Methodology Seminar

The functionalities of the Gretl software

Hassan OUKHOUYA

Instructor-Researcher: M^r.Khalid EL HIMDI
LMSA Lab., FSR, Mohammed V University, Rabat

April 6, 2021

Roadmap

1. What is GRETL?
   - Presentation of the interface
2. Simple Linear Regression in Gretl
   - Importing Data
   - Data Analysis
   - Analysis of Residuals
   - Forecasting
3. Modeling and Forecasting Time Series in Gretl
   - Time series analysis
   - Model selection
   - Diagnostics
   - Forecasting
4. Case Study
5. Interesting Resources

Presentation

THE aim of this presentation is to propose an initiation, in terms applied to econometrics of time series or panel data, using a Gretl econometrics software with illustrative examples.

What is GRETL?

GRETL as per Wikipedia

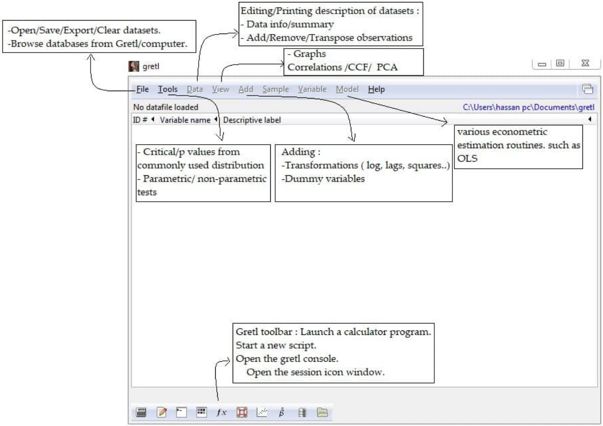
- Gretl is an open-source statistical package, mainly for econometrics.
- The name is an acronym for Gnu Regression, Econometrics and Time-series Library.
- It is written in C, uses GTK+ as widget toolkit for creating its GUI, and calls gnuplot for generating graphs, and can be integrated with other software such as R, Stat, Python...etc.

More on http://gretl.sourceforge.net/

The main window menus

We find:

- File.
- Tools.
- Data.
- View.
- Add.
- Sample.
- Variable.
- Model.
- Help.



-Open/Save/Export/Clear datasets.
-Browse databases from Gretl/computer.

Editing/Printing description of datasets :
- Data info/summary
- Add/Remove/Transpose observations

- Graphs
Correlations /CCF/ PCA

- Critical/p values from commonly used distribution
- Parametric/ non-parametric tests

Adding :
-Transformations (log, lags, squares..)
-Dummy variables

various econometric estimation routines. such as OLS

Gretl toolbar : Launch a calculator program.
Start a new script.
Open the gretl console.
Open the session icon window.

Figure: Interface of gretl.

Simple linear regression Model

$$y = \beta_0 + \beta_1 x + \epsilon$$

- ⚀ y is the dependent variable.
- ⚁ x is the independent variable.
- ⚂ β_0 is the constant or intercept.
- ⚃ β_1 is x's slope or coefficient.
- ⚄ ϵ is the error term.

⚠ Often we suppose that ϵ follows the normal distribution $\mathcal{N}(0, \sigma^2)$.

Two Main Objectives

1. Establish if there is a **relationship** between two variables.
    - More specifically, establish if there is a statistically significant relationship between the two.
    - Examples: Income and spending, wage and gender, student height and exam scores.
2. **Forecast** new observations.
    - Can we use what we know about the relationship to forecast unobserved values?
    - Examples: What will be the sales over the next quarter? What will the ROI (Return On Investment) of a new store opening be contingent on store attributes?

Linear Regression Example

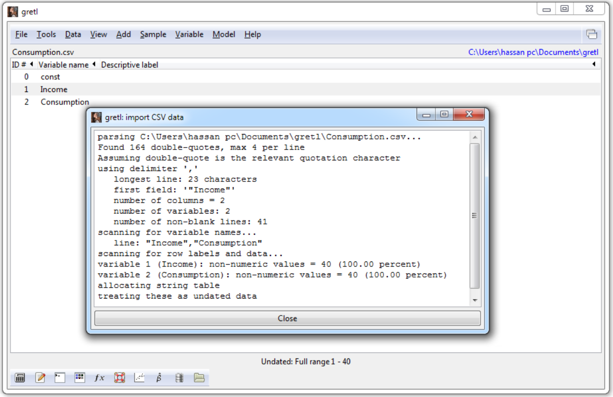
What could explain a family's consumption of a given product?

Table: Data for Linear Regression Example

ID	Income	Consumption
1	119	154
2	85	123
3	97	125
4	95	130
5	120	151
6	92	131
7	105	141
8	110	141
9	98	130
10	98	134
11	81	115
12	81	117
13	91	123
14	105	144
15	100	137
16	107	140
⋮	⋮	⋮
38	96	131
39	82	127
40	114	150

Open and Examine Data in Gretl

In the Gretl program this can be done via the ”File, Open Data, User file”:


gretl
File Tools Data View Add Sample Variable Model Help
Consumption.csv
C:\Users\hassan pc\Documents\gretl
ID # Variable name Descriptive label
0 const
1 Income
2 Consumption
gretl: import CSV data
Close
Undated: Full range 1 - 40

```
parsing C:\Users\hassan pc\Documents\gretl\Consumption.csv...
Found 164 double-quotes, max 4 per line
Assuming double-quote is the relevant quotation character
using delimiter ','
   longest line: 23 characters
   first field: '"Income"'
   number of columns = 2
   number of variables: 2
   number of non-blank lines: 41
scanning for variable names...
   line: "Income","Consumption"
scanning for row labels and data...
variable 1 (Income): non-numeric values = 40 (100.00 percent)
variable 2 (Consumption): non-numeric values = 40 (100.00 percent)
allocating string table
treating these as undated data
```

The supported import formats are as follows:

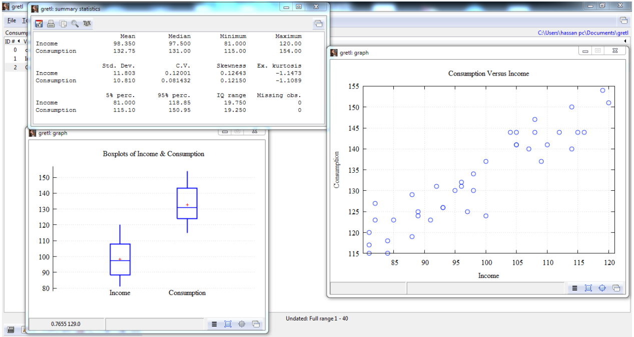
- Stata files(.dta).
- SPSS files(.sav).
- JMulTi files.
- CSV files(.csv).
- Excel fil.(.xlsx).
- ASCII fil.(.txt).

Summary statistics and Graph X-Y

Shows a full set of descriptive statistics and scatter plot for the variables selected in the main window:

You can click on the graph window for a pop-up menu with the following options.

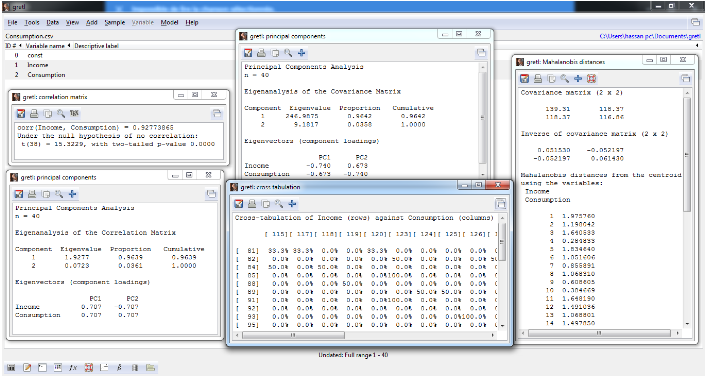
- Save as PNG.
- Zoom.
- Print.
- Edit.
- Close.



```
                     Mean         Median        Minimum        Maximum
Income             98.350         97.500         81.000         120.00
Consumption        132.75         131.00         115.00         154.00

                Std. Dev.           C.V.       Skewness   Ex. kurtosis
Income             11.803        0.12001        0.12643        -1.1473
Consumption        10.810       0.081432        0.12150        -1.1089

                 5% perc.      95% perc.       IQ range   Missing obs.
Income             81.000         118.85         19.750              0
Consumption        115.10         150.95         19.250              0
```

Principal Components

We can go further and produces a Principal Components Analysis.


