HONG KONG CRIME RATE

JUNOES Hans Nathanael

LUI Lam Wai

KO Ming

CHAN Chiu Sing

FU Tsz Ho

CHEN Man Hin

G13

2022 CRIME INDEX

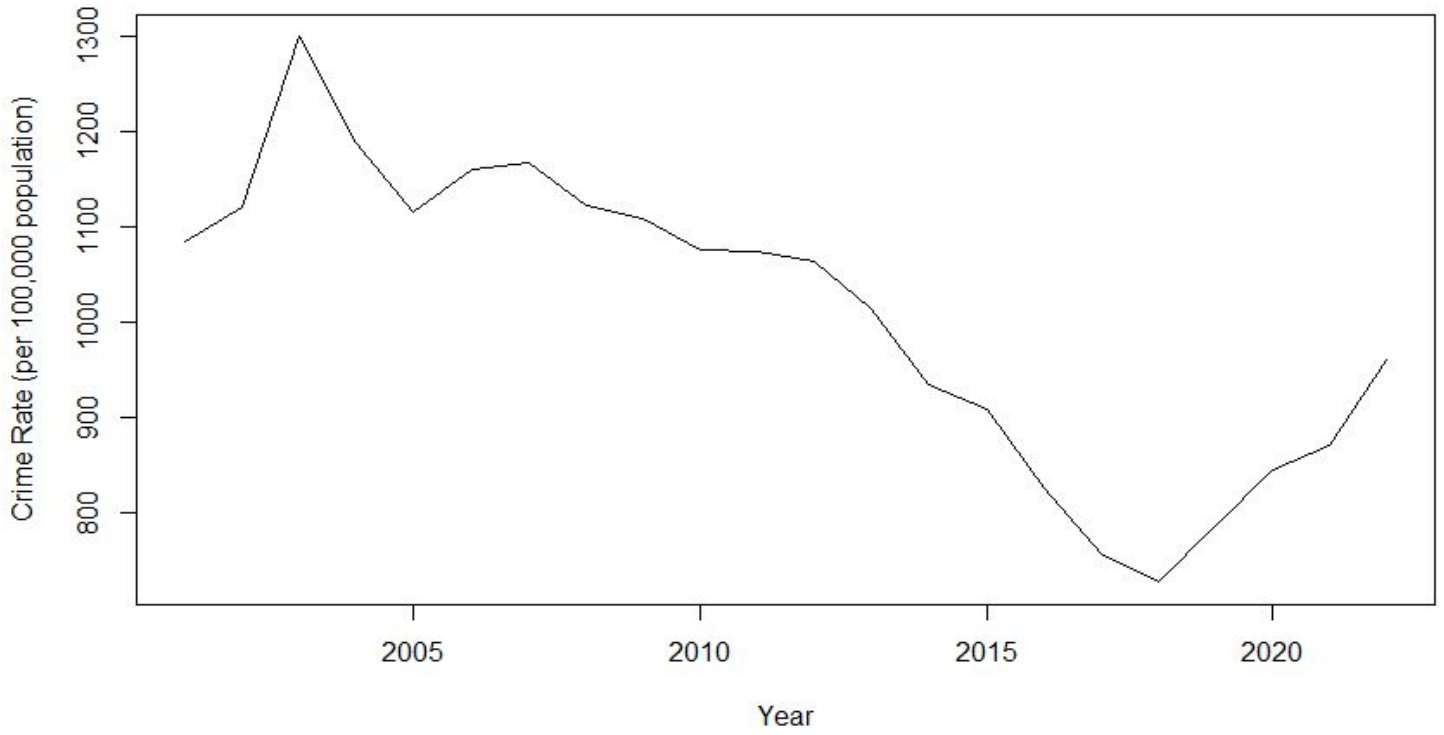
Rank	Country	Crime Index
1	Venezuela	83.6
12	Brazil	62.0
57	United States	48.2
115	China	29.4
119	Singapore	27.6
136	Hong Kong	21.9

Hong Kong □ 7th safest (measurable) country

TABLE OF CONTENTS

1.1 DATA COLLECTION

Crime rate data of HK from 2001-2022
https://www.police.gov.hk/ppp_en/09_statistics/csc.html



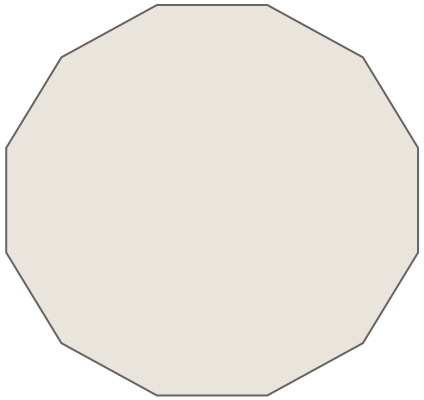
1.2 DATA PREPROCESSING

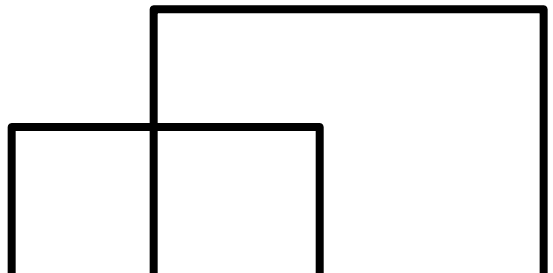
The z-score of each data point is calculated as follows:

$$z = (xi - \bar{x})/s$$

where $\bar{x}$ is the mean of the crime rate data, s is the standard deviation of the crime rate data.

