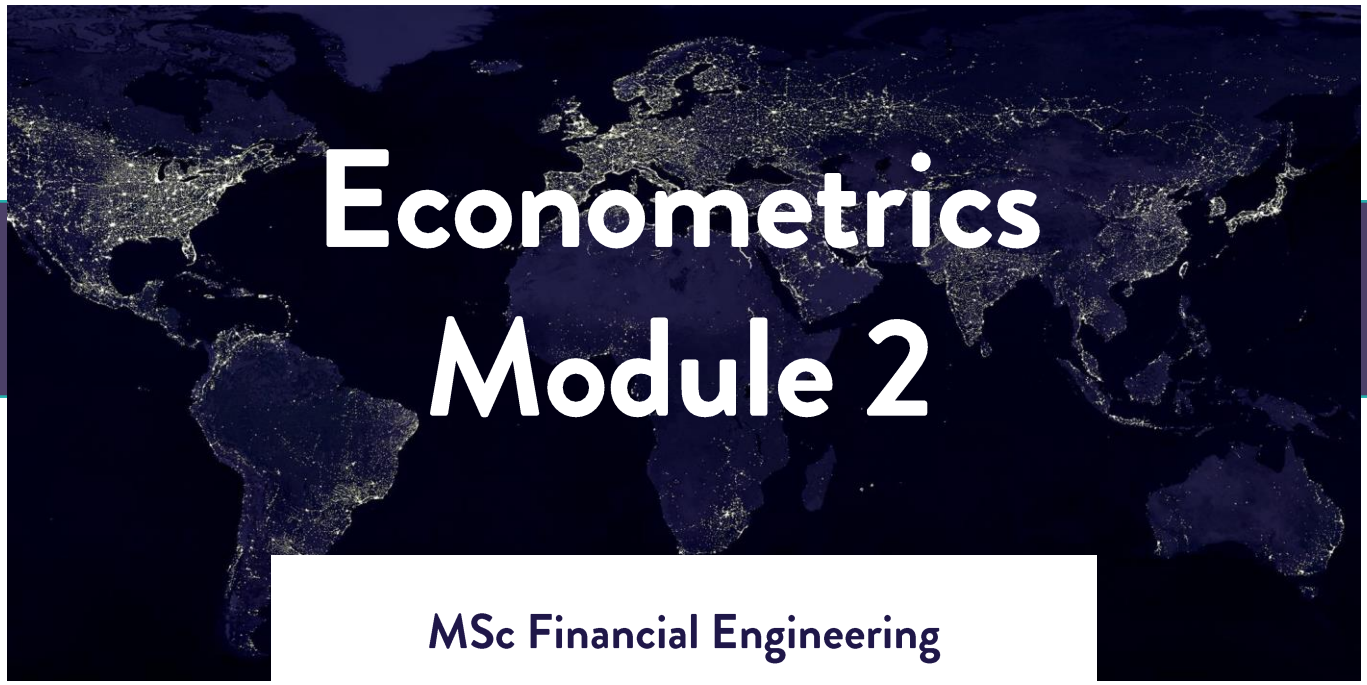
WORLDQUANT UNIVERSITY



Econometrics Module 2

MSc Financial Engineering

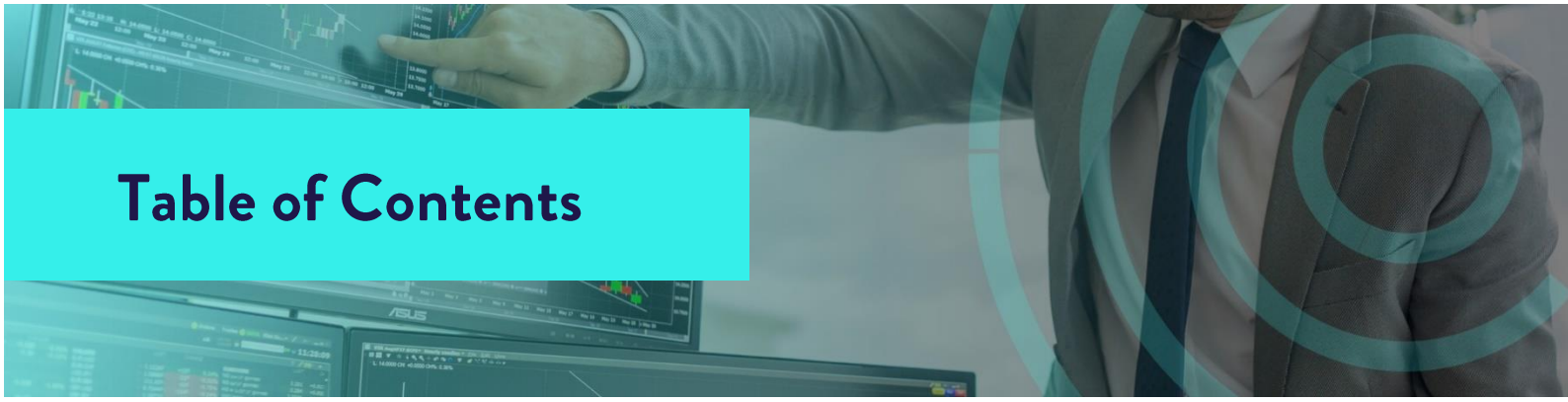


Table of Contents

1. Brief

This document contains the core content for Module 2 of Econometrics, entitled Linear Models. It consists of three lecture video transcripts and three sets of supplementary notes.

2. Course Context

Econometrics is the second course presented in the WorldQuant University (WQU) Master of Science in Financial Engineering (MScFE) program. In this course, you will apply statistical techniques to the analysis of econometric data. The course starts with an introduction to the R statistical programming languages that you will use to build econometric models, including multiple linear regression models, time series models, and stochastic volatility models. You will learn to develop programs using the R language, solve statistical problems, and understand value distributions in modeling extreme portfolio and basic algorithmic trading strategies. The course concludes with a review on applied econometrics in finance and algorithmic trading.

2.1 Course-level Learning Outcomes

Upon completion of the Econometrics course, you will be able to:

1	Write programs using the R language.
2	Use R packages to solve common statistical problems.
3	Formulate a generalized linear model and fit the model to data.
4	Use graphic techniques to visualize multidimensional data.
5	Apply multivariate statistical techniques (PCA, factor analysis, etc.) to analyze multidimensional data.
6	Fit a time series model to data.
7	Fit discrete-time volatility models.
8	Understand and apply filtering techniques to volatility modeling.
9	Understand the use of extreme value distributions in modeling extreme portfolio returns.
10	Define some common risk measures like VaR and Expected Shortfall.
11	Define and use copulas in risk management.
12	Implement basic algorithmic trading strategies.

2.2 Module Breakdown

The Econometrics course consists of the following one-week modules:

1. Basic Statistics
2. Linear Models
3. Univariate Time Series Models
4. Univariate Volatility Modeling
5. Multivariate Time Series Analysis
6. Introduction to Risk Management
7. Algorithmic Trading

3. Module 2:

Linear Models

In Module 2, we extend the aspects of basic statistics and introduce the concepts of stationarity, unit root, and linear model (such as regression and Logit). We will also explore how to implement a basic algorithmic trading strategy using predictions from a linear model.

3.1 Module-level Learning Outcomes

After completing this module, you will be able to:

1. Formulate and fit the linear regression model and its variants.
2. Understand stationarity and unit root.

3.2 Transcripts and Notes

3.2.1 Transcript: The Standard Linear Model: Assumptions

In this video, we will be discussing some of the assumptions that the standard linear model, also called ordinary least squares, has about extracting information about some variable y from a set of variables $x_1, x_2, \ldots, x_k$.

There are four assumptions in total. I will now explain some of the assumptions in detail.

Assumption 1: The true data-generating process is linear in the parameters.
In equation form, this assumption looks like this:

$$y_i = \beta + \beta x_{i1} + \beta_2 x_{i2} + \beta_k \, x_{ik} + \varepsilon_i$$

In this equation y_i is an observation i of the dependent variable, where $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are parameters that we wish to uncover, and ε_i is a positive variance population error.

Before I turn to the other assumptions of the standard model, it is convenient and standard to write the model of interest in matrix form. If we define the row vector: $x_i = [1 \quad X_{i1} \quad x_{i2} \quad \ldots \quad x_{ik}]$ and the column vector $\beta = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \ldots \quad \beta_k]'$ we can rewrite the equation as:

$$y_i = x_i \beta + \varepsilon_i.$$

Next, if we define the column vectors $y = [y_1 \quad y_2 \quad \ldots \quad y_i \quad \ldots y_n]'$ and $\varepsilon = [\varepsilon_1 \quad \varepsilon_2 \quad \ldots \quad \varepsilon_i \quad \ldots \quad \varepsilon_n]'$, and the matrix

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{12} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & x_{i2} & \dots & x_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}$$

We can write the full model of interest as:

$$y = X\beta + \varepsilon.$$

The goal here is to find the β vector. For the moment, we will assume that the relationship is exact, which means that $\varepsilon = 0$. We can use simple linear algebra to recover β. X is an $[n \times (k+1)]$ matrix. First, we pre-multiply by the transpose of X, which gives us:

$$X'y = X'X\,\beta.$$

Now $X'X$ is square with dimensions $[(k+1) \times (k+1)]$. If we pre-multiply by its inverse, this yields the following:

$$(X'X)^{-1}X'y = (X'X)^{-1}X'X\beta$$
$$\beta = (X'X)^{-1}X'y.$$

Thus, we have a closed form expression for the parameters of the model. This is, however, only possible when $(X'X)^{-1}$ is well-defined. This requires that $X'X$ has linearly independent rows and columns, which in turn requires that X has linearly independent columns. This leads to the second assumption.

Assumption 2: there is no perfect multicollinearity in the explanatory variables. Mathematically speaking, the matrix X must have full column rank. We now look in some detail at Assumption 3.

Assumption 3: The explanatory variables are strictly exogenous with respect to the population error: $\mathbb{E}(\varepsilon|X) = 0$

First, we are going to add back the error. In Assumption 1, we assumed that the relationship was exact, so there were no errors. This is, of course, never true in econometrics. If we assume that there is some fundamental noise of the data generating process, we can never uncover the true coefficients. All we can do is obtain an estimate of the parameters. Moreover, for each new sample of data, that estimate will be different. Thus, we define the general expression $(X'X)^{-1}X'y$ as the OLS estimator $\hat{\beta}_{OLS}$ which, from an ex ante perspective, is a random variable.

Following the same steps as above, but now with errors, yields:

$$\hat{\beta}_{OLS} \equiv (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'\,\varepsilon.$$

We want the estimator to be "unbiased".

The definition of unbiasedness is:

$$\mathbb{E}(\hat{\beta}_{OLS}|X) = \beta$$

Applying $\mathbb{E}(\hat{\beta}_{OLS}|X)$ to the estimator:

$$\begin{aligned}\mathbb{E}\left(\hat{\beta}_{OLS}|X\right) &= \mathbb{E}((X'X)^{-1}X'y|X)\\ &= (X'X)^{-1}X'\mathbb{E}(y|X)\\ &= (X'X)^{-1}X'\mathbb{E}(X\beta + \varepsilon|X)\\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\mathbb{E}(\varepsilon|X)\\ &= \beta + (X'X)^{-1}X'\mathbb{E}(\varepsilon|X).\end{aligned}$$

we see that the OLS estimator is unbiased if, and only if, $\mathbb{E}(\varepsilon|X) = 0$. This outcome is called strict exogeneity: we say that X is strictly exogenous with respect to ε. It implies that no element of X is correlated with any element of ε, or more simply, that these two objects share no common information.

The last assumption is:

Assumption 4: The population error process is homoscedastic, as in:

$$\mathbb{E}(\varepsilon_i^2) = \sigma^2 \forall i$$
$$\mathbb{E}(\varepsilon_i \varepsilon_j) = 0 \forall i \neq j.$$

This then brings us to a discussion of homoscedasticity versus serial correction.

Recall that this setup is for a cross-section. Homoscedasticity says that there is no correlation in the errors across observational units (say, stock prices at a given moment).

If this had been a univariate time series, we would have considered the "no serial correlation" condition. That is, each error must be uncorrelated with its own past/future.

In a panel context, we would consider both conditions. To learn more about each of these assumptions, study the provided notes on this subject.

3.2.2 Notes: Regression Analysis

Two-variable regression

Many models in econometrics assume a linear relationship between a dependent variable, Y, and one (or more) independent variables, X. The independent variable can also be referred to as the explanatory variable, regressor, predictor, stimulus, exogenous, covariate or control variable. Similarly, the dependent variable, Y, can also go by many names, including endogenous, predicted, response, outcome and controlled variable:

$$Y = \beta 1 + \beta 2X + u,$$

where u, known as the disturbance term (or error term), is a random stochastic variable that has well-defined probabilistic properties. The disturbance term, u, may well represent all those factors that affect consumption but are not considered explicitly.

A simple business example of a possible linear relationship might be sales, y, and advertising expenditures, x (measured as each additional $10 000 spent on ads). Suppose the relationship is calculated and the following linear relationship is estimated:

$$Y = 200 + 2\,x.$$

For example, if advertising increases by $10 000 (one unit of x), then sales will increase by 2 times $10 000 = $20 000. This is rather a simple model.

Things to bear in mind

1 **Deterministic vs. correlation:**

Econometrics usually entails specifying a functional relationship – such as a linear relationship – between variables. The independent variable is thought to determine the values of the dependent variable. In many instances the relationship may simply reflect a correlation between X and Y. Some

relationships are quite deterministic, for example when we measure the growth of a plant in relation to the amount of sun and water it receives.

2 **Population vs. sampling:**

We are rarely able to observe an entire population of data we want to model. Instead, econometrics typically relies on sampling of data when making estimates.

Linear relationship and sampling

Suppose there is a linear dependence of Y on values of X. For a given value of X, Y is not always the same value but the expected value of Y is given by:

$$E(Y|X) = a + bX \qquad (Y \text{ conditional upon } X).$$

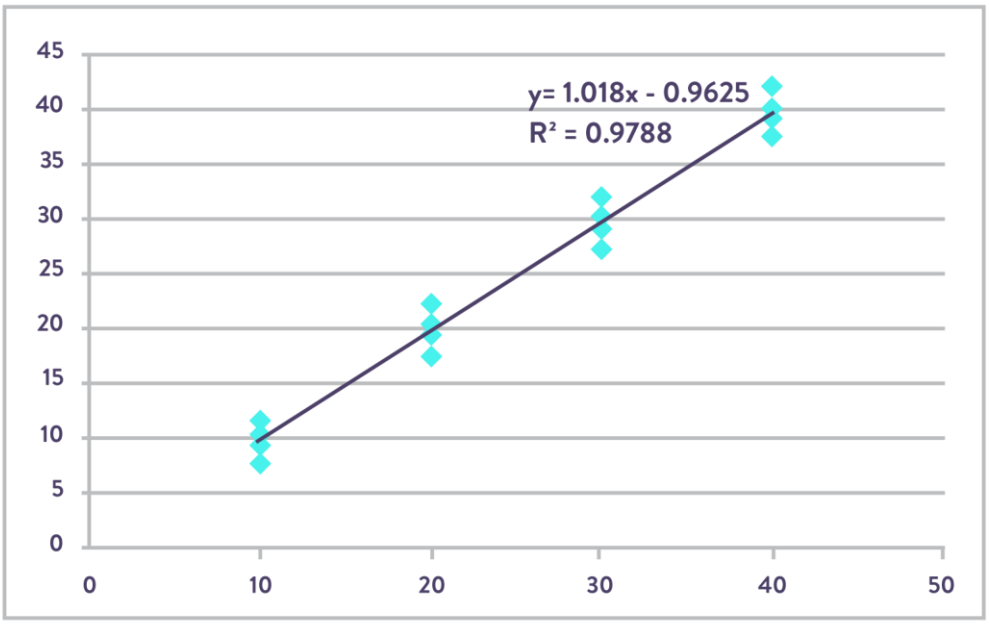
The expected value of Y conditional upon X reflects the population values (infinite sampling) for a and b.

