Time Series Analysis of Central African Republic Economic Data

Angelina Cottone
STA 137
7 December 2025

Table of Contents

Table of Figures

1 Introduction

The Central African Republic (CAR) is a sparsely populated, landlocked nation in Central Africa. Despite abundant natural resources including diamonds, gold, and timber, CAR remains one of the world's least developed nations due to political instability and civil conflicts. (*Central african republic*)

This report analyzes three annual time series from 1960-2017 (n = 58 yearly observations):

- **GDP**: Gross Domestic Product in USD
- **Exports**: Value of goods sold abroad as percentage of GDP
- **Population**: Total population count

Time-series analysis is conducted using the ARIMA (Autoregressive Integrated Moving Average) framework. ARIMA models capture patterns such as trends and seasonality to forecast future values based on historical behavior. (*Model Selection for Arima* 2025) A general ARIMA(p, d, q) model combines autoregressive terms, differencing, and moving average terms:

$$\phi(B)(1 - B)^d Y_t = \theta(B)\varepsilon_t,$$

where B is the backshift operator, $\phi(B)$ is the AR polynomial, $\theta(B)$ is the MA polynomial, and $\varepsilon_t \sim WN(0, \sigma^2)$ is white noise. (Hyndman & Athanasopoulos)

The standard seven-step procedure is followed: plot data, transform if needed, test stationarity, examine ACF/PACF, fit model, diagnose residuals, and, finally, generate forecasts.

2 Data Description

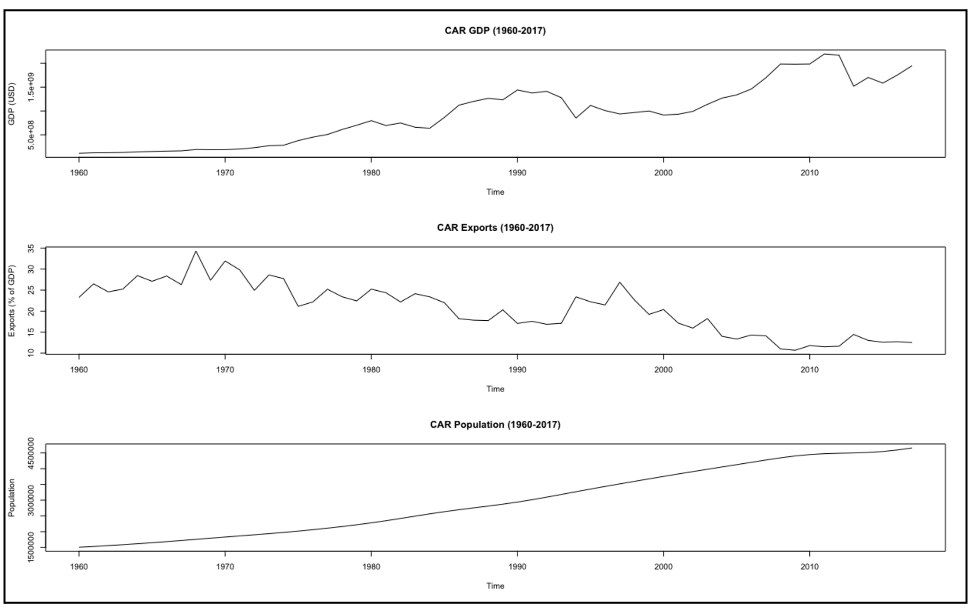


Figure 1: Time Series Plots of GDP (top), Exports (middle), and Population (bottom)

Observations:

- **GDP** shows an overall upward trend with dips around 1980-1985, 1993-2003, and 2013-2015, which may correlate to periods of increased political instability. Increasing variance suggests a log transformation is appropriate.
- **Exports** present considerable fluctuations with an overall decreasing trend, reaching a peak around 1968 and a low around 2009.
- **Population** has a steady, nearly linear growth over time with minimal fluctuations.

All three series display non-stationary behavior as they contain strong trend components.

3 Transformation

To stabilize variance and improve model performance, each series was transformed using a natural logarithm:

$$Y_t^{log} = log(Y_t)$$

Log transformations reduces heteroscedasticity and make linear trends easier to model via ARIMA differencing.

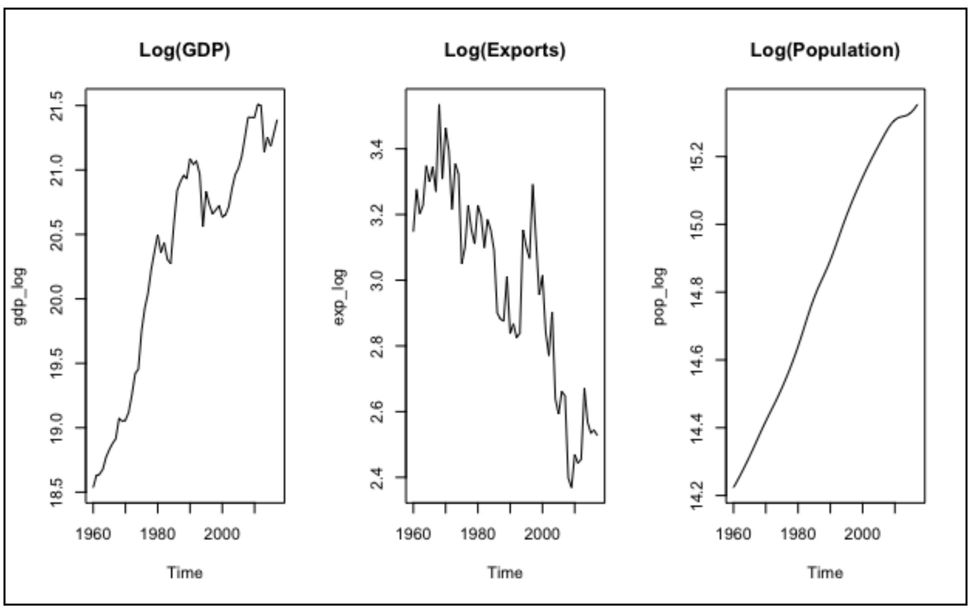


Figure 2: Log Transformation of GDP (left), Exports (middle), and Population (right)

4 Stationarity Analysis

4.1 ACF/PACF Plots

The sample autocorrelation function (ACF) at lag k measures correlation between Y_t and Y_{t-k}:

$$\hat{\rho}(k) = \frac{\sum_{t=k+1}^{n} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}$$

ACF plots measure the linear relationship between a times series and its lagged values, or how much current values rely on historical data. (*Autocorrelation and partial autocorrelation* 2025)

The partial autocorrelation function (PACF) measures the correlation between Y_t and Y_{t-k} after removing the effects of intermediate lags. PACF removes in influence of intermediate lags, showing direct correlation between variables and its lagged values. (*Autocorrelation and partial autocorrelation* 2025)

