Final Assignment

Multivariate Time Series

By Douglas Garcia Torres

August 2017

Hans Rosling, founder of the Gapminder Organisation, discusses in a TED-talk video entitled "the magic washing machine" the impact of technological development on society and its relation with energy consumption in the world. He formulates a conjecture that energy consumption is associated to the stage of economic development of a country. In this assignment, we explore this conjecture, analyzing these indicators for the economy of Brazil:

- The Energy use per capita in tonnes of oil equivalent (toe).
- The Gross Domestic Product (GDP) per capita in constant 2000 US$.

The economy of Brazil has been growing with a relatively steady rate in the period between 1960 to 2008. Like many other South American countries, this economy has proven to be highly dependand on commodity trading and the internal political situation *(see Appendix 2)*. On the other side, the Energy use per capita of a country usually depends of several variables like technological innovations that optimize energy consumption or another external and global macro-economical factors. However, the economy of Brazil has claimed to be one of the world's leading producers of hydroelectric power and to be self-sufficient in oil needs. A fact that makes us assume some degree of independence of the energy consumption from any external factors. Nevertheless, the energy consumption tends to reflect the economic development and growth of any country.

With this in context, and without out performing a deep analysis, we dare to predict that both indicators will continue growing hand to hand with a variable rate influenced by political cycles.

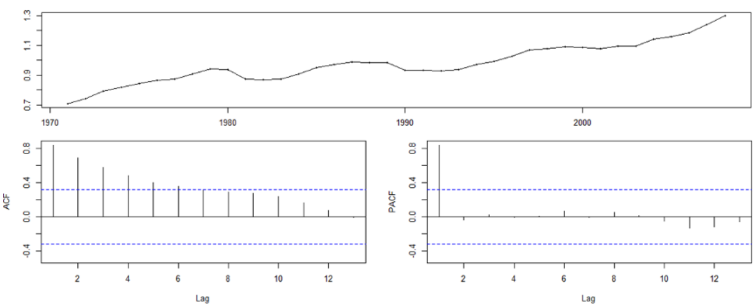
2.1. Exploratory Data Analysis

2.1.1. Brazil's energy use per capita in tonnes of oil equivalent (toe)

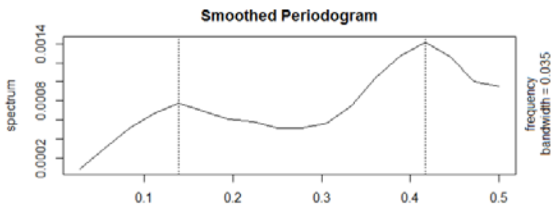
- In comparison with the GDP per capita dataset, the provided data for energy consumption has values for the period 1971-2007. In order to match both series, we updated the missing data with available values from the Gapminder database. Hence, the analyzed period is 1971-2008.

- The time-line shows a clear growing trend in the energy use per capita with three weak waves forming before strong drops that coincide with the debt payment suspensions of 1982 and 1989.

- The plotted auto-correlation function (AFC) is clearly decaying, which indicates the presence of a trend. In addition, there is no strong evidence of a seasonal pattern despite the weak cyclic behavior shown in the time-line plot.

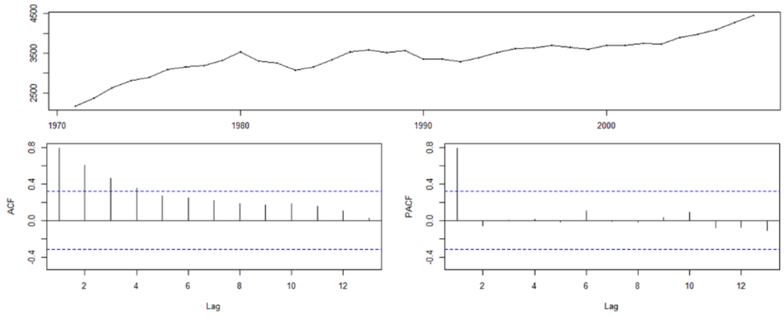


- In the frequency domain, a smoothed periodogram over the de-trended series (with two differentiations), shows two peaks at frequencies 0.416 and 0.1388 (in order of spectrum magnitude) corresponding to periods of 2.4 and 7.2 years respectively. These periods of time may coincide with the average length of political cycles.

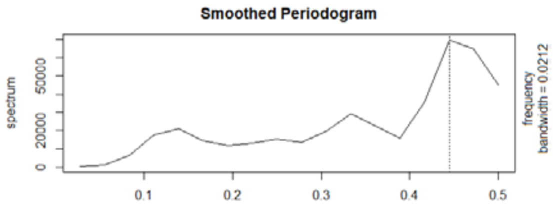


2.1.2. Brazil's Gross Domestic Product (GDP) per capita in constant 2000 US$

- The provided dataset contains values for the period 1960-2008 without any missing data. However, in order to match this time-series with the energy use series, we decided to shorten the dataset to the period 1971-2008.

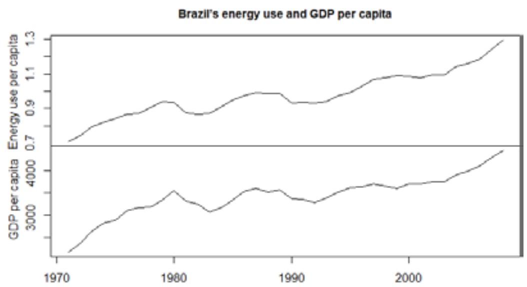


- The time-line shows a growing trend in the GDP per capita with weak cycles forming before considerable drops near to the debt payment suspensions of 1982 and 1989. Similarly to the energy series, the auto-correlation function indicates the presence of a trend and no evident cyclic pattern.



- In the frequency domain, a smoothed periodogram over the de-trended series (with two differentiations), shows one predominant peak at frequency 0.444, which corresponds to a period of 2.25 years.

2.1.3. Relation between Brazil's Gross Domestic Product (GDP) and Energy use per capita

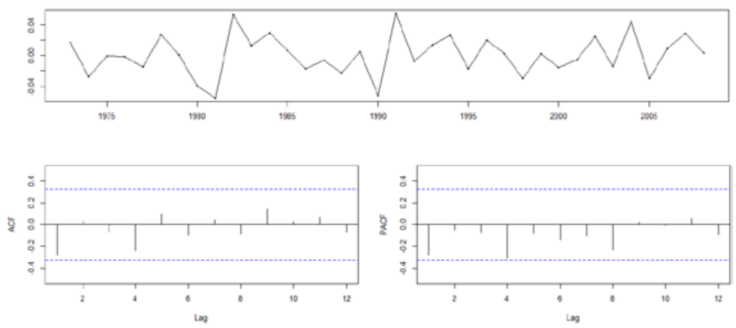


- It is clear that there is a very close relation between both indicators for the Brazilian economy. Although the Energy use time-plot is a smoother line, it shows the same patterns of the GDP per capita line.