gretl
File Tools Data View Add Sample Variable Model Help
Consumption.csv
C:\Users\hassan pc\Documents\gretl
gretl: correlation matrix
gretl: principal components
gretl: cross tabulation
gretl: Mahalanobis distances
Undated: Full range 1 - 40

```
ID #  Variable name  Descriptive label
  0   const
  1   Income
  2   Consumption
```

```
corr(Income, Consumption) = 0.92773865
Under the null hypothesis of no correlation:
 t(38) = 15.3229, with two-tailed p-value 0.0000
```

```
Principal Components Analysis
n = 40

Eigenanalysis of the Correlation Matrix

Component  Eigenvalue  Proportion   Cumulative
    1        1.9277      0.9639       0.9639
    2        0.0723      0.0361       1.0000

Eigenvectors (component loadings)

                 PC1      PC2
Income          0.707   -0.707
Consumption     0.707    0.707
```

```
Principal Components Analysis
n = 40

Eigenanalysis of the Covariance Matrix

Component  Eigenvalue  Proportion   Cumulative
    1      246.9875      0.9642       0.9642
    2        9.1817      0.0358       1.0000

Eigenvectors (component loadings)

                 PC1      PC2
Income         -0.740    0.673
Consumption    -0.673   -0.740
```

```
Cross-tabulation of Income (rows) against Consumption (columns)

       [ 115][ 117][ 118][ 119][ 120][ 123][ 124][ 125][ 126][ 1

[  81]  33.3% 33.3%  0.0%  0.0% 33.3%  0.0%  0.0%  0.0%  0.0%  0
[  82]   0.0%  0.0%  0.0%  0.0%  0.0% 50.0%  0.0%  0.0%  0.0% 50
[  84]  50.0%  0.0% 50.0%  0.0%  0.0%  0.0%  0.0%  0.0%  0.0%  0
[  85]   0.0%  0.0%  0.0%  0.0%  0.0%100.0%  0.0%  0.0%  0.0%  0
[  88]   0.0%  0.0%  0.0% 50.0%  0.0%  0.0%  0.0%  0.0%  0.0%  0
[  89]   0.0%  0.0%  0.0%  0.0%  0.0%  0.0% 50.0% 50.0%  0.0%  0
[  91]   0.0%  0.0%  0.0%  0.0%  0.0%100.0%  0.0%  0.0%  0.0%  0
[  92]   0.0%  0.0%  0.0%  0.0%  0.0%  0.0%  0.0%  0.0%  0.0%  0
[  93]   0.0%  0.0%  0.0%  0.0%  0.0%  0.0%  0.0%  0.0%100.0%  0
[  95]   0.0%  0.0%  0.0%  0.0%  0.0%  0.0%  0.0%  0.0%  0.0%  0
```

```
Covariance matrix (2 x 2)

      139.31       118.37
      118.37       116.86

Inverse of covariance matrix (2 x 2)

    0.051530    -0.052197
   -0.052197     0.061430

Mahalanobis distances from the centroid
using the variables:
 Income
 Consumption

       1  1.975760
       2  1.198042
       3  1.640533
       4  0.284833
       5  1.834640
       6  1.051606
       7  0.855891
       8  1.068310
       9  0.608605
      10  0.384669
      11  1.648190
      12  1.491036
      13  1.068801
      14  1.497850
```

Figure: Principal components; Cross Tabulation; Mahalanobis distances; Correlation matrix.

Simple Linear Regression Model

We can write our model:

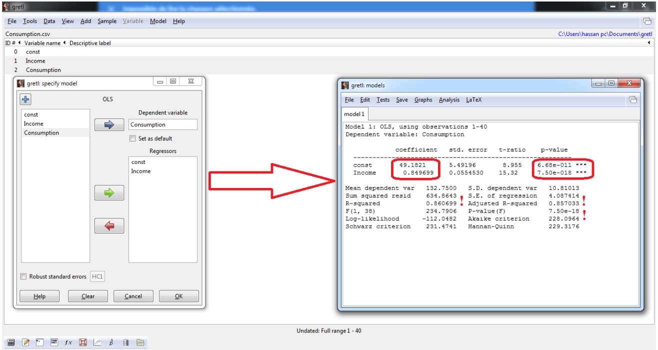
$$y = \beta_0 + \beta_1 x + \epsilon$$

$$\mathit{consumption} \;=\; \beta_0 + \beta_1 \; \mathit{income} + \epsilon$$

? Our assumption, which we will test, is that income explains consumption.

Regression Results

We see that $R^2 = 0.86$. We can therefore say that about 86% of the variability of the dependent variable consumption is explained by the income variable; $\mathit{consumption} = 49.18 + 0.85\ \mathit{income} + \epsilon$


gretl
File Tools Data View Add Sample Variable Model Help
Consumption.csv
C:\Users\hassan pc\Documents\gretl
ID # Variable name Descriptive label
0 const
1 Income
2 Consumption
gretl: specify model
OLS
const
Income
Consumption
Robust standard errors HC1
Help Clear Cancel OK
gretl: models
File Edit Tests Save Graphs Analysis LaTeX
model 1
Undated: Full range 1 - 40

Dependent variable: Consumption

□ Set as default

Regressors: const, Income

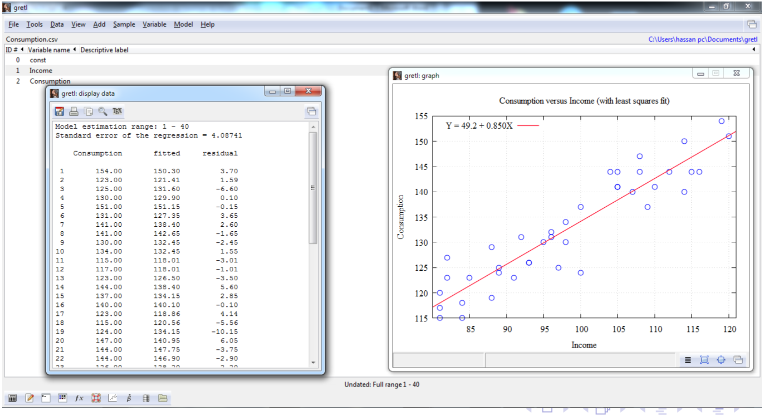
```
Model 1: OLS, using observations 1-40
Dependent variable: Consumption

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       49.1821      5.49196       8.955    6.68e-011 ***
  Income       0.849699    0.0554530    15.32     7.50e-018 ***

Mean dependent var   132.7500   S.D. dependent var   10.81013
Sum squared resid    634.8643   S.E. of regression   4.087414
R-squared            0.860699   Adjusted R-squared   0.857033
F(1, 38)             234.7906   P-value(F)           7.50e-18
Log-likelihood      -112.0482   Akaike criterion     228.0964
Schwarz criterion    231.4741   Hannan-Quinn         229.3176
```

Interpreting the coefficients

- The estimated model is: $\mathit{consumption} = 49.18 + 0.85\ \mathit{income} + \epsilon$.
- 49.18 could be interpreted as the consumption level of a family with 0 income.
- Most generally, intercept doesn't have intuitive interpretation.
- 0.85 is the marginal effect of one unit of income on consumption: for every unit more of income a family has, we estimate its consumption grows by 0.85 units.
- The slope always has an intuitive interpretation.

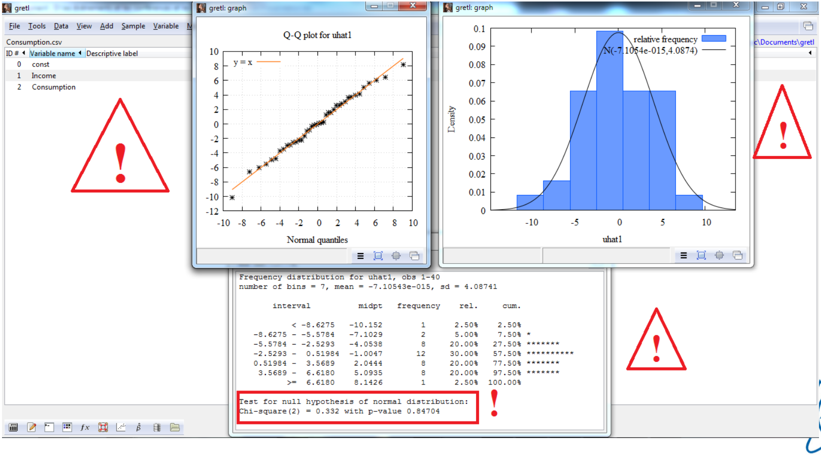
Estimated vs. Actual Values, Residuals

ME	RMSE	MAE	MPE	MAPE	Theil's U
-7.11e-15	3.98	3.24	-0.09	2.46	0.32


gretl
File Tools Data View Add Sample Variable Model Help
Consumption.csv
C:\Users\hassan pc\Documents\gretl
ID # Variable name Descriptive label
0 const
1 Income
2 Consumption
gretl: display data
gretl: graph
Consumption versus Income (with least squares fit)
Y = 49.2 + 0.850X
Consumption
Income
155
150
145
140
135
130
125
120
115
85
90
95
100
105
110
115
120
Undated: Full range 1 - 40

```
Model estimation range: 1 - 40
Standard error of the regression = 4.08741

    Consumption       fitted     residual

 1       154.00       150.30         3.70
 2       123.00       121.41         1.59
 3       125.00       131.60        -6.60
 4       130.00       129.90         0.10
 5       151.00       151.15        -0.15
 6       131.00       127.35         3.65
 7       141.00       138.40         2.60
 8       141.00       142.65        -1.65
 9       130.00       132.45        -2.45
10       134.00       132.45         1.55
11       115.00       118.01        -3.01
12       117.00       118.01        -1.01
13       123.00       126.50        -3.50
14       144.00       138.40         5.60
15       137.00       134.15         2.85
16       140.00       140.10        -0.10
17       123.00       118.86         4.14
18       115.00       120.56        -5.56
19       124.00       134.15       -10.15
20       147.00       140.95         6.05
21       144.00       147.75        -3.75
22       144.00       146.90        -2.90
```

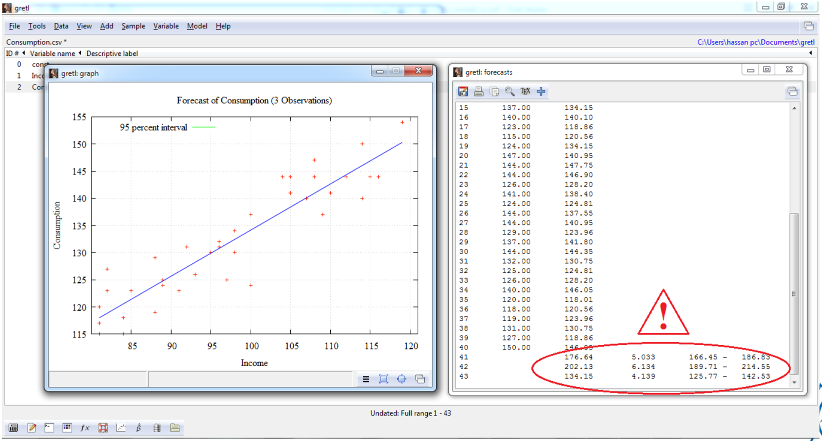
Distribution of residuals



```
Frequency distribution for uhat1, obs 1-40
number of bins = 7, mean = -7.10543e-015, sd = 4.08741

       interval          midpt   frequency    rel.     cum.

           < -8.6275   -10.152        1      2.50%    2.50%
   -8.6275 - -5.5784   -7.1029        2      5.00%    7.50% *
   -5.5784 - -2.5293   -4.0538        8     20.00%   27.50% *******
   -2.5293 -  0.51984  -1.0047       12     30.00%   57.50% **********
   0.51984 -  3.5689    2.0444        8     20.00%   77.50% *******
    3.5689 -  6.6180    5.0935        8     20.00%   97.50% *******
          >=  6.6180    8.1426        1      2.50%  100.00%

Test for null hypothesis of normal distribution:
Chi-square(2) = 0.332 with p-value 0.84704
```

Figure: Residuals.

