Katholieke
Universiteit
Leuven

**Faculty of
Science**

# TIME SERIES ANALYSIS OF WAGES IN THE UK

Advanced Time Series Analysis

**Eugene Ambe Ndamukong (r417710)**

Academic year 2018/2019

# Contents



20

# 1 Data description

The data that was used for this analysis contains information of about wages measured over the years from 1855 until 1987.The focus of this report will be on *GDP* and *Employment*. Missing information was not present in the data hence there was no need for imputation. The data can be found in `https://www.wiley.com/legacy/wileychi/koopdata2ed/dataset.html` which also has the textbook which describes the data. This data consisted of 133 observations. As can be seen from figure 1, logarithmic transformed GDP increases yearly from 1855 until 1987. Also, there seem to be a more or less linear trend between the relationship between log(GDP) and time. However, there seem to be no evidence of stationarity and seasonality. In addition, there seem to be no increase in variance which was due to the logarithmic transformation.
The maximum log(GDP) was 12.52 and minimum was, 10.04. The mean log(GDP) was 11.25. No missing information was noticed in this data. An Augmented Dickey-Fuller (ADF) was used to test for stationaity of the time series. This test was used to test null hypothesis: no stationarity versus stationarity. The result of this test showed that there was not enough evidence to reject the null hypothesis with a p-value of 0.72. Therefore, the time series does not have any evidence of stationarity. Hence, to solve this problem, differencing or integrating of the time series was applied to eliminate the trend leading to a stationary time series. After differencing the time series(figure 2) resulted to a time series that seem stationary.
Another unit root test was performed on the integrated time series to test the stationarity. This test tested the null hypothesis no stationarity versus stationarity. The test resulted in a p-value of 0.01 showing that there was enough evidence to reject the null hypothesis. Therefore the series after differencing or integration is stationary and ready for modelling.

# 2 Univariate Time Series Analysis

In order to specify the model for use, a correlogram was used to specify the order of the model. Judging from the figures 3, Autoregressive model and moving average model of each of orders 1 are possible. However, there is a chance that combination of moving average and autoregressive models with varying orders or lag lengths could lead to a better model.

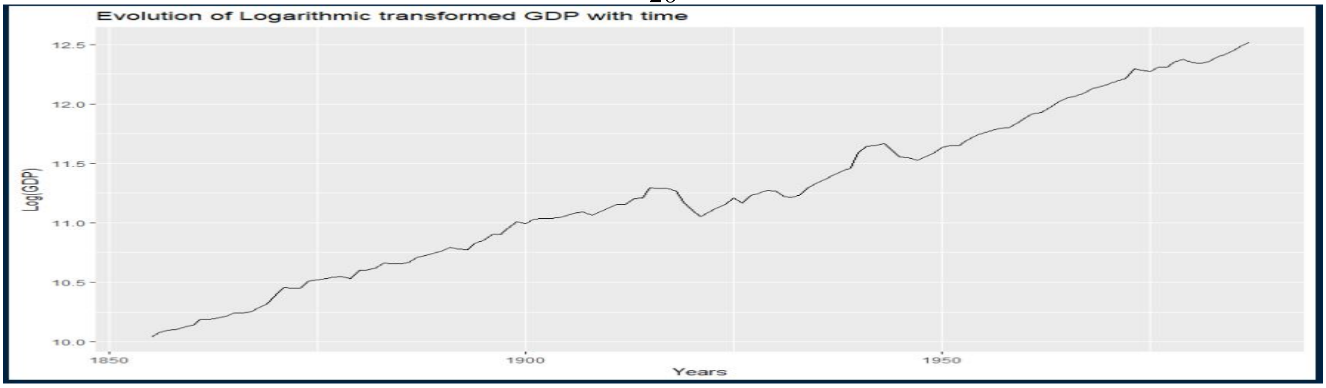


Figure 1: Time series for yearly log transformed GDP

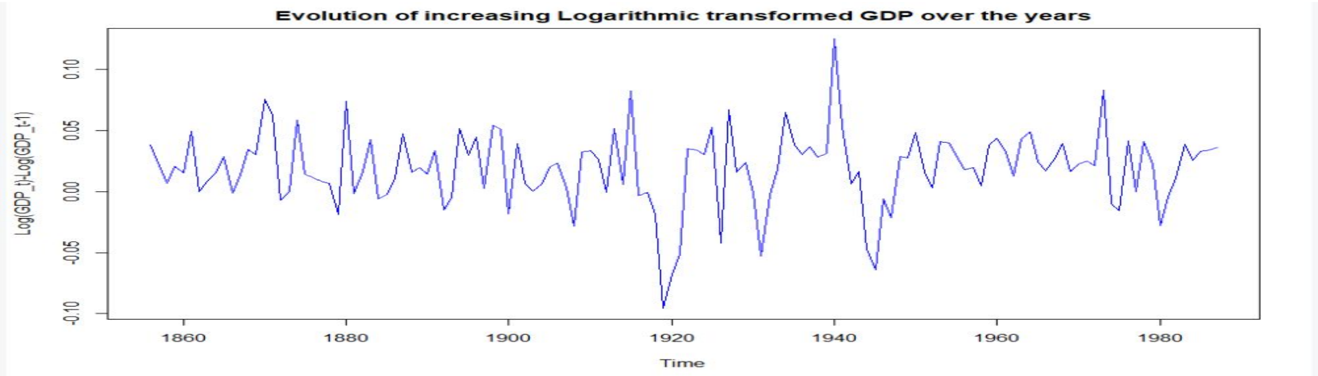


Figure 2: Time series for yearly increase in log transformed GDP

## 2.1 Model estimation

A total of six models were estimated which all had good residual correlograms except one(i.e ARIMA(1,1,0)) as seen in figures 9-12 (in appendix) and insignificant Box-Ljung tests except one(i.e ARIMA(1,1,0)) as seen in table 1. All models had all of their parameters significant except 1 model (ARIMA(2,1,2) table 5 in appendix). ARIMA(2,1,1) seem to be the best model based based on outstanding results from BIC, Box-Ljung test, MAE and RMSE compared those of the other models. Because of the outstanding results from BIC, MAE and RMSE, ARIMA(2,1,1) was used to forcast the log($GDP$).