2.2. Statistical Techniques and Models

2.2.1. An Univariate Box-Jenkins model for Brazil's energy use per capita

- Since the data is not stationary (the ACF is a slowly decreasing function) it is helpful to apply finite differencing to stabilize the mean, stabilize the variance and to remove the trend. However, one can validate that the data still might be non-stationary by applying the augmented Dicker-Fuller test and finding out that the non-stationary null hypothesis does not get rejected (with a p-value = 0.1 > 0.05). Therefore, a second degree differentiation might be necessary.



- Now the data shows to be stationary (it was be validated with the augmented Dicker-Fuller test). Hence, we chose an ARIMA(p,2,q) model without discarding to test the ARIMA(p,1,q) options since the p-value obtained in the first Dicker-Fuller test was somewhat close to the rejection of non-stationarity.

- For the ARIMA(p,1,q) variants, we can see that the correlogram and partial correlogram both tails off to zero after lag 1, leading us to consider the ARMA(0,1) and ARMA(1,0) models with one differentiation. On the other hand, for the ARIMA(p,2,q) options, the correlogram and the partial correlogram tail off to zero from the beginning, leading us to consider an ARMA(0,0) model with two differentiations. In addition, we consider some variations of an ARMA(p,q) mixed model, with p and q greater than zero, since both correlograms tail off to zero for the time series of both first and second differences. The table below show all the considered models with their corresponding Akaike's Information Criterion value.

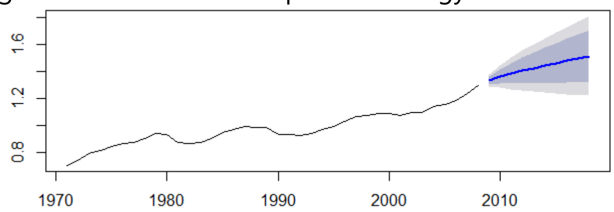
- Based on their AIC values, both the ARIMA(1,1,0) and the ARIMA(1,2,1) seem to be plausible models to fit. However, after validating and comparing both models using the root-mean-square error (RMSE) criterion (*see Appendix 1*), we think that ARIMA(1,2,1) might be a better choice. In addition, we are taking into account that a model with one order of differencing assumes that the original series has a constant average trend, while a model with two orders of differencing assumes a series with a time-varying trend, which might be more suitable for an indicator influenced by political cycles.

Model	AIC
ARIMA(1,1,0)	*-167.08*
ARIMA(0,1,1)	*-161.94*
ARIMA(0,1,0)	*-153.01*
ARIMA(1,1,1)	*-165.22*
ARIMA(2,1,1)	*-163.37*
ARIMA(1,1,2)	*-164.01*
ARIMA(2,1,2)	*-162.11*
ARIMA(0,1,2)	*-164.01*
ARIMA(2,1,0)	*-165.27*
ARIMA(0,2,0)	*-158.4*
ARIMA(1,2,0)	*-159.2*
ARIMA(0,2,1)	*-159.54*
ARIMA(1,2,1)	*-161.72*

- For the validation, we analyzed the residuals of the forecast with the chosen models, applying the following steps *(the details can be seen on Appendix 1)*:
    - Visually verify that the residuals plot look like white noise.
    - Verify the absence of auto-correlations in the residuals by visualizing the ACF and performing a Ljung-Box test.
    - Perform a Shapiro-Wilk test to verify a normal distribution on the residuals.
    - Compare the models with the RMSE measure (the smaller, the better) to determine the final model choice.

- The results of a 10-year forecast with the ARIMA(1,2,1) model predict a constant dampening grow trend. Meanwhile the ARIMA(1,1,0) model results *(in Appendix 1)* predict a higher grow rate in the consumption of energy.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2009	1.334415	1.304903	1.363927	1.289281	1.379549
2010	1.362285	1.308269	1.416301	1.279675	1.444896
2011	1.385072	1.308729	1.461415	1.268315	1.501829
2012	1.405275	1.308745	1.501804	1.257646	1.552904
2013	1.424164	1.309245	1.539084	1.248410	1.599919
2014	1.442387	1.310508	1.574265	1.240696	1.644077
2015	1.460270	1.312558	1.607981	1.234364	1.686175
2016	1.477980	1.315322	1.640638	1.229215	1.726745
2017	1.495603	1.318704	1.672502	1.225059	1.766147
2018	1.513181	1.322610	1.703753	1.221727	1.804635



- The ARIMA(1,2,1) model obtained is expressed in the formula shown below. With the autoregressive part equal to 1, the Z_t variable represents the white noise with a standard deviation of $\sqrt{0.0005171} = 0.0023$. The update levels depends on half the weight of previous values (with a coefficient equal to 0.5), and the values of white noises at time *t* and *t-1* with coefficients of 1 and -1 respectively.

$$W_t = \nabla^2 X_t$$
$$W_t = 0.5W_{t-1} + Z_t - Z_{t-1}$$

2.2.2. A Dynamic Regression model

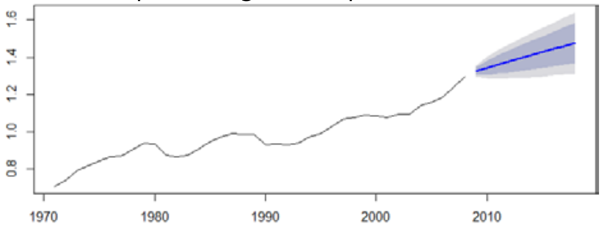
- Taking the GDP per capita of Brazil as a predictor variable for the Energy consumption, the steps executed to fit a multivariate model were as follows *(the details can be seen on Appendix 1)*:
    - Verify that the forecast variable and predictor variable are all stationary.
    - Fit a regression model with AR(2) errors (as we have non-seasonal data)
    - Calculate the errors from the fitted regression model
    - Identify an appropriate ARMA model
    - Re-fit the entire model using the new ARMA model for the errors
    - Verify that the series look like white noise

- We performed the same procedure with *d=1* (one differentiation) and *d=2 (two differentiations)*, taking into account that the tests done on the energy consumption variable slightly rejected stationarity for our models with one degree of differentiation. Hence, after fitting a regression model with AR(2) errors, we tested and identified two models with the lowest AICc value for each case: ARIMA(0,1,1) and ARIMA(0,2,2).

- Then, knowing that the forecasting of the energy consumption with this method is possible having future values of the GDP per capita series, we used the means of a GDP per capita forecast (with the automatic ARIMA function) to generate the prediction intervals of the energy consumption for the next 10 years.