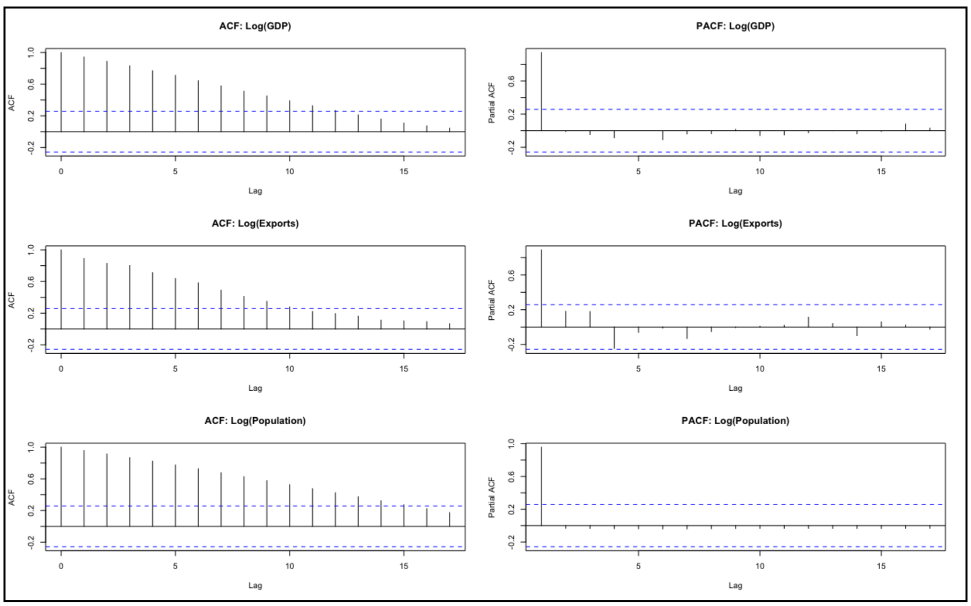


Figure 3: ACF and PACF Plots of GDP (top), Exports (middle), and Population (bottom)

The log-transformed series all show slow ACF decay, which is characteristics of non-stationary processes. A stationary process would show rapid ACF decay. The PACF shows that most spikes (besides lag 0) are within the 95% confidence bounds.

4.2 Augmented Dickey-Fuller Test

The ADF test evaluates whether each series has a unit root. (*Augmented dickey-fuller (ADF)* 2025) The test regression model is:

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta Y_{t-1} + \varepsilon_t$$

- $H_0: \gamma = 0$, series is non-stationary
- $H_1: \gamma < 0$, series is stationary

Results

Series	p-value	Conclusion
GDP	$0.609 > 0.05$	Fail to reject H_0 → non-stationary
Exports	$0.182 > 0.05$	Fail to reject H_0 → non-stationary
Population	$0.526 > 0.05$	Fail to reject H_0 → non-stationary

Table 1: Results of ADF Hypothesis Test

All three series are non-stationary and therefore require differencing.

4.3 Differencing

GDP and Exports require first differencing ($d = 1$):

$$\nabla Y_t = Y_t - Y_{t-1}$$

Population requires second differencing ($d = 2$), due to its strong linear trend:

$$\nabla^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

5 Model Fitting

Model selection is conducted using '$auto.arima()$', which identifies the best ARIMA(p, d, q) model by minimizing the corrected Akaike Information Criterion (Hyndman & Athanasopoulos):

$$AICc = AIC + \frac{2(p+q+k+1)(p+q+k+2)}{T-p-q-k-2}$$

Lower AICc indicates a better model.

5.1 GDP Model

```
Series: gdp_log
ARIMA(0,1,0) with drift

Coefficients:
       drift
      0.0501
s.e.  0.0170

sigma^2 = 0.01683:  log likelihood = 36.05
AIC=-68.11   AICc=-67.88   BIC=-64.02
```

```
z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
drift 0.050095   0.017027  2.9421  0.00326 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 4: GDP ARIMA Model Results and Coefficient Test

GDP follows an ARIMA(0, 1, 0) model with drift, a random walk with drift:

$$Y_t = Y_{t-1} + c + \varepsilon_t$$

This is the simplest model for a trending series. The drift term $c = 0.0501$ represents approximately 5% annual growth on the log scale. The drift term is statistically significant ($z = 2.94$, $p = 0.003 < 0.05$), confirming a stable positive long-term trend. The AICc is -67.88.

5.2 Exports Model

```
Series: exp_log
ARIMA(2,1,2)

Coefficients:
          ar1      ar2     ma1     ma2
      -0.6924  -0.8006  0.3985  0.5571
s.e.   0.1455   0.1879  0.2067  0.2985

sigma^2 = 0.01354:  log likelihood = 43.49
AIC=-76.98   AICc=-75.81   BIC=-66.77
```

```
z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)
ar1 -0.69237    0.14553 -4.7577 1.958e-06 ***
ar2 -0.80060    0.18788 -4.2612 2.034e-05 ***
ma1  0.39845    0.20668  1.9278   0.05387 .
ma2  0.55714    0.29853  1.8663   0.06200 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 5: Exports ARIMA Model Results and Coefficient Test

Exports follows an ARIMA(2, 1, 2) model:

$$\nabla Y_t = \phi_1 \nabla Y_{t-1} + \phi_2 \nabla Y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

with $\phi_1 = -0.69$, $\phi_2 = -0.80$, $\theta_1 = 0.40$, $\theta_2 = 0.56$.

This is a more complex model with two AR terms, one difference, and two MA terms. The AR coefficients are highly significant ($p < 0.001$), indicating strong short-term autocorrelation.

The MA coefficients are marginally significant ($p \approx 0.05 - 0.06$). This complexity reflects the high volatility in CAR's export data. The AICc is -75.81.

