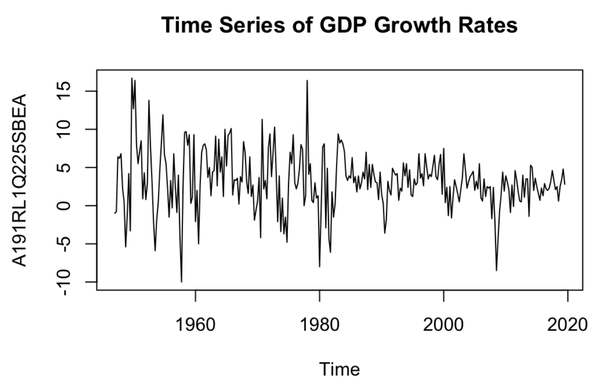
Problem Set 2: Computer Exercises

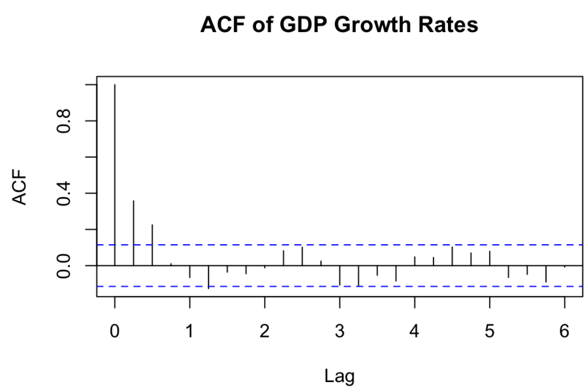
Jenny Petrova

February 20, 2025

6. The following problems examine the US quarterly real GDP growth from 1947Q1 to 2019Q3. This data is accessed using the **quantmod** package in **R**. The following outputs are computed using the **R** programming language.

   (a) Examine the time series and ACF plots of the real GDP growth rates:





The above plots do not show strong serial dependence in the GDP growth rates. This series appears stationary, since the time plot displays a constant mean and the value of the ACF plot rapidly drops within the confidence interval bounds by lag 1, and remains within the confidence interval bounds for the following lags. This suggests there is no dependence between current and past GDP growth rates.

(b) We estimate three candidate models for the GDP growth rates: AR(1), MA(1), and ARMA(1,1). For each time series model, we display the corresponding time plot, ACF, and Ljung-Box test statistic of the residuals.

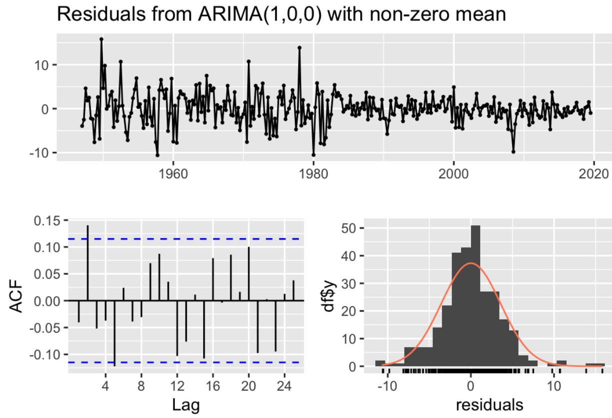


Figure 1: AR(1) Model

For the AR(1) model, the Ljung-Box test results in a p-value $= 0.076 > 0.05$, hence we *reject* the null hypothesis, meaning that the residuals resemble white noise and are hence uncorrelated. The above plots support this result, as the residuals have constant mean and variance and the lags remain within the confidence interval bounds of the ACF plot.

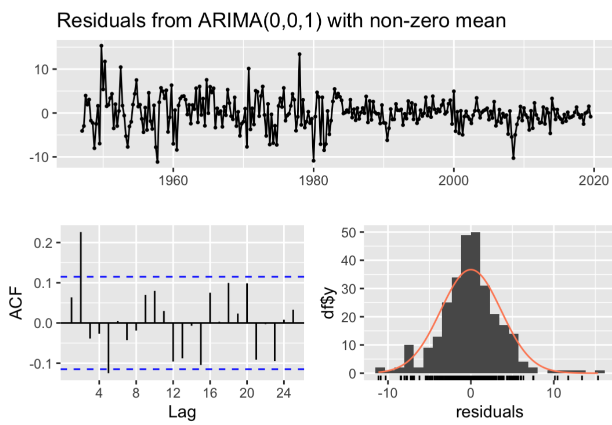


Figure 2: MA(1) Model

For the MA(1) model, the plot of the residuals suggest that the residuals may be uncorrelated. However, lag 5 in the ACF plot is outside the confidence interval, suggesting there exists some correlation. The Ljung-Box test supports this observation. We obtain a p-value $= 0.002302 < 0.05$, hence we *fail to reject* the null hypothesis, meaning there may exist some autocorrelation among the residuals.