There is no data point that has a z-score outside the range of ±3. Therefore, no outliers are removed.





2. TREND PROJECTION

As the time series exhibits a linear trend,a trend line projection is determined for future forecasts using the least squares method. Least squares determines the unique trend line forecast which minimizes the error between the trend line forecasts and the actual observed values for the time series.

The formula for the trend line projection is:

$$T_t = b_o + b_1 t$$

where T_t = trend forecast for time period t

b_1 = slope of the trend line

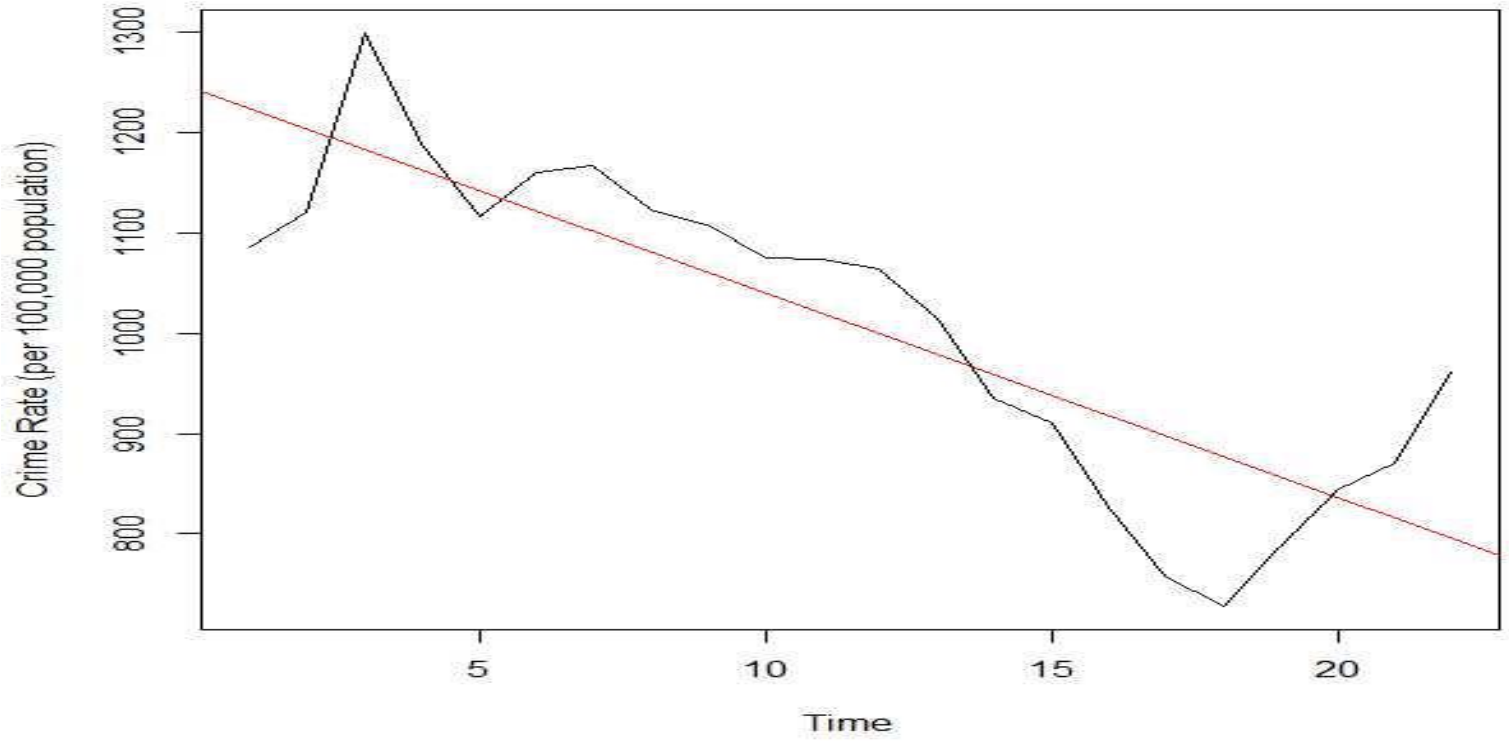
b_0 = trend line projection for time 0 (y-intercept)

$$b_1 = (n\Sigma tY_t - \Sigma t \Sigma Y_t)/(n\Sigma t^2 - (\Sigma t)^2) \qquad b_o = \bar{Y} - b_1 \bar{t}$$