- The results of both models were quite similar in terms of accuracy, with the root mean square error (RMSE) equal to *0.015* in both cases, and with the ARIMA(0,2,2) model predicting more optimistic values:

```
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2009       1.326148 1.305315 1.346981 1.294287 1.358009
2010       1.344129 1.306379 1.381878 1.286396 1.401862
2011       1.361166 1.311466 1.410866 1.285157 1.437175
2012       1.377925 1.318181 1.437670 1.286554 1.469297
2013       1.394603 1.325876 1.463330 1.289494 1.499712
2014       1.411257 1.334234 1.488279 1.293461 1.529052
2015       1.427903 1.343069 1.512737 1.298161 1.557646
2016       1.444548 1.352261 1.536835 1.303407 1.585689
2017       1.461192 1.361728 1.560655 1.309075 1.613308
2018       1.477835 1.371414 1.584256 1.315079 1.640592
```



Model	AICc
ARIMA(0,1,0)	-190,9
ARIMA(0,1,1)	-194,68
ARIMA(1,1,0)	-193,09
ARIMA(1,1,1)	-192,18
ARIMA(1,1,2)	-189,5
ARIMA(2,1,0)	-192,55
ARIMA(2,1,1)	-189,86
ARIMA(2,1,2)	-189,34
ARIMA(0,2,0)	-179,23
ARIMA(0,2,1)	-182,44
ARIMA(0,2,2)	-186,15
ARIMA(1,2,0)	-177,47
ARIMA(1,2,1)	-184,23
ARIMA(1,2,2)	-183,45
ARIMA(2,2,0)	-178,3
ARIMA(2,2,1)	-184,18
ARIMA(2,2,2)	-184,67

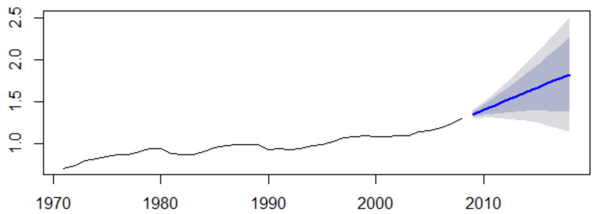
- The ARIMA(0,2,2) model obtained is expressed in the formula shown below where Z_t correspond to the white noise with a standard deviation of $\sqrt{0.00026} = 0.016$.

$$\boldsymbol{Y_t = 0.0004X_t + N_t}$$
$$\boldsymbol{N_t = -0.5N_{t-1} - 0.4N_{t-2} + Z_t}$$

2.2.3. An Exponential Smoothing model: Holt Exponential Smoothing

Considering that the series contains a clear trend and weak seasonality patterns, the suitable model to fit would be a Holts Exponential Smoothing model. Therefore, we made use of the HoltWinters R function (with the Gamma parameter equals to FALSE). In the output, we got an estimated value of Alpha equals to 1 and 0.6 for Beta, forecasting the following:

```
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2009       1.349293 1.316912 1.381675 1.299770 1.398817
2010       1.402046 1.339536 1.464555 1.306446 1.497645
2011       1.454798 1.357506 1.552090 1.306002 1.603593
2012       1.507550 1.371121 1.643979 1.298899 1.716201
2013       1.560302 1.380752 1.739852 1.285704 1.834900
2014       1.613054 1.386712 1.839397 1.266893 1.959215
2015       1.665807 1.389256 1.942357 1.242859 2.088754
2016       1.718559 1.388596 2.048522 1.213924 2.223193
2017       1.771311 1.384910 2.157712 1.180361 2.362261
2018       1.824063 1.378349 2.269777 1.142403 2.505724
```



2.3. Results and Conclusions

The table below show the forecasting of the chosen models for each one of the used techniques. The *RMSE* values correspond to the measures for goodness-of-fit given by the accuracy function, suggesting that the *Dynamic Regression* model could be the more adequate to use. However, the *RMSE (Real)* measure which are the root mean square errors against the real values taken from the World Bank updated dataset, show that the more accurate forecast was generated by the *Box Jenkins ARIMA (1,2,1)* model.

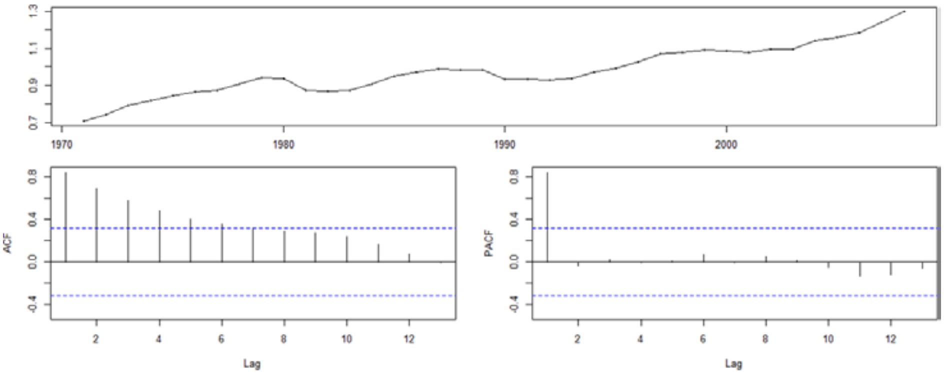
	2008	2009	2010	2011	2012	2013	2014	RMSE	RMSE (Real)
Real Values	1,28812	1,23375	1,35103	1,35912	1,40476	1,45117	1,48493		
Box Jenkins ARIMA(1,2,1)	1,28812	1,33442	1,36229	1,38507	1,40528	1,42416	1,44239	**0,02213297**	**0,116068476**
Dyn.Regression ARIMA(0,2,2)	1,28812	1,32615	1,34413	1,36117	1,37793	1,3946	1,41126	**0,01503537**	**0,133922217**
Exponential Smoothing	1,28812	1,34929	1,40205	1,4548	1,50755	1,5603	1,61305	**0,02492531**	**0,25298136**

One can notice that none of the models could predict the consumption depression of 2009, probably consequence of the financial crisis of 2008. However, putting aside the possible black swans of real life, after finishing this analysis we realize the importance of choosing models that can handle cyclic patterns despite of how weak they are. In addition, we also noticed the advantages of multivariate time series as methods to get more narrow predictions intervals.