Let's look at an example:

Consider the Keynesian concept of the consumption function (from macroeconomics). This says that as the disposable income, X, increases by a dollar, a household will increase its consumption by something less than a dollar (e.g. 0.9). Below is a sample of income and household consumption:

X	Y	X	Y	X	Y	X	Y
10	7	20	17	30	27.1	40	37.5
10	9	20	19	30	29	40	39.3
10	11	20	20	30	30	40	40.1
10	9.7	20	22	30	31.9	40	42.2

We see that for a given X,they tend to regress towards the consumption function. However, there is a dispersion about the true line as depicted in the graph below:



R^2 is the goodness of fit we shall define later. Note that the original graph was produced in Excel, which is handy for creating trend lines in graphs.

Exchange rates and OLS

A forecaster wants to predict the USD/CAD exchange rate over the next year. He believes an econometric model would be a good method to use and has researched the various factors that he thinks affect the exchange rate. From his research and analysis, he concludes the factors that are most influential are:

- The interest rate differential between the U.S. and Canada (INT),
- The difference in GDP growth rates (GDP), and
- The income growth rate (IGR) differences between the two countries.

The econometric model he comes up with is shown as:

$$\frac{USD}{CAD}(1 - year) = z + a(INT) + b(GDP) + c(IGR).$$

We won't go into the detail of how the model is constructed, but after the model is made, the variables INT, GDP, and IGR can be plugged into the model to generate a forecast. The coefficients a, b, and c will determine how much a certain factor affects the exchange rate as well as the direction of the effect – i.e. whether it is positive or negative. You can see that this method is probably the most complex and time-consuming approach of the ones discussed so far. However, once the model is built, new data can be easily acquired and plugged into the model to generate quick forecasts.

Augmented Dickey-Fuller test

The testing procedure for the ADF test is the same as for the Dickey-Fuller test but it is applied to the model:

$$\Delta y_t = \alpha + \gamma y_{t-1} + \beta t + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + u_t,$$

where α is a constant, β the coefficient on a time trend, and p the lag order of the autoregressive process.

The unit root test is then carried out under the null hypothesis $\gamma = 0$ against the alternative hypothesis of $\gamma < 0$. Once a value for the test statistic

$$\mathrm{DF}_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

is computed, it can be compared to the relevant critical value for the Dickey-Fuller test.

If the test statistic is less than[1] the (larger negative) critical value, then the null hypothesis of $\gamma = 0$ is rejected and no unit root is present.

There are different values for the critical DF values depending on the type of test that is being conducted. Some tests do not include a trend, t, for example. DF critical values table below:

			Probability to the Right of Critical Value						
Model	Statistic	N	1%	2.5%	5%	10%	90%	95%	97.5%	99%
Model I (no constant, no trend)										
	ADF_{tr}	25	-2.66	-2.26	-1.95	-1.60	0.92	1.33	1.70	2.16
		50	-2.62	-2.25	-1.95	-1.61	0.91	1.31	1.66	2.08
		100	-2.60	-2.24	-1.95	-1.61	0.90	1.29	1.64	2.03
		250	-2.58	-2.23	-1.95	-1.61	0.89	1.29	1.63	2.01
		500	-2.58	-2.23	-1.95	-1.61	0.89	1.28	1.62	2.00
		>500	-2.58	-2.23	-1.95	-1.61	0.89	1.28	1.62	2.00
Model II (constant, no trend)										
	ADF_{tr}	25	-3.75	-3.33	-3.00	-2.62	-0.37	0.00	0.34	0.72
		50	-3.58	-3.22	-2.93	-2.60	-0.40	-0.03	0.29	0.66
		100	-3.51	-3.17	-2.89	-2.58	-0.42	-0.05	0.26	0.63
		250	-3.46	-3.14	-2.88	-2.57	-0.42	-0.06	0.24	0.62

[1] This test is non-symmetrical so we do not consider an absolute value.

```
               500  -3.44  -3.13  -2.87  -2.57  -0.43  -0.07   0.24
0.61
              >500  -3.43  -3.12  -2.86  -2.57  -0.44  -0.07   0.23
0.60

Model III (constant, trend)
       ADF_tr   25  -4.38  -3.95  -3.60  -3.24  -1.14  -0.80  -0.50  -
0.15
               50  -4.15  -3.80  -3.50  -3.18  -1.19  -0.87  -0.58  -
0.24
              100  -4.04  -3.73  -3.45  -3.15  -1.22  -0.90  -0.62  -
0.28
              250  -3.99  -3.69  -3.43  -3.13  -1.23  -0.92  -0.64  -
0.31
              500  -3.98  -3.68  -3.42  -3.13  -1.24  -0.93  -0.65  -
0.32
             >500  -3.96  -3.66  -3.41  -3.12  -1.25  -0.94  -0.66  -
0.33
```

Notes

- The null hypothesis of the Augmented Dickey-Fuller is that there is a unit root, with the alternative that there is no unit root. If the p-value is above a critical size, then we cannot reject that there is a unit root.
- The p-values are obtained through regression surface approximation from MacKinnon (1994) but using the updated 2010 tables. If the p-value is close to significant, then the critical values should be used to judge whether to accept or reject the null.

In [ ]:

```python
import numpy as np
import pandas as pd
import statsmodels.tsa.stattools
#the following is a list of returns
a=[1,2,-1,2,-3,0,.5,.8,-.21,2,-1,2,-3,0,.5,.8,-2,2,-1,2,-3,0,.5,.8,-.21,2,-1,2,-3,0,.5,.8,-.2]
statsmodels.tsa.stattools.adfuller(a, maxlag=None)
Out[10]:
(-1.7034844119211205,0.42931810315407193,7,25,{'1%': -3.7238633119999998, '10%': -2.6328003999999998, '5%': -2.98648896},
 49.908550173095492)
```

Testing for mean reversion

A *continuous* mean-reverting time series can be represented by an Ornstein-Uhlenbeck stochastic differential equation:

$$dxt = \theta(\mu - xt)dt + \sigma dWt,$$

where θ is the rate of reversion to the mean, μ is the mean value of the process, σ is the variance of the process, and Wt is a Wiener Process or Brownian Motion. In a discrete setting, the equation states that the change of the price series in the next time period is proportional to the difference between the mean price and the current price, with the addition of Gaussian noise. This property motivates the Augmented Dickey-Fuller test, which we will describe below.

Augmented Dickey-Fuller (ADF) is based on the idea of testing for the presence of a unit root in an autoregressive time series sample. It makes use of the fact that if a price series possesses mean reversion, then the next price level will be proportional to the current price level. A linear lag model of order p is used for the time series:

$$\Delta yt = \alpha + \beta t + \gamma yt - 1 + \delta 1 \Delta yt - 1 + \cdots + \delta p - 1 \Delta yt - p + 1 + \epsilon t,$$

where α is a constant, β represents the coefficient of a temporal trend, and $\Delta yt = y(t) - y(t-1)$. The role of the ADF hypothesis test is to consider the null hypothesis that $\gamma = 0$, which would indicate (with $\alpha = \beta = 0$) that the process is a random walk and thus non-mean reverting. If the hypothesis that $\gamma = 0$ can be rejected, then the following movement of the price series is proportional to the current price, thus it is unlikely to be a random walk.

So, how is the ADF test carried out? The first task is to calculate the test statistic (DF_τ), which is given by the sample proportionality constant $\hat{\gamma}$, divided by the standard error of the sample proportionality constant:

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}.$$

Dickey and Fuller have previously calculated the distribution of this test statistic, which allows us to determine the rejection of the hypothesis for any chosen percentage critical value. The test statistic is a negative number and thus in order to be significant beyond the critical values, the number must be more negative than these values – i.e. less than the critical values.

A key practical issue for traders is that any constant long-term drift in a price is of a much smaller magnitude than any short-term fluctuations, and so the drift is often assumed to be zero ($\beta = 0$) for the model. Since we are considering a lag model of order p, we need to actually set p to a particular value. It is usually sufficient, for trading research, to set $p = 1$ to allow us to reject the null hypothesis.

We use `fix_yahoo_finance` module to calculate the Augmented Dickey-Fuller test. We will carry out the ADF test on a sample price series of Google stock, from 1st January 2000 to 1st January 2019.

Python code for ADF test:

In [ ]:

```
# Import the Time Series library
import statsmodels.tsa.stattools as ts

# Import fix_yahoo-finance module
import fix_yahoo_finance as yf

# Download Google data from 1/1/2000 to 1/1/2019
goog=yf.download("GOOG", start="2000-01-01", end="2019-01-01")
# Output the results of the Augmented Dickey-Fuller test for Google
# with a lag order value of 1
ts.adfuller(goog['Adj Close'], 1)
```

Here is the output of the Augmented Dickey-Fuller test for Google over the period. The first value is the calculated test-statistic, while the second value is the p-value. The fourth is the number of data points in the sample. The fifth value, the dictionary, contains the critical values of the test-statistic at the 1, 5, and 10 percent values respectively.

(-0.12114782916582577, 0.94729096304598859, 0, 3616, {'1%': -3.432159720193857, '5%': -2.8623396332879718, '10%': -2.56719565730786}, 25373.287077939662)

Since the calculated value of the test statistic is larger than any of the critical values at the 1, 5, or 10 percent levels, we cannot reject the null hypothesis of $\gamma = 0$ and thus we are unlikely to have found a mean reverting time series. An alternative means of identifying a mean reverting time series is provided by the concept of stationarity, which we will now discuss.

Stationarity

Financial institutions and corporations, as well as individual investors and researchers, often use financial time series data (such as asset prices, exchange rates, GDP, inflation, and other macroeconomic indicators) in economic forecasts, stock market analysis, or studies of the data itself.

However, refining data is crucial in stock analysis. It is possible to isolate the data points that are relevant to your stock reports. **Stationarity** means that the mean, variance, and intertemporal correlation structure remains constant over time. Non-stationarities can either come from deterministic changes (like trend or seasonal fluctuations) or the stochastic properties of the process (if, for example, the autoregressive process has a unit root, that is one of the roots of the lag polynomial is on the unit circle). In the first case, we can remove the deterministic component by de-trending or de-seasonalization.

Cooking raw data

Data points are often non-stationary or have means, variances, and covariances that change over time. Non-stationary behaviors can be trends, cycles, random walks, or combinations of the three.

Non-stationary data, as a rule, are unpredictable and cannot be modeled or

forecasted. The results obtained by using non-stationary time series may be spurious in that they may indicate a relationship between two variables where one does not exist. In order to receive consistent, reliable results, the non-stationary data needs to be transformed into stationary data. In contrast to the non-stationary process that has a variable variance and a mean that does not remain near, or returns to a long-run mean over time, the stationary process reverts around a constant long-term mean and has a constant variance independent of time.

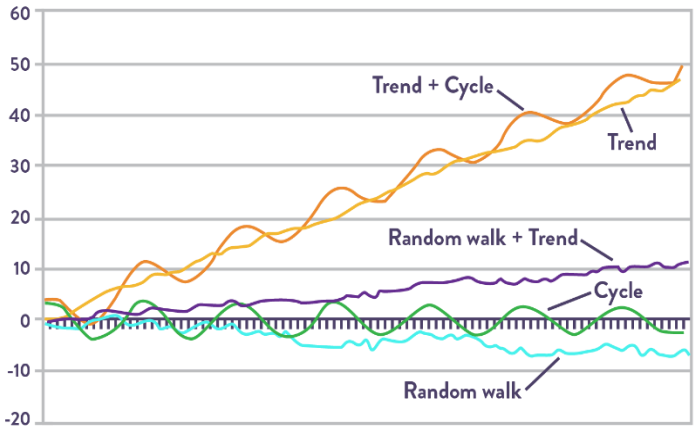


Figure 1: Non-stationary behavior

Types of non-stationary processes

Before we get to the point of transformation for the non-stationary financial time series data, we should distinguish between the different types of the non-stationary processes. This will provide us with a better understanding of the processes and allow us to apply the correct transformation. Examples of non-stationary processes are: random walks with or without a drift (a slow steady change) and deterministic trends. **Deterministic trends** are trends that are constant, positive or negative, independent of time for the whole life of the series.

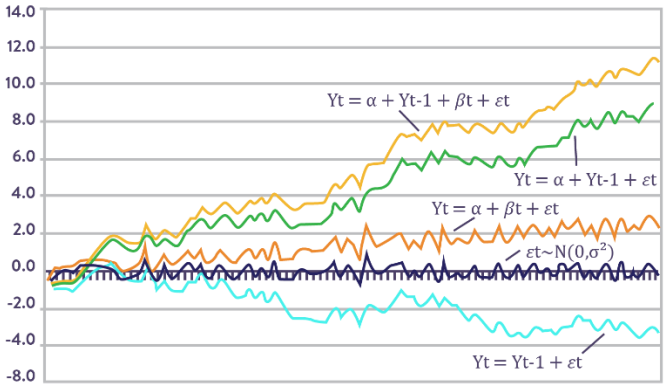


Figure 2: Non-stationary processes

- **Pure random walk** $(Y_t = Y_{t-1} + \varepsilon_t)$

  Random walk predicts that the value at time t will be equal to the last period value plus a stochastic (non-systematic) component that is a white noise. This means that ε_t is independent and identically distributed with mean "0" and variance "σ^2". Random walk can also be called a process integrated of some order, a process with a unit root, or a process with a stochastic trend. It is a non-mean reverting process that can move away from the mean either in a positive or negative direction. Another characteristic of a random walk is that the variance evolves over time and goes to infinity as time goes to infinity; therefore, a random walk cannot be predicted.

- **Random walk with drift** $(Y_t = \alpha + Y_{t-1} + \varepsilon_t)$

  If the random walk model predicts that the value at time t will equal the last period's value plus a constant, or drift (α), and a white noise term (ε_t), then the process is random walk with a drift. It also does not revert to a long-run mean and has variance dependent on time.