We want to predict the consumption of 3 families with following incomes: 90, 95 and 100.


gretl
File Tools Data View Add Sample Variable Model Help
Consumption.csv *
C:\Users\hassan pc\Documents\gretl
ID # Variable name Descriptive label
0 const
1 Inco
2 Con
gretl: graph
Forecast of Consumption (3 Observations)
95 percent interval
Consumption
Income
155
150
145
140
135
130
125
120
115
85
90
95
100
105
110
115
120
gretl: forecasts
Undated: Full range 1 - 43

```
15         137.00        134.15
16         140.00        140.10
17         123.00        118.86
18         115.00        120.56
19         124.00        134.15
20         147.00        140.95
21         144.00        147.75
22         144.00        146.90
23         126.00        128.20
24         141.00        138.40
25         124.00        124.81
26         144.00        137.55
27         144.00        140.95
28         129.00        123.96
29         137.00        141.80
30         144.00        144.35
31         132.00        130.75
32         125.00        124.81
33         126.00        128.20
34         140.00        146.05
35         120.00        118.01
36         118.00        120.56
37         119.00        123.96
38         131.00        130.75
39         127.00        118.86
40         150.00        146.05
41                       176.64        5.033        166.45 -   186.83
42                       202.13        6.134        189.71 -   214.55
43                       134.15        4.139        125.77 -   142.53
```

Figure: Forecasts for 95% confidence intervals.

Stock price analysis "Dow Jones"

Stock price analysis is very popular and important in financial study and time series is widely used to trait this topic. The data we use in this presentaion is the Dow Jones Companies from 1st January 2000 to 6th December 2017.
The Dow Jones is an index that shows how 30 large publicly owned companies (IBM, Apple, Pfizer...etc.) based in the United States have traded during a standard trading session in the stock market.

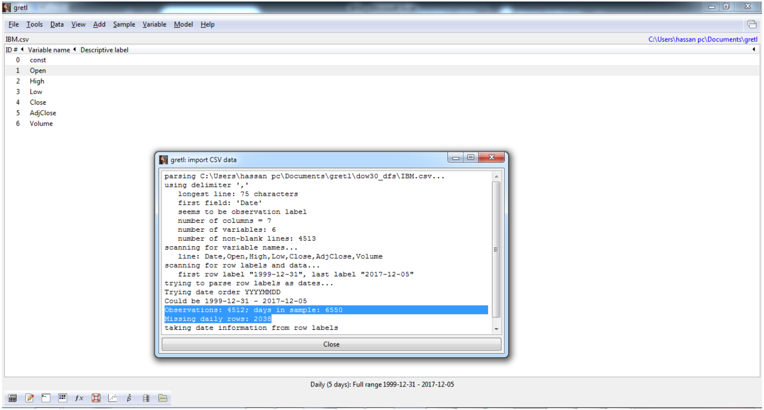
Data and Methodology

The data we use in this presentation is the daily stock price of IBM Holdings:

- Daily (5 days) IBM from 1999-12-31 to 2017-12-05.
- 4512 Days of data.
- In sample training: 1999-12-31 - 2016-02-23 (90%).
- Run Model ARIMA/GARCH.
- Out-of-sample forecast: 2016-02-24 - 2017-12-05 (10%).

Load time series dataset in Gretl.

The dataset contains Open, High, Low, Close and Adjusted Close prices of IBM stock each day of this period.



gretl
File Tools Data View Add Sample Variable Model Help
IBM.csv
C:\Users\hassan pc\Documents\gretl
gretl: import CSV data
Close

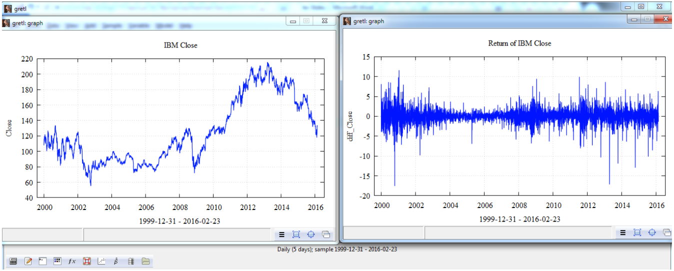
ID #	Variable name	Descriptive label
0	const	
1	Open	
2	High	
3	Low	
4	Close	
5	AdjClose	
6	Volume	

```
parsing C:\Users\hassan pc\Documents\gretl\dow30_dfs\IBM.csv...
using delimiter ','
   longest line: 75 characters
   first field: 'Date'
   seems to be observation label
   number of columns = 7
   number of variables: 6
   number of non-blank lines: 4513
scanning for variable names...
   line: Date,Open,High,Low,Close,AdjClose,Volume
scanning for row labels and data...
   first row label "1999-12-31", last label "2017-12-05"
trying to parse row labels as dates...
Trying date order YYYYMMDD
Could be 1999-12-31 - 2017-12-05
Observations: 4512; days in sample: 6550
Missing daily rows: 2038
taking date information from row labels
```

Daily (5 days): Full range 1999-12-31 - 2017-12-05

Plot Time Series of IBM Close and Return of IBM

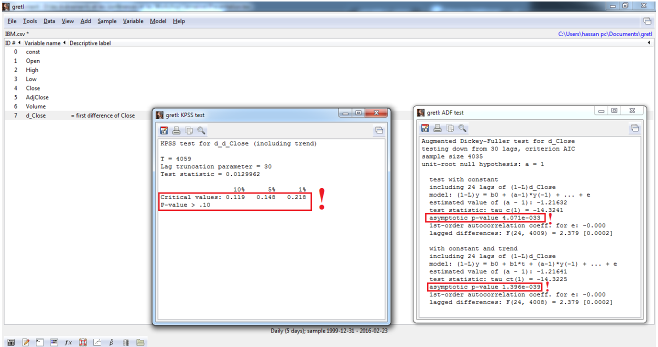
We can find that close price have a general increasing trend over time, and there is a significant drop at the beginning of 2003, 2009 and the beginning of 2016. We decide to analyse returns of IBM which is the differenced close price so that we get a more stationary time serie. We can see that There are periods of persistently high volatility, which shows significantly fluctuations.



Daily (5 days); sample 1999-12-31 - 2016-02-23

Stationarity tests (Augmented Dickey-Fuller test, KPSS test)

After the differenced of the series, we see that the series is stationary in differences. Using test for unit roots (ADF test, KPSS test). I chose the constant and trend version, setting the max lag to 30 and testing down.


gretl
File Tools Data View Add Sample Variable Model Help
C:\Users\hassan pc\Documents\gretl
gretl: KPSS test
gretl: ADF test
!
!
!

IBM.csv *

ID #	Variable name	Descriptive label
0	const	
1	Open	
2	High	
3	Low	
4	Close	
5	AdjClose	
6	Volume	
7	d_Close	= first difference of Close

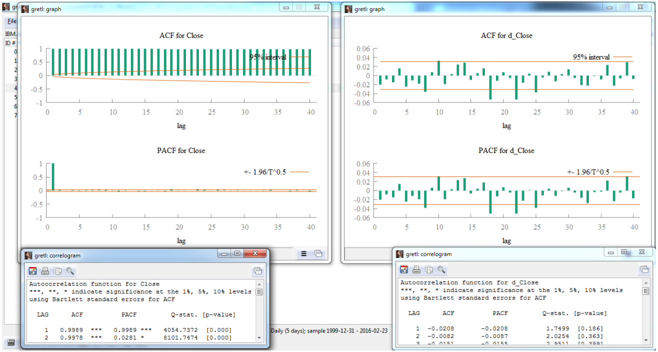
```
KPSS test for d_d_Close (including trend)

T = 4059
Lag truncation parameter = 30
Test statistic = 0.0129962

                   10%      5%      1%
Critical values: 0.119   0.148   0.218
P-value > .10
```

```
Augmented Dickey-Fuller test for d_Close
testing down from 30 lags, criterion AIC
sample size 4035
unit-root null hypothesis: a = 1

  test with constant
  including 24 lags of (1-L)d_Close
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -1.21632
  test statistic: tau_c(1) = -14.3241
  asymptotic p-value 4.071e-033
  1st-order autocorrelation coeff. for e: -0.000
  lagged differences: F(24, 4009) = 2.379 [0.0002]

  with constant and trend
  including 24 lags of (1-L)d_Close
  model: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -1.21641
  test statistic: tau_ct(1) = -14.3225
  asymptotic p-value 1.396e-039
  1st-order autocorrelation coeff. for e: -0.000
  lagged differences: F(24, 4008) = 2.379 [0.0002]
```