where Y_t = observed value at time t

$\bar{Y}$ = average of the values for Y_t

$\bar{t}$ = average time period for the n observations



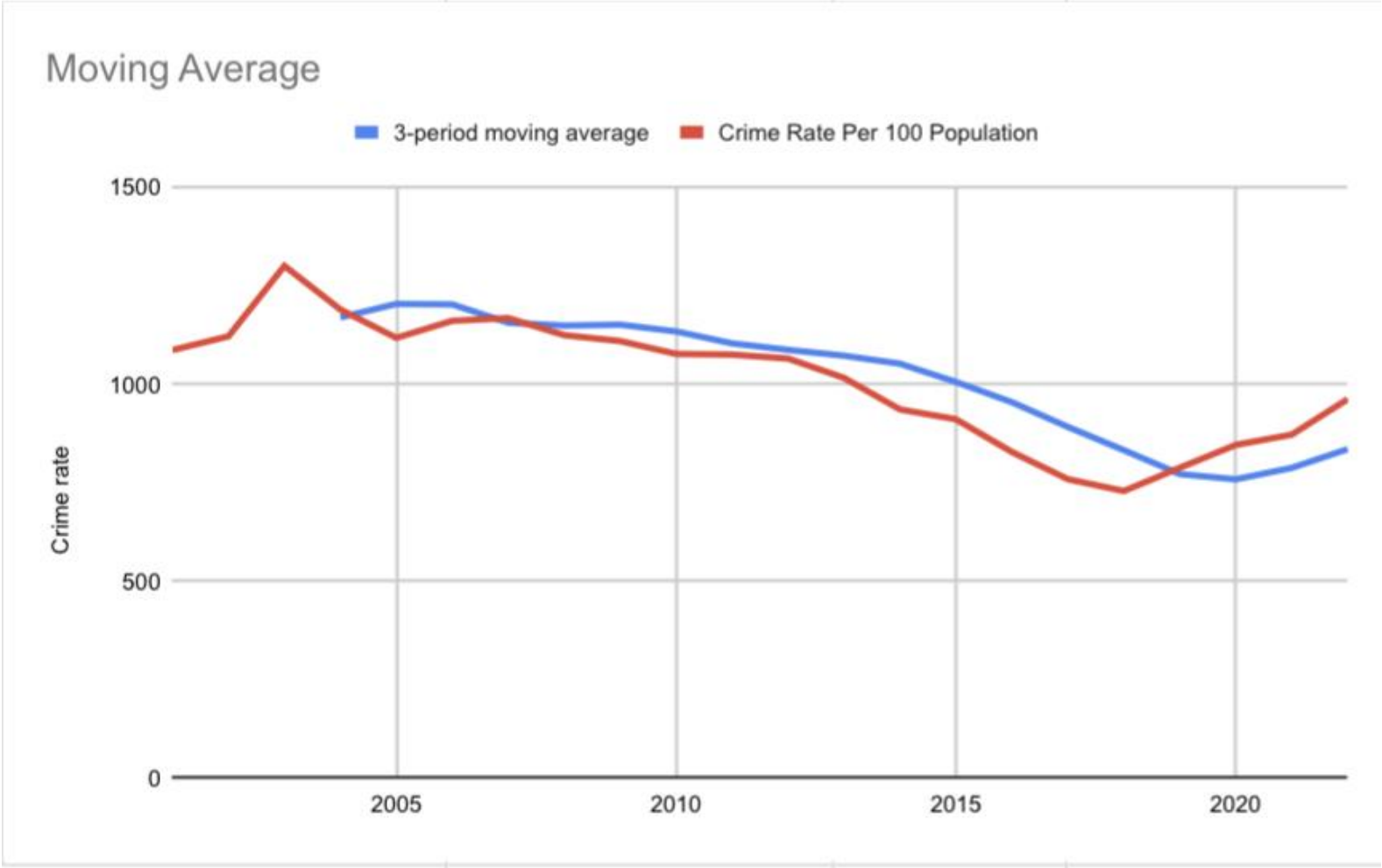
The slope and intercept are -20.42654 and 1244.86 respectively. Mean absolute error (MSE) and Mean squared error (MAE) are 6676.635 and 68.24527 respectively. The forecast value for 2023 is 775.0494. $T_t=b_0+b_1t$