## 2.2 Forcast

The best model from table 1, ARIMA(2,1,1) was used to forcast the log(GDP) for the next 10 years (horrizon=10). This resulted in the time series indicated in blue as seen in figure 4. The forcasted time series seem to show that log(GDP) will increase for the next 10 years. The prediction intervals(dark grey: 80% and light grey:95%) of the forcast seem to increase over the next 10 years. The black line corresponds to the time series of the original data while the red line is the time series of the predicted(fitted) in-sample data. The closeness fitted and actual time series' according to figure 4 seem to be a good indication that the model was good.

# 3 Multivariate Time Series Analysis

It has been noted that GDP has an impact on employment(Basnett & Sen, 2013). This relationship was modelled using several multivariate time series models.

## 3.1 Test for cointergration

Figure 14 in the appendix shows the time series of log($GDP$) and log($employment$). Judging from the plots, Log($GDP$) and Log($Employment$) don't seem to be cointergrated. A cointergration test was made on $Employment$ versus Log($GDP$). The Engle-Granger approach was used to do this. In order to perform this test the following regression model was made:

$$Log(Employment) = c + \beta Log(GSP)_t + \epsilon_t \qquad (1)$$

A unit root test was performed on the residuals from the regression model above. It was concluded from this test that there was not enough evidence to reject the null hypothesis (i.e null hypothesis: $\beta$=0) of no cointegration between both time series. The unit root test resulted

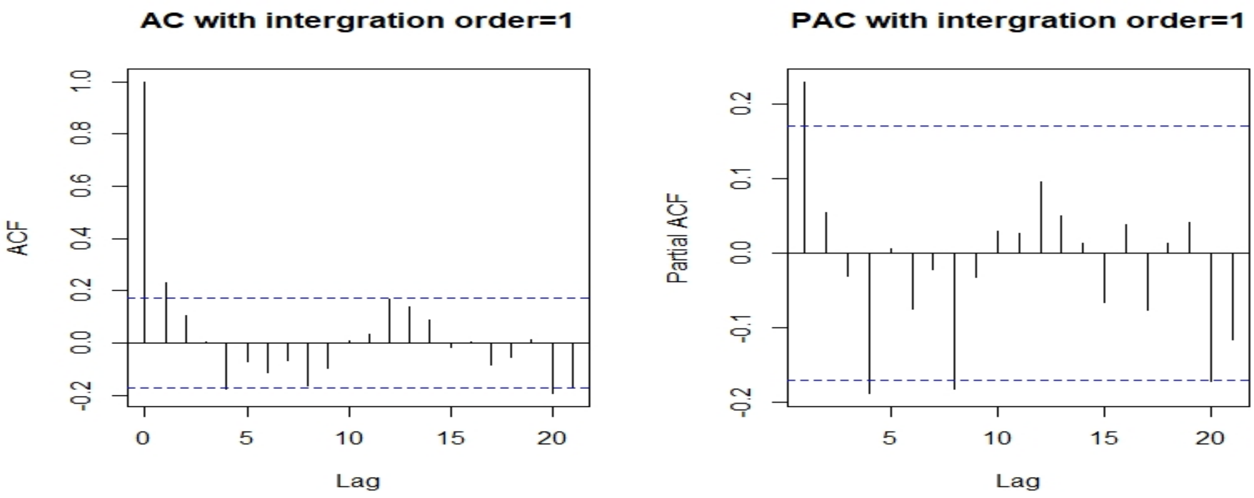


Figure 3: Autocorrelations(Left) and Partial autocorrelations(Right) of time GDP time series of intergration=1

| Model | BIC | Residual $\rho$-plot | Box-Ljung test | Sig parameters | MAE | RMSE |
|---|---|---|---|---|---|---|
| ARIMA(1,1,0) | -521.68 | not good | p-value=0.01445 | all | 0.0247 | 0.0322 |
| ARIMA(0,1,1) | -513.12 | ok | p-value = 0.1424 | all | 0.0261048 | 0.03325843 |
| ARIMA(1,1,1) | -522.63 | ok | p-value = 0.6174 | all | 0.02398 | 0.0315 |
| ARIMA(2,1,1) | -528.96 | ok | p-value = 0.9827 | all | 0.0222 | 0.02990414 |
| ARIMA(1,1,2) | -527.85 | ok | p-value = 0.8573 | all | 0.0224 | 0.02998964 |
| ARIMA(2,1,2) | -524.25 | ok | p-value = 0.9188 | Only 2 | 0.0223 | 0.02992035 |

Table 1: Comparison of different univariate models

in a p-value of 0.9213. Therefore, an error-correcting model was not necessary to model the relationship between $\Delta$Log($GDP$) and $\Delta$Log($employment$).

### 3.2 Dynamic lag(DL) model )

A DL model was used to model the relationship between $\Delta$Log($GDP$) and $\Delta$(Employment). To do this, their respective series were ensured to be stationary by applying a difference(intergrated). $\Delta$Log($GDP$) was already proven to be stationary before. The series of Log($Employment$) was proven not to be stationary through the ADF test (p-value=0.738) while the $\Delta Employment$ was proven to be stationary(p-value=0.01) also by the ADF test. The relationship being modelled with lag length of 2 was as follows:

$$\Delta Log(Employment) = c + \beta_1 \Delta Log(GDP)_t + \quad (2)$$
$$\beta_2 \Delta Log(GDP)_{t-1} + \beta_3 \Delta Log(GDP)_{t-2} + \epsilon_t \quad (3)$$

The output of this DL(2) model can be seen in table 2. The F statistic was significant p-value of 1.181e$^{-15}$ suggesting that there is at least one effect of $\Delta$Log($GDP$)in the model that is not equal to zero i.e at least one significant relationship. The residuals were tested using the Box-Ljung test to know if there were white noise. The results of the Box-Ljung test concluded that there is not enough evidence to reject that the residuals are white noise(P-value=0.602). It can be seen that as the $\Delta$Log($GDP$) increases by one unit at time $t$, the short run effect on growth in $\Delta$Log($Employment$) increases significantly by 0.3698 on log scale. Meanwhile, the long run effect on $\Delta$Log($Employment$) increases by 0.6948($\beta_1 + \beta_2 + \beta_3$) on log scale when $\Delta$Log($GDP$) increases by 1 unit at time $t$.