Daily (5 days); sample 1999-12-31 - 2016-02-23

ACF and PACF

The ACF don't die fast which indacates that the serie are not stationary (below left). After differencing data, things seem pretty nice.



gretl: correlogram

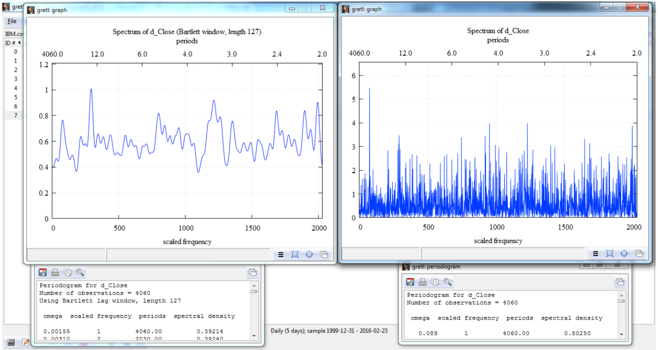
```
Autocorrelation function for Close
***, **, * indicate significance at the 1%, 5%, 10% levels
using Bartlett standard errors for ACF

  LAG      ACF          PACF         Q-stat. [p-value]

    1   0.9989  ***   0.9989 ***   4054.7372  [0.000]
    2   0.9978  ***   0.0281 *     8101.7474  [0.000]
```

gretl: correlogram

```
Autocorrelation function for d_Close
***, **, * indicate significance at the 1%, 5%, 10% levels
using Bartlett standard errors for ACF

  LAG      ACF          PACF         Q-stat. [p-value]

    1  -0.0208       -0.0208          1.7499  [0.186]
    2  -0.0082       -0.0087          2.0254  [0.363]
    3  -0.0151       -0.0155          2.9511  [0.399]
```

Check seasonal effect

I plot the frequency domain of original data and smoothed data. we could see that there's no significant cycle in a year, which means there's no seasonal effect in this problem. Therefore, we choose to use the ARMA model.



```
Periodogram for d_Close
Number of observations = 4060
Using Bartlett lag window, length 127

 omega  scaled frequency  periods  spectral density

 0.00155       1          4060.00          0.39214
 0.00310       2          2030.00          0.39240
```

Daily (5 days); sample 1999-12-31 - 2016-02-23

```
Periodogram for d_Close
Number of observations = 4060

 omega  scaled frequency  periods  spectral density

   0.089       1          4060.00          0.80250
```

ARMA Model

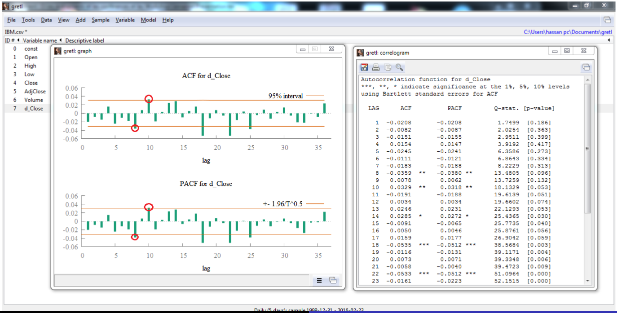
I will start by fitting an ARMA(p,q) model of the form:

$$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1} + \ldots + \theta_q\epsilon_{t-q}$$

- $\{\epsilon_t\}$ is a white noise process with distribution $\mathcal{N}(0, \sigma^2)$.
- $\phi_1, \ldots, \phi_p$ are the coefficients for the autoregressive part of the model.
- $\theta_1, \ldots, \theta_q$ are the coefficients for moving average part of the model.
- μ is the population mean.
- σ^2 is error variance.

Determine 'p' and 'q'

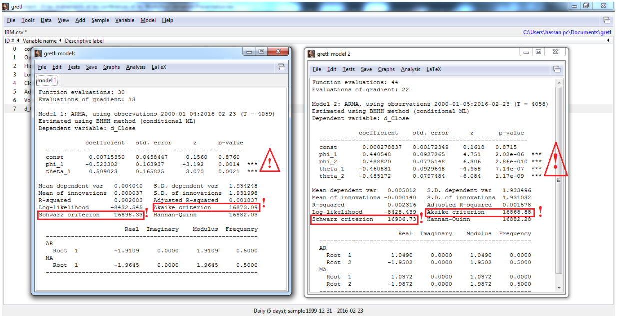
The orders p and q of the AR and MA models are then obtained by looking at the autocorrelations and partial autocorrelations. Very often, these orders are not obvious. In this case, we can obtain upper bounds for p and q and then select a model by minimizing a penalized criterion of type AIC or BIC.



```
Autocorrelation function for d_Close
***, **, * indicate significance at the 1%, 5%, 10% levels
using Bartlett standard errors for ACF

  LAG      ACF          PACF         Q-stat. [p-value]

    1  -0.0208       -0.0208          1.7499  [0.186]
    2  -0.0082       -0.0087          2.0254  [0.363]
    3  -0.0151       -0.0155          2.9511  [0.399]
    4   0.0154        0.0147          3.9192  [0.417]
    5  -0.0245       -0.0241          6.3586  [0.273]
    6  -0.0111       -0.0121          6.8643  [0.334]
    7  -0.0183       -0.0188          8.2229  [0.313]
    8  -0.0359  **   -0.0380 **      13.4805  [0.096]
    9   0.0078        0.0062         13.7259  [0.132]
   10   0.0329  **    0.0318 **      18.1329  [0.053]
   11  -0.0191       -0.0188         19.6139  [0.051]
   12   0.0034        0.0034         19.6602  [0.074]
   13   0.0246        0.0231         22.1293  [0.053]
   14   0.0285  *     0.0272 *       25.4365  [0.030]
   15  -0.0091       -0.0065         25.7735  [0.040]
   16   0.0050        0.0046         25.8761  [0.056]
   17   0.0159        0.0177         26.9042  [0.059]
   18  -0.0535  ***  -0.0512 ***     38.5684  [0.003]
   19  -0.0116       -0.0131         39.1171  [0.004]
   20   0.0073        0.0071         39.3348  [0.006]
   21  -0.0058       -0.0040         39.4723  [0.009]
   22  -0.0533  ***  -0.0512 ***     51.0964  [0.000]
   23  -0.0161       -0.0223         52.1515  [0.000]
```

Estimate and interpretation of coefficients ARIMA(p,d,q)

From the result, the autoregressive and moving average term has a p-value that is less than the significance level of 5%. You can conclude that the coefficient for the AR and MA is statistically significant, and you can choose model ARMA(1,1) with a minimum of AIC.

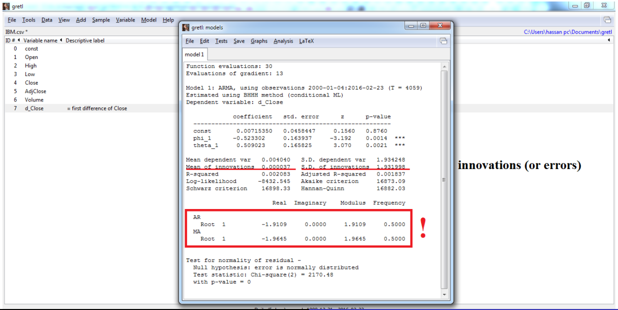

gretl
File Tools Data View Add Sample Variable Model Help
IBM.csv *
C:\Users\hassan pc\Documents\gretl
ID # Variable name Descriptive label
gretl: models
File Edit Tests Save Graphs Analysis LaTeX
model 1
gretl: model 2
File Edit Tests Save Graphs Analysis LaTeX
Daily (5 days); sample 1999-12-31 - 2016-02-23

```
Function evaluations: 30
Evaluations of gradient: 13

Model 1: ARMA, using observations 2000-01-04:2016-02-23 (T = 4059)
Estimated using BHHH method (conditional ML)
Dependent variable: d_Close

             coefficient   std. error      z      p-value
  -------------------------------------------------------
  const       0.00715350   0.0458447     0.1560   0.8760
  phi_1      -0.523302     0.163937     -3.192    0.0014   ***
  theta_1     0.509023     0.165825      3.070    0.0021   ***

Mean dependent var   0.004040   S.D. dependent var   1.934248
Mean of innovations  0.000037   S.D. of innovations  1.931998
R-squared            0.002083   Adjusted R-squared   0.001837
Log-likelihood      -8432.545   Akaike criterion     16873.09
Schwarz criterion    16898.33   Hannan-Quinn         16882.03

                        Real  Imaginary    Modulus  Frequency
  -----------------------------------------------------------
  AR
    Root  1          -1.9109     0.0000     1.9109     0.5000
  MA
    Root  1          -1.9645     0.0000     1.9645     0.5000
  -----------------------------------------------------------
```

```
Function evaluations: 44
Evaluations of gradient: 22

Model 2: ARMA, using observations 2000-01-05:2016-02-23 (T = 4058)
Estimated using BHHH method (conditional ML)
Dependent variable: d_Close

             coefficient    std. error      z       p-value
  ---------------------------------------------------------
  const       0.000278837   0.00172349    0.1618   0.8715
  phi_1       0.440548      0.0927265     4.751    2.02e-06  ***
  phi_2       0.488820      0.0775148     6.306    2.86e-010 ***
  theta_1    -0.460881      0.0929648    -4.958    7.14e-07  ***
  theta_2    -0.485172      0.0797484    -6.084    1.17e-09  ***

Mean dependent var   0.005012   S.D. dependent var   1.933496
Mean of innovations -0.000140   S.D. of innovations  1.931032
R-squared            0.002316   Adjusted R-squared   0.001578
Log-likelihood      -8428.439   Akaike criterion     16868.88
Schwarz criterion    16906.73   Hannan-Quinn         16882.28

                        Real  Imaginary    Modulus  Frequency
  -----------------------------------------------------------
  AR
    Root  1           1.0490     0.0000     1.0490     0.0000
    Root  2          -1.9502     0.0000     1.9502     0.5000
  MA
    Root  1           1.0372     0.0000     1.0372     0.0000
    Root  2          -1.9872     0.0000     1.9872     0.5000
```