3. SMOOTHING

- 3.1 Moving Average
- 3.2 Center Moving Average
- 3.3 Exponential Smoothing

3.1 MOVING AVERAGE

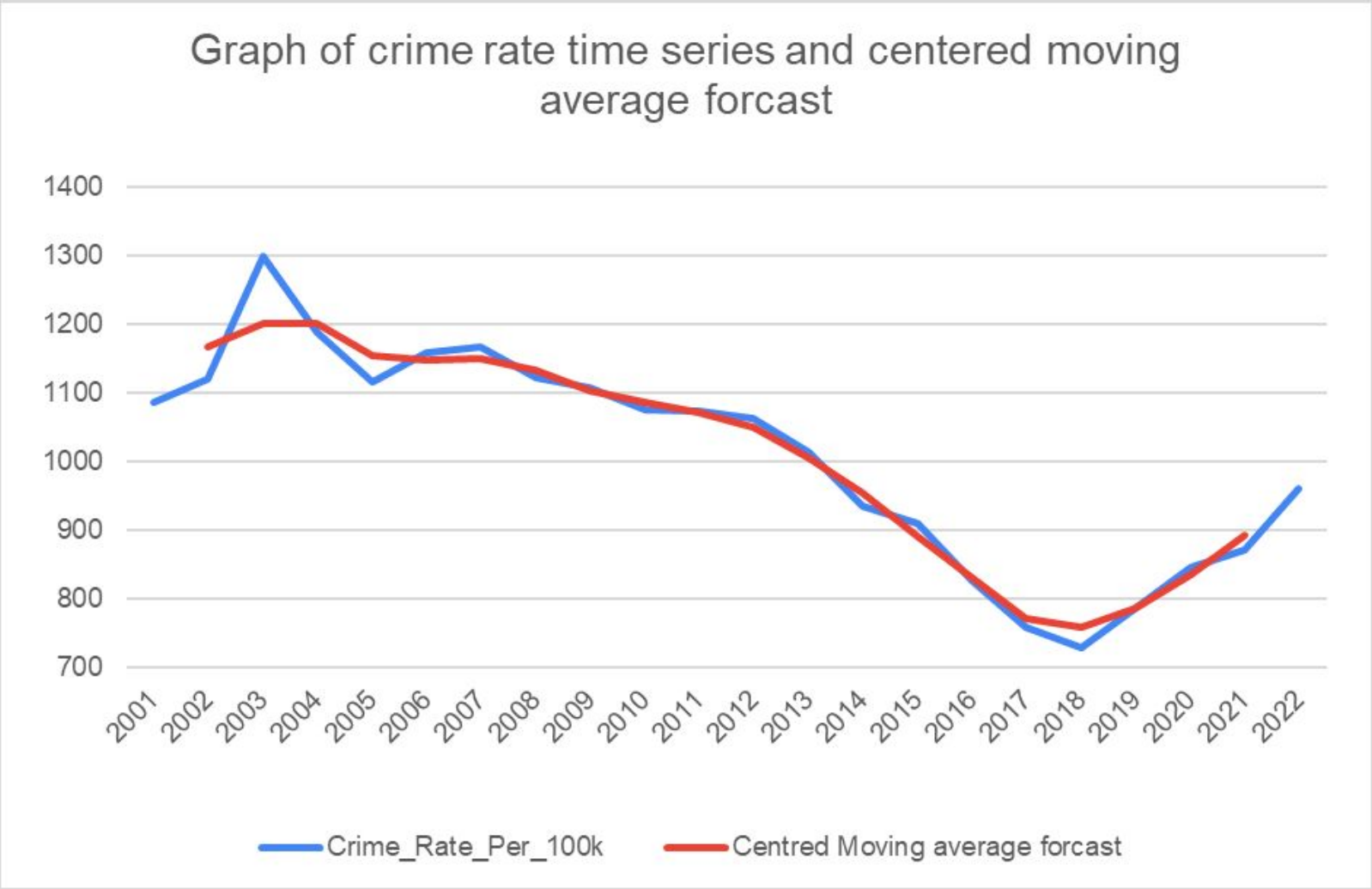
A simple moving average (SMA), is calculated by taking the arithmetic mean of a given set of values over a specified period. A set of numbers, or prices of stocks, are added together and then divided by the number of prices in the set. In this case we choose 3-Period to forecast the crime rate.



For example, the sum of the crime rate from 2001 to 2003 is 3505, so we predict that in 2004 the crime rate will be 3505/3=1168.333. The squared forecast error is 406.6957886. The rest is shown above. The total MSE and MAE are 14346.44 and 60.219 respectively , and the forecast of crime rate per 100 population for 2023 would be 892.333 .

3.2 CENTER MOVING AVERAGE

The centered moving average method consists of computing an average of n periods' data and associating it with the midpoint of the periods. For this case, we choose 3 time period to make the predictions of the data the most applicable and have the smallest error.



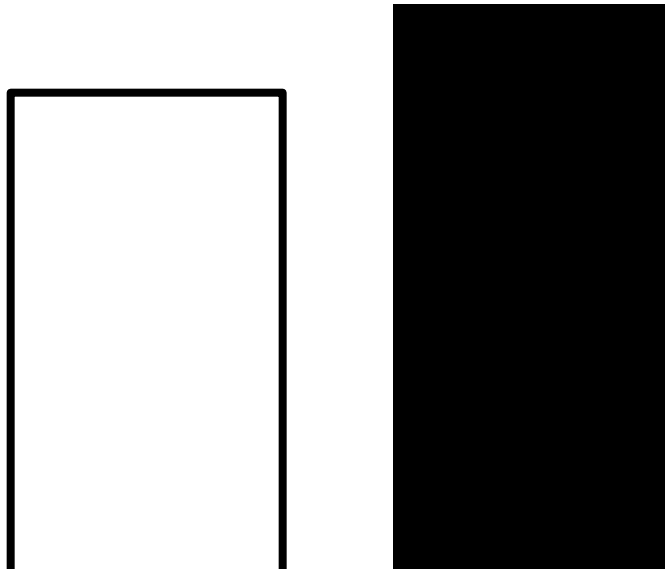
For example, in 2011,2012,2013, the crime rate per 100k is 1074,1064,1015, so the forecast crime rate of 2012 is (11074+1064+1015)/3=1051, forecast error is 1064-1051=13. Total MSE (crime rate, from 2001 to 2022) is 827, MAE is 319.7.