The DL(2) model was reported to explain 43.68% of the variance in the change in Log($Employment$) and AIC of -673.1729 as fitness criterion. Whereas, regressing Log($GDP$) on Log($Employment$) would of explained only 1.279% of $GDP$ and a higher AIC of -529.6719.

### 3.3 Autoregressive distributed lag model

An autoregressive distributed lag model was fitted as follows:

$$\Delta Log(Employment) = c + \phi_1 \Delta Log(Employment)_{t-1} + \phi_2 \Delta Log(Employment)_{t-2} + \beta_1 \Delta Log(GDP)_{t-1} + \beta_2 \Delta Log(GDP)_{t-2} + \epsilon_t$$

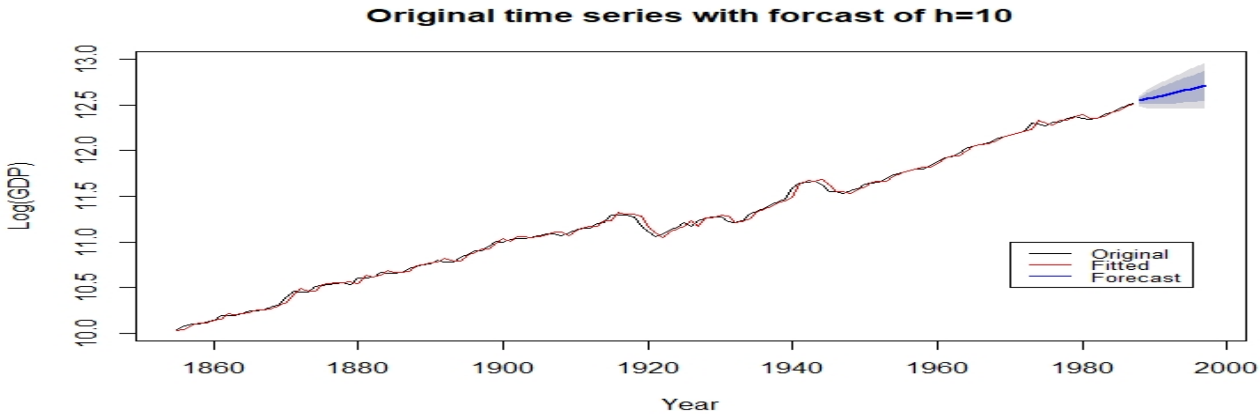


Figure 4: Original and time series and forcasted time series

| Predictors | Estimate | SE | t value | Pr(> $\|t\|$) |
|---|---|---|---|---|
| Intercept | -0.007441 | 0.002089 | -3.561 | 0.000521 |
| Δ Log(GDP) | 0.3698 | 0.052 | 7.104 | 7.89e$^{-11}$ |
| Δ Log(GDP) at lag 1 | 0.211 | 0.053 | 3.964 | 0.000123 |
| Δ Log(GDP) at lag 2 | 0.114 | 0.052 | 2.198 | 0.0299 |

Table 2: **Output of Dynamic lag 2 model**

However, test for no granger causality was made to see if ΔLog(GDP) actually contributes incrementally to the predictive power of ΔLog(*Employment*). It was concluded from this test that, ΔLog(*GDP)* at lag 1 and 2 significantly contribute incrementally to the predictive power of ΔLog(*Employment)* with an F-statistic =9.844 and p-value of 5.89e$^{-07}$. However, on the contrary, ΔLog(*Employment)* does not contribute incrementally to the predictive power of ΔLog(*GDP)* as the F-statistic isn't significant(p-value of 0.3896). The output of the model of the above equation can be seen in table 3.

| Predictors | Estimate | SE | t value | Pr(> $\|t\|$) |
|---|---|---|---|---|
| Intercept | -0.002359 | 0.002374 | -0.994 | 0.322 |
| ΔLog(Employment) at lag 1 | 0.081 | 0.103 | 0.787 | 0.433 |
| ΔLog(Employment) at lag 2 | -0.203 | 0.096 | -2.111 | 0.037 |
| ΔLog(GDP) at lag 1 | 0.255 | 0.072 | 3.518 | 0.000607 |
| ΔLog(GDP) at lag 2 | 0.205 | 0.076 | 2.707 | 0.0077 |

Table 3: Output of Autoregressive lag model

This model had a fitness criterion, AIC of -632.1435 and was able to explain only 23.96% of the variance in growth in Log(*Employment*). If the Log(*GDP*) was regressed on the Log(*Employment*), only 6.994% of the variance would have been explained and the model would have had a worse AIC of -533.4244. This model was also validated by testing the residuals if they come from a white noise distribution. The Box-Ljung test confirm that the residuals indeed are white noise(p-value=0.743) hence the model was well validated.

### 3.4 VAR(Vector Autoregresive) model

The model system to be fitted was as follows:

$$\begin{pmatrix} \Delta Log(Employment) \\ \Delta Log(GDP) \end{pmatrix} = \begin{pmatrix} c1 \\ c2 \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} \Delta Log(Employment)_{t-1} \\ \Delta Log(GDP)_{t-1} \end{pmatrix} + \begin{pmatrix} U_{Employment,t} \\ U_{GDP,t} \end{pmatrix}$$

An order or lag length of 1 was used to fit the final model as it had the smallest Schwarz information criterion among the other models. The F-test of both equations confirm that at least one effect was not equal to zero i.e p-value=1.582e$^{-06}$ for equation with response ΔLog(*Employment* and p-value=0.02618 for equation with response ΔLog(*GDP*. Both the univariate and multivariate residuals have been observed and confirmed to come from a white noise distribution as seen in correlograms of figure 13 in appendix. The results of this VAR model can be seen in table 4.