Testing ARMA(1,1) Model Root

From the result, we can see that ARMA(1,1) model is causal (or stationary) and invertible because the AR and MA polynomial has all its roots outside of the unit circle. So my final model is ARMA(1,1): $Y_{1:t}$ is defined by $(1-\phi_1 B)(Y_t-\mu) = (1+\theta_1 B)\epsilon_t$ i.e.: $(1+0.523B)(Y_t-0.007) = (1+0.509B)\epsilon_t$ where $\epsilon_t \sim \mathcal{N}(0, 1.932^2)$


gretl
File Tools Data View Add Sample Variable Model Help
IBM.csv *
C:\Users\hassan pc\Documents\gretl
gretl: models
File Edit Tests Save Graphs Analysis LaTeX
model 1

```
ID #  Variable name   Descriptive label
  0   const
  1   Open
  2   High
  3   Low
  4   Close
  5   AdjClose
  6   Volume
  7   d_Close         = first difference of Close
```

```
Function evaluations: 30
Evaluations of gradient: 13

Model 1: ARMA, using observations 2000-01-04:2016-02-23 (T = 4059)
Estimated using BHHH method (conditional ML)
Dependent variable: d_Close

             coefficient   std. error      z      p-value
  -------------------------------------------------------
  const       0.00715350   0.0458447     0.1560   0.8760
  phi_1      -0.523302     0.163937     -3.192    0.0014  ***
  theta_1     0.509023     0.165825      3.070    0.0021  ***

Mean dependent var   0.004040   S.D. dependent var   1.934248
Mean of innovations  0.000037   S.D. of innovations  1.931998
R-squared            0.002083   Adjusted R-squared   0.001837
Log-likelihood      -8432.545   Akaike criterion     16873.09
Schwarz criterion    16898.33   Hannan-Quinn         16882.03

                        Real  Imaginary    Modulus  Frequency
  -----------------------------------------------------------
  AR
    Root  1          -1.9109     0.0000     1.9109     0.5000
  MA
    Root  1          -1.9645     0.0000     1.9645     0.5000

Test for normality of residual -
  Null hypothesis: error is normally distributed
  Test statistic: Chi-square(2) = 2170.48
  with p-value = 0
```

!

innovations (or errors)

GARCH model

A GARCH(p,q) model can be defined as follows:

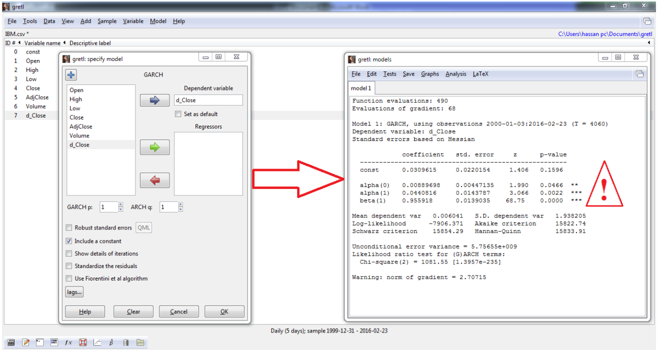
$$\epsilon_t = \sigma_t z_t$$

- $z_t \sim$ i.i.d $\mathcal{N}(0,1)$ independent of the past $\epsilon_{t-1}, \epsilon_{t-2}, \dots \forall t$
- $\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_p \epsilon_{t-p}^2 + \beta_1 \epsilon_{t-1}^2 + \ldots + \beta_q \epsilon_{t-q}^2.$
- Where $\omega > 0, \alpha_i, \beta_j \geq 0,\ i = 1, \dots, p,\ j = 1, \dots, q$ and $\sum_i^p \alpha_i + \sum_j^q \beta_j < 1.$

We can model the trajectory of return by a stationary GARCH(1,1) process.

GARCH in Gretl

All coefficients of Volatility (alpha(0), alpha(1), beta(1)) are significant. But the omega (const) isn't significant.


gretl
File Tools Data View Add Sample Variable Model Help
IBM.csv *
C:\Users\hassan pc\Documents\gretl
gretl: specify model
GARCH
Help
Clear
Cancel
OK
lags...
QML
gretl: models
File Edit Tests Save Graphs Analysis LaTeX
model 1
!
Daily (5 days); sample 1999-12-31 - 2016-02-23

ID #	Variable name	Descriptive label
0	const	
1	Open	
2	High	
3	Low	
4	Close	
5	AdjClose	
6	Volume	
7	d_Close	

Open
High
Low
Close
AdjClose
Volume
d_Close

Dependent variable: d_Close

□ Set as default

Regressors

GARCH p: 1 ARCH q: 1

□ Robust standard errors

☑ Include a constant

□ Show details of iterations

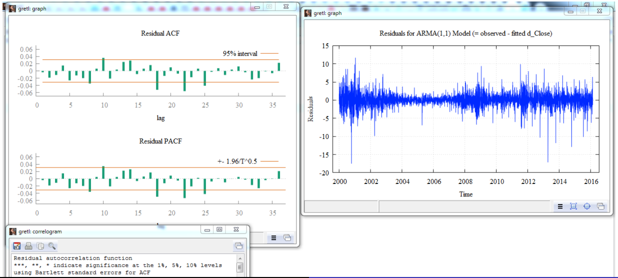
□ Standardize the residuals

□ Use Fiorentini et al algorithm

```
Function evaluations: 490
Evaluations of gradient: 68

Model 1: GARCH, using observations 2000-01-03:2016-02-23 (T = 4060)
Dependent variable: d_Close
Standard errors based on Hessian

             coefficient   std. error      z      p-value
  --------------------------------------------------------
  const       0.0309615    0.0220154     1.406   0.1596

  alpha(0)    0.00889698   0.00447135    1.990   0.0466  **
  alpha(1)    0.0440816    0.0143787     3.066   0.0022  ***
  beta(1)     0.955918     0.0139035    68.75    0.0000  ***

Mean dependent var   0.006041   S.D. dependent var   1.938205
Log-likelihood      -7906.371   Akaike criterion     15822.74
Schwarz criterion    15854.29   Hannan-Quinn         15833.91

Unconditional error variance = 5.75655e+009
Likelihood ratio test for (G)ARCH terms:
  Chi-square(2) = 1081.55 [1.3957e-235]

Warning: norm of gradient = 2.70715
```

Residuals for ARMA(1,1) Model

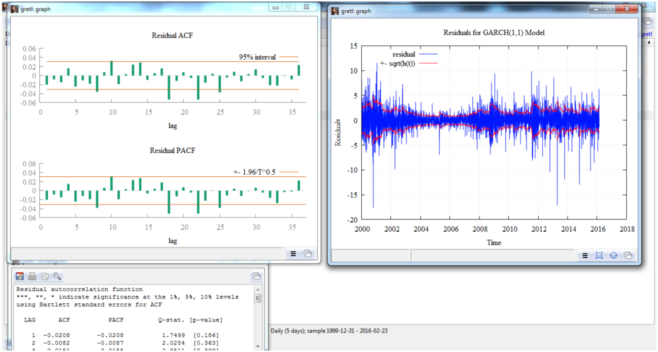
- We can see the mean of residuals is almost zero. However, we can see the not a constant of variance.
- The ACF of the residuals behave like the ACF of a white noise process. The values of ACF almost fall inside the dashed lines except lag 8, 10, 18, 22, 25. So we conclude that the autocorelation of residuals are zero.



```
Residual autocorrelation function
***, **, * indicate significance at the 1%, 5%, 10% levels
using Bartlett standard errors for ACF
```

Residuals for GARCH(1,1) Model

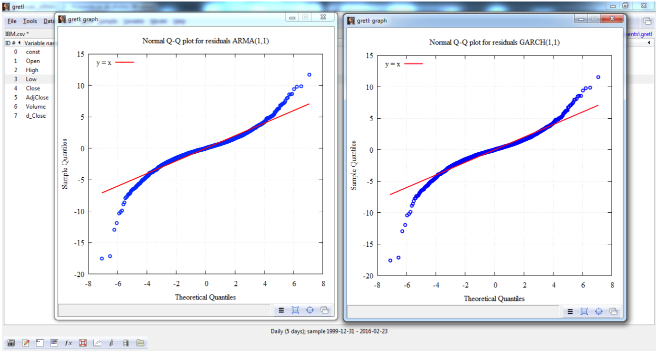
Same thing for GARCH model...!



```
Residual autocorrelation function
***, **, * indicate significance at the 1%, 5%, 10% levels
using Bartlett standard errors for ACF

  LAG      ACF          PACF         Q-stat. [p-value]

    1  -0.0208       -0.0208          1.7499  [0.186]
    2  -0.0082       -0.0087          2.0254  [0.363]
    3   0.0151        0.0155          2.9511  [0.399]
```

Testing for Normality

From QQ-Plot, we can see that the distribution of residuals are not normally distributed. There are too many extreme positive and negative residuals. We say the distribution is ”**heavy tailed**”.