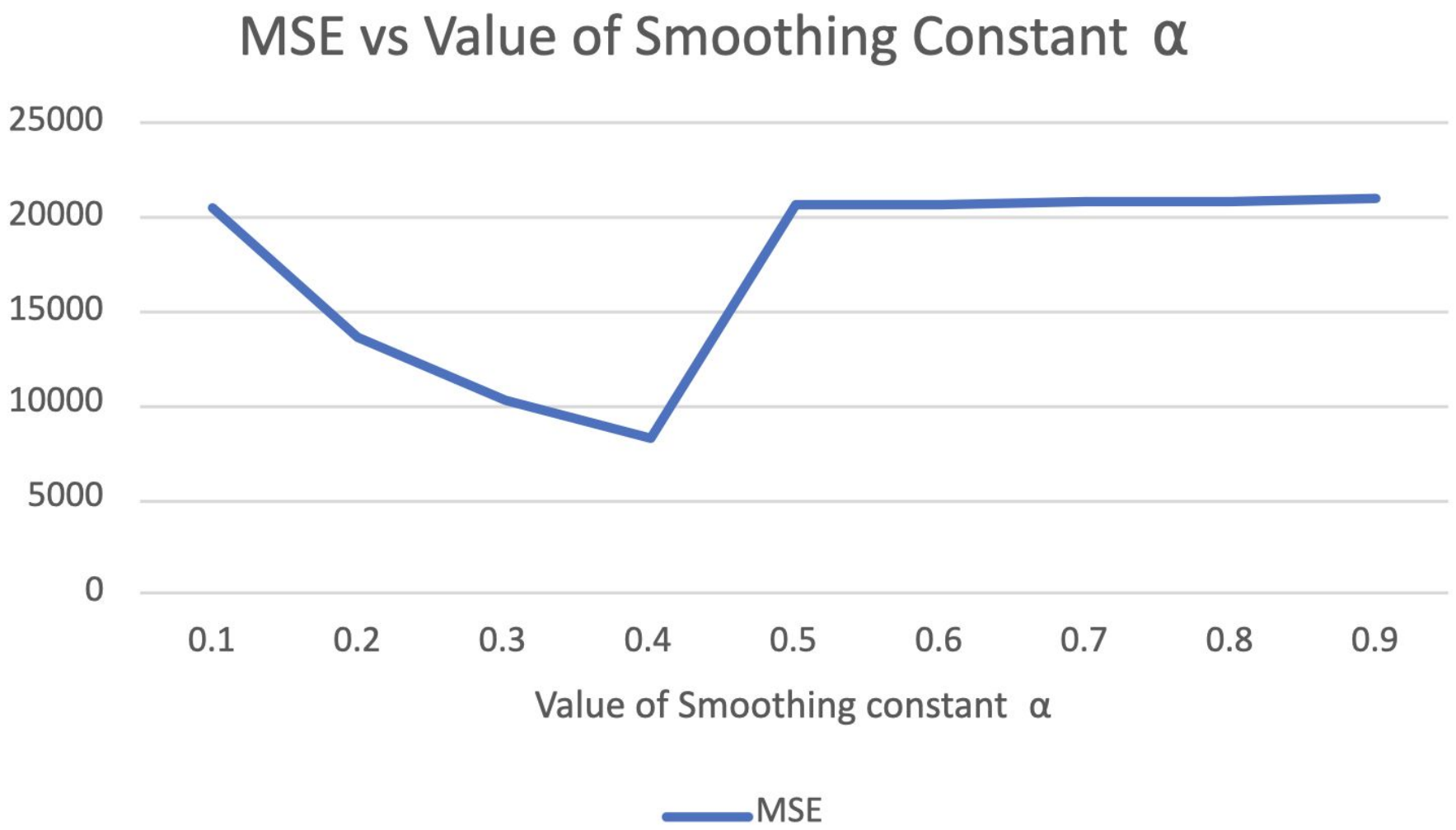
3.3 Exponential Smoothing

Exponential smoothing is a technique used in time series analysis to forecast future values based on past data. In this case, we calculated the mean squared error (MSE) by using different smoothing constants α, ranging from 0.1 to 0.9. A smaller MSE indicates better forecasts.