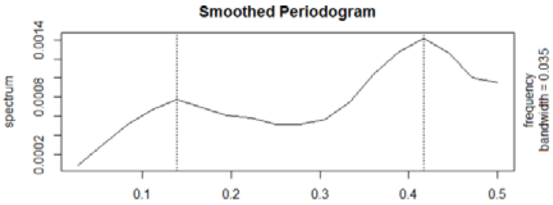
Appendix 1

This section contains the R scripts used for the analyses with the relevant statistical output (for part 2).

```
> #Exploratory Analysis on Brazil's energy use per capita
> brazil <- group15_data_2
> brazil$enerbrazil[as.Date(brazil$date,'%Y') == as.Date("2008",'%Y')] <- 1.29654123
> enerbrazil <- brazil$enerbrazil
> enerbrazil <- na.omit(enerbrazil)
> enerbrazil.ts <- ts(enerbrazil, start=c(1971))
> tsdisplay(enerbrazil.ts)
```

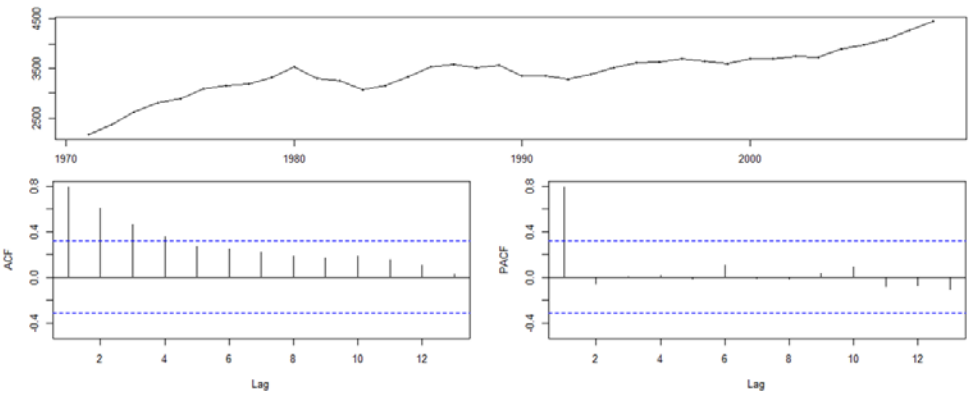



```
> enerbrazil.d2 <- diff(enerbrazil.ts,differences=2) #de-trending for spectral analysis
> enerbrazil.spec <- spectrum(enerbrazil.d2,log="no",span=5) #smoothed periodogram
> enerbrazil.spec
```



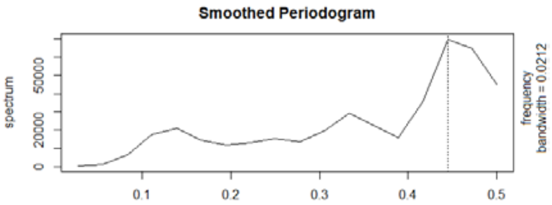

```
$freq
 [1] 0.02777778 0.05555556 0.08333333 0.11111111 0.13888889 0.16666667 0.19444444 0.22222222 0.25000000 0.27777778
[11] 0.30555556 0.33333333 0.36111111 0.38888889 0.41666667 0.44444444 0.47222222 0.50000000

$spec
 [1] 8.527873e-05 3.095303e-04 5.140333e-04 6.645010e-04 7.755261e-04 6.891439e-04 6.080707e-04 5.810537e-04 5.157841e-04
[10] 5.155813e-04 5.685499e-04 7.522996e-04 1.043811e-03 1.276194e-03 1.419132e-03 1.261666e-03 1.006788e-03 9.560744e-04
```

```
> #Exploratory Analysis on Brazil's GDP per capita
> gdpbrazilcomplete <- brazil$gdpbrazil
> gdpbrazil <- na.omit(gdpbrazilcomplete)
> gdpbrazil <- gdpbrazil[-(1:11)]
> gdpbrazil.ts <- ts(gdpbrazil, start=c(1971))
> tsdisplay(gdpbrazil.ts)
```




```
> gdpbrazil.d2 <- diff(gdpbrazil.ts,differences=2) #de-trending for spectral analysis
```

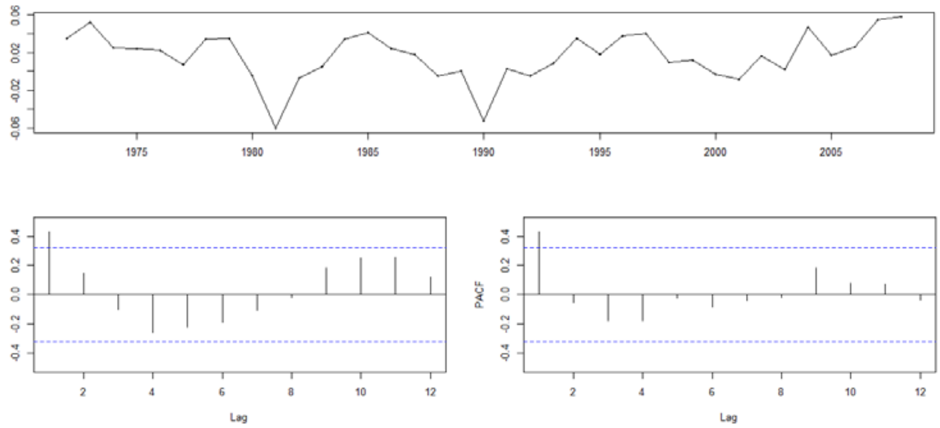
```
> gdpbrazil.spec <- spectrum(gdpbrazil.d2,log="no",span=3) #smoothed periodogram
> gdpbrazil.spec
```



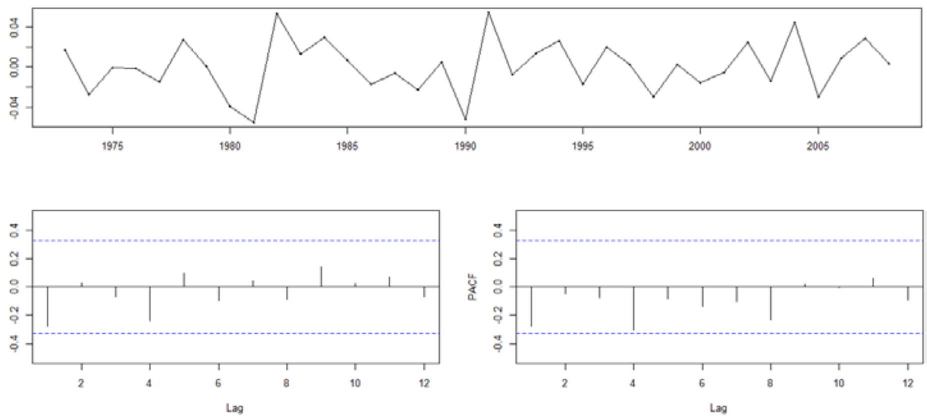

```
$freq
 [1] 0.02777778 0.05555556 0.08333333 0.11111111 0.13888889 0.16666667 0.19444444 0.22222222 0.25000000 0.27777778
[11] 0.30555556 0.33333333 0.36111111 0.38888889 0.41666667 0.44444444 0.47222222 0.50000000

$spec
 [1]   227.5336  1018.1113  6432.3101 17495.9652 20781.4898 14492.4643 11811.6505 13167.9148 15292.9373 13501.8388
[11] 19532.5853 29497.7112 22657.8825 15565.7987 35730.6505 69947.2292 64978.3447 45367.5066
```