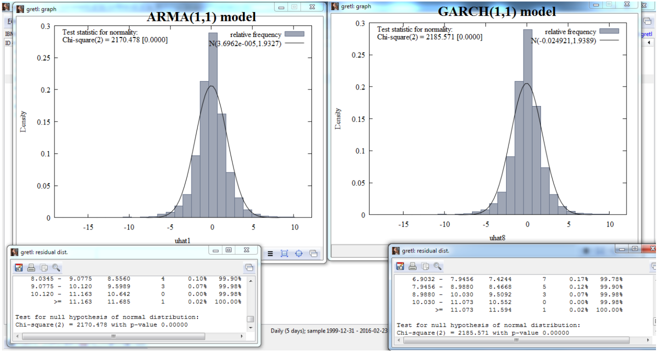


ID #	Variable nar
0	const
1	Open
2	High
3	Low
4	Close
5	AdjClose
6	Volume
7	d_Close

Daily (5 days); sample 1999-12-31 - 2016-02-23

Histogram of residuals

Graphically, histogram of residuals suggests that the residuals (and hence the error terms) are normally distributed:

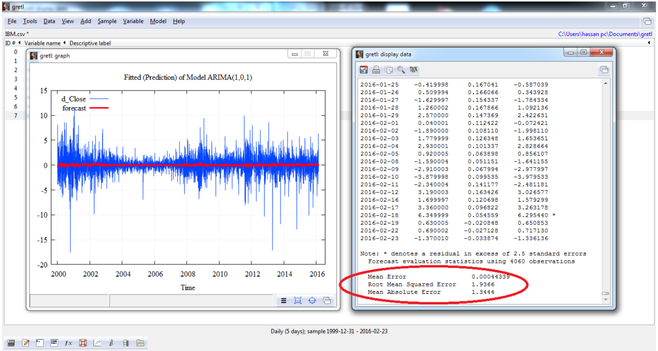


```
    8.0345 -  9.0775    8.5560         4       0.10%   99.90%
    9.0775 -  10.120    9.5989         3       0.07%   99.98%
    10.120 -  11.163    10.642         0       0.00%   99.98%
          >=  11.163    11.685         1       0.02%  100.00%

Test for null hypothesis of normal distribution:
Chi-square(2) = 2170.478 with p-value 0.00000
```

```
    6.9032 -  7.9456    7.4244         7       0.17%   99.78%
    7.9456 -  8.9880    8.4668         5       0.12%   99.90%
    8.9880 -  10.030    9.5092         3       0.07%   99.98%
    10.030 -  11.073    10.552         0       0.00%   99.98%
          >=  11.073    11.594         1       0.02%  100.00%

Test for null hypothesis of normal distribution:
Chi-square(2) = 2185.571 with p-value 0.00000
```

Forecasting

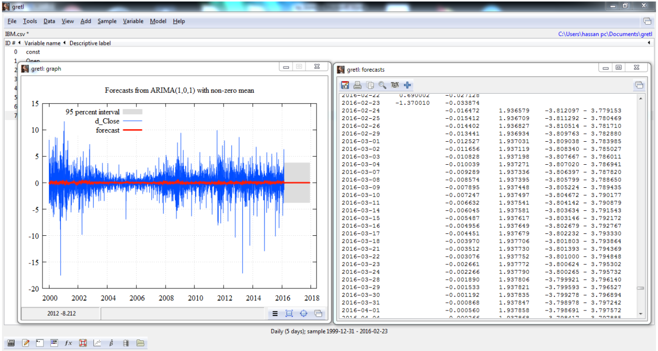
The value of RMSE is 1.9366, which is pretty small. And this may indicate the ARMA(1,1) model is reasonable although we need to further study this for better analysis.


gretl
File Tools Data View Add Sample Variable Model Help
IBM.csv *
C:\Users\hassan pc\Documents\gretl
ID # Variable name Descriptive label
gretl: graph
Fitted (Prediction) of Model ARIMA(1,0,1)
d_Close
forecast
15
10
5
0
-5
-10
-15
-20
2000
2002
2004
2006
2008
2010
2012
2014
2016
Time
gretl: display data
Daily (5 days); sample 1999-12-31 - 2016-02-23

```
2016-01-25    -0.419998     0.167041    -0.587039
2016-01-26     0.509994     0.166066     0.343928
2016-01-27    -1.629997     0.154337    -1.784334
2016-01-28     1.260002     0.167866     1.092136
2016-01-29     2.570000     0.147369     2.422631
2016-02-01     0.040001     0.112422    -0.072421
2016-02-02    -1.890000     0.108110    -1.998110
2016-02-03     1.779999     0.126348     1.653651
2016-02-04     2.930001     0.101337     2.828664
2016-02-05     0.920005     0.063898     0.856107
2016-02-08    -1.590004     0.051151    -1.641155
2016-02-09    -2.910003     0.067994    -2.977997
2016-02-10    -3.879998     0.099535    -3.979533
2016-02-11    -2.340004     0.141177    -2.481181
2016-02-12     3.190003     0.163426     3.026577
2016-02-16     1.699997     0.120698     1.579299
2016-02-17     3.360000     0.096822     3.263178
2016-02-18     6.349999     0.054559     6.295440 *
2016-02-19     0.630005    -0.020848     0.650853
2016-02-22     0.690002    -0.027128     0.717130
2016-02-23    -1.370010    -0.033874    -1.336136

Note: * denotes a residual in excess of 2.5 standard errors
  Forecast evaluation statistics using 4060 observations

  Mean Error                 0.00044339
  Root Mean Squared Error    1.9366
  Mean Absolute Error        1.3444
```

Forecasting

The heavy light gray bar seperately represent the 95% confidence interval for the forecast. We can see that there will not be a large volatility by forecasting.


gretl
File Tools Data View Add Sample Variable Model Help
IBM.csv *
C:\Users\hassan pc\Documents\gretl
ID # Variable name Descriptive label
0 const
1 Open
2
3
4
5
6
7
gretl: graph
Forecasts from ARIMA(1,0,1) with non-zero mean
95 percent interval
d_Close
forecast
15
10
5
0
-5
-10
-15
-20
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2012 -8.212
gretl: forecasts
Daily (5 days); sample 1999-12-31 - 2016-02-23

```
2016-02-22     0.690002    -0.027128
2016-02-23    -1.370010    -0.033874
2016-02-24                 -0.016472     1.936579    -3.812097 - 3.779153
2016-02-25                 -0.015412     1.936709    -3.811292 - 3.780469
2016-02-26                 -0.014402     1.936827    -3.810514 - 3.781710
2016-02-29                 -0.013441     1.936934    -3.809763 - 3.782880
2016-03-01                 -0.012527     1.937031    -3.809038 - 3.783985
2016-03-02                 -0.011656     1.937119    -3.808340 - 3.785027
2016-03-03                 -0.010828     1.937198    -3.807667 - 3.786011
2016-03-04                 -0.010039     1.937271    -3.807020 - 3.786941
2016-03-07                 -0.009289     1.937336    -3.806397 - 3.787820
2016-03-08                 -0.008574     1.937395    -3.805799 - 3.788650
2016-03-09                 -0.007895     1.937448    -3.805224 - 3.789435
2016-03-10                 -0.007247     1.937497    -3.804672 - 3.790177
2016-03-11                 -0.006632     1.937541    -3.804142 - 3.790879
2016-03-14                 -0.006045     1.937581    -3.803634 - 3.791543
2016-03-15                 -0.005487     1.937617    -3.803146 - 3.792172
2016-03-16                 -0.004956     1.937649    -3.802679 - 3.792767
2016-03-17                 -0.004451     1.937679    -3.802232 - 3.793330
2016-03-18                 -0.003970     1.937706    -3.801803 - 3.793864
2016-03-21                 -0.003512     1.937730    -3.801393 - 3.794369
2016-03-22                 -0.003076     1.937752    -3.801000 - 3.794848
2016-03-23                 -0.002661     1.937772    -3.800624 - 3.795302
2016-03-24                 -0.002266     1.937790    -3.800265 - 3.795732
2016-03-28                 -0.001890     1.937806    -3.799921 - 3.796140
2016-03-29                 -0.001533     1.937821    -3.799593 - 3.796527
2016-03-30                 -0.001192     1.937835    -3.799278 - 3.796894
2016-03-31                 -0.000868     1.937847    -3.798978 - 3.797242
2016-04-01                 -0.000560     1.937858    -3.798691 - 3.797572
2016-04-04                 -0.000266     1.937868    -3.798417 - 3.797885
```

Conclusion

We find that ARMA(1,1) model with white noise fits the IBM stock price data. During the analysis, we use difference function to Returns. And then, the AIC suggests that ARMA(1,1) is the proper choice. After diagnosing the result using residual plot and ACF, we further made a forecast using the ARMA(1,1) model, which needs more effort for better analysis.

Gross domestic product (GDP) of U.S.

GDP data

Gross domestic product (GDP), the featured measure of U.S. output, is the market value of the goods and services produced by labor and property located in the United States.
For more information, see the Guide to the National Income and Product Accounts of the United States (NIPA) and the Bureau of Economic Analysis.

- Seasonal Adjustment: Seasonally Adjusted Annual Rate.
- Frequency: Quarterly.
- Units: Billions of Dollars.
- Date Range: 1990-01-01 to 2020-10-01.

The Box-Jenkins Univariate ARIMA Approach:

- Identification.
- Estimation.
- Diagnostic Checking.
- Forecast.