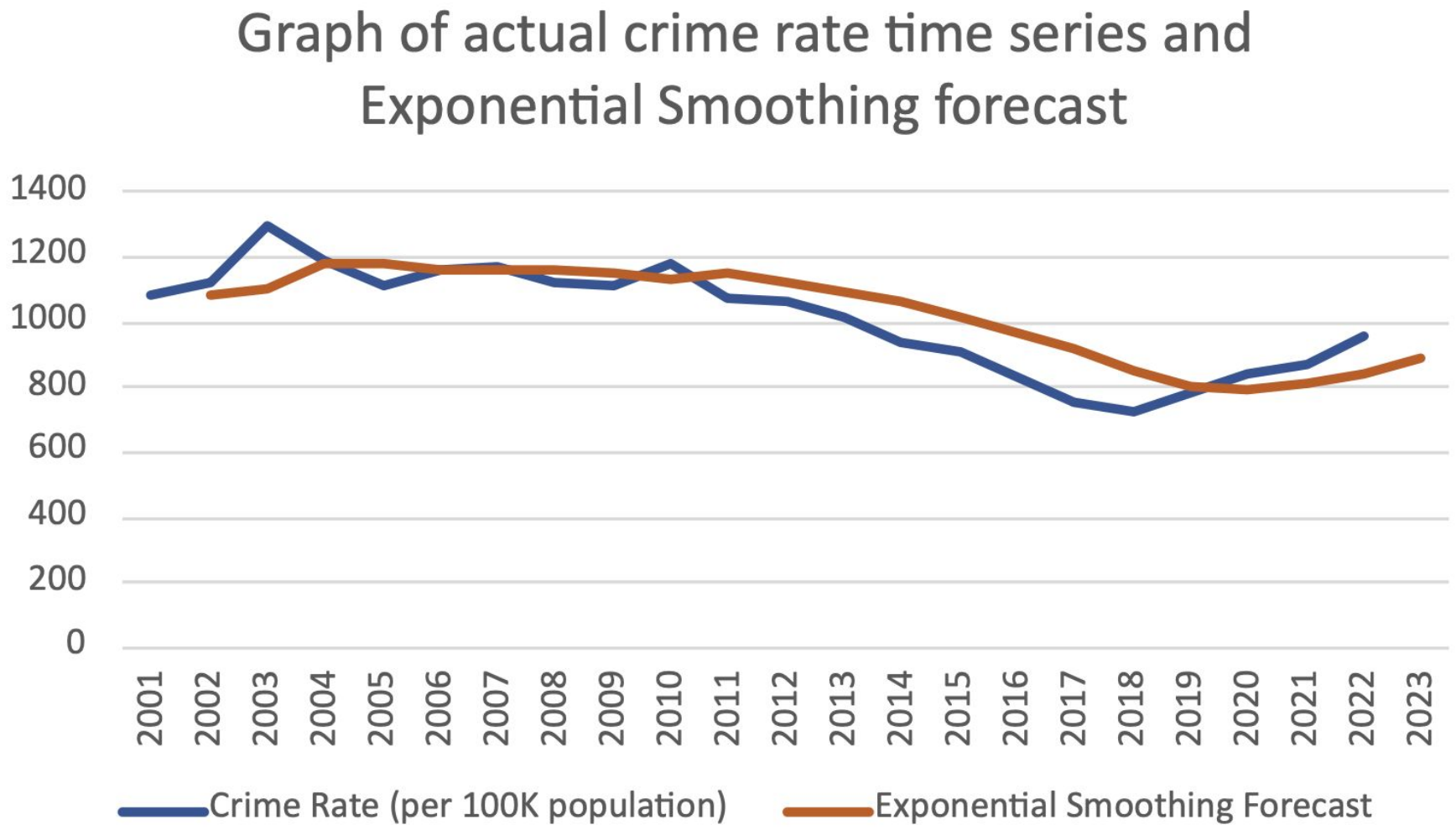
The equation is: Forecast for next period (t+1) =

α x actual value for current period(t) + (1-α) x forecast for current period t





The graph illustrates the relationship between the value of the smoothing constant α and the mean squared error (MSE). The line plot indicates that the MSE decreases initially as the value of α increases, reaching a minimum at α=0.4 with a value of 8296.8. Therefore, the optimum value of α that results in the smallest MSE is 0.4. This indicates that α=0.4 better forecasts compared to other values of α.



In the graph, the Exponential Smoothing forecast is computed using a smoothing constant of α=0.4.

The total MSE and MAE are 8296.8 and 73.9 respectively , and the forecast of crime rate per 100K population for 2023 would be 887.1 .

4. ARIMA

$$\underbrace{\phi(B)}_{AR}(1-B)^d Y_t = \underbrace{\Theta(B)}_{MA} Z_t$$

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - .... - \phi_p B^p$ *(AR component)*

$\Theta(B) = 1 - \Theta_1 B - \Theta_2 B^2 - .... - \Theta_q B^p$ *(MA component)*