> #An Univariate Box-Jenkins model for Brazil’s energy use per capita

```
> enerbrazil.d1 <- diff(enerbrazil.ts,differences=1)
> tsdisplay(enerbrazil.d1)
```




```
> adf.test(enerbrazil.d1) #Dicker-Fuller test for 1st differenciation

        Augmented Dickey-Fuller Test

data:  enerbrazil.d1
Dickey-Fuller = -3.2104, Lag order = 3, p-value = 0.1013
alternative hypothesis: stationary

> enerbrazil.d2 <- diff(enerbrazil.ts,differences=2)
> tsdisplay(enerbrazil.d2)
```

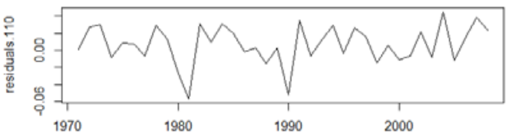



```
> adf.test(enerbrazil.d2) #Dicker-Fuller test for 2nd differenciation

        Augmented Dickey-Fuller Test

data:  enerbrazil.d2
Dickey-Fuller = -4.2752, Lag order = 3, p-value = 0.01048
alternative hypothesis: stationary
```

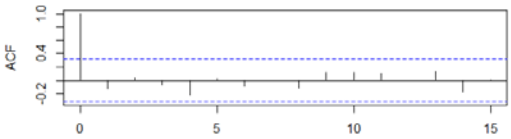
> #Model determination according to AIC values

```
> enerbrazil.arima.110 <- arima(enerbrazil.ts, order=c(1,1,0))
```

```
> enerbrazil.arima.011 <- arima(enerbrazil.ts, order=c(0,1,1))
> enerbrazil.arima.010 <- arima(enerbrazil.ts, order=c(0,1,0))
> enerbrazil.arima.111 <- arima(enerbrazil.ts, order=c(1,1,1))
> enerbrazil.arima.211 <- arima(enerbrazil.ts, order=c(2,1,1))
> enerbrazil.arima.112 <- arima(enerbrazil.ts, order=c(1,1,2))
> enerbrazil.arima.212 <- arima(enerbrazil.ts, order=c(2,1,2))
> enerbrazil.arima.012 <- arima(enerbrazil.ts, order=c(0,1,2))
> enerbrazil.arima.210 <- arima(enerbrazil.ts, order=c(2,1,0))
> enerbrazil.arima.020 <- arima(enerbrazil.ts, order=c(0,2,0))
> enerbrazil.arima.120 <- arima(enerbrazil.ts, order=c(1,2,0))
> enerbrazil.arima.021 <- arima(enerbrazil.ts, order=c(0,2,1))
> enerbrazil.arima.121 <- arima(enerbrazil.ts, order=c(1,2,1))

> #Forecasts with ARIMA(1,1,0) and ARIMA(1,2,1)
> enerbrazil.arima.110.forecast <- forecast.Arima(enerbrazil.arima.110, h=10)
> enerbrazil.arima.121.forecast <- forecast.Arima(enerbrazil.arima.121, h=10)

> #Validation of model: ARIMA(1,1,0)
> residuals.110 <- enerbrazil.arima.110.forecast$residuals
> plot.ts(residuals.110) #verify that the residuals look like white noise
```

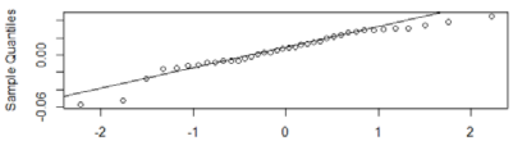



```
> acf(residuals.110) #correlogram of the forecast errors
```



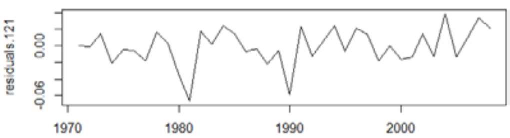

```
> Box.test(residuals.110, type="Ljung-Box") #test to verify the absence of auto-correlations

        Box-Ljung test

data:  residuals.110
X-squared = 0.58418, df = 1, p-value = 0.4447
```

```
> qqnorm(residuals.110); qqline(residuals.110) #QQ-Plot to verify normality of the residuals
```

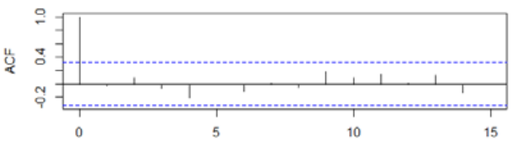



```
> shapiro.test(residuals.110) #test to verify normality of the residuals

        Shapiro-Wilk normality test

data:  residuals.110
W = 0.94579, p-value = 0.06475
```

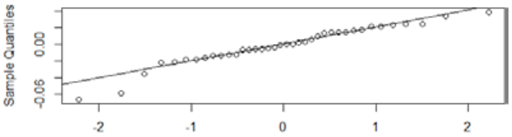
```
> #Validation of model: ARIMA(1,2,1)
> residuals.121 <- enerbrazil.arima.121.forecast$residuals
> plot.ts(residuals.121) #verify that the residuals look like white noise
```




```
> acf(residuals.121) #correlogram of the forecast errors
```



```
> Box.test(residuals.121, type="Ljung-Box") #test to verify the absence of auto-correlations

        Box-Ljung test

data:  residuals.121
X-squared = 0.027289, df = 1, p-value = 0.8688

> qqnorm(residuals.121); qqline(residuals.121) #QQ-Plot to verify normality of the residuals
```



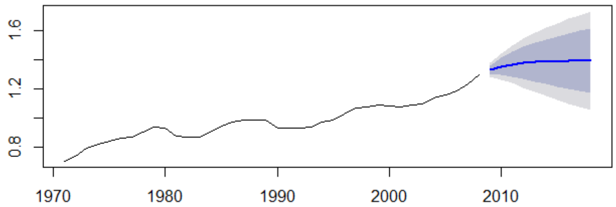

```
> shapiro.test(residuals.121) #test to verify normality of the residuals

        Shapiro-Wilk normality test

data:  residuals.121
W = 0.94378, p-value = 0.05564
```

> #Compare ARIMA(1,1,0) with ARIMA(1,2,1) based on the RMSE (the smaller, the better)

```
> accuracy(enerbrazil.arima.110)

                      ME       RMSE        MAE       MPE     MAPE      MASE       ACF1
Training set 0.006545409 0.02349731 0.01881268 0.6574957 1.950777 0.7872481 -0.1192482

> accuracy(enerbrazil.arima.121)

                       ME       RMSE        MAE        MPE     MAPE      MASE        ACF1
Training set -0.001160853 0.02213297 0.01682172 -0.1862645 1.721094 0.7039328 -0.02577355
```