- **Deterministic trend** $(Y_t = \alpha + \beta t + \varepsilon_t)$

  Often a random walk with a drift is confused for a deterministic trend. This is because they both include a drift and a white noise component. However, the value at time t in the case of a random walk is regressed on the

last period's value (Y_{t-1}), while in the case of a deterministic trend it is regressed on a time trend (βt). A non-stationary process with a deterministic trend has a mean that grows around a fixed trend, which is constant and independent of time.

- **Random walk with drift and deterministic trend** ($\boldsymbol{Y_t = \alpha + Y_{t-1} + \beta t + \varepsilon_t}$)
  Another example is a non-stationary process that combines a random walk with a drift component (α) and a deterministic trend (βt). It specifies the value at time t by the last period's value, a drift, a trend, and a stochastic component. (To learn more about random walks and trends, see the *Financial Concepts* tutorial.)

Trend and difference stationary

A random walk with or without a drift can be transformed to a stationary process by differencing (subtracting Y_{t-1} from Y_t, taking the difference $Y_t - Y_{t-1}$) correspondingly to $Y_t - Y_{t-1} = \varepsilon_t$ or $Y_t - Y_{t-1} = \alpha + \varepsilon_t$ and then the process becomes difference-stationary. The disadvantage of differencing is that the process loses one observation each time the difference is taken.

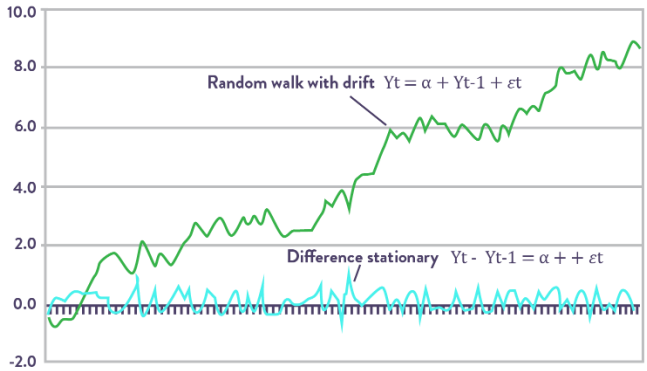


Figure 3: Differencing

A non-stationary process with a deterministic trend becomes stationary after removing the trend, or detrending. For example, $Yt = \alpha + \beta t + \varepsilon t$ is transformed into a stationary process by subtracting the trend βt: $Yt - \beta t = \alpha + \varepsilon t$, as shown in

Figure 4 below. No observation is lost when detrending is used to transform a non-stationary process to a stationary one.

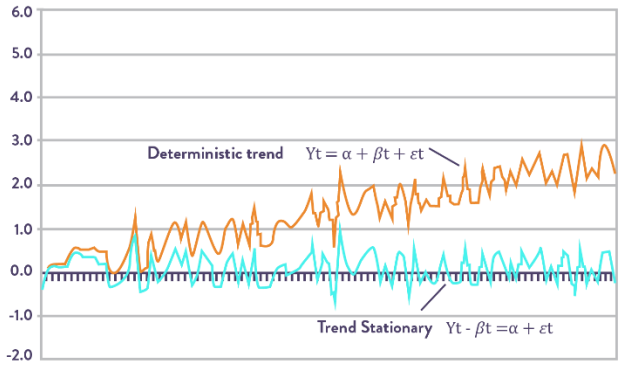


Figure 4: Detrending

In the case of a random walk with a drift and deterministic trend, detrending can remove the deterministic trend and the drift, but the variance will continue to go to infinity. As a result, differencing must also be applied to remove the stochastic trend.

Using non-stationary time series data in financial models produces unreliable and spurious results and leads to poor understanding and forecasting. The solution to the problem is to transform the time series data so that it becomes stationary. If the non-stationary process is a random walk with or without a drift, it is transformed to stationary process by differencing. On the other hand, if the time series data analyzed exhibits a deterministic trend, the spurious results can be avoided by detrending. Sometimes, the non-stationary series may combine a stochastic and deterministic trend at the same time. To avoid obtaining misleading results, both differencing and detrending should be applied, as differencing will remove the trend in the variance and detrending will remove the deterministic trend.

3.2.3 Notes: Regression Analysis (cont.)

Auto-correlation

In finance, serial correlation is used by technical analysts to determine how well the past price of a security predicts the future price.

Durbin Watson test

The Durbin Watson d statistic is given as:

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2},$$

where e is the residual from the regression at time t, and T is the number of observations in a time series. The critical values of the DW table can be found at a number of websites, such as http://www.stat.ufl.edu/~winner/tables/DW_05.pdf. Note that most DW tables assume a constant term in the regression and no lagged dependent variables.

The DW statistic, d, is approximately equal to $2(1 - r)$, where r is the sample autocorrelation of the residuals. If $d = 2$, it indicates no auto-correlation. The value of d lies between 0 and 4. If $d > 2$, successive error terms are, on average, much different in value from one another – i.e. negatively correlated.

To test for positive autocorrelation at significance α, the test statistic d is compared to lower and upper critical values ($d_{L,a}$ and $d_{U,a}$):

- If $d < d_{L,a}$ there is statistical evidence that the error terms are positively auto-correlated.
- If $d > d_{U,a}$ there is no statistical evidence that the error terms are positively auto-correlated.
- If $d_{L,a} < d < d_{U,a}$ the test is inconclusive.

Positive serial correlation is the serial correlation in which a positive error for one observation increases the chances of a positive error for another observation.

To test for negative auto-correlation at significance α, the test statistic $(4 - d)$ is compared to lower and upper critical values ($d_{L,a}$ and $d_{U,a}$):

- If $(4 - d) < d_{L,a}$ there is statistical evidence that the error terms are negatively auto-correlated.
- If $(4 - d) > d_{U,a}$ there is no statistical evidence that the error terms are negatively auto-correlated.
- If $d_{L,a} < (4 - d) < d_{U,a}$ the test is inconclusive.

Positive serial correlation is a time series process in which positive residuals tend to be followed over time by positive error terms and negative residuals tend to be followed over time by negative residuals. (Positive serial correlation is thought to be more common than the negative case.)

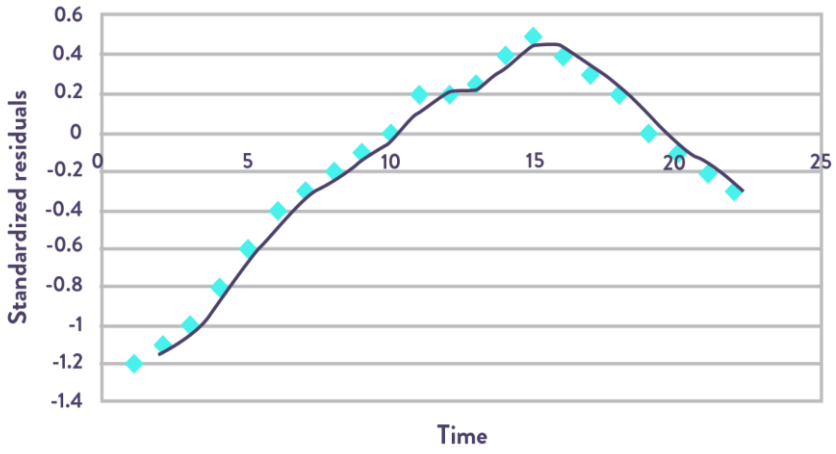


Figure 5: Positive serial correlation

Negative serial correlation example:

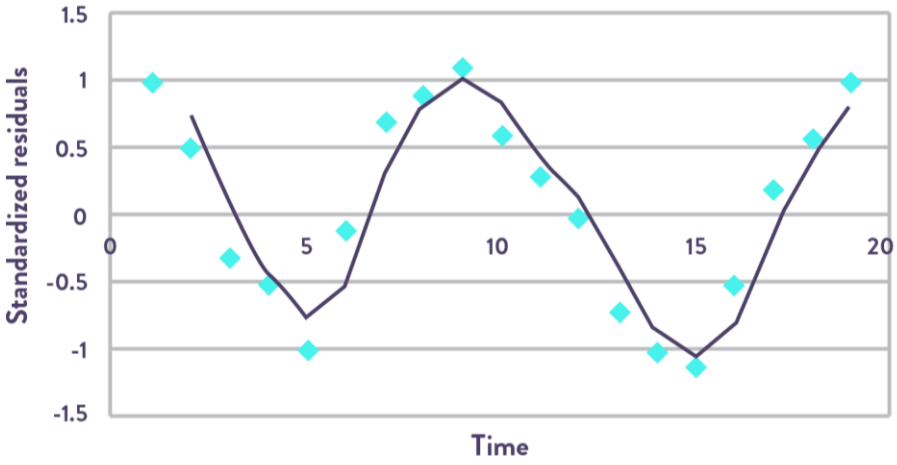


Figure 6: Standardized residuals

Hypothesis testing with the DW

Suppose you had 30 observations, ($n = 30$), and two independent variables in the regression. You use the 5% level of confidence. From the DW table you should find the lower bound (d_L) of the table equal to 1.28 and the upper bound (d_U) equal to 1.57. If the statistic that is produced by the regression is.4 (below the lower bound) we would reject the null hypothesis below in favor of the alternative:

- H_0: no auto-correlation,
- H_1: positive auto-correlation.

Let's say instead that the computed regression DW stat was 1.45. This is the "inconclusive" area of the DW analysis and we cannot say one way or the other. Suppose the DW statistic from the regression was 1.90; then we do not reject the null hypothesis that there exists no serial correlation.

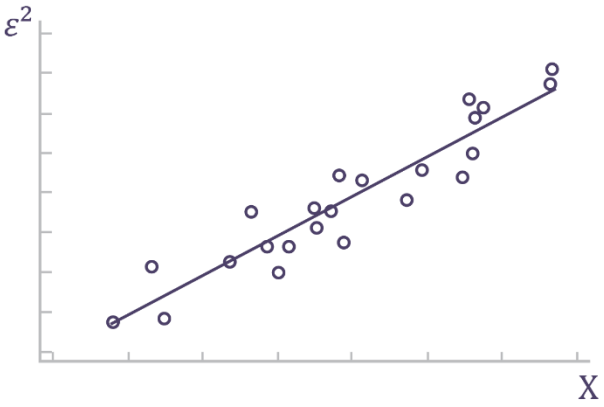
Heteroscedasticity

One of the assumptions of the classic linear regression model is homoscedasticity (same dispersion of errors). The error variances are constant. We have

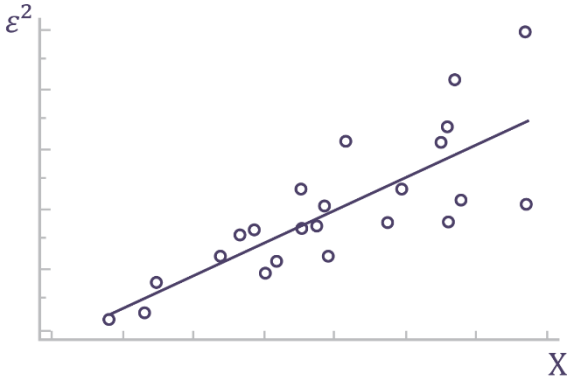
heteroscedasticity when unequal dispersion occurs. Unequal variances of the error terms ε_i, occurs when

$$\text{var}(\varepsilon_i) = \text{E}[(\varepsilon_i)^2\,] - [E(\varepsilon_i)]^2 \;= E[(\varepsilon_i)^2\,] = (\sigma_i)^2.$$

Homoscedasticity:



Heteroscedasticity:



The model we are regressing is

$$\text{Model: } y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon^2.$$

Homoscedasticity occurs when:

$$\text{Var}\,(\varepsilon|x_1, \ldots, x_k) = \sigma^2.$$

For heteroscedasticity:

$$\mathrm{Var}(\varepsilon|x_1, \ldots, x_k) = \sigma^2 \mathrm{f}(x_1, \ldots., x_k).$$

For example:

$$\mathrm{f} = \delta_0 + \delta_1 \mathrm{x}_1 + \delta_2 \mathrm{x}_4.$$

Note that only the first and forth regressors are implicated in the heteroscedasticity.

In some instances, only one regressor may be responsible for the heteroscedasticity so we can observe the relationship graphically.

For example, consider a 4x4 matrix:

$$\begin{matrix} \sigma, & 0, & 0, & 0 \\ 0, & \sigma, & 0, & 0 \\ 0, & 0, & \sigma, & 0 \\ 0, & 0, & 0, & \sigma \end{matrix}$$

Sigma is constant (homoscedasticity or constant variance).

For heteroscedasticity,

$$\mathrm{Var}(\mathrm{u_i}|\,\mathrm{X_i}) = \sigma_\mathrm{i}^2.$$

If heteroscedasticity is present, the variance is not stationary so the above matrix would be written as:

$$\begin{matrix} \sigma_1, & 0, & 0, & 0 \\ 0, & \sigma_2, & 0, & 0 \\ 0, & 0, & \sigma_3, & 0 \\ 0, & 0, & 0, & \sigma_4 \end{matrix}$$

In the bivariate relationship in the graph below, the error terms increase as x increases.

Tests for heteroscedasticity

Graph the residuals against suspected explanatory variables as below:

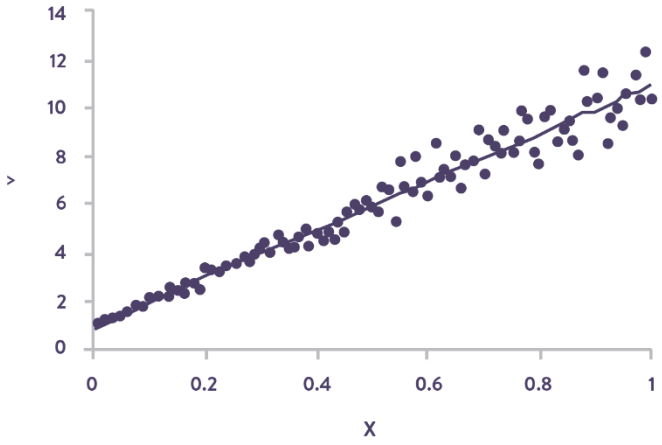


Figure 7: Detecting heteroscedasticity

In this instance the variance increases as x increases.

The problem here is that the heteroscedasticity may not be the result of one variable but instead from a combination of the regressors:

$$\text{Var}\,(\varepsilon|x_1, \ldots, x_k) = \sigma^2 f(x_1, \ldots., x_k).$$

For example:

$$f = \delta_0 + \delta_1 x_1 + \delta_2 x_4 + \delta_3 x_5 + \delta_4 x_7.$$

The White test states that if disturbances are homoscedastic then squared errors are, on average, constant. Thus, regressing the squared residuals against explanatory variables should result in a low R squared.