5.3 Population Model

```
Series: pop_log
ARIMA(3,2,2)

Coefficients:
        ar1     ar2      ar3     ma1     ma2
      0.835  0.2800  -0.5642  1.6471  0.9780
s.e.  0.135  0.2208   0.1423  0.0856  0.0856

sigma^2 = 7.961e-06:  log likelihood = 383.68
AIC=-755.37   AICc=-753.65   BIC=-743.21
```

```
z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)
ar1  0.835012   0.135042   6.1833 6.276e-10 ***
ar2  0.280019   0.220848   1.2679    0.2048
ar3 -0.564208   0.142331  -3.9641 7.369e-05 ***
ma1  1.647137   0.085604  19.2414 < 2.2e-16 ***
ma2  0.977982   0.085593  11.4260 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 6: Population ARIMA Model Results and Coefficient Test

Population follows an ARIMA(3, 2, 2):

$$\nabla^2 Y_t = \phi_1 \nabla^2 Y_{t-1} + \phi_3 \nabla^2 Y_{t-3} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

This is the most complex model of the three series, with three AR terms, two differences, and two MA terms. The significant coefficients (ϕ_1, ϕ_2, θ_1, θ_2; all $p < 0.001$) capture the smooth, predictable growth pattern. The very small $\sigma^2 = 7.96 \times 10^{-6}$ indicates extremely low residual variance, indicating population is highly predictable. The AICc is -753.65.

6 Model Diagnostics

We examine residual ACF plots and QQ plots to verify model adequacy. For a good model, residuals should behavior like white noise, with no significant autocorrelation (ACF spikes should stay within 95% bounds) and approximately normal distribution (points should follow the QQ line).

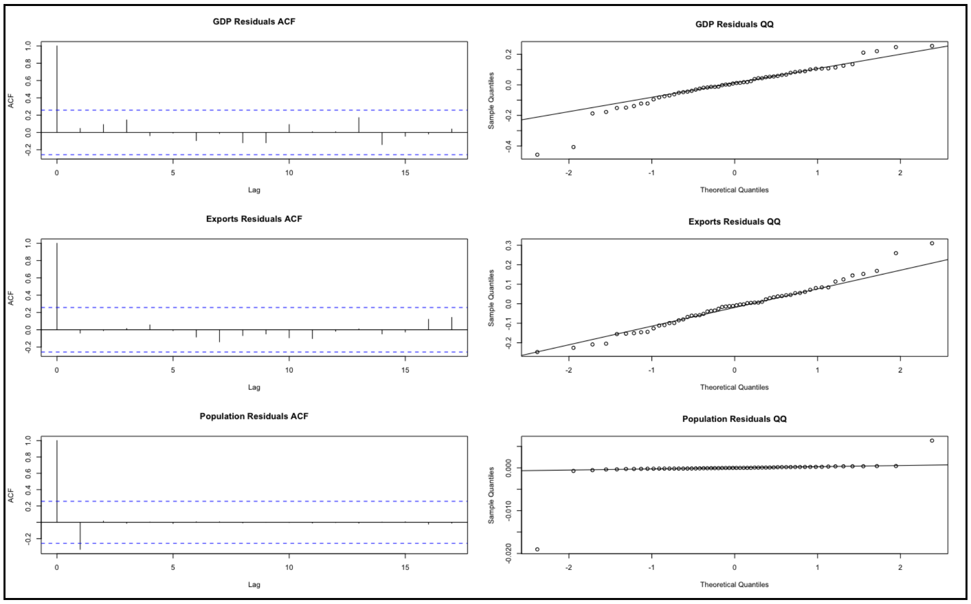


Figure 7: Residuals ACF and QQ-Plots of GDP (top), Exports (middle), and Population (bottom)

Results

- **GDP**: All ACF spikes within bounds, QQ plot shows slight tails but generally follows the line.
- **Exports**: All ACF spikes within bounds, QQ plot shows slight tails but acceptable fit.
- **Population**: One spike marginally outside bounds (acceptable at 5% level), QQ plot follows line closely.

Overall, residual diagnostics confirm adequate model fit with no significant autocorrelation remaining and approximate normality.

7 Joint Analysis: Cross-Correlation

To examine the three series jointly, we use the cross-correlation function (CCF). The CCF at lag k measures:

$$\rho_{XY}(k) = Corr(X_t, Y_{t+k})$$

- Significant spike at positive lag k: X leads Y by k periods
- Significant spike at negative lag k: Y leads X by k periods