The effect of ΔLog(Employment) at lag 1 on ΔLog(*GDP*) was decreasing insignificantly meanwhile that of ΔLog(*GDP*) at lag 1 was increasing significantly when either of them independently increases by 1 unit. Also, effect of ΔLog(Employment) at lag 1 on ΔLog(Employment) was increasing insignificantly mean-

| Response | Predictors | Estimate | SE | t value | Pr(> \|t\|) |
|---|---|---|---|---|---|
| Δ Log(GDP) | ΔEmployment at lag 1 | -0.077826 | 0.137953 | -0.564 | 0.5736 |
| | ΔLog(GDP) at lag 1 | 0.263582 | 0.104398 | 2.525 | 0.0128 |
| | constant | 0.014177 | 0.003107 | 4.564 | $1.16e^{-05}$ |
| Response | Predictors | Estimate | SE | t value | Pr(> \|t\|) |
| ΔEmployment | ΔEmployment at lag 1 | 0.1176698 | 0.0967777 | 1.216 | 0.226269 |
| | ΔLog(GDP) at lag 1 | 0.2694554 | 0.0732381 | 3.679 | 0.000343 |
| | constant | -0.0002414 | 0.0021793 | -0.111 | 0.911975 |

Table 4: Output of VAR model

while that of $\Delta\text{Log}(GDP)$ at lag 1 was increasing significantly. However, the effect of $\Delta\text{Log}(GDP)$ at lag 1 on $\Delta\text{Log}(Employment)$ is more significant than the effect $\Delta\text{Log}(Employment)$ on $\Delta\text{Log}(GDP$. It was also seen that the average $\Delta\text{Log}(GDP)$ was increasing significantly while the average $\Delta\text{Log}(employment$ was decreasing insignificantly given there are no predictors affecting them. 5.533% of variance was explained in growth of the *GDP* on log scale and 18.84% of variance in the change in *Employment* on log scale with AIC of -15.07015. The impulse response plots can be seen in figures 5-6. According to figure 5, each time the innovation $U_{Employment}$ increases by one unit, there was a significant decrease in the $\Delta\text{Log}(Employment)$(cEmp) till the first period (year) and then an insignificant decrease till the second period. Meanwhile, there was an insignificant change in the $\Delta\text{Log}(GDP)$(cGDP) through out the periods.

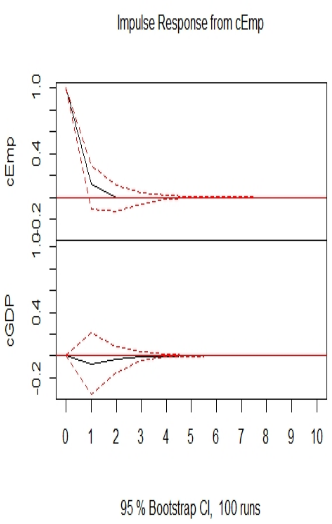


Figure 5: Impulse response after shock from change in *Employment*(cEMP)

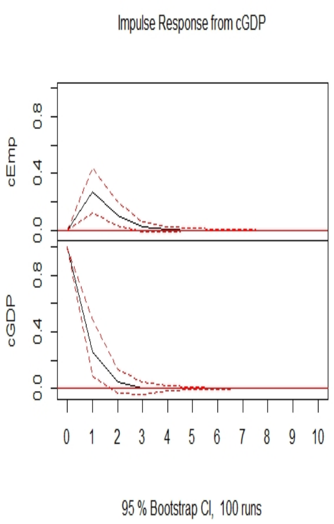


Figure 6: Impulse response after shock from growth in $\log(GDP)$(cGDP)

According to figure 6, an increase in the innovation $U_{GDP}$ by one unit, the $\Delta\text{Log}(Employment)$ increases significantly till the first period and then decreases significantly till the $3^{rd}$ period. Meanwhile, the $\Delta\text{Log}(GDP)$ decreases significantly till approximately the $2^{nd}$ period.

## 4 Conclusion

From the univariate analysis, it can be concluded that the most appropriate model cannot be specified solely from the correlogram but could also be a combination of both autoregressive and moving average models with varying lag lengths. ARIMA(1,1,0) and ARIMA(0,1,1) were not better than their respective combinations after comparing their different validation and prediction measures. Also, varying the orders or lag lengths of the different univariate models will lead to different model qualities.
From the multivariate analysis, it can be seen from all model outputs of the 3 models used that *Employment* does depend on GDP and not the other way round. This was evident in the DL(2) model as the short and long run effects of $\Delta Log(GDP)$ have a significant effect on the $\Delta Log(Employment)$. Additionally, the granger causality test concludes that $\Delta Log(GDP)$ significantly contributes incrementally to the predictive power of $\Delta Log(Employment)$. Also, from the VAR(1) model, it was noticed that the variance explained in growth of $GDP$(5.533%) was lower than that of variance explained in change in *Employment*. In addition, the percentage of variance explained in *GDP* was much smaller than the percentage of of variance explained in *Employment*. Therefore, this seem to conclude that *Employment* does depend on the *GDP* of the UK.

# 5 Appendix

## 5.1 Figures

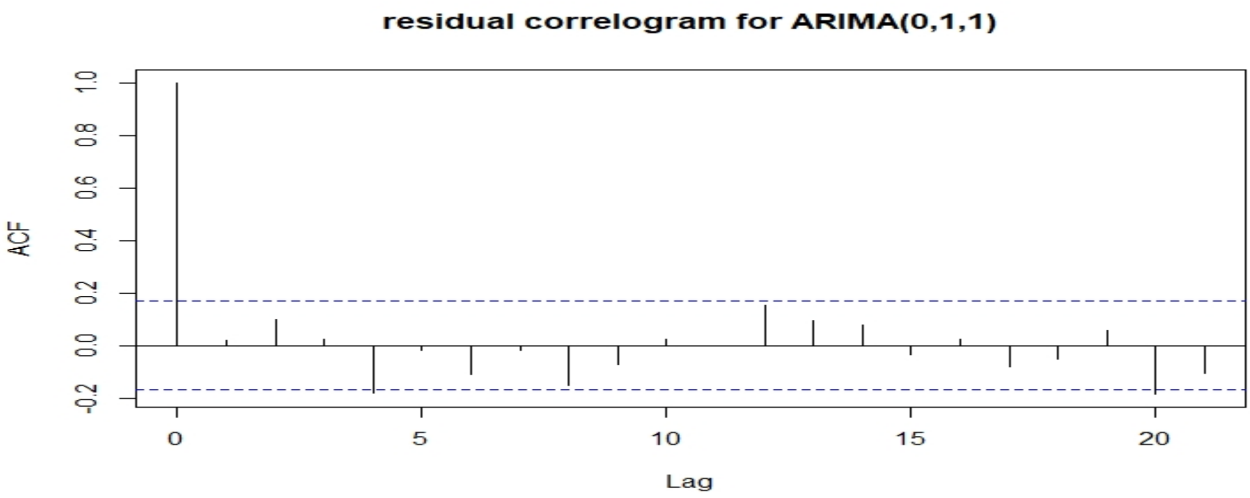


Figure 7: Residual correlogram for ARIMA(0,1,1)