Steps of the White test:

a. Regress Y against your various explanatory variables using OLS.
b. Compute the OLS residuals, $e_1 \ldots e_n$, and square them.
c. Regress squared residuals e_1^2 against a constant, all of the explanatory variables, their squares, and possible interactions (x_1 times x_2) between the explanatory variables (p slopes total).
d. Compute R^2 from (c)
e. Compare $\mathrm{n}R^2$ to the critical value from the Chi-squared distribution with p degrees of freedom.

Breusch–Pagan test

The Breusch-Pagan Lagrange-Multiplier (LM) test is the same as the White test except that the econometrician selects the explanatory variables to include in the auxiliary equation. This may, or may not, result in a more powerful test than the White test.

For example, suppose the form chosen is:

$$\text{Squared residuals} = \beta_0 + \beta_1 \mathrm{X} + \beta_2 \mathrm{X}^2 + \beta_3 \mathrm{Z}^2 + \beta_4 \mathrm{XZ}.$$

Again, apply the Chi-squared test but $p = 4$. Compare $\mathrm{n}R^2$ to the critical value from the Chi-squared distribution with p degrees of freedom.
The Breusch-Pagan F test is as follows: Suppose the auxiliary equation being tested is

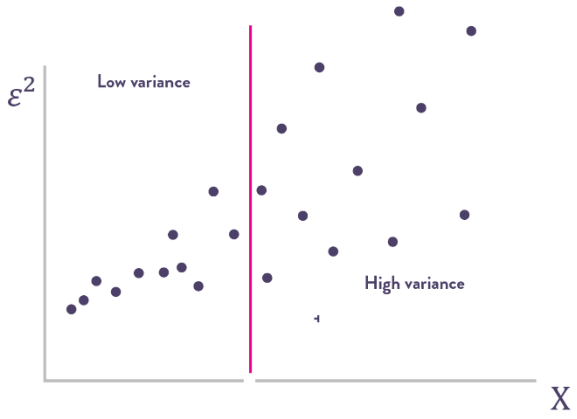
$$\text{Aux: } \varepsilon^2 = \delta_0 + \delta_1 \mathrm{x}_1 + \ldots + \delta_k \mathrm{x}_\mathrm{p} + \varepsilon \qquad (p \text{ slope terms})$$

Using this equation, calculate the following F:

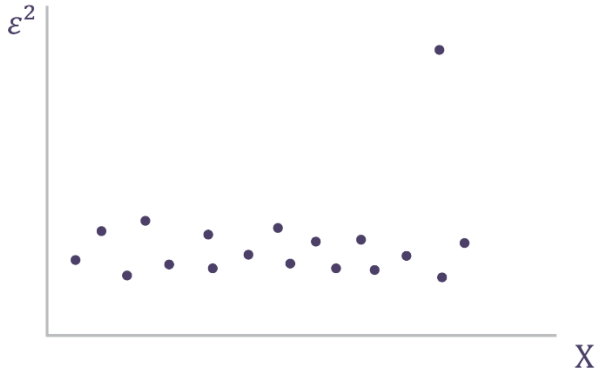
$$F_c = \frac{\frac{R^2}{p}}{\frac{(1-R^2)}{n-p-1}}$$

If $F_c > F_{p,\ n-p-1}$, reject the null hypothesis of homoscedasticity.

Graphing the squared errors against a suspected cause of the heteroscedasticity may reveal the nature of the heteroscedasticity. In the case below, the variance of the error terms increases as the exogenous variable, x, increases:



Graphical inspection may help to identify situations in which an outlier causes a test to show erroneously the presence of heterogeneity as in the graph below:



In this case, the squares of the errors are fairly constant, except for the one outlier.

Multicollinearity

Multicollinearity is a condition where two or more independent variables are strongly correlated with each other. In an extreme case, called perfect multicollinearity, one may be a multiple of the other $z = 3 * x$ or variable z is exactly 3 times variable x. OLS cannot estimate both z and x.

- Problem: $x1 = 2 + 3 * x2$. Does perfect multicollinearity exist?
- Problem: you have two measures of gold produced in Chile over a span of years. In one year, data assert that 1 000 kg were produced. Another data source states that for the same year 2 204.62 pounds were produced. (There are 2.20462 pounds in a kilogram.)
- The variables may measure the same concepts.
- The existence of multicollinearity is not a violation of the OLS assumption.

Symptoms of multicollinearity may be observed in the following situations:

- Small changes in the data cause wide swings in the parameter estimates. This is a big problem.
- Coefficients may have very high standard errors and low significance levels.
- Coefficients may have the "wrong" sign or implausible magnitude.

Where does it occur in finance?

Imagine you are trying to estimate the amount of company investment. You have two variables: $x =$ income of the company, and $z =$corporate income tax paid by the company. Are they likely to be multicollinear? What about the current and quick ratios of firms?

Detecting multicollinearity

1 Check for the correlation between the predictor variables. If it is high, this may be a MC warning.

2 Construct VIF (variance inflation factor) by regressing a predictor variable against all the other predictor variables you are using. If it is greater than 5, this may be another sign of MC.

Correcting

1 Remove one of the offending variables if the other can stand in for it. You have to do this if there exists perfect multicollinearity.

2 Standardize the predictor variables.

Variance inflation factor (VIF)

To compute VIF, the auxiliary regressions of each independent variable on all the other K-independent variables (regressors) are performed. For the variable i, the R-squared is R_i. Then the VIF for independent variable j is defined as:

$$VIF_i = \frac{1}{(1-R_i^2)}.$$

For example, the regression:

$$y = \beta + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.$$

Regressions are:

$$x_1 = \gamma_1 + \gamma_{12} x_2 + \gamma_{13} x_3$$
$$x_2 = \gamma_2 + \gamma_{21} x_1 + \gamma_{23} x_3$$
$$x_3 = \gamma_3 + \gamma_{31} x_1 + \gamma_{32} x_2$$

Example:

$$R^2 = .2$$
$$\text{VIF} = 1/(1-.2) = 1/.8 = 1.25.$$

In contrast, if the Pearson product moment correlation coefficient is high, say .
95, VIF is

$$1/(1-.9) = 1/(.05) = 20.$$

Diagnostics

- As a rule of thumb, collinearity is potentially a problem for values of VIF > 10 (sometimes VIF > 5).
- The average VIF is the average VIF_i across the K independent variables. If the VIFs are .2, .4, .1, and .8, the average VIF is $(.2 + .4 + .1 + .8)/4 = .375$.
- Rule of thumb: an average VIF > 1 indicates multicollinearity.
- The square root of the VIF considers how much larger the standard error is as compared to the situation when variables were not correlated with other predictor variables,. For example, if $VIF_i = 6$ then sqrt(VIF_i) = 2.446. So, the standard error for the coefficient of that predictor variable is 2.446 times as large as it would be if that (i^{th}) predictor variable were uncorrelated with the other predictor variables.
- Tolerance: Tolerance $(\beta i) = 1/VIF_i = 1 - R_i^2$.
- If a variable has a VIF = 10 then the (low) tolerance = 1/10 = .1, indicates MC, whereas VIF = 1 indicates a high tolerance of 1/1 = 1.

Condition number

The condition number (often called kappa) of a data matrix is $\kappa(X)$ and measures the sensitivity of the parameter estimates to small changes in the data. It is calculated by taking the ratio of the largest to the smallest singular values from the singular value decomposition of X. Kappa is

$$\sqrt{\left(\frac{maxeigenvalue(X^T X)}{mineigenvalue(X^T X)}\right)}.$$

A condition number above 30 is considered to be an indication of multicollinearity but some say as low as 15. (Again, these are rules of thumb).

Natural logarithm of financial variables

Many research papers use the natural logarithm (*ln*) of macroeconomic and financial variables instead of the simple value in the analysis. I will present the benefits of using *ln* in data analysis below.

Note that ≈ means approximately equal.

1. Change in *ln* ≈ percentage change.
   $ln(1 + x) \approx x$
   When $x < 1$.
2. *ln* transformation converts the exponential growth pattern to a linear growth pattern. It is easier to work with linear relations in finance as compared to non-linear relations. The advantage of *ln* transformation is that it converts non-linear sophisticated relations into linear and easier to model connections between the variables.
3. Trend measured in *ln* units ≈ percentage growth.
4. Coefficients in log-log regressions ≈ proportional percentage changes.

Example 1

We want to see the relation between GDP and M1 money supply in the U.S.

Variables:

M1 for the United States, Index 2009=100 of (National Currency), Annual, Not Seasonally Adjusted

Real Gross Domestic Product, Index 2009=100 of (Index 2009=100), Annual, Not Seasonally Adjusted

Frequency: Annual data

Period: 1980 – 2014

Source: FRED database

According to Monetary Theory:

When M1 increases, the GDP increases;

When M1 decreases, the GDP decreases.

Now, we calculate this connection using a linear regression in R. GDP will be the dependent variable while M1 will be the independent variable.

GDP = coefficient*M1.

Coefficient must be estimated in the linear model.

We use *ln* values in the model as they have a much smoother evolution as compared to raw variables.

R code:

```
#gdp and m1 are introduced as arrays
linear_regression=lm(gdp~m1+0)
linear_regression
summary(linear_regression)
```

File: **click here** to download *R_code_GDP_M1 that can be found in the additional files folder*

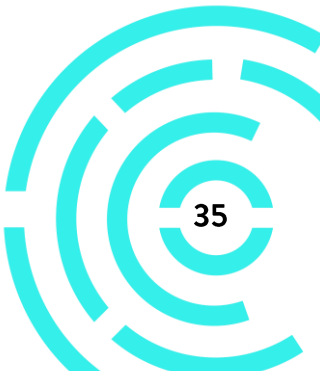
Regression results provided by R code:

```
Call:
lm(formula = gdp ~ m1 + 0)
Coefficients:
  m1
1.031
```

This means that GDP=1.031*M1.

Constant is assumed to be zero.

This result shows a direct correlation between M1 and GDP.

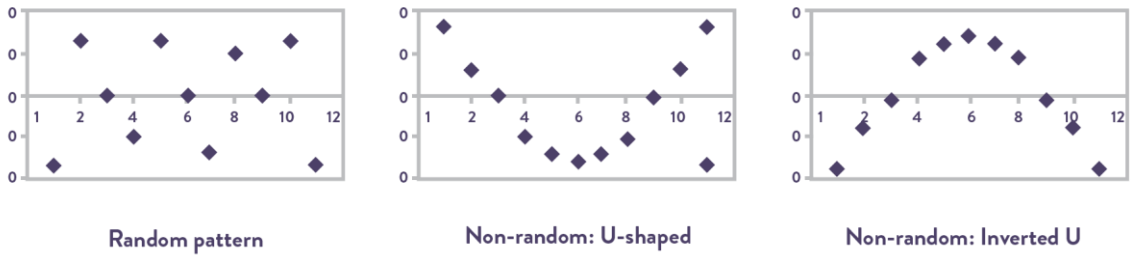


Please note that macroeconomic variables are usually introduced using first difference in models $(x(t) - x(t-1))$ as they are not generally stationary variables. We used them without differencing in order to simplify the model. The concept of stationarity was already presented in detail in previous chapters.

Residuals and predicted values

It is very frequently the case that we use residuals and predicted values in data analysis. The analysis of regression residuals may give us some information regarding the accuracy of predictions, or the type of model used. If the regression residuals are normally distributed, then the prediction intervals are considered to be accurate. If the regression residuals are not normal, then the prediction intervals are inaccurate.

If the residuals have a random pattern, a linear regression is a good fit. If the residuals are non-random and have a U-shape or an inverted U shape, the variables can be modeled better using a non-linear model.



Python code:

In [ ]:

```
import numpy as np
import statsmodels.api as sm

# New Zealand's economic growth

y=[3.4, 3.0, 2.6, 2.8, 2.8, 3.2, 3.5, 3.4, 3.0, 2.0, 0.8, -0.5, -1.6, -2.0, -2.0, -1.4,

   -0.3, 0.8, 1.6, 1.6, 1.4, 1.0, 1.1, 1.8, 2.2, 2.6, 2.5, 2.4, 2.2, 2.1, 2.4, 2.2, 2.5, 2.8, 3.0, 3.3, 3.2]

# house price evolution in New Zeeland

x=[12.3, 10.4, 9.9, 9.6, 11.2, 13.4, 11.4, 7.7, 2.7, -4.5, -6.7, -9.0, -9.1, -3.1, 1.1, 5.4,
   6.3, 3.1, 0.2, -1.6, -1.2, 0.8, 2.6, 2.8, 3.2, 4.0, 4.8, 6.9, 7.7, 9.0, 10.2, 9.2, 7.9, 7.1, 4.8, 6.3, 8.4]

x=sm.add_constant(x)                #add constant in the model
results=sm.OLS(y,x).fit()           #implement regression estimated by OLS method
results.params
residuals=sm.OLS(y,x).fit().resid   #obtain residuals
residuals
y_pr=results.predict()              #obtain predicted values
y_pr

```

3.2.4 Transcript: Homoscedasticity vs Heteroscedasticity

In this video, we will be comparing homoscedastic and heteroscedastic data generating processes. Homoscedasticity means that the variance-covariance matrix of the unexplained part of your variable of interest, ε, is a constant diagonal matrix in the population. One of the assumptions of the classic linear regression model is homoscedasticity, or in other words, the same dispersion of errors. Heteroscedasticity, on the other hand, is when unequal dispersion of errors occurs.

Consider a very simple case:

Process 1 is homoscedastic:

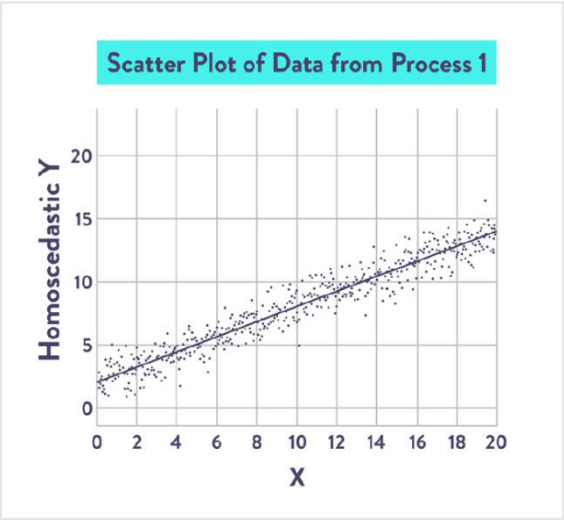
$$y_i = 2 + 0.6x_i + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

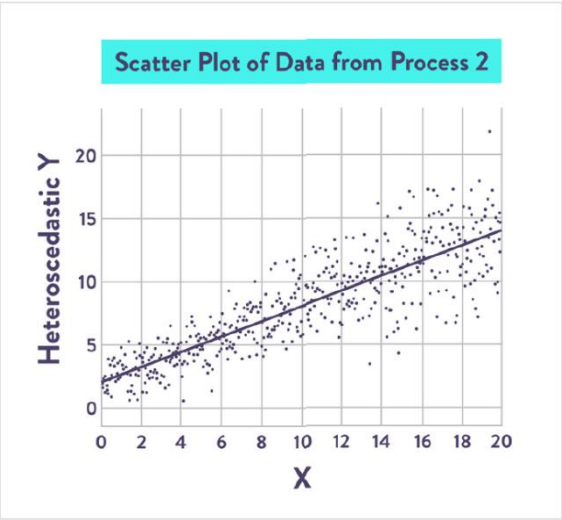
Process 2 is heteroscedastic:

$$y_i = 2 + 0.6x_i + \varepsilon_i$$
$$\varepsilon_i \sim N(0, (1 + x_i)\sigma^2)$$

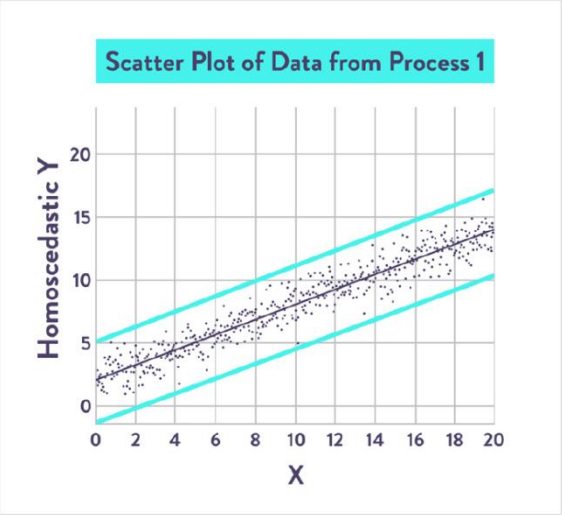
Note that the variance of the error increases with the size of x_i.