The differenced (stationary) series was used for valid CCF inference. (Shumway & Stoffer)

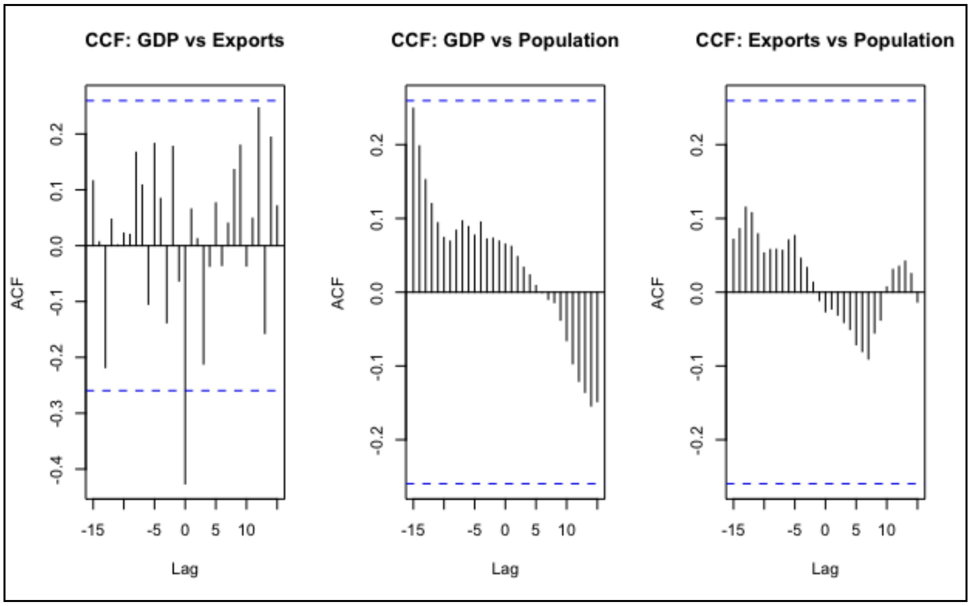


Figure 8: Cross-Correlation Analysis of GDP, Exports, and Population

Results

- **GDP vs Exports**: Significantly correlated at lag 0, indicating simultaneous movement. When exports increase, GDP increases in the same year, which is intuitive since exports contribute directly to GDP. There is no lead-lag relationship.
- **GDP vs Population**: All spikes fall within interval, so there is no significant lead-lag movement between the series.
- **Exports vs. Population**: All spikes fall within interval, so there is no significant lead-lag movement between the series.

Population is independent of both GDP and Exports, indicating that demographic trends evolve independently of economic fluctuations.

8 Forecasting

We generate 10-year forecasts (2018-2027) for each series.

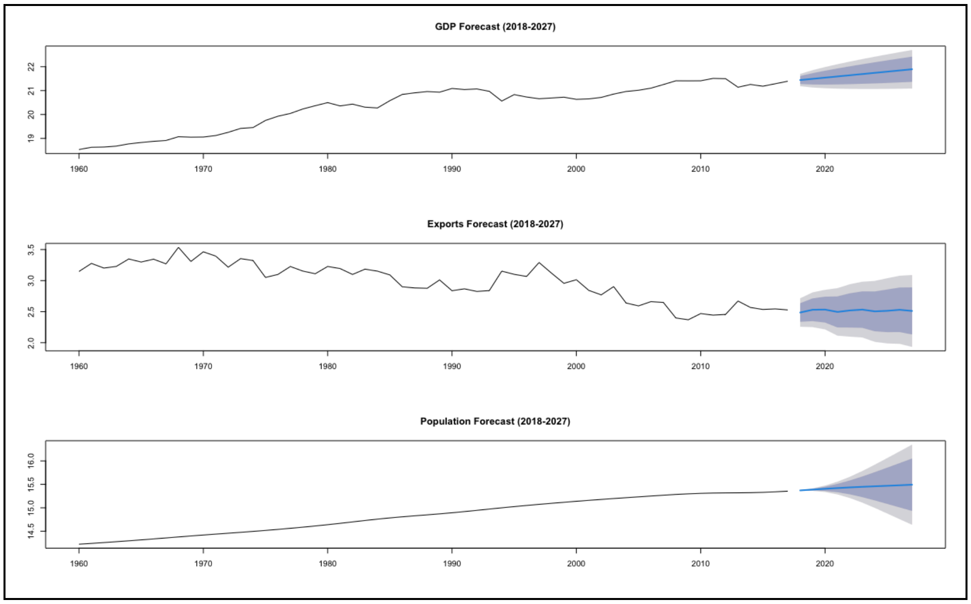


Figure 8: 10-Year Forecasts for GDP (top), Exports (middle), and Population (bottom)

The shaded regions show 80% (dark) and 95% (light) prediction intervals.

- **GDP**: Continued upward trend with moderate uncertainty, the drift term drives the forecast upward.
- **Exports**: Relatively flat forecast with wide prediction intervals reflecting high volatility and uncertainty.
- **Population**: Continued steady growth with intervals that go from narrow to wide.

9 Comparison and Conclusion

9.1 Model Summary

Series	Model	Differencing	AICc	Key Findings
GDP	ARIMA(0, 1, 0)	$d = 1$	-67.88	Random walk, ~5% annual growth
Exports	ARIMA(2, 1, 2)	$d = 1$	-75.81	Complex, high volatility
Population	ARIMA(3, 2, 2)	$d = 2$	-753.65	Strong trend, highly predictable

Table 2: ARIMA Model Summary for GDP, Exports, and Population

9.2 Key Findings

1. **GDP** is best described by a random walk with drift, the simplest non-stationary model. The significant drift term ($p = 0.003$) confirms CAR's long-term economic growth despite periods of instability.

2. **Exports** requires the most complex ARIMA structure (4 parameters), reflecting its sensitivity to global commodity prices and CAR's political situation. The wide forecast intervals indicate high uncertainty in future predictions.
3. **Population** is the most predictable series with the lowest residual variance ($\sigma^2 = 7.96 \times 10^{-6}$). Demographic trends are inherently more stable than economic variables, though it requires the highest differencing order ($d = 2$) due to strong persistent growth.

