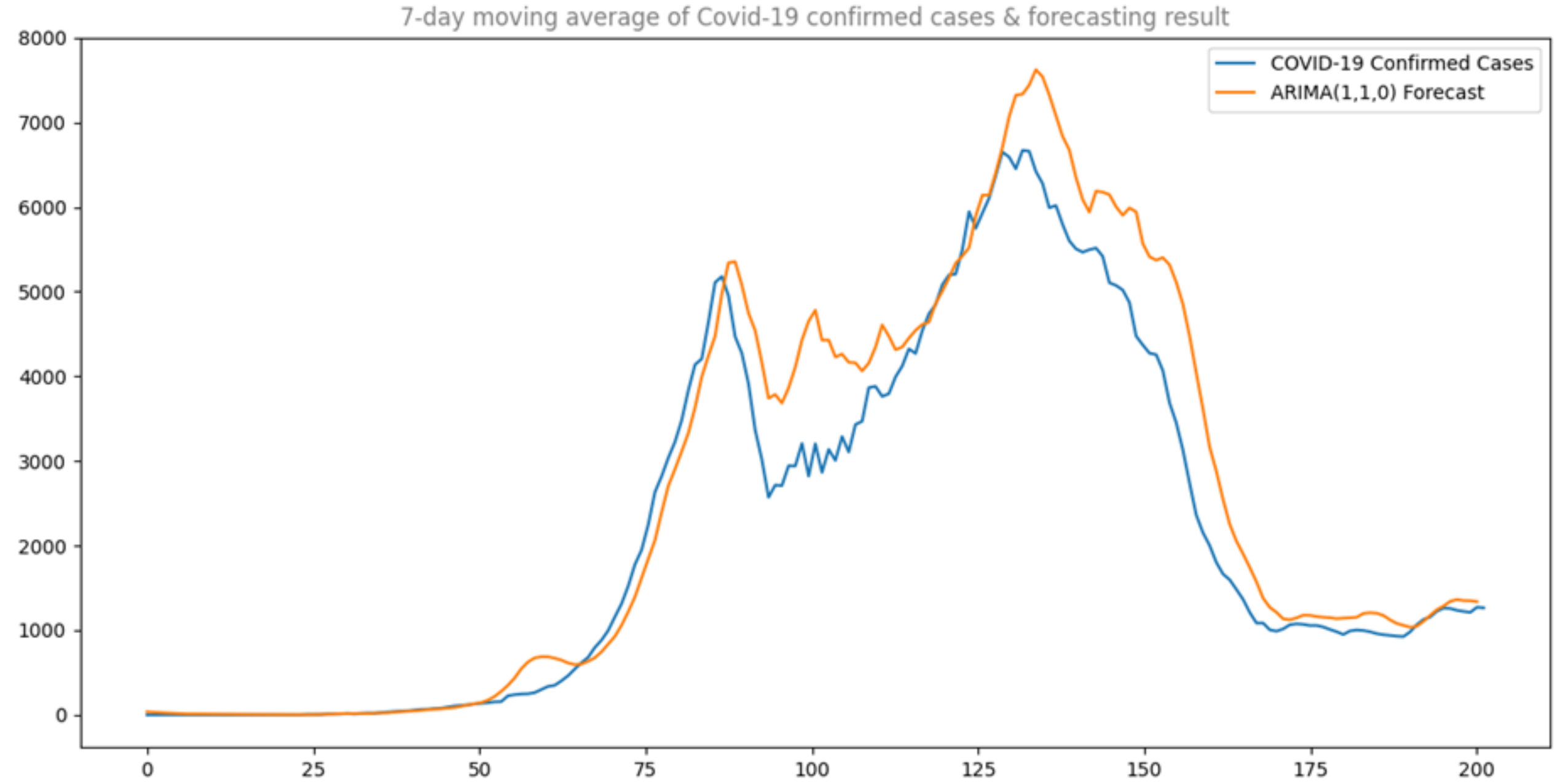
$$= 10.4 + Y_{t-1} - 0.628(Y_{t-1} - Y_{t-2}) + 0.361\, x_1(i-2) + 0.577\, x_2(i-2) - 0.03\, x_3(i-2) + 0.092\, x_4(i-2)$$

V. FORECASTING RESULTS & ANALYSIS

Part 1: Forecasting Results



Part 2: 7-Day Moving Average



VI. CONCLUSION & DISCUSSION

How can our model benefit the society?

- Assist organizations in recognizing the possible epidemic outbreak in different scenarios, where different levels of control policies and the characteristics of covid-variants are applied.

- The dynamic of the upcoming outbreak can be visualized.

- It is worth understanding that forecasting can be considered ineffective when the cost for forecasting is too high, its uncertainty is too critical to accept.

Effects of COVID-19 forecast on Economic and Social aspects

ECONOMIC ASPECTS: The main purpose of this study for the economy is to adapt with changes and maintain economic development under these changes.

- Not interrupt the logistics and supply chains: ensure enough supply for the population during pandemic
- Appropriately and timely distribution of the money to strengthen the healthcare system and resources
- Minimize the negative impacts the pandemic has on GDP growth of HCMC
- Minimize inflation rates

Effects of COVID-19 forecast on Economic and Social aspects

SOCIAL ASPECTS:

- Education: Helps school and educational organisations to prepare better plans & Ensure students still receive proper educational programs while keeping them safe
- Healthcare: strengthen healthcare system with enough supply and medical resources through forecasting
- Employment: Companies and organizations can provide better and more suitable production and business strategies to minimize layoffs and suspension of production and business operations

THANK YOU

This is the end of our presentation.
Thank you for listening