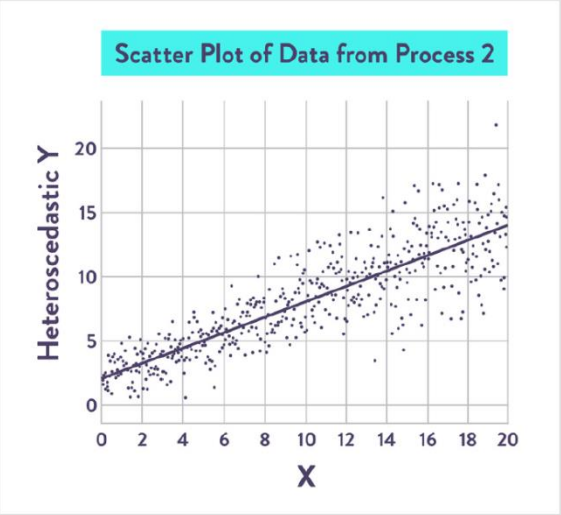
When we compare scatter plots of data from Process 1 to Process 2, there can clearly be seen that Process 2 as an unequal dispersion of errors.



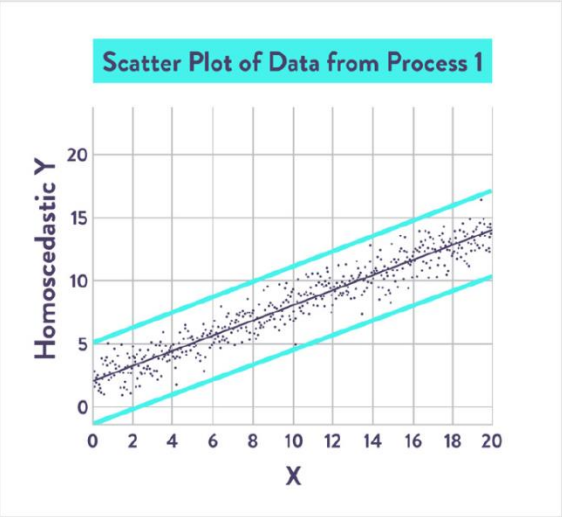


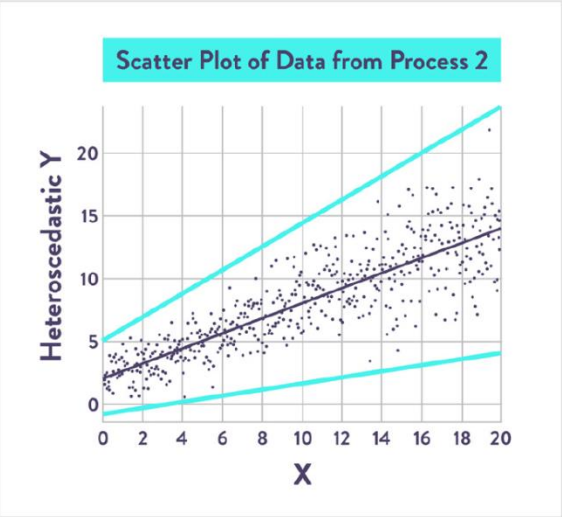
Notice in the left-hand graph how the variability of the error around the true expected value (grey line) is constant across X. Most of the points are close to the line and the dispersion around the line does not obviously vary with X.



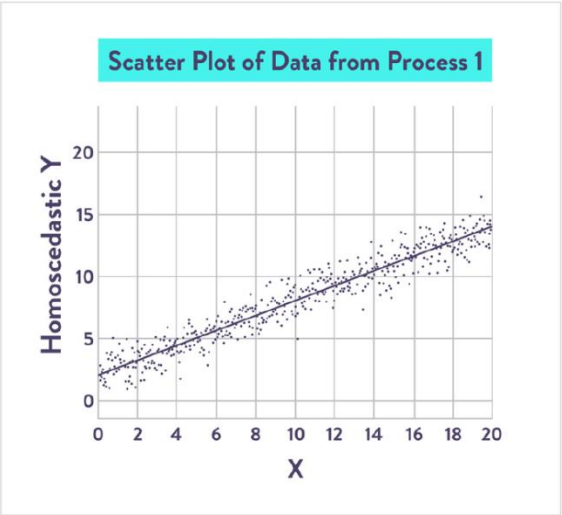


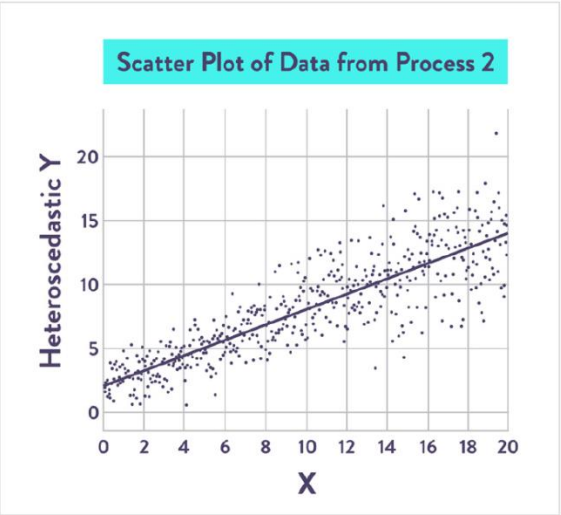
Notice in the right-hand graph how the variability of the error around the true expected value (grey line) increases with X. Most of the points are still close to the line, but the dispersion around the line obviously increases with X.



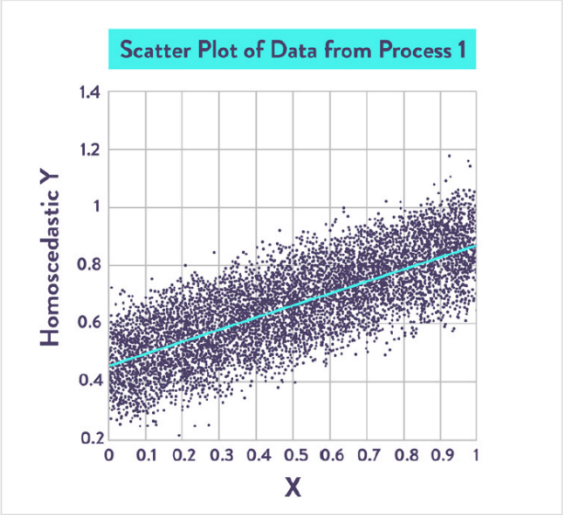


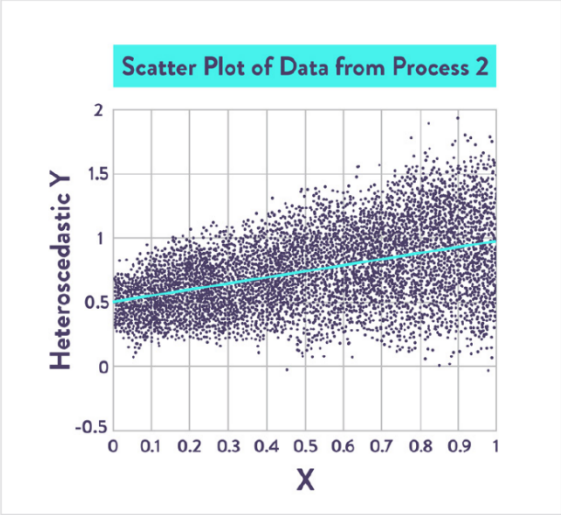
In this simulated example it is clear to see that there is enough data to accurately identify the true relationship.



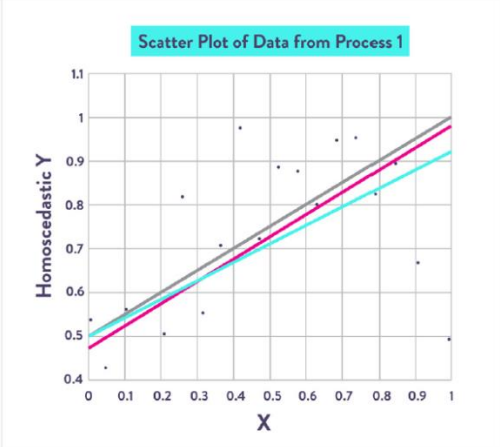


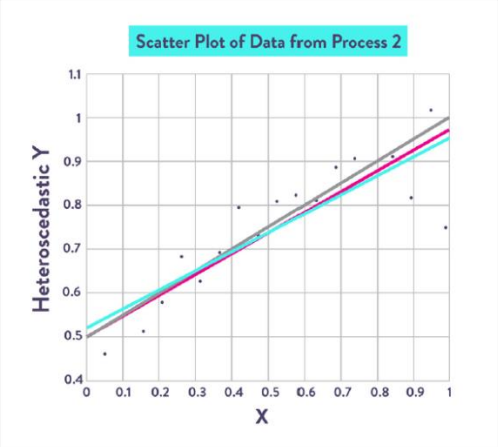
When there is a very large data set ($n = 1000$) it is obvious that there is enough information to accurately infer. Here you see an OLS estimate (cyan) and a GLS estimate (magenta) of the true slope, but they are so similar you can only see the top one. Both are consistent, so with this amount of data they will give equally reliable inference.



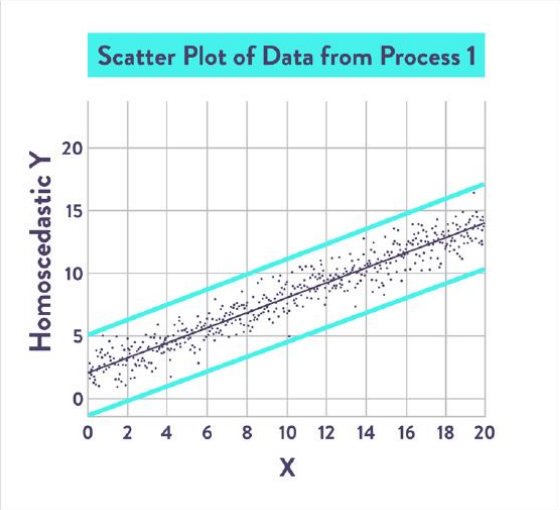


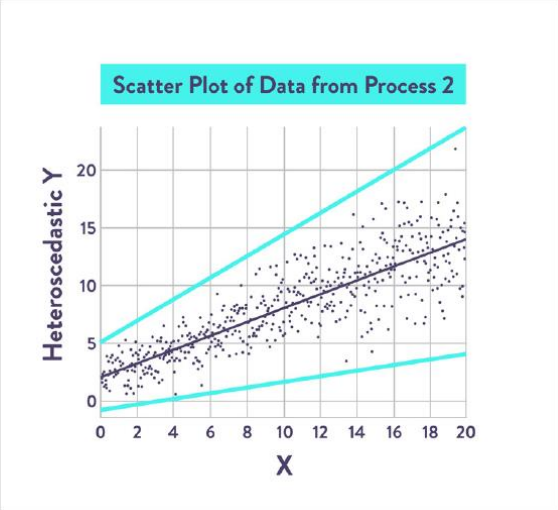
When we have a very small data set ($n = 20$), the estimates can be quite different (OLS estimates are cyan, GLS estimates are magenta). However, at such a small sample, one would be unwilling to trust either model.



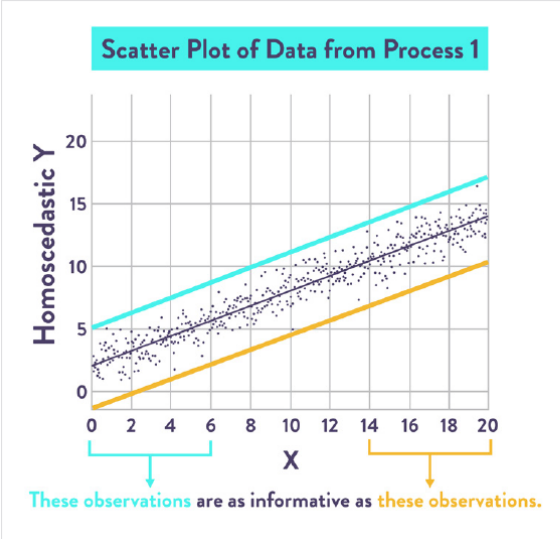


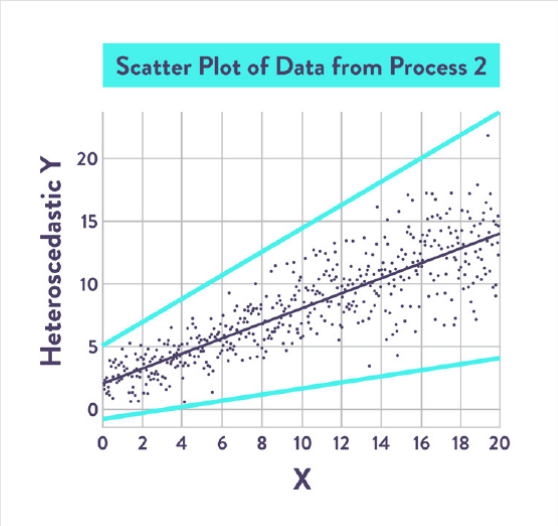
You might ask: Why does heteroscedasticity matter? It is about how much information there is in a specific observation.