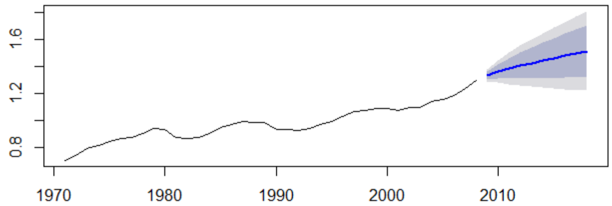
> #Forecast with ARIMA(1,1,0)

```
> plot.forecast(enerbrazil.arima.110.forecast)
> enerbrazil.arima.110.forecast

     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2009       1.332627 1.302110 1.363144 1.285956 1.379299
2010       1.355253 1.296974 1.413532 1.266123 1.444383
2011       1.369438 1.284605 1.454272 1.239696 1.499181
2012       1.378333 1.268875 1.487791 1.210931 1.545734
2013       1.383909 1.251849 1.515970 1.181940 1.585878
2014       1.387406 1.234617 1.540195 1.153735 1.621076
2015       1.389598 1.217733 1.561462 1.126754 1.652442
2016       1.390972 1.201461 1.580483 1.101140 1.680805
2017       1.391834 1.185900 1.597767 1.076886 1.706782
2018       1.392374 1.171068 1.613681 1.053915 1.730833
```



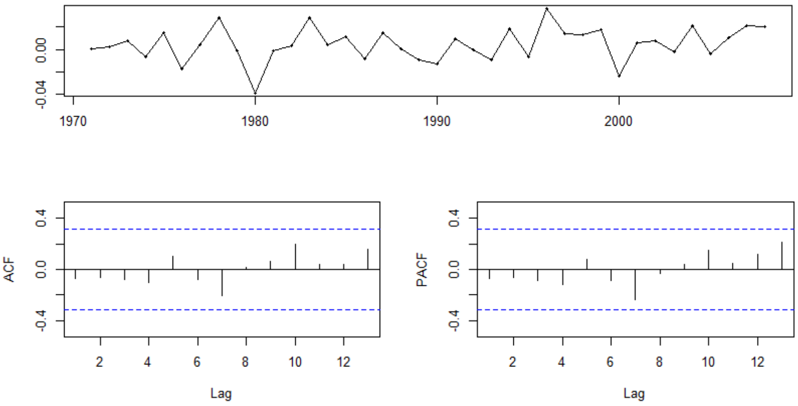
> #Forecast with ARIMA(1,2,1)

```
> enerbrazil.arima.121.forecast
> plot.forecast(enerbrazil.arima.121.forecast)

     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2009       1.334415 1.304903 1.363927 1.289281 1.379549
2010       1.362285 1.308269 1.416301 1.279675 1.444896
2011       1.385072 1.308729 1.461415 1.268315 1.501829
2012       1.405275 1.308745 1.501804 1.257646 1.552904
2013       1.424164 1.309245 1.539084 1.248410 1.599919
2014       1.442387 1.310508 1.574265 1.240696 1.644077
2015       1.460270 1.312558 1.607981 1.234364 1.686175
2016       1.477980 1.315322 1.640638 1.229215 1.726745
2017       1.495603 1.318704 1.672502 1.225059 1.766147
2018       1.513181 1.322610 1.703753 1.221727 1.804635
```



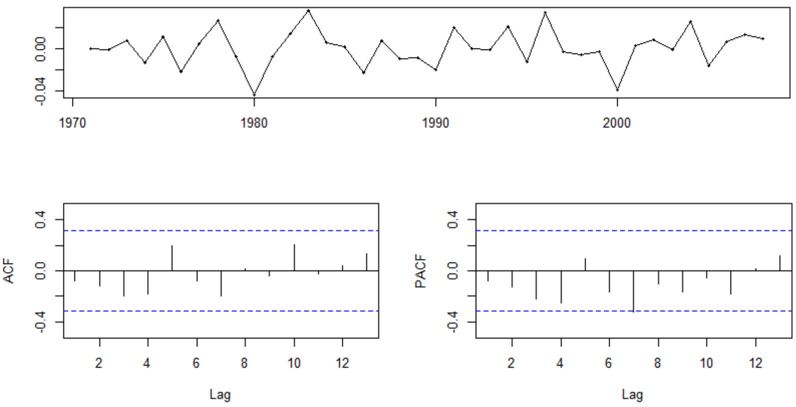
> #Dynamic Regression Model: fit the regression model with AR(2) errors

```
> enerbrazil.proxy.210 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(2,1,0))
> enerbrazil.proxy.210
> tsdisplay(residuals(enerbrazil.proxy.210))

Coefficients:
         ar1      ar2  gdpbrazil.ts
      0.4565  -0.2371         2e-04
s.e.  0.0532   0.1593           NaN

sigma^2 estimated as 0.0002707:  log likelihood=100.9
AIC=-193.8   AICc=-192.55   BIC=-187.35
```




```
>#Check model with 2 differenciations
> enerbrazil.proxy.220 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(2,2,0))
> enerbrazil.proxy.220
> tsdisplay(residuals(enerbrazil.proxy.220))
```

```
Coefficients:
          ar1      ar2  gdpbrazil.ts
      -0.1751  -0.3023         1e-04
s.e.   0.1810   0.1992         2e-04

sigma^2 estimated as 0.0003465:  log likelihood=93.8
AIC=-179.6   AICc=-178.3   BIC=-173.26
```

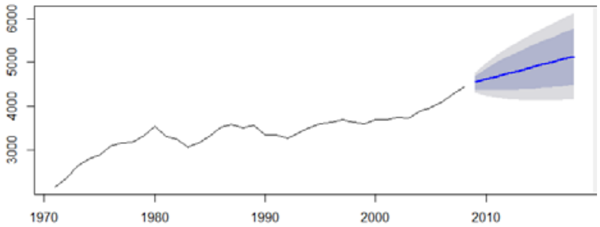



```
> #Check other model candidates
> enerbrazil.reg.010 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(0,1,0))
> enerbrazil.reg.011 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(0,1,1))
> enerbrazil.reg.110 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(1,1,0))
> enerbrazil.reg.111 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(1,1,1))
> enerbrazil.reg.112 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(1,1,2))
> enerbrazil.reg.210 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(2,1,0))
> enerbrazil.reg.211 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(2,1,1))
> enerbrazil.reg.212 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(2,1,2))
> enerbrazil.reg.020 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(0,2,0))
> enerbrazil.reg.021 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(0,2,1))
> enerbrazil.reg.022 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(0,2,2))
> enerbrazil.reg.120 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(1,2,0))
> enerbrazil.reg.121 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(1,2,1))
> enerbrazil.reg.122 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(1,2,2))
> enerbrazil.reg.220 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(2,2,0))
> enerbrazil.reg.221 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(2,2,1))
> enerbrazil.reg.222 <- Arima(enerbrazil.ts, xreg=gdpbrazil.ts, order=c(2,2,2))

> #Verify the automatic choice for the dynamic regression: ARIMA(0,1,1)
> enerbrazil.auto.reg <- auto.arima(enerbrazil.ts,xreg=gdpbrazil.ts)
> enerbrazil.auto.reg
```

```
Regression with ARIMA(0,1,1) errors

Coefficients:
         ma1   drift  gdpbrazil.ts
      0.4662  0.0073         1e-04
s.e.  0.1582  0.0011           NaN

sigma^2 estimated as 0.0002496:  log likelihood=102.41
AIC=-196.82   AICc=-195.57   BIC=-190.37
```