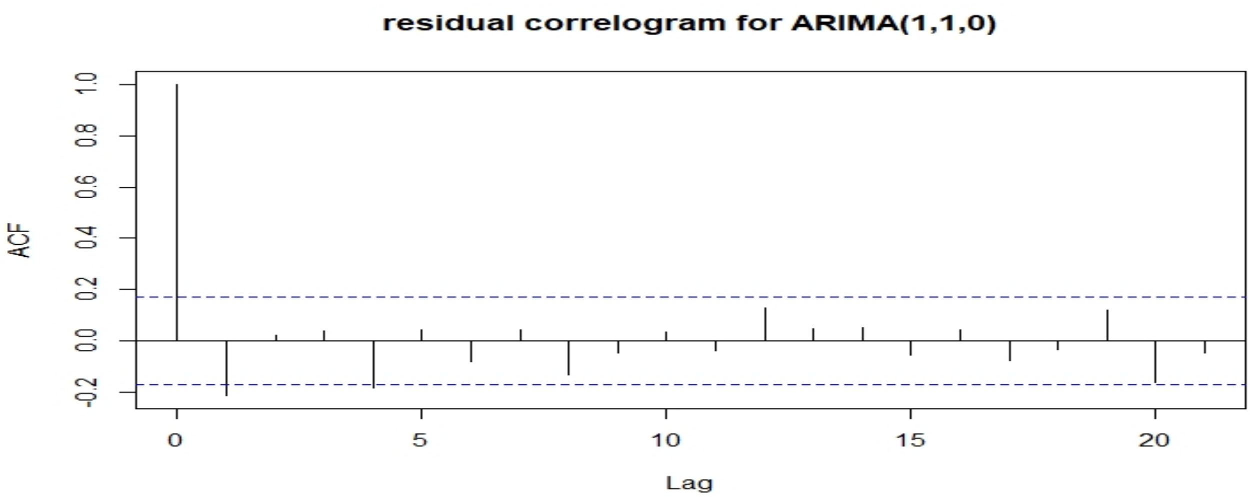


Figure 8: Residual correlogram for ARIMA(1,1,0)

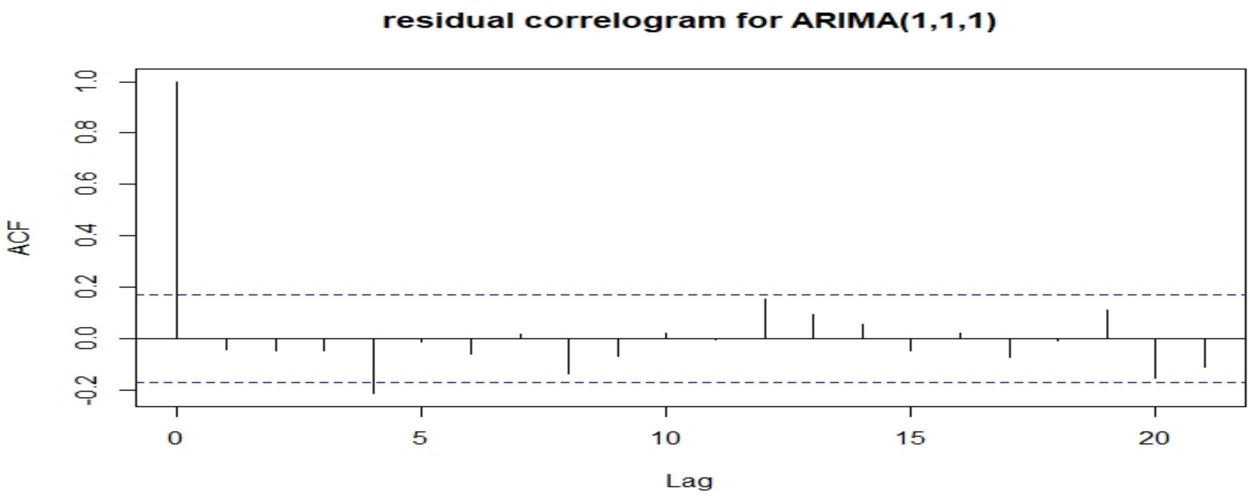


Figure 9: Residual correlogram for ARIMA(1,1,1)

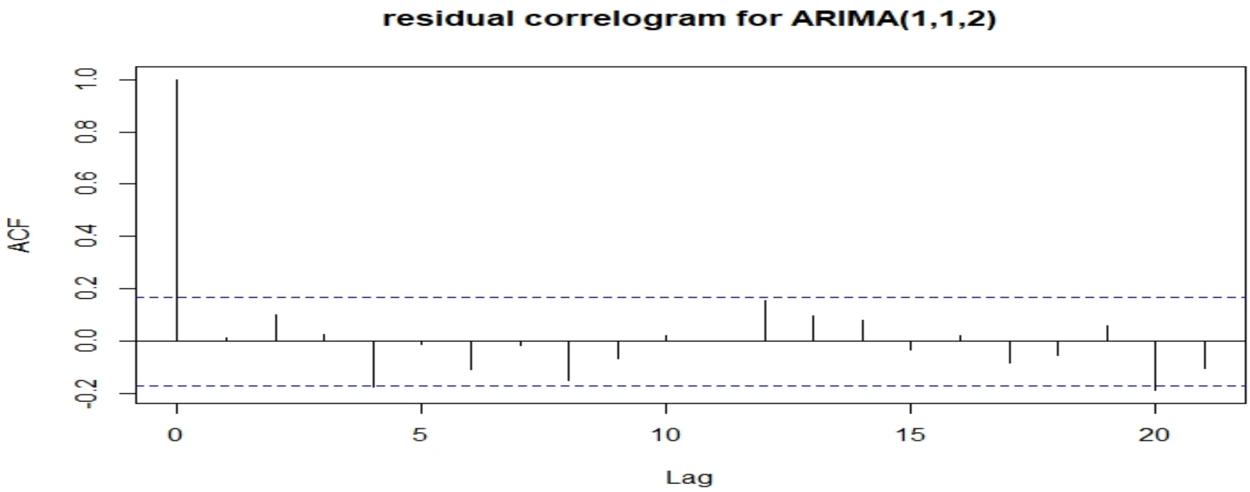


Figure 10: Residual correlogram for ARIMA(1,1,2)