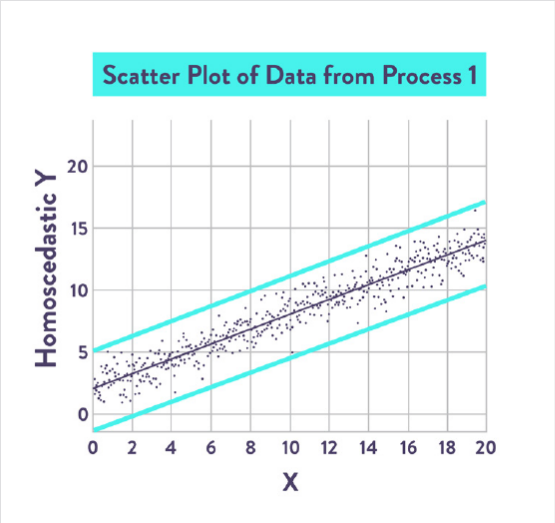


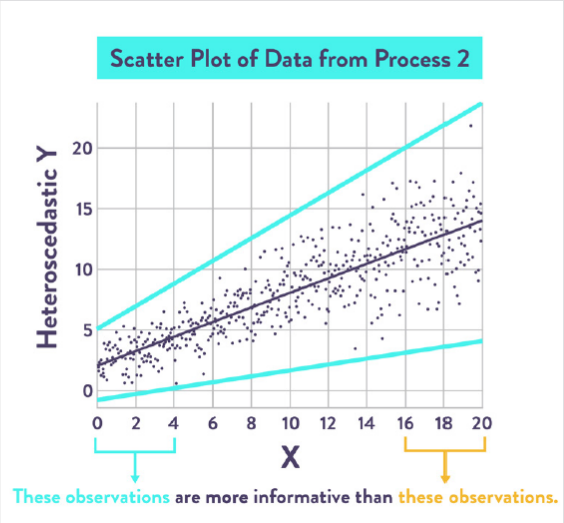
Under homoscedasticity:





Under heteroscedasticity:





In conclusion, the information content of an observation is important. Doing so statistically is very difficult. GLS estimators tend to have poor small sample properties – they only work well when the sample is very large. The standard practice is to use OLS estimators for the slope but robust estimators for the variance-covariance matrix of the slope coefficients.

This video compared homoscedasticity to heteroscedasticity. Study the notes to learn more about homoscedasticity.

3.2.5 Notes: Forecasting Market Trend Using Logit Regression

Logit regression is used when estimating a binary, or dummy, dependent variable. The dependent variable, y, can assume values of 1 or 0. Note that OLS is not considered appropriate in this case. The functional form is

$$\mathrm{P} = \Pr(y = 1|x) = \frac{e^{\alpha+\beta \mathrm{x}}}{1 + e^{\alpha+\beta \mathrm{x}}}$$

which assume one independent variable x. This is usually referred to as a success if $\mathrm{y} = 1$ and a failure if $\mathrm{y} = 0$. The probability of a failure is

$$1 - \mathrm{P} = \Pr(\mathrm{y} = 0|\mathrm{x}) = \frac{1}{1 + e^{\alpha+\beta \mathrm{x}}}$$

To derive the logit, the odds of success to failure are taken as follows:

$$\mathrm{Odds} = \frac{\mathrm{P}}{1 - \mathrm{P}} = \mathrm{e}^{\alpha+\beta \mathrm{x}}.$$

The logit uses the log of the odds, or

$$\ln(\mathrm{odds}) = \ln \frac{\mathrm{P}}{1 - \mathrm{P}} = \alpha + \beta \mathrm{x}.$$

Financial assets have a stochastic behaviour. Predicting their trend is a challenge for quantitative analysts. The logistic regression can be used for predicting market trends as it provides a binary result. This type of regression is a probabilistic model which assigns probability to each possible event.

Dow Jones Index (DJI) trend is predicted using logistic regression implemented in R.

Algorithm:

1. DJI data is extracted from Yahoo Finance

2. Different indicators are calculated (moving average, standard deviation, RSI and MACD, Bollinger band)

3. Create variable direction
   Up(1) or down (0)
   Current price > 20 days previous price -> Up direction
   Current price < 20 days previous price -> Down direction

4. Data is divided in two parts
   In-sample data -> Model building process
   Out-sample data -> Evaluation
   In-sample and out-sample start and end dates are indicated

5. Data is standardized in order to avoid the higher scaled variables which have a higher impact on the results
   Standardized data = (X-Mean(X))/Std(X)
   Mean and standard deviation is calculated for each column

6. Logit regression is implemented

#FORECASTING MARKET TREND USING LOGISTIC REGRESSION

#Copy paste the code in R

```
library("quantmod")
getSymbols("^DJI",src="yahoo")

dow_jones<- DJI[,"DJI.Close"]
dow_jones
average10<- rollapply(dow_jones,10,mean)
```

```
average10
average20<- rollapply(dow_jones,20,mean)
average20
std10<- rollapply(dow_jones,10,sd)
std20<- rollapply(dow_jones,20,sd)
rsi5<- RSI(dow_jones,5,"SMA")
rsi14<- RSI(dow_jones,14,"SMA")
macd12269<- MACD(dow_jones,12,26,9,"SMA")
macd7205<- MACD(dow_jones,7,20,5,"SMA")
bollinger_bands<- BBands(dow_jones,20,"SMA",2)
direction<- NULL
direction[dow_jones> Lag(dow_jones,20)]<- 1
direction[dow_jones< Lag(dow_jones,20)]<- 0

dow_jones<-
cbind(dow_jones,average10,average20,std10,std20,rsi5,rsi14,macd12269,m
acd7205,bollinger_bands,direction)

dimension<- dim(dow_jones)
dimension
issd<- "2010-01-01"
ised<- "2014-12-31"
ossd<- "2015-01-01"
osed<- "2015-12-31"
isrow<- which(index(dow_jones) >= issd& index(dow_jones) <= ised)
osrow<- which(index(dow_jones) >= ossd& index(dow_jones) <= osed)
```

```
isdji<- dow_jones[isrow,]
osdji<- dow_jones[osrow,]
isme<- apply(isdji,2,mean)
isstd<- apply(isdji,2,sd)
isidn<- matrix(1,dim(isdji)[1],dim(isdji)[2])
norm_isdji<- (isdji - t(isme*t(isidn))) / t(isstd*t(isidn))
dm<- dim(isdji)
norm_isdji[,dm[2]] <- direction[isrow]
```

```
formula<- paste("direction ~ .",sep="")
model<- glm(formula,family="binomial",norm_isdji)
```

The Linear Probability Model

The **Linear Probability Model** (LPM) is an easier alternative to Logit or Probit. Suppose the dependent variable can only assume values of 0 and 1. What if we still want to use a multiple linear regression model?

$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + u_i$

Y_i can take values of 0 and 1.

$b_j = \overline{1,2}$

b_j - coefficient

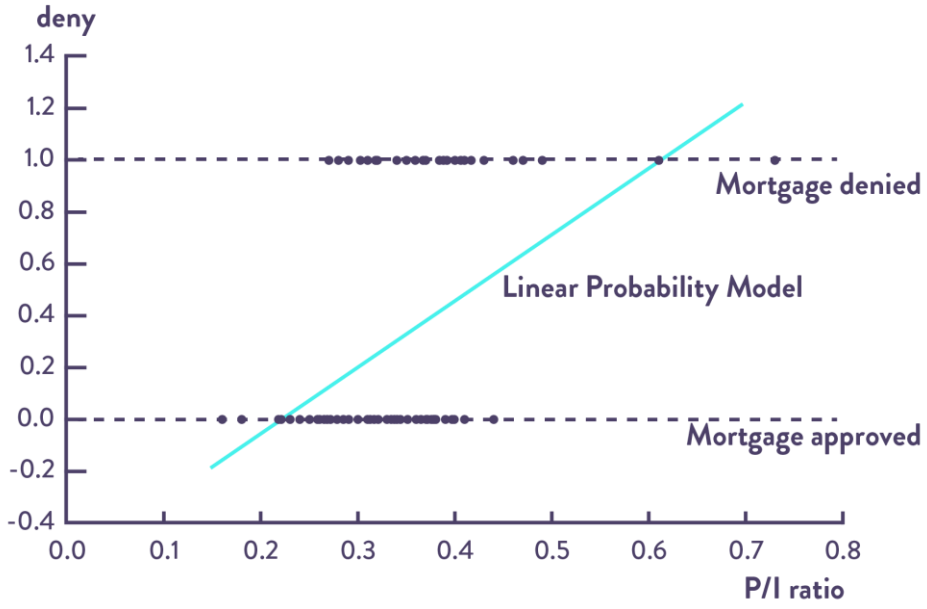
$E(Y_i|X) = b_0 + b_1X_{1i} + b_2X_{2i} = P(Y_i = 1)$ Response Probability

Predicted probability for $b_0^* + b_1^*X_{1i} + b_2^*X_{2i}$ can be outside of the range $0, 1$.

Example:

Linear probability model, Home Mortgage Disclosure Act (HMDA) data

Mortgage denial v. ratio of debt payments to income (P/I ratio) in a subset of the HMDA data set $(n = 127)$



Source: people.ku.edu/~p112m883/pdf/Econ526/Ch11Part1_slides.docx

Example:

$$Y = -0.40 + 0.09X_1 + 0.14X_2$$

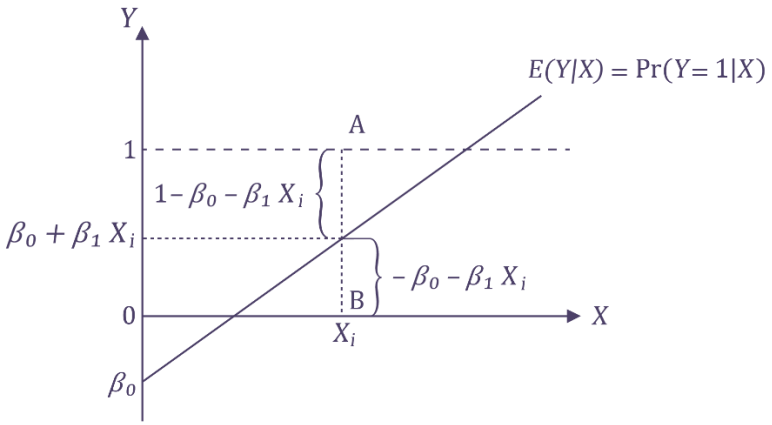
One unit increase of X_2 increases the probability of occurrence of Y by 0.14.

Advantages of the LPM:

- It is computationally simpler.
- It is easier to interpret the "marginal effects".
- It avoids the risk of mis-specification of the "link function".
- There are complications with Logit or Probit if you have endogenous dummy regressors.
- The estimated marginal effects from the LPM, Logit, and Probit models are usually very similar, especially if you have a large sample size.

Disadvantage of the LPM:

- LPM yields biased and inconsistent estimates in most situations. See Horrace and Oaxaca (2006), and Amemiya (1977).



3.2.6 Transcript: Interpreting and Evaluating Regression Output

In this video, we will be interpreting and evaluating regression output from the true model

$$y_i = 0.5 + 0.5x_{1,i} - 0.3x_{2,i} + 0.4x_{3,i} + \varepsilon_i$$

with: $\varepsilon_i \sim N(0, (0.8)^2)$.

You can reproduce a version of the results with the following code. Since we are using random number generators, your results will differ slightly. This is a good time to consider the impact that sampling variation may have, even under the best of circumstances. All variables are normally distributed and independent. Note that we add a fourth unrelated random variable to the regression to show an example of an insignificant estimate.

```
# generate variables:
x1 <- rnorm(100)
x2 <- rnorm(100)
x3 <- rnorm(100)
x4 <- rnorm(100)
epsilon <- 0.8*rnorm(100)
y = 0.5 + 0.5*x1 - 0.3*x2 + 0.4*x3 + epsilon
```

```
regression = lm(formula = y ~ x1 + x2 + x3 + x4)
summary(regression)
```

We will be interpreting and evaluating the different sections of the estimation output to reach a conclusion about the model.

Consider the following ordinary least squares estimation output:

```
Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept)  0.41131    0.07502   5.483 3.45e-07 ***

x1           0.54475    0.08293   6.569 2.70e-09 ***

x2          -0.37683    0.09643  -3.908 0.000175 ***

x3           0.38048    0.11046   3.444 0.000854 ***

x4          -0.02449    0.05117  -0.479 0.633380

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1



Residual standard error: 0.7446 on 95 degrees of
freedom

Multiple R-squared:  0.4241,  Adjusted R-squared:
0.3999

F-statistic: 17.49 on 4 and 95 DF,  p-value: 8.709e-11
```

These are the estimates of the parameters. Notice that the constant and the slope parameters on the first three variables are all close to their true values.

```
Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept)  0.41131    0.07502   5.483 3.45e-07 ***

x1           0.54475    0.08293   6.569 2.70e-09 ***

x2          -0.37683    0.09643  -3.908 0.000175 ***

x3           0.38048    0.11046   3.444 0.000854 ***

x4          -0.02449    0.05117  -0.479 0.633380

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1



Residual standard error: 0.7446 on 95 degrees of
freedom

Multiple R-squared:  0.4241,  Adjusted R-squared:
0.3999

F-statistic: 17.49 on 4 and 95 DF,  p-value: 8.709e-11
```

These are the standard errors of the estimated coefficients. Notice that the difference between the estimated coefficients of the variables that belong in the model are less than two times the standard error. This means that any hypothesis at the true values will not be rejected.

```
 Coefficients:

             Estimate Std. Error t value Pr(>|t|)

(Intercept)  0.41131    0.07502   5.483 3.45e-07 ***

x1            0.54475    0.08293   6.569 2.70e-09 ***

x2           -0.37683    0.09643  -3.908 0.000175 ***

x3            0.38048    0.11046   3.444 0.000854 ***

x4           -0.02449    0.05117  -0.479 0.633380

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1



Residual standard error: 0.7446 on 95 degrees of
freedom

Multiple R-squared:  0.4241,  Adjusted R-squared:
0.3999

F-statistic: 17.49 on 4 and 95 DF,  p-value: 8.709e-11
```

These are the t-statistics of the test for individual significance for each of the estimates. Each is simply the ratio of the estimate and its standard error.

```
 Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept)  0.41131    0.07502   5.483 3.45e-07 ***

x1            0.54475    0.08293   6.569 2.70e-09 ***

x2           -0.37683    0.09643  -3.908 0.000175 ***

x3            0.38048    0.11046   3.444 0.000854 ***

x4           -0. 02449    0.05117  -0.479 0.633380

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1



Residual standard error: 0.7446 on 95 degrees of
freedom

Multiple R-squared:  0.4241,  Adjusted R-squared:
0.3999

F-statistic: 17.49 on 4 and 95 DF,  p-value: 8.709e-11
```

These are the probability values of the t-tests for individual significance for each of the estimates. These are each the answer to the question: If the true coefficient was 0, how likely is it that we would obtain an estimate of the given size purely from sample variation?

```
 Coefficients:

             Estimate Std. Error t value Pr(>|t|)

(Intercept)  0.41131    0.07502   5.483 3.45e-07 ***

x1            0.54475    0.08293   6.569 2.70e-09 ***

x2           -0.37683    0.09643  -3.908 0.000175 ***

x3            0.38048    0.11046   3.444 0.000854 ***

x4           -0. 02449     0.05117  -0.479 0.633380

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1



Residual standard error: 0.7446 on 95 degrees of
freedom

Multiple R-squared:  0.4241,  Adjusted R-squared:
0.3999

F-statistic: 17.49 on 4 and 95 DF,  p-value: 8.709e-11
```

The R-squared measures how much of the variation in y is explained by the 4 explanatory variables. The adjusted R-squared is the same but penalized for the number of explanatory variables. This is a better model diagnostic, as the R-squared is always increasing in the number of explanatory variables, so it may lead a researcher to include irrelevant variables.

```
 Coefficients:

             Estimate Std. Error t value Pr(>|t|)

(Intercept)  0.41131    0.07502   5.483 3.45e-07 ***

x1            0.54475    0.08293   6.569 2.70e-09 ***

x2           -0.37683    0.09643  -3.908 0.000175 ***

x3            0.38048    0.11046   3.444 0.000854 ***

x4           -0. 02449    0.05117  -0.479 0.633380

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1



Residual standard error: 0.7446 on 95 degrees of
freedom

Multiple R-squared:  0.4241,  Adjusted R-squared:
0.3999

F-statistic: 17.49 on 4 and 95 DF,  p-value: 8.709e-11
```

Is 42% of the variation explained enough? It depends on the context. In the cross section, it is reasonably acceptable. In time series it would be far too low in many applications.