$Z_t \sim WN(0, \sigma^2)$ *(White Noise)*

Y_t is observed value at time t

Modeling Process

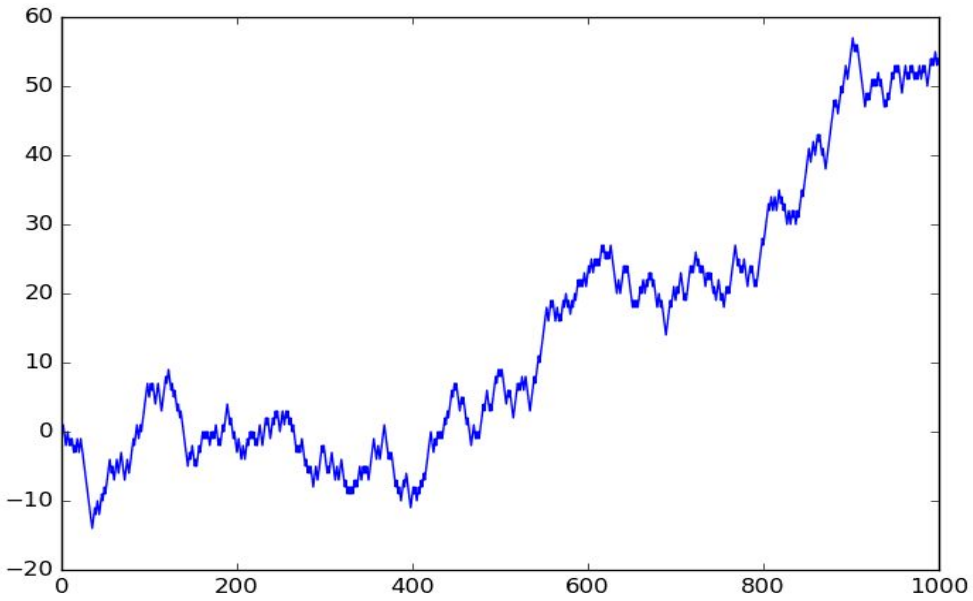
```
> fit_arima = auto.arima(crime, stepwise = F, approximation = F)
> print(summary(fit_arima))
Series: crime
ARIMA(0,1,0)
```

ARIMA(0,1,0) → Random Walk

$$Y_t = Y_{t-1} + Z_t$$

(random)

→ Not good for forecasting!



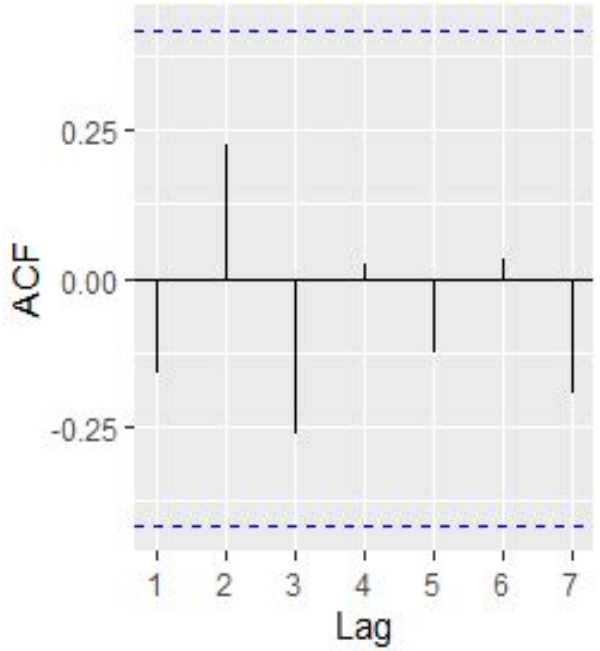
Modeling Selection

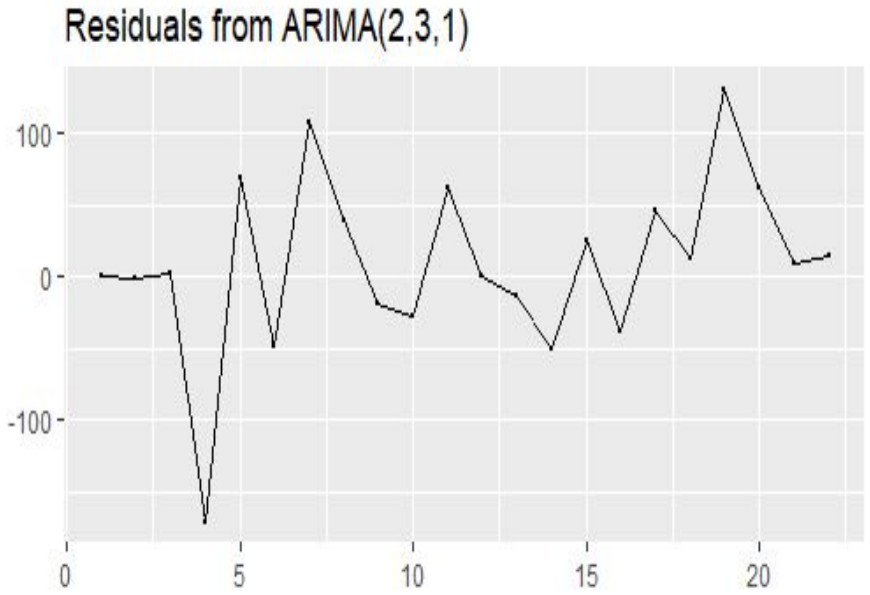
Brute Force

- Loop all combination of p, d, q from 0 to 3 in ARIMA(p, d, q)
- Check for lowest AIC and BIC value and update

```
> bestAIC = c(1e8, NA, NA, NA)
> bestBIC = c(1e8, NA, NA, NA)
> for(d in 0:3) for(p in 0:3) for(q in 0:3){
+       m = arima(crime, order = c(p, d, q), optim.control = list(maxit=1000))
+       if(AIC(m) < bestAIC[1]){
+         bestAIC = c(AIC(m), p, d, q) }
+       if(BIC(m) < bestBIC[1]){
+         bestBIC = c(BIC(m), p, d, q) } }
```