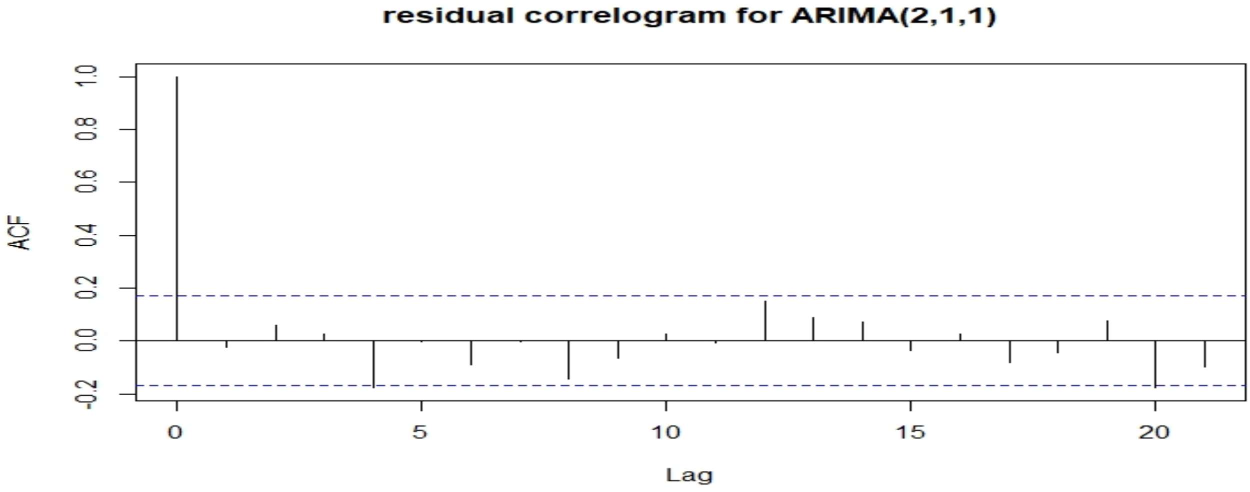


Figure 11: Residual correlogram for ARIMA(2,1,1)

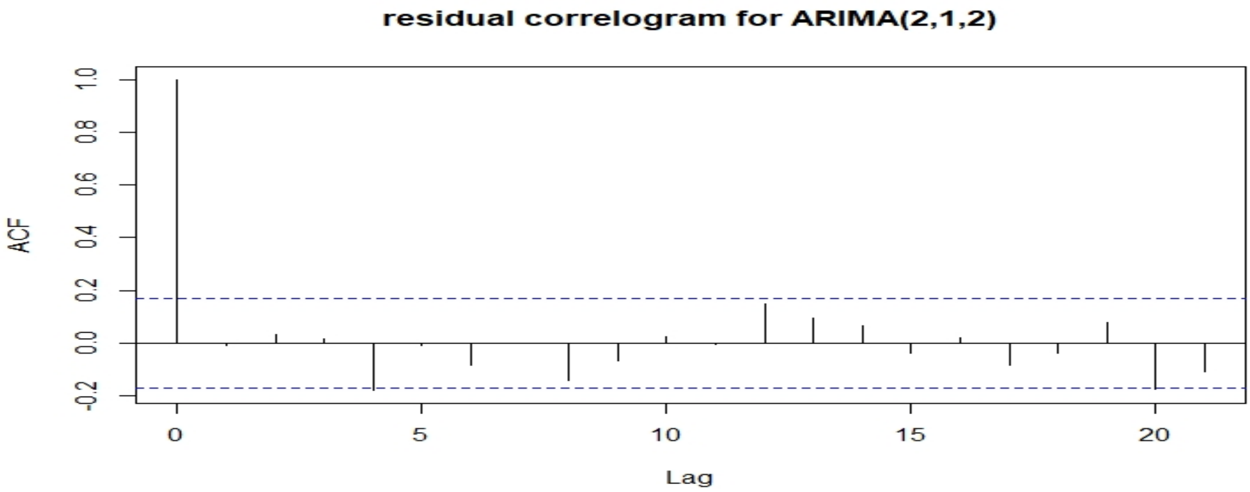


Figure 12: Residual correlogram for ARIMA(2,1,2)