Just like the t-tests evaluate the statistical significance of each explanatory variable individually, the F-test evaluates the joint statistical significance of all explanatory variables (other than the constant). It has a degree of freedom, namely 4,95.

The 4 is the 4 zero restrictions we place on the model, that is the slope coefficients are all zero. The 95 refers to the sample size of 100 minus the 5 parameters estimated in the model – 4 slope coefficients and 1 constant. The p-value is the probability of obtaining an F-statistic as large as we did if the true F-value is zero. We clearly reject this and the hypothesis that the slope coefficients are all zero.

```
 Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept)  0.41131    0.07502   5.483 3.45e-07 ***

x1            0.54475    0.08293   6.569 2.70e-09 ***

x2           -0.37683    0.09643  -3.908 0.000175 ***

x3            0.38048    0.11046   3.444 0.000854 ***

x4           -0. 02449     0.05117  -0.479 0.633380

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1



Residual standard error: 0.7446 on 95 degrees of
freedom

Multiple R-squared:  0.4241,  Adjusted R-squared:
0.3999

F-statistic: 17.49 on 4 and 95 DF,  p-value: 8.709e-11
```

In conclusion: Is the model acceptable?

Almost. While most of the variables are statistically significant, x4 is not. In order to get a model as parsimonious as possible, we should re-estimate the model without x4. If all the diagnostic tests remain acceptable, that would be an improved model for forecasting.

This video showed you how to evaluate regression output. In the next section we will be looking at the generalized linear model.

3.2.7 Notes: Taylor Rule and Fed's Monetary Policy

Introduction

A **Taylor rule** is a monetary-policy rule that stipulates how much the central bank should change the nominal interest rate in response to changes in inflation, output, or other economic conditions.

Several decades ago, John Taylor proposed an equation for indicating how Fed should adjust its interest rate considering the evolution of inflation and output. Nowadays, the so-called "Taylor rule" has dozens of versions that show the connection between the Central Bank interest rate and different macroeconomic indicators.
A Taylor rule can be written as following:

$$i_t = r^* + \pi^T + \alpha_x x_t + \alpha_\pi(\pi_t - \pi^T),$$

where:
i_t – federal funds interest rate in U.S.
π^T – targeted level of average inflation (Taylor assumed it to be 2%)
r^* – equilibrium interest rate (Taylor assumed it to be 2%)
x_t – output gap (the difference between GDP and potential GDP)
Potential GDP – the output level that does not generate inflationary pressures.

Implementing a Taylor rule

We want to implement a Taylor rule for the U.S. economy using a multiple regression.
r - Effective Federal Funds Rate, Percent, Quarterly, Not Seasonally Adjusted
x – matrix which contains 2 elements
The first component:
The output gap

- Log(Gross Domestic Product, Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate)
- Log(Real Potential Gross Domestic Product, Billions of Chained 2009 Dollars, Quarterly, Not Seasonally Adjusted)

The second component:
Consumer Price Index for All Urban Consumers: All Items, Percent Change from Year Ago, Quarterly, Seasonally Adjusted – desired inflation rate (assumed 2%).

Source: FRED database

Timeframe: 1999Q4 – 2014Q2

There are two ways of obtaining U.S. potential GDP:

1. Download it from FRED Database as it is already calculated. http://research.stlouisfed.org/fred2/series/GDPPOT/
2. We determine it.

Example 1:
We use the potential GDP from FRED database in order to calculate the output gap.

In [ ]:

```
>>> import numpy as np
>>> import statsmodels.api as sm
>>> r=[0.0568, 0.0627, 0.0652, 0.0647, 0.0559, 0.0433, 0.035, 0.0213, 0.0173, 0.0175, 0.0174, 0.0144, 0.0125, 0.0125, 0.0102,
       0.01, 0.01, 0.0101, 0.0143, 0.0195, 0.0247, 0.0294, 0.0346, 0.0398, 0.0446, 0.0491, 0.0525, 0.0525, 0.0526, 0.0525,
       0.0507, 0.045, 0.0318, 0.0209, 0.0194, 0.0051, 0.0018, 0.0018, 0.0016, 0.0012, 0.0013, 0.0019, 0.0019, 0.0019, 0.0016,
       0.0009, 0.0008, 0.0007, 0.001, 0.0015, 0.0014, 0.0016, 0.0014, 0.0012, 0.0008, 0.0009, 0.0007, 0.0009, 0.0009]
>>> x=[[0.012580, 0.012934, 0.014689, 0.014435, 0.014098, 0.013249, 0.006780, -0.001249, -0.007681, -0.006824, -0.004237,
        0.002535, 0.009764, 0.000059, 0.002169, 0.000017, -0.001823, 0.007859, 0.006752, 0.013851, 0.010354, 0.009229,
        0.018196, 0.016745, 0.016909, 0.019243, 0.013403, -0.000346, 0.004315, 0.006651, 0.003488, 0.020315, 0.021372,
        0.023106, 0.032525, -0.004042, -0.021842, -0.029423, -0.036070, -0.005125, 0.003369, -0.002141, -0.007713,
        -0.007845, 0.001211, 0.013752, 0.017318, 0.013364, 0.008060, -0.000820, -0.003133, -0.000982, -0.003263, -0.005736,
        -0.004618, -0.007800, -0.005996, 0.000583, -0.002092], [-0.184380, -0.169038, -0.170417, -0.168430, -0.174054,
        -0.170678, -0.179419, -0.182332, -0.178408, -0.177581, -0.176467, -0.178422, -0.174842, -0.169737, -0.154665,
        -0.145081, -0.137048, -0.127197, -0.117958, -0.108186, -0.094240, -0.087621, -0.075752, -0.068300, -0.054381,
        -0.049295, -0.047370, -0.042002, -0.036140, -0.028975, -0.024763, -0.022689, -0.029463, -0.025047, -0.028159,
        -0.052932, -0.068797, -0.075744, -0.076354, -0.067206, -0.062676, -0.051854, -0.043744, -0.035571, -0.038480,
        -0.027533, -0.023047, -0.014141, -0.007181, -0.002513, 0.004238, 0.004080, 0.010361, 0.013399, 0.024485, 0.032571,
        0.026528, 0.039007, 0.047878]]
>>> def rule(r, x):
...  ones = np.ones(len(x[0]))
...  X = sm.add_constant(np.column_stack((x[0], ones)))
...  for ele in x[1:]:
...   X = sm.add_constant(np.column_stack((ele, X)))
...  results = sm.OLS(r, X).fit()
...  return results
...
>>> print rule(r,x).summary()
```

OLS Regression Results

Dep. Variable:	y	R-squared:	0.403
Model:	OLS	Adj. R-squared:	0.382
Method:	Least Squares	F-statistic:	18.90
Date:	Mon, 22 Dec 2014	Prob (F-statistic):	5.34e-07
Time:	18:08:10	Log-Likelihood:	159.18
No. Observations:	59	AIC:	-312.4
Df Residuals:	56	BIC:	-306.1
Df Model:	2		

	coef	std err	t	P>|t|	[95.0% Conf.	Int.]
x1	-0.1064	0.031	-3.428	0.001	-0.169	-0.044
x2	0.8681	0.180	4.822	0.000	0.507	1.229
const	0.0090	0.003	2.809	0.007	0.003	0.015

Omnibus:	5.258	Durbin-Watson:	0.231
Prob(Omnibus):	0.072	Jarque-Bera (JB):	5.211
Skew:	0.717	Prob(JB):	0.0739
Kurtosis:	2.748	Cond. No.	82.9

Results:

Dependent variable: y (Federal Funds rate from U.S.)

Independent variables:

`x1` – the difference between registered inflation and the desired inflation

Consumer Price Index for All Urban Consumers: All Items, Percent Change from Year Ago, Quarterly, Seasonally Adjusted – desired inflation rate (assumed 2%)

x2 – output gap

Log(Gross Domestic Product, Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate)

-Log(Real Potential Gross Domestic Product, Billions of Chained 2009 Dollars, Quarterly, Not Seasonally Adjusted)

$$y = -0.1064x1 + 0.8681x2 + 0.0090$$

Method of estimation: Ordinary Least Squares (OLS) Method

Number of observations: 59

Degrees of freedom: 2

Coefficient of determination (R-squared)=0.403. The output gap and the difference between registered inflation and desired inflation explain 40.3 % of Federal Funds rate evolution.

Adjusted coefficient of determination (adjusted R-squared)=0.382

Akaike Information Criterion (AIC): -312.4

Bayesian Information Criterion (BIC): -306.1

t-test

Null hypothesis: The estimated coefficient is zero

Alternative hypothesis: The estimated coefficient is not zero

x1 coefficient: t is -3.428

x2 coefficient: t is 4.822

constant: 2.809

The low probability values indicate that estimated coefficients are statistically different from zero.

Skew: 0.717

Kurtosis: 2.748

An assumption of the linear regression is that the residuals are normally distributed. Skewness should be close to zero while kurtosis should be close to 3 when the residuals are normally distributed. JB test indicates whether the residuals are normally distributed or not. A JB close to zero indicates that residuals come from a normal distribution. In our case, kurtosis is close to the desired value while skewness is not close to the zero value. JB is not zero. p-value is higher than 0.05, therefore we cannot reject the null hypothesis of normally distributed residuals.

Durbin-Watson statistics: 0.231
Therefore, there is autocorrelation of residuals. This means that we have omitted some variables in our analysis.

Example 2:
We calculate potential GDP using a Hodrick-Prescott (HP) statistical filter.

HP filter:

- GDP is separated into a cyclical and a trend component by minimizing a loss function
- Takes into consideration time series data (U.S. GDP) and a parameter lambda – Hodrick-Prescott smoothing parameter. We have to indicate lambda's value before the filter calculates the new values. A value of 1600 is suggested for quarterly data. Ravn and Uhlig suggest using a value of 6.25 (1600/256) for annual data and 129600 (1600*81) for monthly data.
  >>> gdp2=[9.213436, 9.237790, 9.245457, 9.256489, 9.259902, 9.272225, 9.272329, 9.278121, 9.290482, 9.299706, 9.309018, 9.315043, 9.326353, 9.338795, 9.360922, 9.377278, 9.391695, 9.407665, 9.422844, 9.438448, 9.458270, 9.470710, 9.488381, 9.501636, 9.521414, 9.532409, 9.540255, 9.551544, 9.563333, 9.576531, 9.586699, 9.594602, 9.593451, 9.603260, 9.605284, 9.585339, 9.573865, 9.570836, 9.573879, 9.586480, 9.594316, 9.608351, 9.619645, 9.631036, 9.631574, 9.646070,

9.654199, 9.666834, 9.677622, 9.686245, 9.697011, 9.700912, 9.711261, 9.718314,

9.733429, 9.745564, 9.743554, 9.760091, 9.773105]

>>> import statsmodels.tsa.filters as filter

>>> filter.hpfilter(gdp2, 1600)

```
>>> filter.hpfilter(gdp2, 1600)
(array([ 0.00320611,  0.01638169,  0.01286827,  0.01270561,  0.00488941,
         0.00592046, -0.00536153, -0.01108455, -0.01039922, -0.01303624,
        -0.01578882, -0.02204202, -0.02322403, -0.02347425, -0.01421158,
        -0.01084925, -0.00950361, -0.00662421, -0.00449068, -0.0018185 ,
         0.00525666,  0.00520725,  0.01071545,  0.01220017,  0.02065966,
         0.02083949,  0.01841236,  0.01999594,  0.02266136,  0.02731929,
         0.02951823,  0.02999359,  0.02190834,  0.0252097 ,  0.0210722 ,
        -0.00478438, -0.02202943, -0.03079537, -0.03354782, -0.02687418,
        -0.02514988, -0.01744653, -0.01272404, -0.00815343, -0.01468573,
        -0.0075059 , -0.00692572, -0.00205428,  0.00077768,  0.00127567,
         0.00377073, -0.00072288,  0.00112772, -0.00040113,  0.00606821,
         0.00951361, -0.00121383,  0.00658804,  0.01086012]), array([ 9.21022989,  9.22140831,  9.23258873,  9.24378339,  9.25501259,
         9.26630454,  9.27769053,  9.28920555,  9.30088122,  9.31274224,
         9.32480682,  9.33708502,  9.34957703,  9.36226925,  9.37513358,
         9.38812725,  9.40119861,  9.41428921,  9.42733468,  9.4402665 ,
         9.45301334,  9.46550275,  9.47766555,  9.48943583,  9.50075434,
         9.51156951,  9.52184264,  9.53154806,  9.54067164,  9.54921171,
         9.55718077,  9.56460841,  9.57154266,  9.5780503 ,  9.5842118 ,
         9.59012338,  9.59589443,  9.60163137,  9.60742682,  9.61335418,
         9.61946588,  9.62579753,  9.63236904,  9.63918943,  9.64625973,
         9.6535759 ,  9.66112472,  9.66888828,  9.67684432,  9.68496933,
         9.69324027,  9.70163488,  9.71013328,  9.71871513,  9.72736079,
         9.73605039,  9.74476783,  9.75350296,  9.76224488]))
```

hpfilter function returns array one and array two:

array one – is the first and represents the cycle array;

array two – is the trend array (estimated potential GDP in our case). Potential GDP could also be seen as GDP's long-term trend.