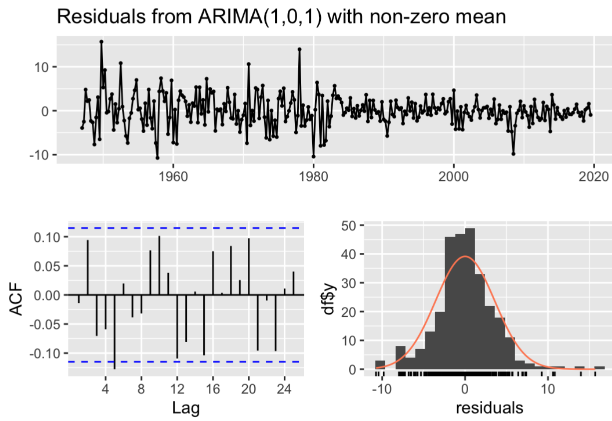


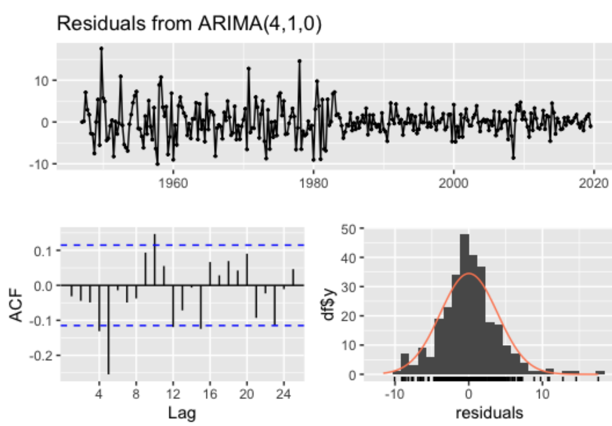
Figure 3: ARMA(1,1) Model

For the ARMA(1,1) model, the lags in the ACF plot remain within confidence interval bounds, suggesting no significant autocorrelation in the data at those lags. The residuals display a constant mean and variance in the residual plot, and the Ljung-Box test gives a p-value $= 0.09149 > 0.05$, meaning we *reject* the null hypothesis. We conclude that the residuals are not significantly correlated.

From the model outputs, we estimate the following equations:

Model	General Equation	Result	White Noise
AR(1)	$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$	$y_t = 3.2217 + 0.3583 y_{t-1} + \varepsilon_t$	$\varepsilon_t \sim WN(0, 12.75)$
MA(1)	$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$	$y_t = 3.2268 + \varepsilon_t + 0.2670 \varepsilon_{t-1}$	$\varepsilon_t \sim WN(0, 13.23)$
ARMA(1,1)	$y_t = \mu + \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$	$y_t = 3.2188 + 0.5091 y_{t-1} + \varepsilon_t - 0.1697 \varepsilon_{t-1}$	$\varepsilon_t \sim WN(0, 12.65)$

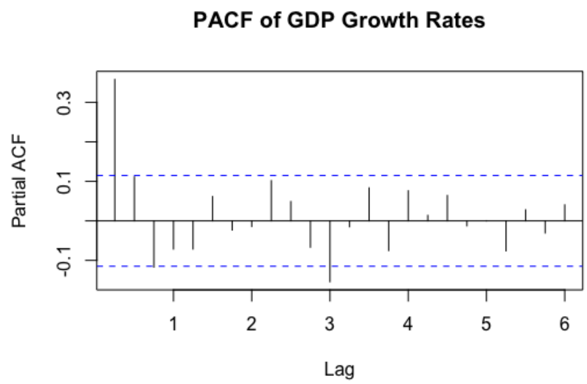
(c) To find the best AR model, we examine the **auto.arima**() function in **R** with maximum lags $p_{max} = 4$ and $q_{max} = 4$, and assume no seasonality (set seasonality=FALSE). The result yields the following plot:



which suggests the best AR model for our data is an AR(4) model. We obtain the coefficients for this model:

ar1	ar2	ar3	ar4
-0.4737280	-0.1937334	-0.2060667	-0.1320296

Examine the PACF plot of GDP growth rates:

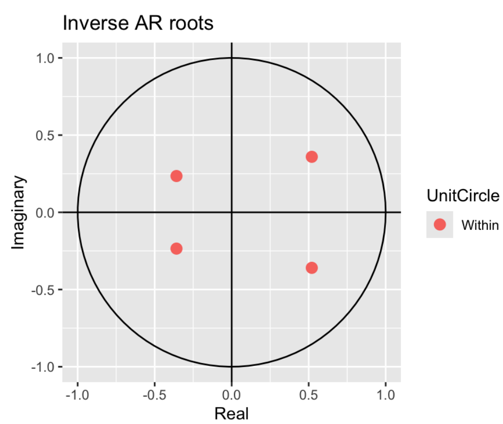


The PACF plot is useful in determining the order of an AR model. This plot supports the conclusion of an AR(4) model for our data, since the lag-h PACF is zero for all $h > 3$. Further, we fit our data to an AR(4) process and compare the AIC and BIC rates between the AR(1) and AR(2) models:

	AR(1)	AR(4)
AIC	1572.615	1569.154
BIC	1583.635	1591.194

The AIC values penalize a model by the number of parameters used. The AIC estimate for the AR(4) model is lower than the AR(1) model, meaning the AR(4) is preferred, since it has a better trade-off between goodness-of-fit and number of parameters. The BIC values penalize a model's complexity. Since AR(4) has a higher BIC estimate than AR(1), we note that the BIC favors the AR(1) model.
From the above data, we conclude that the AR(4) model is the best-fit model for *GDP* growth rates. For the following analyses, we assume an AR(4) process.

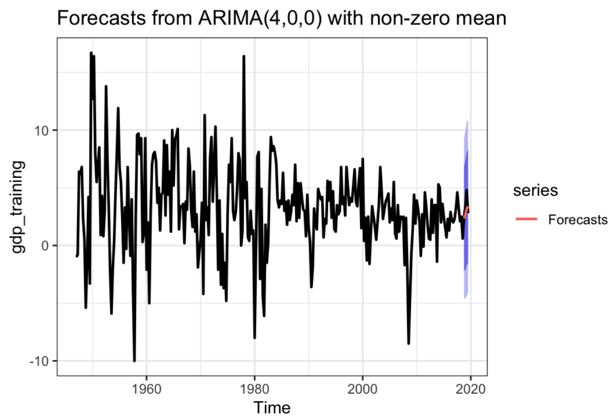
(d) The roots of the characteristic equation for the AR(4) model are found using the **polyroot**() function and plotted using the **autoplot**() function in **R**:



Note that the plot shows the inverse AR(4) roots, meaning that the true roots fall *outside* the unit circle, allowing us to conclude that this is a stationary process. Our result yielded complex-valued characteristic roots, implying the existence of business cycles. We calculate the business cycles and display our results:

Polynomial Roots	Stationarity	Average Length of Business Cycles
$\lambda_{1,2} = 0.6886766 \pm 1.4550262i$	$\lvert\lambda_{1,2}\rvert \approx 1.6098 > 1$	5.566603
$\lambda_{3,4} = -1.4690572 \pm 0.8744506i$	$\lvert\lambda_{1,2}\rvert \approx 1.7096 > 1$	2.412271

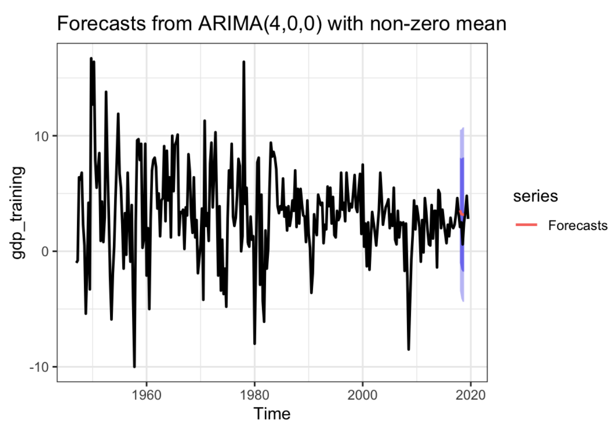
(e) We predict the 4-step ahead GDP growth rates at forecast origin 2019Q3:



The 95% prediction intervals for these forecast points are

2018Q4	2019Q1	2019Q2	2019Q3
-4.576312	-4.621408	-4.318387	-4.126090

(f) We predict the 4-step ahead GDP growth rates at forecast origin 2018Q4:



The 95% prediction intervals for these forecast points are

2018Q1	2018Q2	2018Q3	2018Q4
-3.407919	-4.091989	-4.370897	-4.373424

Corresponding Code

```r
library(forecast)
library(quantmod)
library(tseries)
library(ggplot2)

gdp.full <- getSymbols("A191RL1Q225SBEA", src = "FRED", auto.assign = FALSE)
gdp <- gdp.full["/2019"]
(start(gdp))
(end(gdp))
gdp <- ts(gdp, frequency=4, start=c(1947,1))
(start(gdp))
(end(gdp))

# (a)
# Plot time series
plot.ts(gdp, main="Time Series of GDP Growth Rates")

# Plot ACF
acf(gdp, main="ACF of GDP Growth Rates")


# (b)
# Estimate AR(1)
fit_ar1 <- arima(gdp, order=c(1,0,0))
(fit_ar1)
checkresiduals(fit_ar1)

# Estimate MA(1)
fit_ma1 <- arima(gdp, order=c(0,0,1))
(fit_ma1)
checkresiduals(fit_ma1)

# Estimate ARMA(1,1)
fit_arma11 <- arima(gdp, order=c(1,0,1))
(fit_arma11)
checkresiduals(fit_arma11)

# (c) Find best AR model
# Best AR model
best_ar = auto.arima(gdp, max.p=4, max.q=0, seasonal=FALSE)
best_ar$coef
checkresiduals(best_ar)

# Checking best model - AR(4)
fit_ar4 <- arima(gdp, order=c(4,0,0))
checkresiduals(fit_ar4)

# AIC
AIC(fit_ar1)
AIC(fit_ar4)

# BIC
BIC(fit_ar1)
BIC(fit_ar4)

# PACF
(pacf(gdp, main="PACF of GDP Growth Rates"))
```

```r
# (d)
# Find polynomial zeros
ar_coeffs <- coef(best_ar)[grep("ar", names(coef(best_ar)))]
roots <- polyroot(c(1, -ar_coeffs))
print(roots)

# Compute business cycle length
if (any(Im(roots) != 0)) {
  cycle_length <- 2 * pi / abs(Arg(roots[Im(roots) != 0]))
  print(cycle_length)
} else {
  print("No business cycles detected.")
}

autoplot(fit_ar4)

# (e)
# 4-step ahead forecast (time origin 2019Q4)
gdp_training <- window(gdp, end = c(2018,3))
gdp_test <- window(gdp, start = c(2018,4))
ar4 <- arima(gdp_training, order=c(4,0,0))
autoplot(ar4)
auto <-auto.arima(gdp_training)
auto
fcast_ar4 <- forecast(ar4, h = 4)
autoplot(fcast_ar4) +
  autolayer(gdp, color = 1, lwd = 0.8) +
  autolayer(fcast_ar4$mean, series = "Forecasts", lwd = 0.8) +
  theme_bw()
accuracy(fcast_ar4, gdp_test)

# 95% confidence interval bounds
lower_95 <- fcast_ar4$lower[,2]
upper_95 <- fcast_ar4$upper[,2]
print(lower_95)
print(upper_95)

# (f)
# 4-step ahead forecast (time origin 2018Q4)
gdp_training <- window(gdp, end = c(2017,4))
gdp_test <- window(gdp, start = c(2018,1))
ar4_2 <- arima(gdp_training, order=c(4,0,0))
autoplot(ar4_2)
auto2 <-auto.arima(gdp_training)
auto2
fcast_ar4_2 <- forecast(ar4_2, h = 4)
autoplot(fcast_ar4_2) +
  autolayer(gdp, color = 1, lwd = 0.8) +
  autolayer(fcast_ar4_2$mean, series = "Forecasts", lwd = 0.8) +
  theme_bw()
accuracy(fcast_ar4_2, gdp_test)

# 95% confidence interval bounds
lower_95_2 <- fcast_ar4_2$lower[,2]
upper_95_2 <- fcast_ar4_2$upper[,2]
print(lower_95_2)
print(upper_95_2)
```