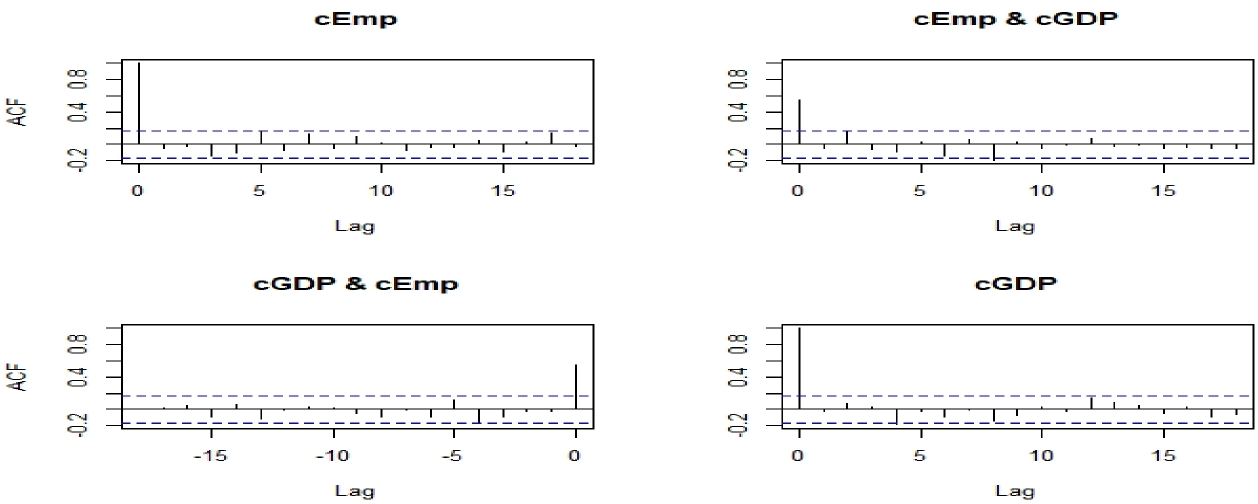


Figure 13: Residual correlogram for VAR(1) model

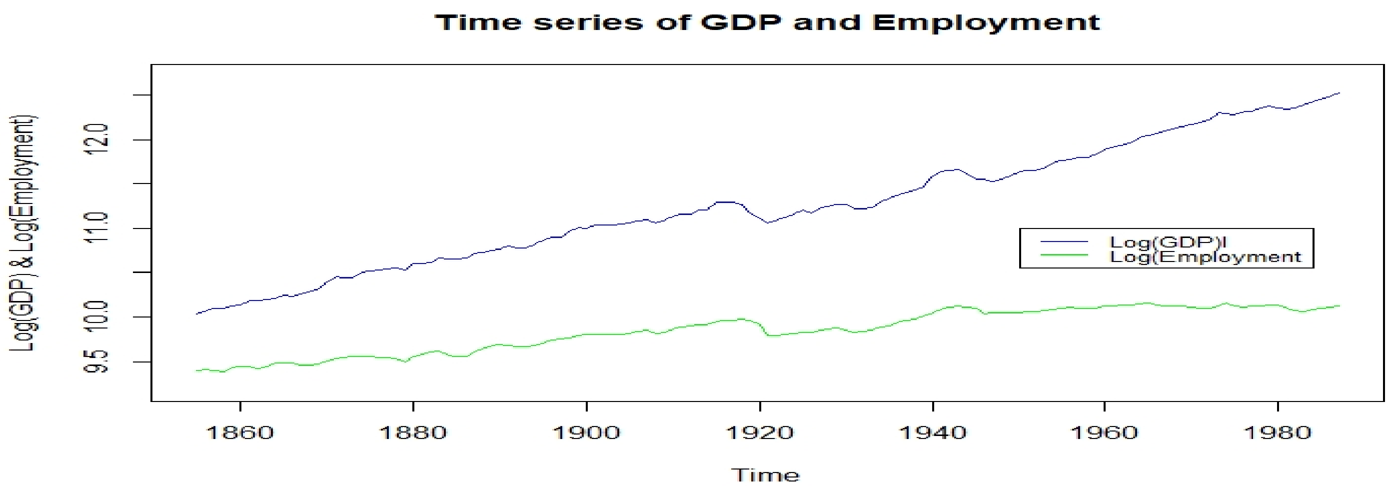


Figure 14: Time series of other variables

### 5.2 Tables

| Model | AR1($\|t - value\|$) | AR2($\|t - value\|$) | MA1($\|t - value\|$) | MA2($\|t - value\|$) |
|---|---|---|---|---|
| ARIMA(1,1,0) | 5.596939 | | | |
| ARIMA(0,1,1) | | | 4.536685 | |
| ARIMA(1,1,1)) | 6.442926 | | 2.445647 | |
| ARIMA(2,1,1) | 14.52453 | 2.843458 | 87.88496 | |
| ARIMA(1,1,2) | 3333 | | 9.759006 | 2.600251 |
| ARIMA(2,1,2) | 5.591259 | 1.545123 | 4.479371 | 0.5884219 |