9.3 Overall Conclusions

All three series are non-stationary and require differencing to achieve stationarity. The model complexity increases from GDP (simplest) to Exports to Population (most complex), the population has the smallest prediction uncertainty.

Economic variables (GDP, Exports) are hard to forecast than demographic trends due to their sensitivity to political and global market conditions. These models provide reasonable short-term forecasts, but long-term predictions should be interpreted cautiously due to CAR’s history of economic instability.

10 References

Central african republic: World Bank Group. World Bank Group. (n.d.). https://www.worldbank.org/ext/en/country/centralafricanrepublic

GeeksforGeeks. (2025a, July 23). *Augmented dickey-fuller (ADF)*. GeeksforGeeks. https://www.geeksforgeeks.org/machine-learning/augmented-dickey-fuller-adf/

GeeksforGeeks. (2025b, July 23). *Autocorrelation and partial autocorrelation*. GeeksforGeeks. https://www.geeksforgeeks.org/r-machine-learning/autocorrelation-and-partial-autocorrelation/

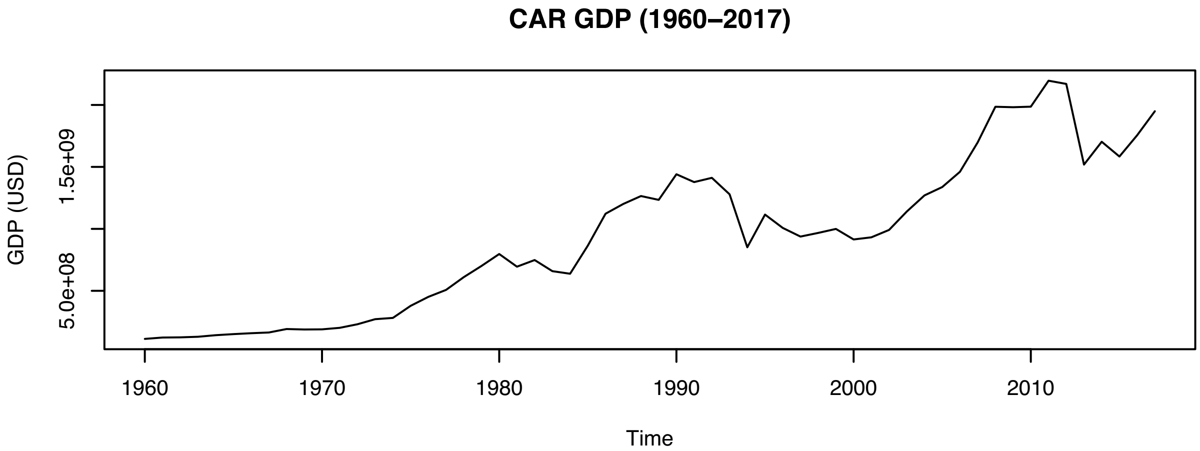
GeeksforGeeks. (2025c, August 19). *Model Selection for Arima*. GeeksforGeeks. https://www.geeksforgeeks.org/data-science/model-selection-for-arima/

Hyndman, R., & Athanasopoulos, G. (n.d.-a). *Forecasting: Principles and practice (2nd ed)*. ARIMA modelling in R. https://otexts.com/fpp2/arima-r.html

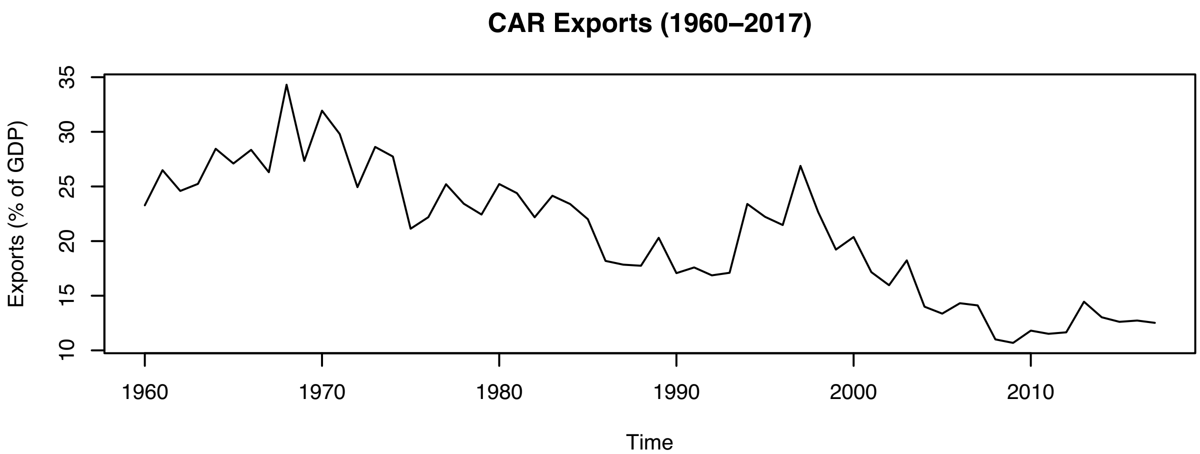
Hyndman, R., & Athanasopoulos, G. (n.d.-b). *Forecasting: Principles and practice (2nd ed)*. Estimation and order selection. https://otexts.com/fpp2/arima-estimation.html

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: With R examples* (4th ed.). Springer.

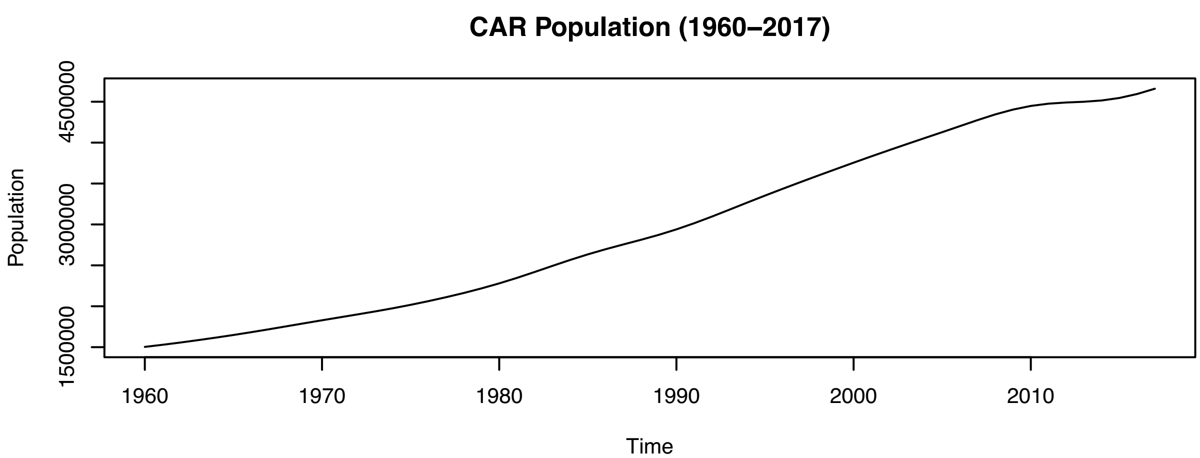
Contents



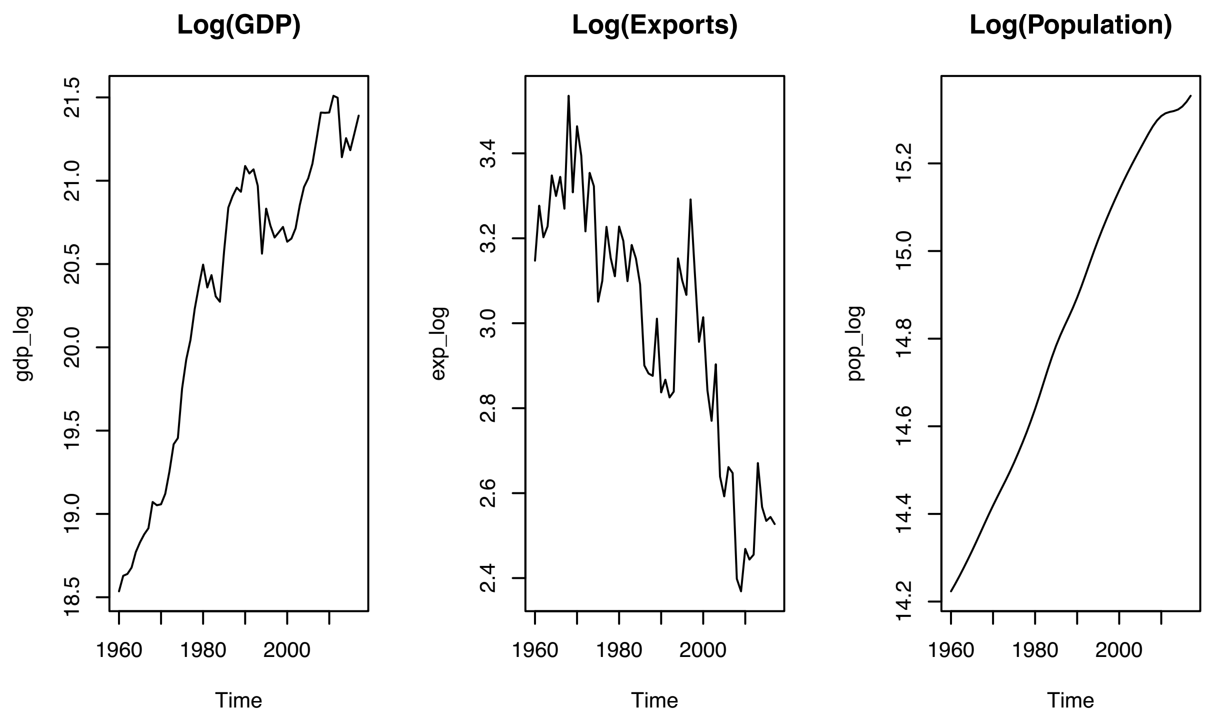
CAR GDP (1960–2017)
GDP (USD)
5.0e+08
1.5e+09
1960
1970
1980
1990
2000
2010
Time

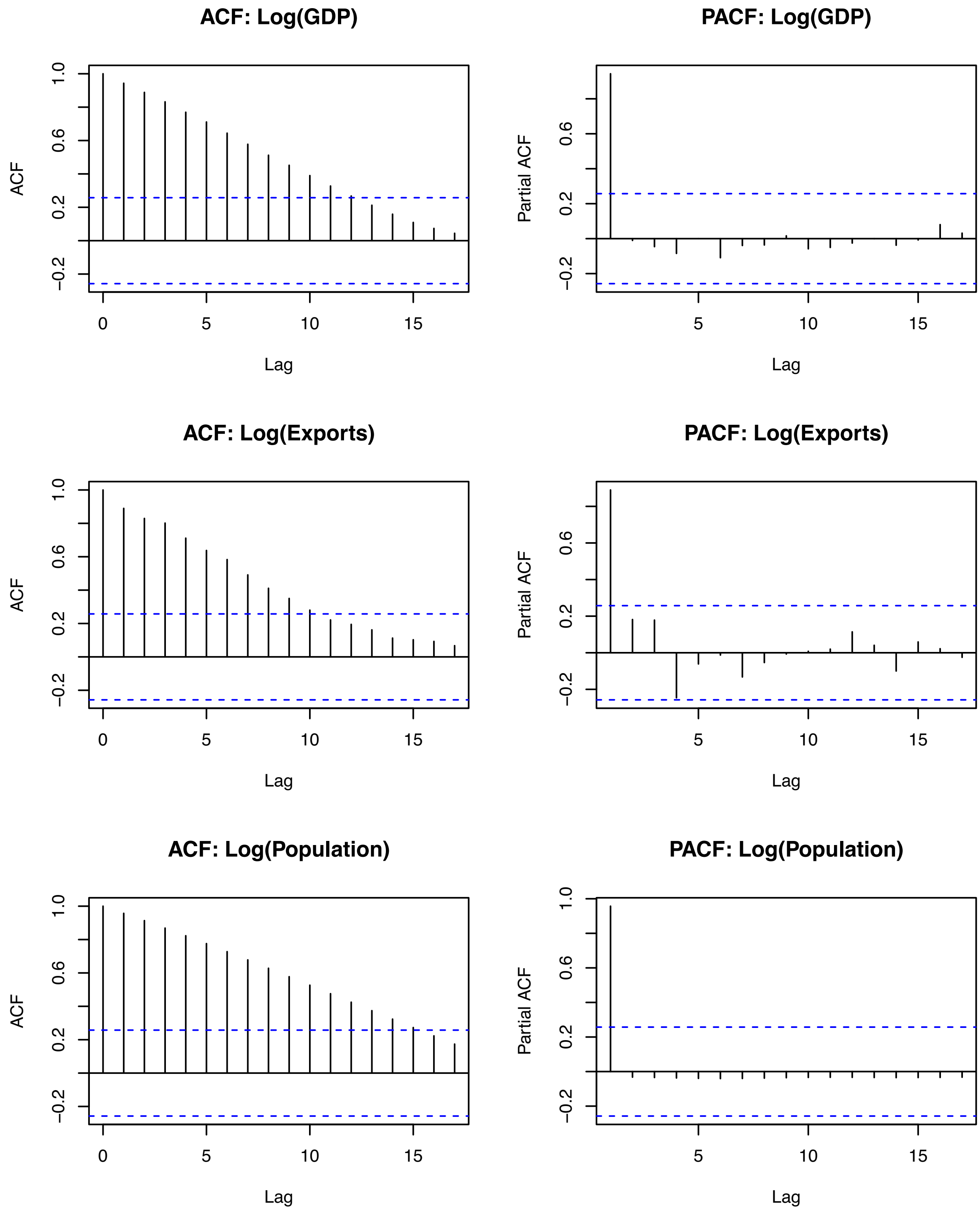


CAR Exports (1960–2017)
Exports (% of GDP)
10
15
20
25
30
35
1960
1970
1980
1990
2000
2010
Time



CAR Population (1960–2017)
Population
1500000
3000000
4500000
1960
1970
1980
1990
2000
2010
Time


Log(GDP)
gdp_log
18.5
19.0
19.5
20.0
20.5
21.0
21.5
1960
1980
2000
Time
Log(Exports)
exp_log
2.4
2.6
2.8
3.0
3.2
3.4
1960
1980
2000
Time
Log(Population)
pop_log
14.2
14.4
14.6
14.8
15.0
15.2
1960
1980
2000
Time