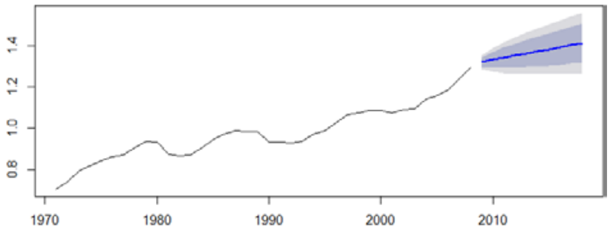
```
> #Check that there are no autocorrelation in the residuals for the chosen models
> Box.test(residuals(enerbrazil.reg.011), type="Ljung-Box")

        Box-Ljung test

data:  residuals(enerbrazil.reg.011)
X-squared = 0.10803, df = 1, p-value = 0.7424

> Box.test(residuals(enerbrazil.reg.022), type="Ljung-Box")

        Box-Ljung test

data:  residuals(enerbrazil.reg.022)
X-squared = 0.035756, df = 1, p-value = 0.85

> #Forecast future values for GDP per capita using auto arima
> gdpbrazil.auto <- auto.arima(gdpbrazil.ts)
> gdpbrazil.auto
> gdpbrazil.auto.fore <- forecast(gdpbrazil.auto, h=10)
> gdpbrazil.auto.fore
> plot(gdpbrazil.auto.fore)
```

```
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2009       4544.281 4394.969 4693.593 4315.928 4772.634
2010       4618.291 4374.068 4862.514 4244.785 4991.797
2011       4685.632 4366.002 5005.262 4196.800 5174.464
2012       4751.010 4368.561 5133.458 4166.105 5335.914
2013       4815.810 4378.954 5252.665 4147.697 5483.922
2014       4880.439 4395.097 5365.782 4138.172 5622.707
2015       4945.019 4415.572 5474.466 4135.300 5754.738
2016       5009.584 4439.424 5579.744 4137.600 5881.568
2017       5074.145 4465.988 5682.301 4144.050 6004.239
2018       5138.704 4494.789 5782.619 4153.921 6123.487
```



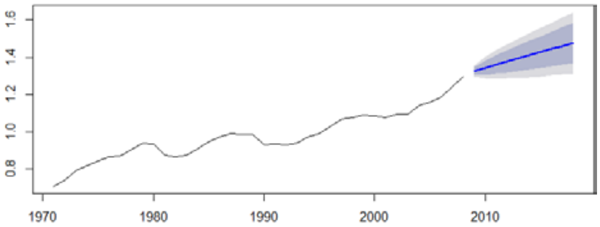

```
> #Forecast with ARIMA(0,1,1) errors
> enerbrazil.011.fore <- forecast(enerbrazil.reg.011,xreg=gdpbrazil.auto.fore$mean, h=10)
> plot(enerbrazil.011.fore, main="Forecasts from regression with ARIMA(0,1,1) errors")
> enerbrazil.011.fore
```

```
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2009       1.321583 1.300690 1.342476 1.289630 1.353536
2010       1.333014 1.295707 1.370320 1.275958 1.390069
2011       1.343414 1.294968 1.391861 1.269322 1.417507
2012       1.353512 1.296046 1.410978 1.265625 1.441398
2013       1.363520 1.298270 1.428770 1.263729 1.463311
2014       1.373502 1.301302 1.445702 1.263082 1.483922
2015       1.383476 1.304939 1.462014 1.263364 1.503589
2016       1.393448 1.309048 1.477848 1.264369 1.522527
2017       1.403419 1.313538 1.493301 1.265958 1.540881
2018       1.413391 1.318344 1.508437 1.268029 1.558752
```




```
> #Forecast with ARIMA(0,2,2) errors
> enerbrazil.022.fore <- forecast(enerbrazil.reg.022,xreg=gdpbrazil.auto.fore$mean, h=10)
> plot(enerbrazil.022.fore, main="Forecasts from regression with ARIMA(0,2,2) errors")
> enerbrazil.022.fore
```

```
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2009       1.326148 1.305315 1.346981 1.294287 1.358009
2010       1.344129 1.306379 1.381878 1.286396 1.401862
2011       1.361166 1.311466 1.410866 1.285157 1.437175
2012       1.377925 1.318181 1.437670 1.286554 1.469297
2013       1.394603 1.325876 1.463330 1.289494 1.499712
2014       1.411257 1.334234 1.488279 1.293461 1.529052
2015       1.427903 1.343069 1.512737 1.298161 1.557646
2016       1.444548 1.352261 1.536835 1.303407 1.585689
2017       1.461192 1.361728 1.560655 1.309075 1.613308
2018       1.477835 1.371414 1.584256 1.315079 1.640592
```




```
> #Compare ARIMA(0,1,1) with ARIMA(0,2,2) based on the RMSE (the smaller, the better)
> summary(enerbrazil.011.fore)
```