Table 5: t-values of different univariate models

### 5.3 R-code

```
#############################################################################
################Time Series analysis#########################################
##################################################################

#Description
names(WAGE)
#W=log(wage); OBS=time; P=log(Price index); L=total labour
#Identifying GDP as a time series data
W.GDP.ts=ts(WAGE$GDP, start = c(min(WAGE$OBS.),1))
#Checking data
summary(W.GDP.ts)
summary(WAGE)
class(W.GDP.ts)
start(W.GDP.ts)
frequency(W.GDP.ts)
plot(WAGE$GDP, type = 'l')
#Plot of log(GDP) time series
plot.ts(W.GDP.ts,ylab="Log(GDP)",col="blue",
        main='Evolution of GDP over the years')
library(ggplot2)
ggplot(data=WAGE, aes(x=OBS.,y=W.GDP.ts,group=1))+
  geom_line()+ylab("Log(GDP)")+xlab("Years")+
  ggtitle("Evolution of Logarithmic transformed GDP with time")

#Autocorrelation plot of log(GDP)
acf(W.GDP.ts, main="Correlogram for Wage time series")

#Test for stationarity
library(CADFtest)
max.lag<-round(sqrt(length(W.GDP.ts)))
#CADFtest(W.GDP.ts, type= "drift", criterion= "BIC", max.lag.y=max.lag)
adf.test(W.GDP.ts,alternative = "stationary",k=0)
#Plot of differenced time series
plot.ts(diff(W.GDP.ts),ylab="Log(GDP_t)-Log(GDP_t-1)",col="blue",
        main='Evolution of increasing Logarithmic transformed GDP over the years')
monthplot(diff(W.GDP.ts))

#Test for stationarity in differenced time series
#max.lag<-round(sqrt(length(diff(W.GDP.ts))))
#CADFtest(diff(W.GDP.ts), type= "drift", criterion= "BIC", max.lag.y=max.lag)
adf.test(diff(W.GDP.ts),alternative = "stationary",k=0)

#####Univariate Analysis#######
par(mfrow=c(1,2))
acf(diff(W.GDP.ts), main="AC with intergration order=1")

pacf(diff(W.GDP.ts),main="PAC with intergration order=1")
#library(forecast)
#seasonplot(diff(WAGE$GDP),frequency=4, start=c(1900))
```

```
library(forecast)
model1=Arima(W.GDP.ts, order = c(1,1,1))
model1
summary(model1)
acf(model1$residuals,main="residual correlogram for ARIMA(1,1,1)")
Box.test(model1$residuals)

model2=Arima(W.GDP.ts, order = c(0,1,1))
summary(model2)
acf(model2$residuals,main="residual correlogram for ARIMA(0,1,1)")
Box.test(model2$residuals)

model3=Arima(W.GDP.ts, order = c(1,1,0))
summary(model3)
acf(model3$residuals,main="residual correlogram for ARIMA(1,1,0)")
Box.test(model3$residuals)

model4=Arima(W.GDP.ts, order = c(2,1,2))
summary(model4)
acf(model4$residuals,main="residual correlogram for ARIMA(2,1,2)")
Box.test(model4$residuals)

model5=Arima(W.GDP.ts, order = c(1,1,2))
summary(model5)
acf(model5$residuals,main="residual correlogram for ARIMA(1,1,2)")
Box.test(model5$residuals)

model6=Arima(W.GDP.ts, order = c(2,1,1),method = "ML")
summary(model6)
acf(model6$residuals,main="residual correlogram for ARIMA(2,1,1)")
Box.test(model6$residuals)

auto.arima(W.GDP.ts)
#Forcast
library(forecast)
uni.forc=predict(model6,n.ahead = 20)
names(uni.forc)
uni.forc$pred[60]
ts.plot(uni.forc$pred)
plot(forecast(model6,h=10),
      main="Original time series with forcast of h=10",
      xlab="Year",ylab="Log(GDP)")
lines(model4$fitted ,col="red")
legend(1970, 11, legend = c("Original", "Fitted", "Forecast"),
        col = c("black", "red", "blue"), lty=1, cex=0.8)

#####Multivariate Analysis########
#Plots of time series
names(WAGE)
W.employ.ts=ts(WAGE$E,start = c(min(WAGE$OBS.),1))
ts.plot(W.employ.ts,W.GDP.ts)
```

```
par(mfrow=c(1,2))
plot.ts(W.GDP.ts, ylab="Log(GDP) & Log(Employment)", col="blue",
        main='Time series of GDP and Employment',
        ylim=c(9.2,12.7))
lines(W.employ.ts, ylab="Log(Employment)", col="green",
        main='Evolution of Employment over the years')
legend(1950, 11, legend = c("Log(GDP)l", "Log(Employment")),
       col = c("blue", "green"), lty=1, cex=0.8)

#plot.ts(W.wage.ts, ylab="Wage", col="blue",
#        main='Evolution of Wage over the years')

#Cointegration test
mult.mod4 <- lm(E~GDP, data=WAGE)
CADFtest(mult.mod4$residuals)
acf(mult.mod4$residuals)

max.lag<-round(sqrt(length(W.Labor.ts)))
CADFtest(diff(W.Labor.ts), type= "drift", criterion= "BIC", max.lag.y=max.lag)

max.lag<-round(sqrt(length(W.employ.ts)))
CADFtest(diff(W.employ.ts), type= "drift", criterion= "BIC", max.lag.y=max.lag)


#Distributed lag model
adf.test(WAGE$E, alternative = "stationary",k=0)
adf.test(diff(WAGE$E),alternative = "stationary",k=0)
lag=2
T=length(diff(WAGE$E))
x=diff(WAGE$GDP)[(lag+1):T]
x.1=diff(WAGE$GDP)[(lag):(T-1)]
x.2=diff(WAGE$GDP)[(lag-1):(T-2)]
DL.mod1=lm(diff(E)[-c(1,2)]~x+x.1+x.2,data=WAGE)
summary(DL.mod1)
AIC(DL.mod1)
acf(DL.mod1$residuals)
Box.test(DL.mod1$residuals)



y=diff(WAGE$E)[lag:(T-1)]
y.1=diff(WAGE$E)[lag:(T-1)]
y.2=diff(WAGE$E)[(lag-1):(T-2)]
DL.mod2=lm(diff(GDP)[-c(1,2)]~y+y.1+y.2,data = WAGE)
summary(DL.mod2)
AIC(DL.mod2)

#Autoregressive model
AR.mod1=lm(diff(E)[-c(1,2)]~y.1+y.2+x.1+x.2,data=WAGE)
summary(AR.mod1)
AIC(AR.mod1)
```

```
BIC(AR.mod1)
AR.mod2=lm(diff(E)[-c(1,2)]~y.1+y.2,data=WAGE)
summary(AR.mod2)
anova(AR.mod1,AR.mod2)

Box.test(AR.mod1$residuals)


AR.mod1.2=lm(diff(GDP)[-c(1,2)]~x.1+x.2+y.1+y.2,data=WAGE)
AIC(AR.mod1.2)
summary(AR.mod1.2)
AR.mod2.2=lm(diff(GDP)[-c(1,2)]~x.1+x.2,data = WAGE)
anova(AR.mod1.2,AR.mod2.2)


#VAR model
install.packages("vars")
library(vars)
cGDP=diff(WAGE$GDP)
cEmp=diff(WAGE$E)
new.wage=cbind(cEmp,cGDP)
VARselect(new.wage)$criteria
VARselect(new.wage)$selection
var.mod1=VAR(new.wage,p=1,type="const")
summary(var.mod1)
VARselect(new.wage)$criteria[1]
acf(resid(var.mod1))

irf_var<-irf(var.mod1,ortho=FALSE,boot=TRUE)
par(mfrow=c(2,2))
plot(irf_var)
names(irf_var)
vec_yrs=1:nrow(forcast$fcst$cGDP)
plot(forcast$fcst$cGDP~,
      main="Original time series with forcast of h=10",
      xlab="Year",ylab="Log(GDP)",type="l")
forcast=predict(var.mod1,n.ahead=5)
as.data.frame(forcast)
is.data.frame(forcast$fcst$cGDP)
is.matrix(forcast$fcst$cGDP)
dim(forcast$fcst$cGDP)
```