```
##
##  Augmented Dickey-Fuller Test
##
## data:  gdp_log
## Dickey-Fuller = -1.9149, Lag order = 3, p-value = 0.609
```

```
## alternative hypothesis: stationary
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  exp_log
## Dickey-Fuller = -2.9724, Lag order = 3, p-value = 0.1821
## alternative hypothesis: stationary
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  pop_log
## Dickey-Fuller = -2.1218, Lag order = 3, p-value = 0.5255
## alternative hypothesis: stationary
```

```
## Recommended differencing:
```

```
## GDP: 1
```

```
## Exports: 1
```

```
## Population: 2
```

```
## Series: gdp_log
## ARIMA(0,1,0) with drift
##
## Coefficients:
##        drift
##       0.0501
## s.e.  0.0170
##
## sigma^2 = 0.01683:  log likelihood = 36.05
## AIC=-68.11   AICc=-67.88   BIC=-64.02
```

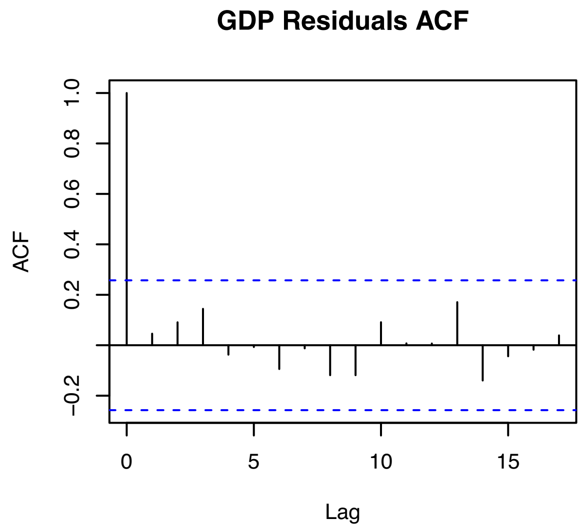
```
##
## z test of coefficients:
##
##       Estimate Std. Error z value Pr(>|z|)
## drift 0.050095   0.017027  2.9421  0.00326 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

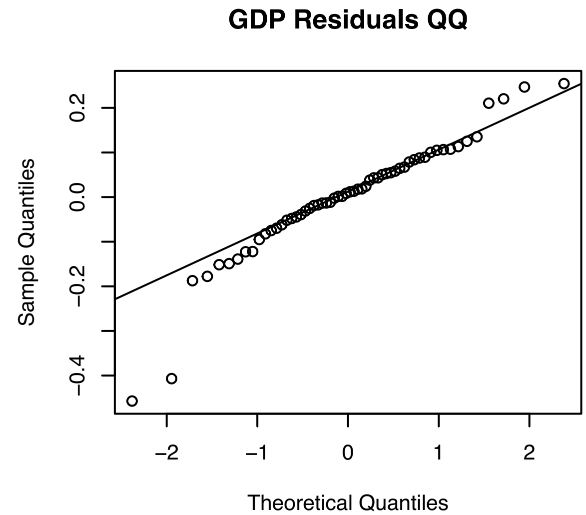
```
## Series: exp_log
## ARIMA(2,1,2)
##
## Coefficients:
##           ar1      ar2     ma1     ma2
##       -0.6924  -0.8006  0.3985  0.5571
## s.e.   0.1455   0.1879  0.2067  0.2985
##
## sigma^2 = 0.01354:  log likelihood = 43.49
## AIC=-76.98   AICc=-75.81   BIC=-66.77
```

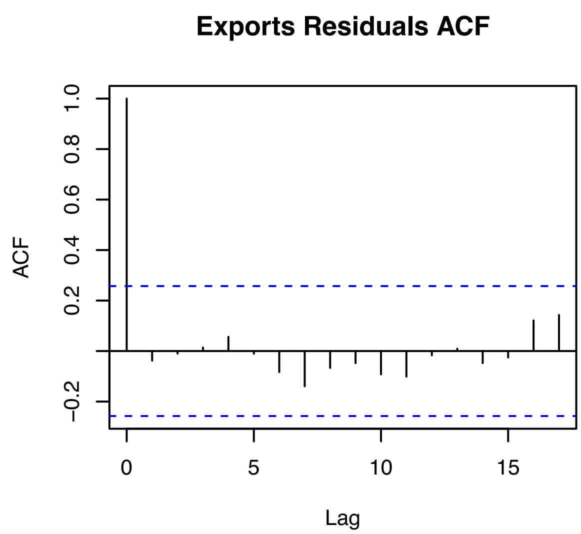
```
##
## z test of coefficients:
##
##     Estimate Std. Error z value  Pr(>|z|)
## ar1 -0.69237    0.14553 -4.7577 1.958e-06 ***
## ar2 -0.80060    0.18788 -4.2612 2.034e-05 ***
## ma1  0.39845    0.20668  1.9278   0.05387 .
## ma2  0.55714    0.29853  1.8663   0.06200 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

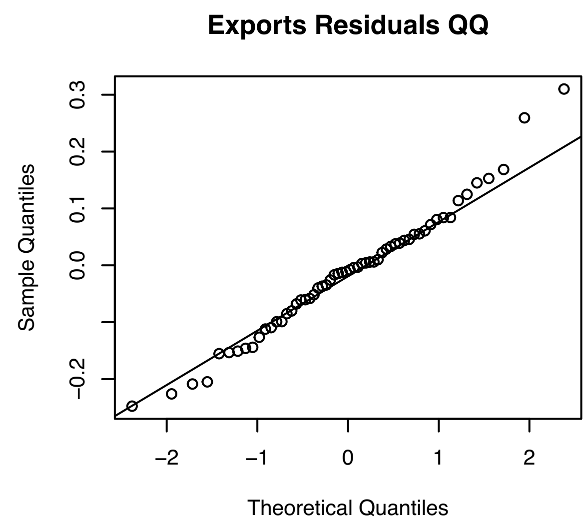
```
## Series: pop_log
## ARIMA(3,2,2)
##
## Coefficients:
##         ar1     ar2      ar3     ma1     ma2
##       0.835  0.2800  -0.5642  1.6471  0.9780
## s.e.  0.135  0.2208   0.1423  0.0856  0.0856
##
## sigma^2 = 7.961e-06:  log likelihood = 383.68
## AIC=-755.37   AICc=-753.65   BIC=-743.21
```

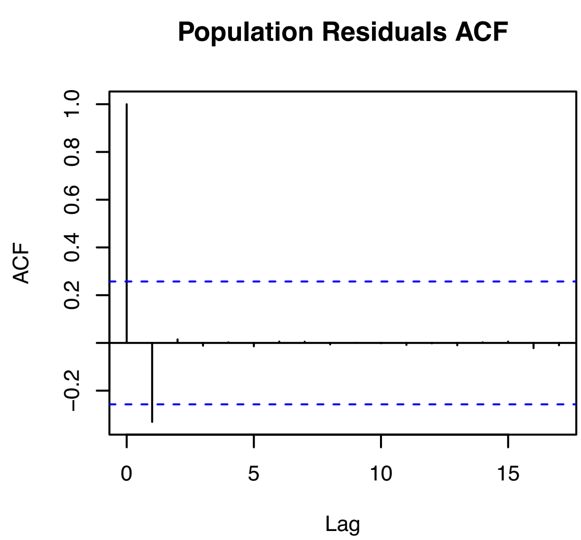
```
##
## z test of coefficients:
##
##      Estimate Std. Error z value  Pr(>|z|)
## ar1  0.835012   0.135042  6.1833 6.276e-10 ***
## ar2  0.280019   0.220848  1.2679    0.2048
## ar3 -0.564208   0.142331 -3.9641 7.369e-05 ***
## ma1  1.647137   0.085604 19.2414 < 2.2e-16 ***
## ma2  0.977982   0.085593 11.4260 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