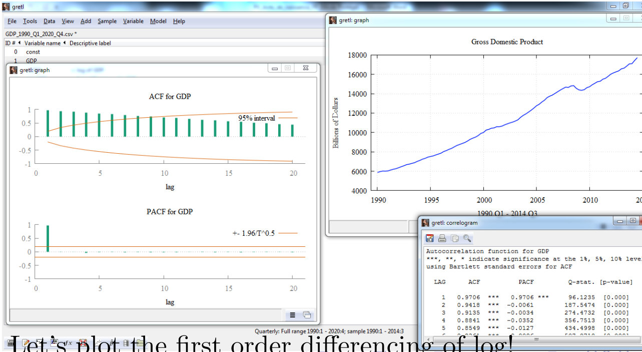
Data collected:

1. Use for model construction (Sample Range): 1990 Q1 - 2014 Q3 (80%).
2. Use for forecast performance assessment (Forecast Range): 2014 Q4 - 2020 Q4 (20%).

Sample size for each observation: 124.

Identification

- There is a trend (perhaps exponential).
- The ACF graph for GDP dies out slowly (exponentially decaying), with one spike in PACF that cuts off after lag 1. Data is nonstationary.



GDP_1990_Q1_2020_Q4.csv *

ID #	Variable name	Descriptive label
0	const	
1	GDP	

Quarterly: Full range 1990:1 - 2020:4; sample 1990:1 - 2014:3

```
Autocorrelation function for GDP
***, **, * indicate significance at the 1%, 5%, 10% leve
using Bartlett standard errors for ACF

  LAG      ACF          PACF         Q-stat. [p-value]

    1   0.9706  ***   0.9706 ***     96.1235  [0.000]
    2   0.9418  ***  -0.0061        187.5474  [0.000]
    3   0.9135  ***  -0.0034        274.4732  [0.000]
    4   0.8841  ***  -0.0352        356.7513  [0.000]
    5   0.8549  ***  -0.0127        434.4998  [0.000]
```

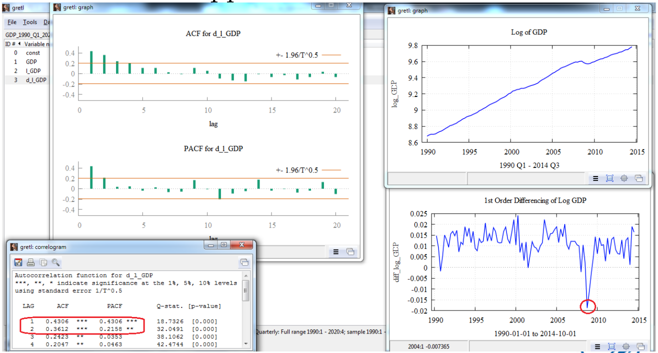
It is possible to remove the trend since differencing should remove any deterministic(linear) trend from the series. Let's plot the first order differencing of log!

Identification

📊 The difference of logs GDP (below right) may show a slight upward trend until the bottom dropped out in late 2008.

From ACF & PACF we can choose a models:

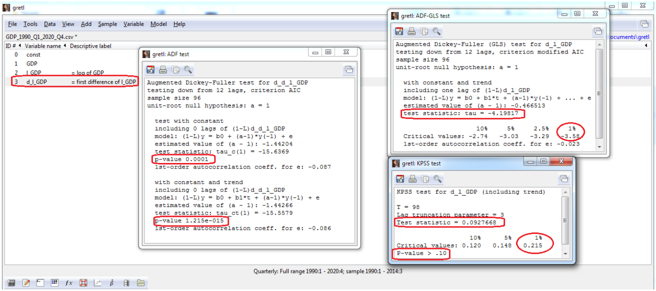
- $ARIMA(1,1,1)$
- $ARIMA(1,1,0)$
- $ARIMA(2,1,0)$



```
Autocorrelation function for d_l_GDP
***, **, * indicate significance at the 1%, 5%, 10% levels
using standard error 1/T^0.5

  LAG      ACF          PACF         Q-stat. [p-value]

    1   0.4306  ***   0.4306 ***    18.7326  [0.000]
    2   0.3612  ***   0.2158 **     32.0491  [0.000]
    3   0.2423  **    0.0353        38.1062  [0.000]
    4   0.2047  **    0.0463        42.4744  [0.000]
```

Identification

Now test for unit roots using the ADF-GLS test. I chose the constant and trend version, setting the maximum lag to 12 and testing down.



GDP_1990_Q1_2020_Q4.csv *

ID #	Variable name	Descriptive label
0	const	
1	GDP	
2	l_GDP	= log of GDP
3	d_l_GDP	= first difference of l_GDP

```
Augmented Dickey-Fuller test for d_d_l_GDP
testing down from 12 lags, criterion AIC
sample size 96
unit-root null hypothesis: a = 1

  test with constant
  including 0 lags of (1-L)d_d_l_GDP
  model: (1-L)y = b0 + (a-1)*y(-1) + e
  estimated value of (a - 1): -1.44204
  test statistic: tau_c(1) = -15.6369
  p-value 0.0001
  1st-order autocorrelation coeff. for e: -0.087

  with constant and trend
  including 0 lags of (1-L)d_d_l_GDP
  model: (1-L)y = b0 + b1*t + (a-1)*y(-1) + e
  estimated value of (a - 1): -1.44266
  test statistic: tau_ct(1) = -15.5579
  p-value 1.215e-015
  1st-order autocorrelation coeff. for e: -0.086
```

```
Augmented Dickey-Fuller (GLS) test for d_l_GDP
testing down from 12 lags, criterion modified AIC
sample size 96
unit-root null hypothesis: a = 1

  with constant and trend
  including one lag of (1-L)d_l_GDP
  model: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.466513
  test statistic: tau = -4.19817

                   10%     5%     2.5%     1%
  Critical values: -2.74   -3.03   -3.29   -3.58
  1st-order autocorrelation coeff. for e: -0.023
```

```
KPSS test for d_l_GDP (including trend)

T = 98
Lag truncation parameter = 3
Test statistic = 0.0927668

                   10%     5%      1%
Critical values: 0.120   0.148   0.215
P-value > .10
```

Quarterly: Full range 1990:1 - 2020:4; sample 1990:1 - 2014:3

The test results indicate that d_l_GDP has not unit root with significantly asymptotic p-value $< 10\%$. This tells us that serie is stationary.

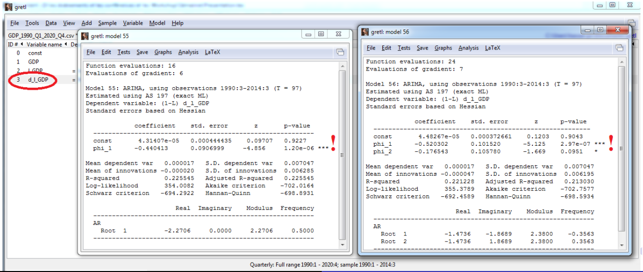
Estimation

Below are the estimated results of ARIMA (1, 1, 0)/ARIMA (2, 1, 0) of model d_l_GDP (Y_{1t}).

Replace the symbol with estimated parameters, we have:

$$\Delta^{1d} Y_{1t} = 4.314 \times 10^{-5} - 0.44 \Delta^{1d} Y_{1,t-1} + \epsilon_{1t} \quad \text{where} \quad \Delta^{1d} Y_{1t} = Y_{1t} - Y_{1,t-1}$$

$$\Delta^{1d} Y_{1t} = 4.482 \times 10^{-5} - 0.52 \Delta^{1d} Y_{1,t-1} - 0.176 \Delta^{1d} Y_{1,t-2} + \epsilon_{1t}$$


gretl
File Tools Data View Add Sample Variable Model Help
GDP_1990_Q1_2020_Q4.csv
ID # Variable name De
0 const
1 GDP
2 l_GDP
3 d_l_GDP
gretl: model 55
File Edit Tests Save Graphs Analysis LaTeX
gretl: model 56
File Edit Tests Save Graphs Analysis LaTeX
!
!
Quarterly: Full range 1990:1 - 2020:4; sample 1990:1 - 2014:3

```
Function evaluations: 16
Evaluations of gradient: 6

Model 55: ARIMA, using observations 1990:3-2014:3 (T = 97)
Estimated using AS 197 (exact ML)
Dependent variable: (1-L) d_l_GDP
Standard errors based on Hessian

             coefficient    std. error       z       p-value
  -----------------------------------------------------------
  const       4.31407e-05   0.000444435    0.09707   0.9227
  phi_1      -0.440413      0.0906999     -4.856     1.20e-06 ***

Mean dependent var   0.000017   S.D. dependent var   0.007047
Mean of innovations -0.000020   S.D. of innovations  0.006285
R-squared            0.225545   Adjusted R-squared   0.225545
Log-likelihood       354.0082   Akaike criterion    -702.0164
Schwarz criterion   -694.2922   Hannan-Quinn        -698.8931

                        Real  Imaginary    Modulus  Frequency
  -----------------------------------------------------------
  AR
    Root  1          -2.2706     0.0000     2.2706     0.5000
  -----------------------------------------------------------
```

```
Function evaluations: 24
Evaluations of gradient: 7

Model 56: ARIMA, using observations 1990:3-2014:3 (T = 97)
Estimated using AS 197 (exact ML)
Dependent variable: (1-L) d_l_GDP
Standard errors based on Hessian

             coefficient    std. error      z       p-value
  ----------------------------------------------------------
  const       4.48267e-05   0.000372661    0.1203   0.9043
  phi_1      -0.520302      0.101520      -5.125    2.97e-07 ***
  phi_2      -0.176543      0.105780      -1.669    0.0951   *

Mean dependent var   0.000017   S.D. dependent var   0.007047
Mean of innovations -0.000047   S.D. of innovations  0.006195
R-squared            0.221228   Adjusted R-squared   0.213030
Log-likelihood       355.3789   Akaike criterion    -702.7577
Schwarz criterion   -692.4589   Hannan-Quinn        -698.5934

                        Real  Imaginary    Modulus  Frequency
  -----------------------------------------------------------
  AR
    Root  1          -1.4736    -1.8689     2.3800    -0.3563
    Root  2          -1.4736     1.8689     2.3800     0.3563
  -----------------------------------------------------------
```

Diagnostic Checking

For diagnostic checking, I will perform additional tests on all the models passing the normality test, for instance Ljung-Box Q test for autocorrelation and ARCH test for heteroskedasticity (lag order 4). The selection technique for the best fitted model is to have the best overall white noise, which is normally distributed, independently distributed (NO autocorrelation), and homoskedastic (NO ARCH effect).