Higher order ARIMA models lose their simplicity/interpretability

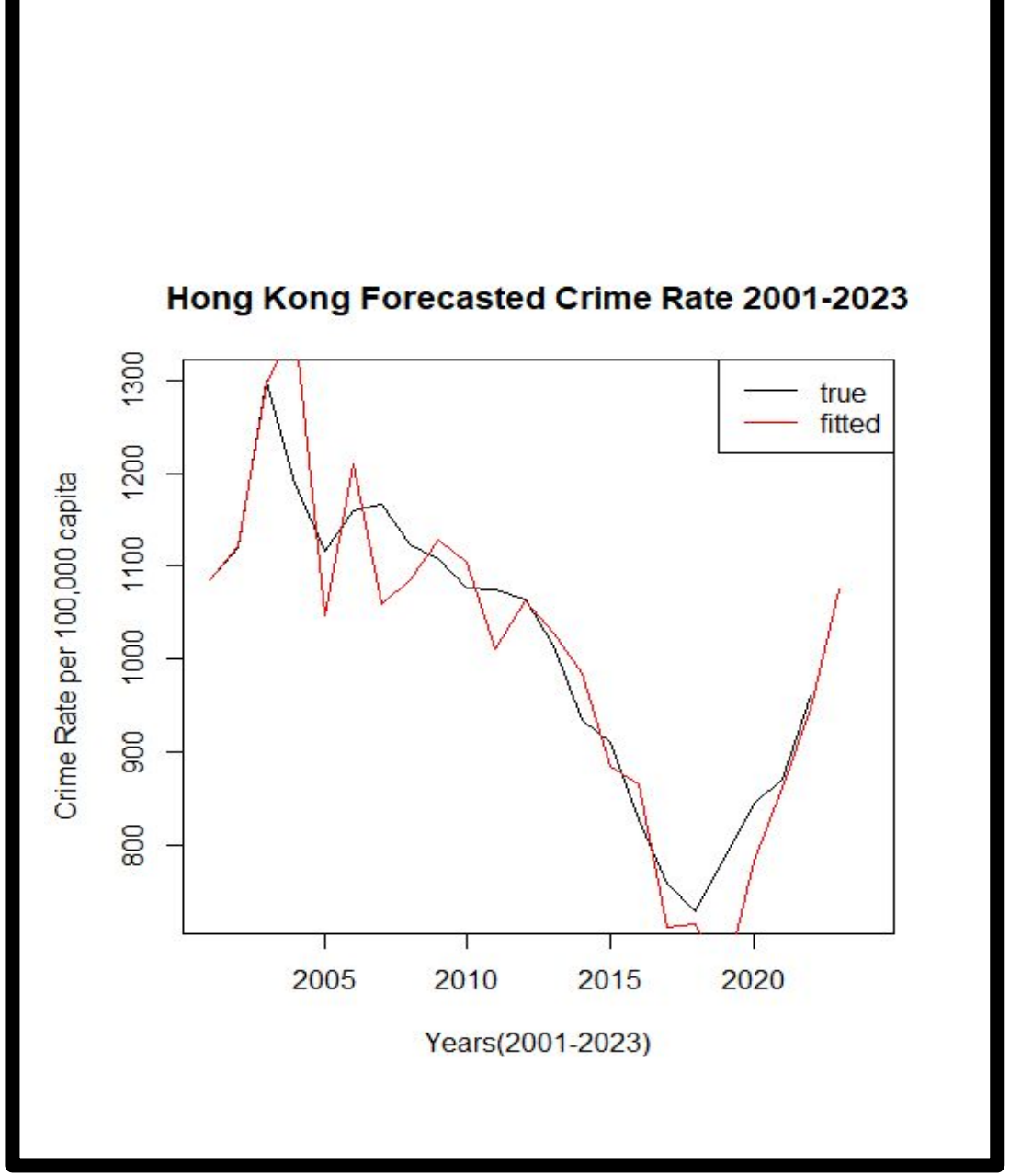




Result : ARIMA(2, 3, 1)

from AIC and BIC

- Within $2/\sqrt{n}$ (Stationary)
- MSE : 3798.107 (random walk model : 4243.929)
- MAE : 43.47104 (random walk model : 50.85844)



Hong Kong Crime Rate Prediction (2023) :

1074.6

(113.6005 increase)

5. CONCLUSION

Method	MSE	MAE	Forecast
Trend Projection	6676.635	68.24527	775.0494
Moving Average	14346.44	60.2192768	892.333
Center Moving Average	827	319.7	1051
Exponential Smoothing	8296.8	73.9	887.1
ARIMA (2, 3, 1)	3798.107	43.47104	1074.6