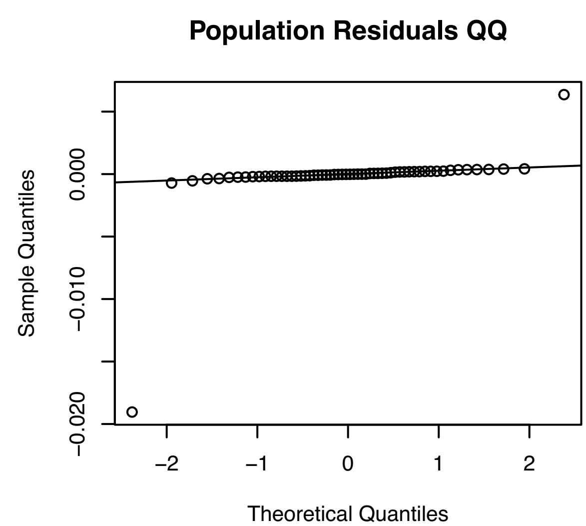

GDP Residuals ACF
ACF
Lag

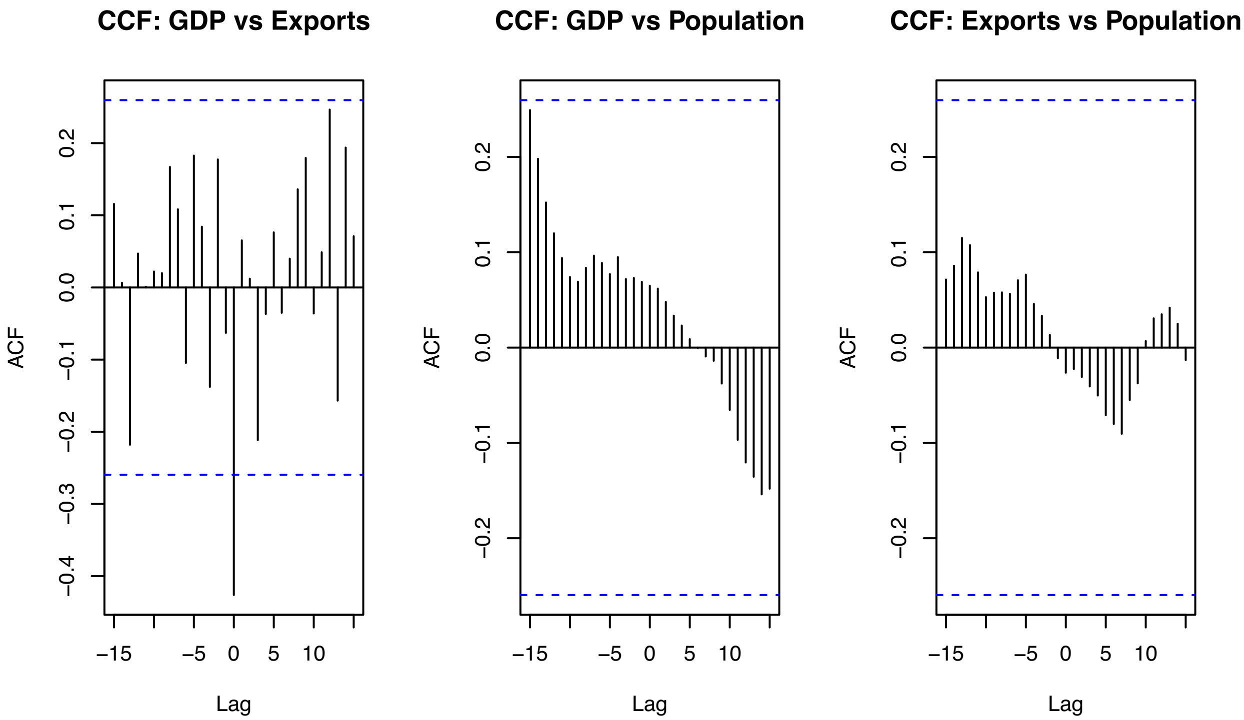

GDP Residuals QQ
Sample Quantiles
Theoretical Quantiles


Exports Residuals ACF
ACF
Lag

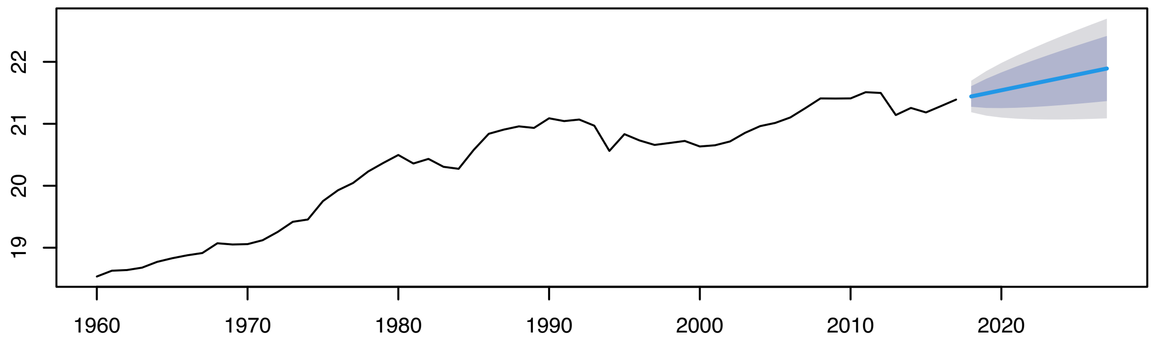

Exports Residuals QQ
Sample Quantiles
Theoretical Quantiles


Population Residuals ACF
ACF
Lag

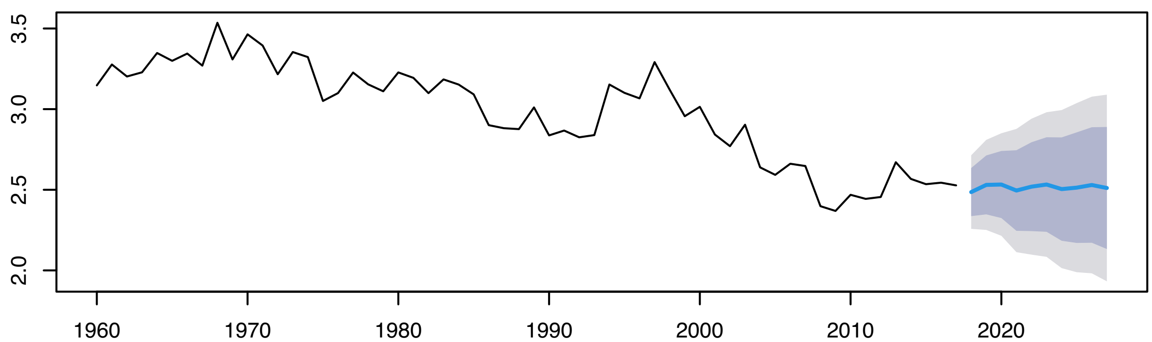

Population Residuals QQ
Sample Quantiles
Theoretical Quantiles


CCF: GDP vs Exports
CCF: GDP vs Population
CCF: Exports vs Population
ACF
Lag
−15
−5
0
5
10
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4

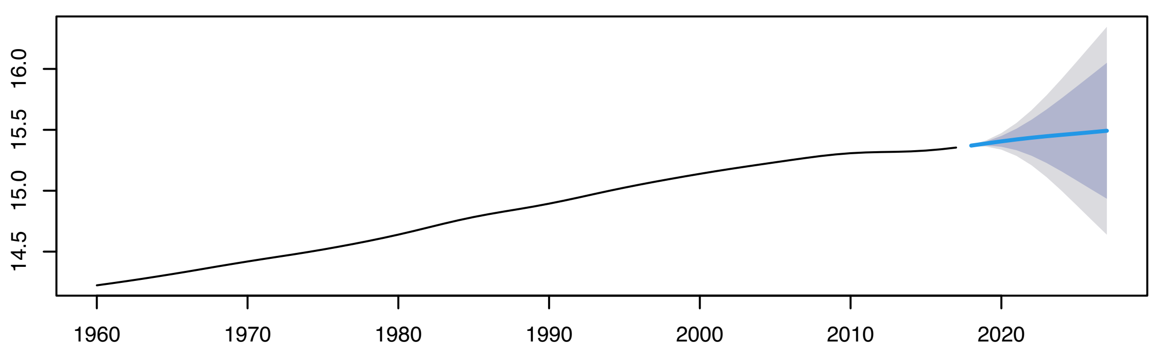
GDP Forecast (2018–2027)



Exports Forecast (2018–2027)



Population Forecast (2018–2027)




```
## GDP: ARIMA 0 1 0 | AICc = -67.88

## Exports: ARIMA 2 1 2 | AICc = -75.81

## Population: ARIMA 3 2 2 | AICc = -753.65
```

```
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE)
library(forecast)
library(tseries)
library(lmtest)
load("Data_File_finalproject.Rdata")
gdp_ts <- ts(finalPro_data$GDP, start = 1960, frequency = 1)
exp_ts <- ts(finalPro_data$Exports, start = 1960, frequency = 1)
pop_ts <- ts(finalPro_data$Population, start = 1960, frequency = 1)
par(mfrow = c(3, 1))
plot(gdp_ts, main = "CAR GDP (1960-2017)", ylab = "GDP (USD)")
plot(exp_ts, main = "CAR Exports (1960-2017)", ylab = "Exports (% of GDP)")
plot(pop_ts, main = "CAR Population (1960-2017)", ylab = "Population")
gdp_log <- log(gdp_ts)
exp_log <- log(exp_ts)
pop_log <- log(pop_ts)

par(mfrow = c(1, 3))
plot(gdp_log, main = "Log(GDP)")
plot(exp_log, main = "Log(Exports)")
plot(pop_log, main = "Log(Population)")
par(mfrow = c(3, 2))
acf(gdp_log, main = "ACF: Log(GDP)")
pacf(gdp_log, main = "PACF: Log(GDP)")
acf(exp_log, main = "ACF: Log(Exports)")