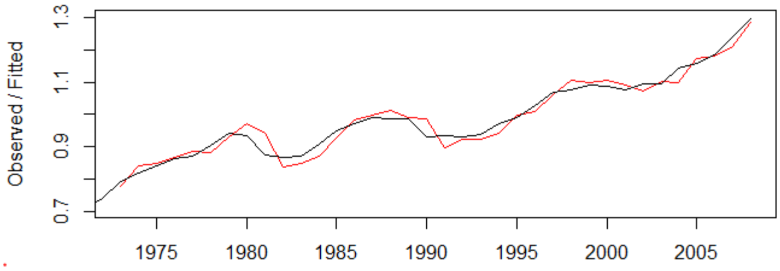
```
Coefficients:
         ma1  gdpbrazil.ts
      0.4793         2e-04
s.e.     NaN           NaN

sigma^2 estimated as 0.0002658:  log likelihood=100.7
AIC=-195.41   AICc=-194.68   BIC=-190.58

Error measures:
                      ME       RMSE        MAE       MPE     MAPE      MASE        ACF1
Training set 0.004383804 0.01564612 0.01216362 0.4202208 1.235265 0.5090072 -0.05128012
```

```
> summary(enerbrazil.022.fore)
```

```
Coefficients:
          ma1      ma2  gdpbrazil.ts
      -0.5034  -0.4768         1e-04
s.e.   0.2798      NaN           NaN

sigma^2 estimated as 0.0002603:  log likelihood=97.72
AIC=-187.44   AICc=-186.15   BIC=-181.1

Error measures:
                      ME       RMSE        MAE       MPE     MAPE      MASE        ACF1
Training set 0.001674025 0.01503537 0.01180951 0.1413726 1.202466 0.4941887 -0.02950209
```

```
#Holt Exponential Smoothing
> enerbrazil.es <- HoltWinters(enerbrazil.ts,gamma=FALSE)
> enerbrazil.es
> plot(enerbrazil.es)
```

```
Holt-Winters exponential smoothing with trend and without seasonal component.

Call:
HoltWinters(x = enerbrazil.ts, gamma = FALSE)

Smoothing parameters:
 alpha: 1
 beta : 0.6511855
 gamma: FALSE

Coefficients:
        [,1]
a 1.29654123
b 0.05275218
```

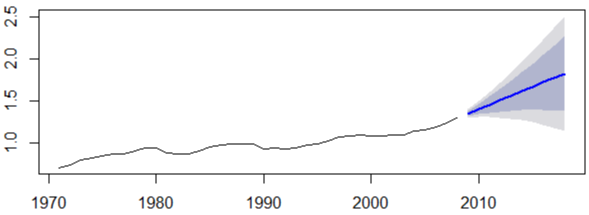



```
#Holt Exponential Smoothing accuracy
> with(enerbrazil.es,accuracy(fitted,x))
```

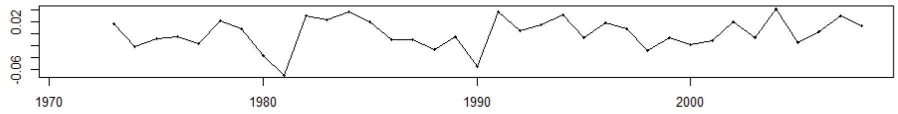
```
                   ME       RMSE       MAE        MPE     MAPE       ACF1 Theil's U
Test set 0.0007471921 0.02492531 0.0202881 0.04087523 2.101678 0.03770895 0.9062354
```

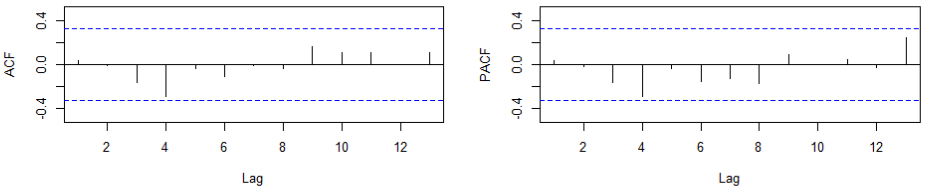
```
#Holt Winters forecast
enerbrazil.es.fore <- forecast.HoltWinters(enerbrazil.es,h=10)
plot.forecast(enerbrazil.es.fore)
```

```
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2009       1.349293 1.316912 1.381675 1.299770 1.398817
2010       1.402046 1.339536 1.464555 1.306446 1.497645
2011       1.454798 1.357506 1.552090 1.306002 1.603593
2012       1.507550 1.371121 1.643979 1.298899 1.716201
2013       1.560302 1.380752 1.739852 1.285704 1.834900
2014       1.613054 1.386712 1.839397 1.266893 1.959215
2015       1.665807 1.389256 1.942357 1.242859 2.088754
2016       1.718559 1.388596 2.048522 1.213924 2.223193
2017       1.771311 1.384910 2.157712 1.180361 2.362261
2018       1.824063 1.378349 2.269777 1.142403 2.505724
```




```
#Holt Winters forecast: diagnostic
tsdisplay(enerbrazil.es.fore$residuals)
```






```
Box.test(enerbrazil.es.fore$residuals,lag=20,type="Ljung-Box")
```

```
        Box-Ljung test

data:  enerbrazil.es.fore$residuals
X-squared = 11.833, df = 20, p-value = 0.9217
```

```
shapiro.test(enerbrazil.es.fore$residuals)
```

```
        Shapiro-Wilk normality test

data:  enerbrazil.es.fore$residuals
W = 0.9637, p-value = 0.2788
```

Appendix 2

The fluctuations in the GDP of the Brazilian economy can be explained by the following political and economical events:

- **1960**: the country was gaining some political stability and economic growth with president Juscelino Kubitschek after a volatile period where the former president commited suicide.
- **1964:** the democratic government is ousted and military start ruling the country with repression but also with rapid economic growth based on state-ownership of key sectors.
- **1974:** a new military president introduces reforms allowing political activity and elections.
- **1982**: the country halts payment of its main foreign debt (which was among the world's biggest in the moment).
- **1985**: during an economic crisis, the first civilian president is elected after 21 years, and introduces economic policies to freeze prices, wages and the local currency exchange rate in an effort to control inflation. However, the plan was a failure and the inflation rate explodes in the next years when the freeze is lifted.
- **1989**: Fernando Collor de Mello the first directly elected president since 1960 is elected by the people, immediately introducing radical reforms promising economic improvement which fails to materialize. However, inflation remains out of control and the foreign debt payments are suspended again. In 1992, Collor de Mello resigns after being accused of corruption.
- **1994**: Fernando Henrique Cardozo is elected president after helping to bring inflation under control. He makes controversial moves on land issue, seizing land for distribution among poor and indigenous claims.
- **1998**: the IMF provides a rescue package after the economy was hit by the collapse of Asian stock markets.
- **2002**: the local currency, the Real, is stronger than ever. For the first time in more than 40 years, a left-wing populist movement wins the presidencial elections. The new president, Luiz Inacio Lula da Silva, promises popular political and economical reforms.
- **2004**: Brazil launches its first space rocket.
- **2005**: Voters in a referendum reject a proposal to ban the sale of firearms.
- **2006**: Thanks to an ambitious program to reduce the dependence on imported oil, the country became self-sufficient in oil consumption.
- **2006**: Luiz Inacio Lula da Silva is re-elected.
- **2008**: The EU halts all imports of Brazilian beef, arguing that its foot-and-mouth diseases checks are unacceptable.
- **2008:** Brazil turns down an invitation from Iran to join the international Oil cartel, OPEC.

Source: http://www.bbc.com/news/world-latin-america-19359111