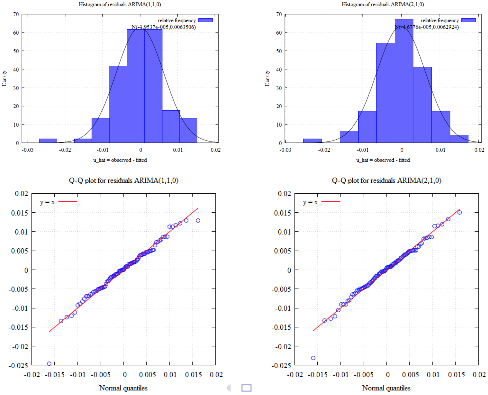
Below is the summarized testing results:

	Normality (χ^2) Test	Autocorrelation Test	ARCH Effect Test
ARIMA(1,1,0)	8.398 (0.015)	4.461 (0.216)	7.859 (0.097)
ARIMA(2,1,0)	6.821 (0.033)	3.466 (0.177)	10.516 (0.033)

p-value for the testing statistics is in the parenthesis. 0.02 is used as significance level for all the hypothesis testing.

Diagnostic Checking

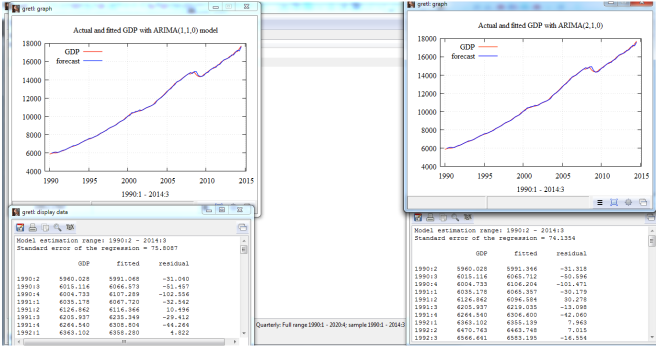
On a Q-Q plot normally distributed residuls appears as roughly a straight line.



The normal distribution is symmetric, so it has no skew (the mean is equal to the median).

Diagnostic Checking

We will utilize as model 2 (ARIMA(2,1,0)) based on their forecasting performance and overall the goodness of fit for forecasting.



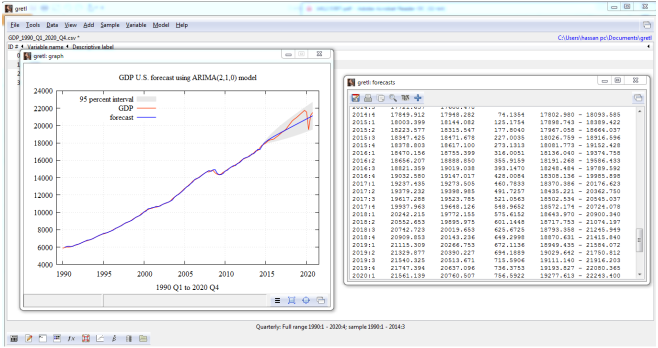
```
Model estimation range: 1990:2 - 2014:3
Standard error of the regression = 75.8087

                 GDP         fitted      residual

1990:2     5960.028      5991.068       -31.040
1990:3     6015.116      6066.573       -51.457
1990:4     6004.733      6107.289      -102.556
1991:1     6035.178      6067.720       -32.542
1991:2     6126.862      6116.366        10.496
1991:3     6205.937      6235.349       -29.412
1991:4     6264.540      6308.804       -44.264
1992:1     6363.102      6358.280         4.822
```

Quarterly: Full range 1990:1 - 2020:4; sample 1990:1 - 2014:3

```
Model estimation range: 1990:2 - 2014:3
Standard error of the regression = 74.1354

                 GDP         fitted      residual

1990:2     5960.028      5991.346       -31.318
1990:3     6015.116      6065.712       -50.596
1990:4     6004.733      6106.204      -101.471
1991:1     6035.178      6065.357       -30.179
1991:2     6126.862      6096.584        30.278
1991:3     6205.937      6219.035       -13.098
1991:4     6264.540      6306.600       -42.060
1992:1     6363.102      6355.139         7.963
1992:2     6470.763      6463.748         7.015
1992:3     6566.641      6583.195       -16.554
```

	ME	RMSE	MAE	MPE	MAPE	Theil's U
ARIMA(1,1,0)	0.188	75.809	53.845	-0.041	0.465	0.467
ARIMA(2,1,0)	0.453	74.135	52.789	-0.032	0.454	0.453

Forecast


gretl
File Tools Data View Add Sample Variable Model Help
GDP_1990_Q1_2020_Q4.csv *
C:\Users\hassan pc\Documents\gretl
ID # Variable name Descriptive label
gretl: graph
GDP U.S. forecast using ARIMA(2,1,0) model
95 percent interval
GDP
forecast
24000
22000
20000
18000
16000
14000
12000
10000
8000
6000
4000
1990
1995
2000
2005
2010
2015
2020
1990 Q1 to 2020 Q4
gretl: forecasts

```
2014:3    17721.657   17608.478
2014:4    17849.912   17948.282     74.1354    17802.980 - 18093.585
2015:1    18003.399   18144.082    125.1754    17898.743 - 18389.422
2015:2    18223.577   18315.547    177.8040    17967.058 - 18664.037
2015:3    18347.425   18471.678    227.0035    18026.759 - 18916.596
2015:4    18378.803   18617.100    273.1313    18081.773 - 19152.428
2016:1    18470.156   18755.399    316.0051    18136.040 - 19374.758
2016:2    18656.207   18888.850    355.9159    18191.268 - 19586.433
2016:3    18821.359   19019.038    393.1470    18248.484 - 19789.592
2016:4    19032.580   19147.017    428.0084    18308.136 - 19985.898
2017:1    19237.435   19273.505    460.7833    18370.386 - 20176.623
2017:2    19379.232   19398.985    491.7257    18435.221 - 20362.750
2017:3    19617.288   19523.785    521.0563    18502.534 - 20545.037
2017:4    19937.963   19648.126    548.9652    18572.174 - 20724.078
2018:1    20242.215   19772.155    575.6152    18643.970 - 20900.340
2018:2    20552.653   19895.975    601.1448    18717.753 - 21074.197
2018:3    20742.723   20019.653    625.6725    18793.358 - 21245.949
2018:4    20909.853   20143.236    649.2998    18870.631 - 21415.840
2019:1    21115.309   20266.753    672.1136    18949.435 - 21584.072
2019:2    21329.877   20390.227    694.1889    19029.642 - 21750.812
2019:3    21540.325   20513.671    715.5906    19111.140 - 21916.203
2019:4    21747.394   20637.096    736.3753    19193.827 - 22080.365
2020:1    21561.139   20760.507    756.5922    19277.613 - 22243.400
```

Quarterly: Full range 1990:1 - 2020:4; sample 1990:1 - 2014:3

Forecast evaluation statistics using 25 observations:

	ME	RMSE	MAE	MPE	MAPE	Theil's U
ARIMA(2,1,0)	212.37	590.08	447.83	0.942	2.184	1.027

Conclusion

Overall, we can see that ARIMA (2, 1, 0) provides a good fit for GDP data. It gives a fairly accurate forecasting. However, although forecasts from 2014 Q3 to 2020 Q4 are within the 95% percent interval, the graph shows that the red line of actual data has in of the confidence interval starting from 2020 Q1 and climbing back up toward the interval from 2020 Q2. Such trend exactly coincides with the way of how the economy has evolved since great recession of 2020 (COVID-19). But, the weakness of ARIMA model is that it could not predict such trend but rather assume the same pattern from 1990 Q1-2014 Q3, that is, the ARIMA model is not good for volatility analysis, but GARCH is. Finally, GDP does strive back after a deep dip in 2020 Q1.

Sources

- Gretl Tutorial 1: Simple Linear Regression
- Time Series Analysis for Stock Data
- Data of Dow Jones
- Time Series Analysis for Log Returns of Nasdaq
- ARIMA(p,d,q) Models (Video 6 of 7 in the gretl Instructional Video Series)
- Les processus ARIMA 2020-2021
- Time Series Analysis for Log Returns of S&P500

Sources

Further reading:

- Interpret the key results for ARIMA

External links:

- Gretl Tutorial 6: Modeling and Forecasting Time Series Data
- ARIMA and its Diagnostics in gretl
- ARCH & GARCH in gretl

Guide d'utilisation de GRETL

Gross Domestic Product Data

Sources

- Comprehensive Time-Series Regression Models Using Gretl-U.S. GDP and Government Consumption Expenditures & Gross Investment from 1980 to 2013
- Using gretl for Principles of Econometrics, 5th Edition
- Introduction to Gretl