pacf(exp_log, main = "PACF: Log(Exports)")
acf(pop_log, main = "ACF: Log(Population)")
pacf(pop_log, main = "PACF: Log(Population)")
adf.test(gdp_log)
adf.test(exp_log)
adf.test(pop_log)
cat("Recommended differencing:\n")
cat("GDP:", ndiffs(gdp_log), "\n")
cat("Exports:", ndiffs(exp_log), "\n")
cat("Population:", ndiffs(pop_log), "\n")
fit_gdp <- auto.arima(gdp_log)
fit_gdp
coeftest(fit_gdp)
fit_exp <- auto.arima(exp_log)
fit_exp
coeftest(fit_exp)
fit_pop <- auto.arima(pop_log)
fit_pop
coeftest(fit_pop)
par(mfrow = c(3, 2))
acf(residuals(fit_gdp), main = "GDP Residuals ACF")
qqnorm(residuals(fit_gdp), main = "GDP Residuals QQ")
qqline(residuals(fit_gdp))
acf(residuals(fit_exp), main = "Exports Residuals ACF")
qqnorm(residuals(fit_exp), main = "Exports Residuals QQ")
qqline(residuals(fit_exp))
acf(residuals(fit_pop), main = "Population Residuals ACF")
qqnorm(residuals(fit_pop), main = "Population Residuals QQ")
qqline(residuals(fit_pop))
```

```r
par(mfrow = c(1, 3))
ccf(diff(gdp_log), diff(exp_log), main = "CCF: GDP vs Exports", lag.max = 15)
ccf(diff(gdp_log), diff(pop_log), main = "CCF: GDP vs Population", lag.max = 15)
ccf(diff(exp_log), diff(pop_log), main = "CCF: Exports vs Population", lag.max = 15)
fc_gdp <- forecast(fit_gdp, h = 10)
fc_exp <- forecast(fit_exp, h = 10)
fc_pop <- forecast(fit_pop, h = 10)

par(mfrow = c(3, 1))
plot(fc_gdp, main = "GDP Forecast (2018-2027)")
plot(fc_exp, main = "Exports Forecast (2018-2027)")
plot(fc_pop, main = "Population Forecast (2018-2027)")
cat("GDP: ARIMA", arimaorder(fit_gdp), "| AICc =", round(fit_gdp$aicc, 2), "\n")
cat("Exports: ARIMA", arimaorder(fit_exp), "| AICc =", round(fit_exp$aicc, 2), "\n")
cat("Population: ARIMA", arimaorder(fit_pop), "| AICc =", round(fit_pop$aicc, 2), "\n")
```