>>> import numpy as np

>>> import statsmodels.api as sm

>>> r=[0.0568, 0.0627, 0.0652, 0.0647, 0.0559, 0.0433, 0.035, 0.0213, 0.0173, 0.0175, 0.0174, 0.0144, 0.0125, 0.0125, 0.0102, 0.01, 0.01, 0.0101, 0.0143, 0.0195, 0.0247, 0.0294, 0.0346, 0.0398, 0.0446, 0.0491, 0.0525, 0.0525, 0.0526, 0.0525, 0.0507, 0.045, 0.0318, 0.0209, 0.0194, 0.0051, 0.0018, 0.0018, 0.0016, 0.0012, 0.0013, 0.0019, 0.0019, 0.0019, 0.0016, 0.0009, 0.0008, 0.0007, 0.001, 0.0015, 0.0014, 0.0016, 0.0014, 0.0012, 0.0008, 0.0009, 0.0007, 0.0009, 0.0009]

>>> x=[[0.012580, 0.012934, 0.014689, 0.014435, 0.014098, 0.013249, 0.006780, -0.001249, -0.007681, -0.006824, -0.004237, 0.002535, 0.009764, 0.000059, 0.002169, 0.000017, -0.001823, 0.007859, 0.006752, 0.013851, 0.010354, 0.009229, 0.018196, 0.016745, 0.016909, 0.019243, 0.013403, -0.000346, 0.004315, 0.006651, 0.003488, 0.020315, 0.021372, 0.023106, 0.032525, -0.004042, -0.021842, -0.029423, -0.036070, -0.005125, 0.003369, -0.002141, -0.007713, -0.007845, 0.001211, 0.013752, 0.017318,

0.013364, 0.008060, -0.000820, -0.003133, -0.000982, -0.003263, -0.005736, -0.004618, -0.007800, -0.005996, 0.000583, -0.002092], [0.003206, 0.016382, 0.012868, 0.012706, 0.004889, 0.005921, -0.005362, -0.011085, -0.010400, -0.013037, -0.015789, -0.022042, -0.023224, -0.023474, -0.014212, -0.010850, -0.009504, -0.006624, -0.004491, -0.001819, 0.005257, 0.005208, 0.010716, 0.012200, 0.020660, 0.020840, 0.018413, 0.019996, 0.022661, 0.027319, 0.029519, 0.029994, 0.021908, 0.025210, 0.021072, -0.004784, -0.022030, -0.030795, -0.033548, -0.026875, -0.025150, -0.017446, -0.012724, -0.015224, -0.007616, -0.007506, -0.006926, -0.002054, 0.000837, 0.001276, 0.003770, -0.000723, 0.001128, -0.000401, 0.006068, 0.009514, -0.001214, 0.006588, 0.010861]]

```
>>> def rule(r, x):
... ones = np.ones(len(x[0]))
... X = sm.add_constant(np.column_stack((x[0], ones)))
... for ele in x[1:]:
...  X = sm.add_constant(np.column_stack((ele, X)))
... results = sm.OLS(r, X).fit()
... return results
...
>>> print rule(r, x).summary()
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.436
Model:                            OLS   Adj. R-squared:                  0.415
Method:                 Least Squares   F-statistic:                     21.61
Date:                Tue, 23 Dec 2014   Prob (F-statistic):           1.11e-07
Time:                        11:29:25   Log-Likelihood:                160.84
No. Observations:                  59   AIC:                            -315.7
Df Residuals:                      56   BIC:                            -309.4
Df Model:                           2
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
x1             0.6859      0.173      3.958      0.000         0.339     1.033
x2             0.3360      0.228      1.474      0.146        -0.121     0.793
const          0.0191      0.002      8.293      0.000         0.014     0.024
==============================================================================
Omnibus:                        3.513   Durbin-Watson:                   0.138
Prob(Omnibus):                  0.173   Jarque-Bera (JB):                1.990
Skew:                           0.184   Prob(JB):                        0.370
Kurtosis:                       2.179   Cond. No.                         124.
==============================================================================
>>>
```

Results:

Dependent variable: y (Federal Funds rate from U.S.)

Independent variables:

x1 – the difference between registered inflation and the desired inflation

Consumer Price Index for All Urban Consumers: All Items, Percent Change from Year Ago, Quarterly, Seasonally Adjusted – desired inflation rate (assumed 2%).

x2 – output gap

Log(Gross Domestic Product, Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate)

-Log(Potential Gross Domestic Product estimated by HP filter)

$$y = 0.6859x1 + 0.3360x2 + 0.0191$$

Method of estimation: Ordinary Least Squares (OLS) Method

Number of observations: 59

Degrees of freedom: 2

Coefficient of determination (R-squared)=0.436 . The output gap and the difference between registered inflation and desired inflation explains 43.6 % of Federal Funds rate evolution.

Adjusted coefficient of determination (Adjusted R-squared=0.415).

Akaike Information Criterion (AIC): -315.7

Bayesian Information Criterion (BIC): -309.4

t-test

Null hypothesis: The estimated coefficient is zero

Alternative hypothesis: The estimated coefficient is not zero

x1 coefficient: t is 3.958

x2 coefficient: t is 1.474

constant: t is 8.293

For x1 coefficient and the constant p-value is very small (0.00) which indicates that the coefficients are statistically different from zero. As for x2 coefficient p-value is also small (0.146) although higher as compared to the first example.

Durbin-Watson statistic: 0.138
The value is not approximately 2, which means that there is autocorrelation of residuals.

Skew: 0.184
Kurtosis: 2.179

Jarque-Bera (JB): 1.990
Prob(JB): 0.370

Skew and kurtosis are close to the desired values while the probability is high which means that we cannot reject the hypothesis that the residuals are normally distributed.

Calculating federal funds rate using Taylor rules

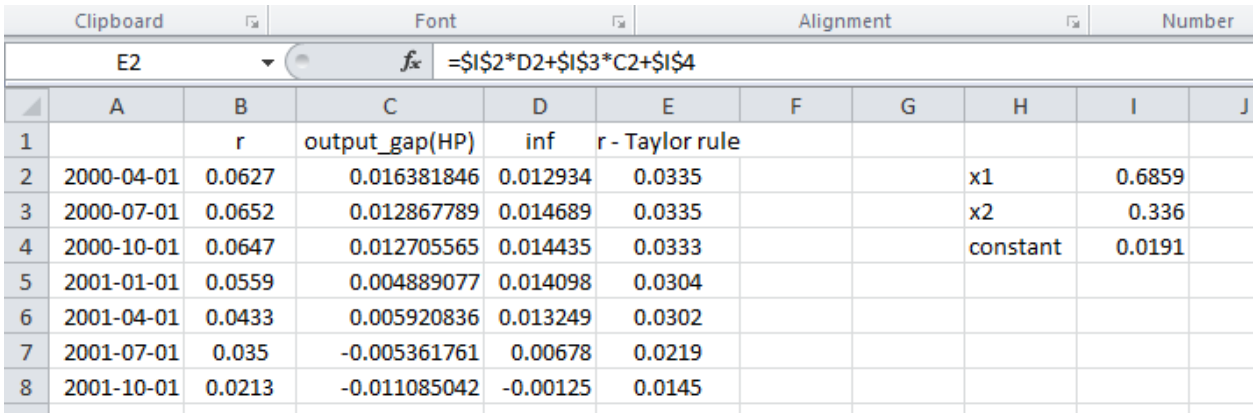
Taylor rules show if the federal funds rate in a certain period is calibrated adequately considering the inflationary pressures in the economy. We calculate the federal funds rate values indicated by the Taylor rules determined in this chapter.

Example 1:

`r_Taylor_rule1=0.6859x1+0.3360x2+0.0191`

We know `x1`, `x2`, and the constant and we can estimate the interest rate. In this case the output gap(`x2`) is determined considering potential GDP obtained from HP filter.

We calculate `r_Taylor_rule1` using Excel.



Clipboard | Font | Alignment | Number

E2 | =$I$2*D2+$I$3*C2+$I$4

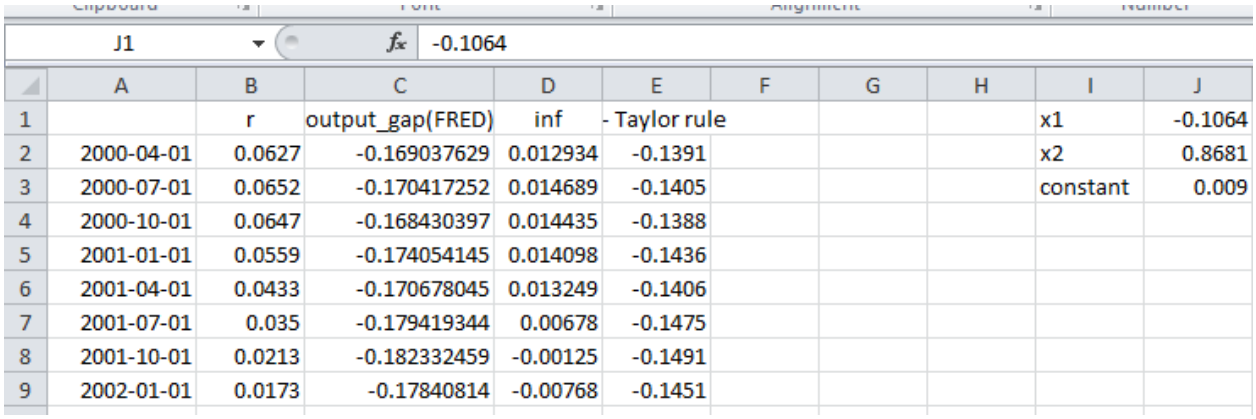
	A	B	C	D	E	F	G	H	I	J
1		r	output_gap(HP)	inf	r - Taylor rule					
2	2000-04-01	0.0627	0.016381846	0.012934	0.0335			x1	0.6859	
3	2000-07-01	0.0652	0.012867789	0.014689	0.0335			x2	0.336	
4	2000-10-01	0.0647	0.012705565	0.014435	0.0333			constant	0.0191	
5	2001-01-01	0.0559	0.004889077	0.014098	0.0304					
6	2001-04-01	0.0433	0.005920836	0.013249	0.0302					
7	2001-07-01	0.035	-0.005361761	0.00678	0.0219					
8	2001-10-01	0.0213	-0.011085042	-0.00125	0.0145					

Example 2:

`r_Taylor_rule2=-0.1064x1+0.8681x2+0.0090`

We know `x1`, `x2`, and the constant and we can estimate the interest rate. In this case, the output gap(`x2`) is determined considering potential GDP provided by FRED.

We calculate `r_Taylor_rule2` using Excel.



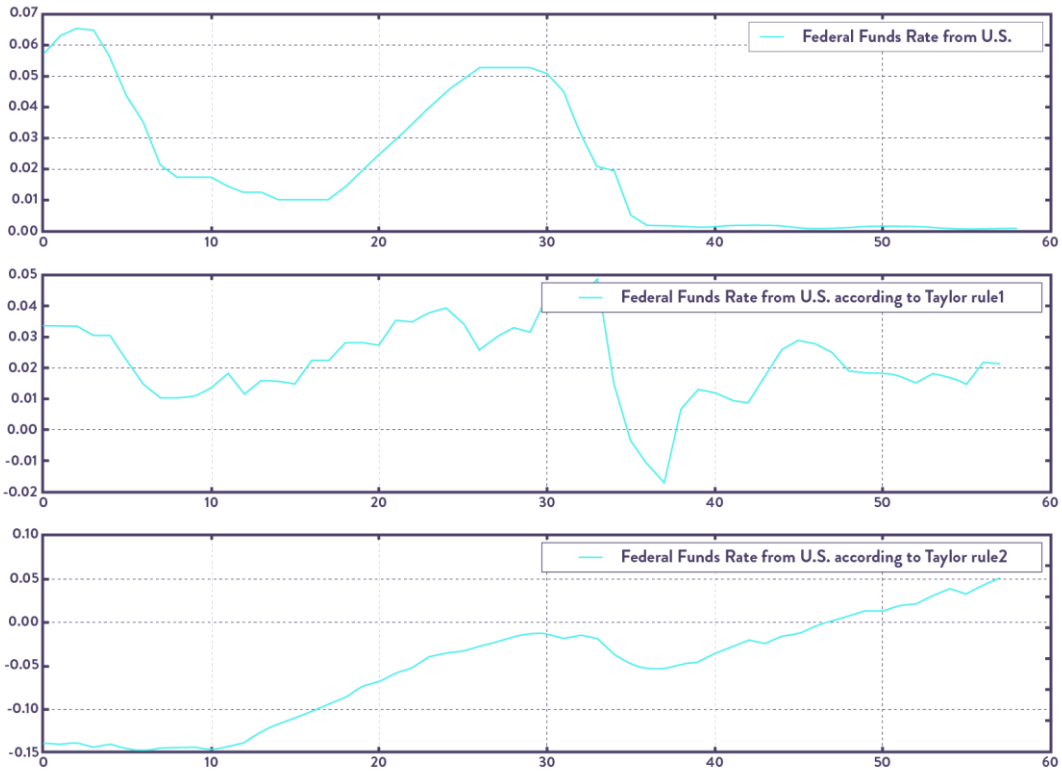
J1 | -0.1064

	A	B	C	D	E	F	G	H	I	J
1		r	output_gap(FRED)	inf	- Taylor rule				x1	-0.1064
2	2000-04-01	0.0627	-0.169037629	0.012934	-0.1391				x2	0.8681
3	2000-07-01	0.0652	-0.170417252	0.014689	-0.1405				constant	0.009
4	2000-10-01	0.0647	-0.168430397	0.014435	-0.1388					
5	2001-01-01	0.0559	-0.174054145	0.014098	-0.1436					
6	2001-04-01	0.0433	-0.170678045	0.013249	-0.1406					
7	2001-07-01	0.035	-0.179419344	0.00678	-0.1475					
8	2001-10-01	0.0213	-0.182332459	-0.00125	-0.1491					
9	2002-01-01	0.0173	-0.17840814	-0.00768	-0.1451					

Federal funds rate and Taylor rule 1 and 2 are introduced in Python.

In [ ]:

```
>>> import numpy as np
>>> import matplotlib.pyplot as plt
>>> r=[0.0568, 0.0627, 0.0652, 0.0647, 0.0559, 0.0433, 0.035, 0.0213, 0.0173, 0.0175, 0.0174, 0.0144, 0.0125, 0.0125, 0.0102,
       0.01, 0.01, 0.0101, 0.0143, 0.0195, 0.0247, 0.0294, 0.0346, 0.0398, 0.0446, 0.0491, 0.0525, 0.0525, 0.0526, 0.0525,
       0.0507, 0.045, 0.0318, 0.0209, 0.0194, 0.0051, 0.0018, 0.0018, 0.0016, 0.0012, 0.0013, 0.0019, 0.0019, 0.0019, 0.0016,
       0.0009, 0.0008, 0.0007, 0.001, 0.0015, 0.0014, 0.0016, 0.0014, 0.0012, 0.0008, 0.0009, 0.0007, 0.0009, 0.0009]
#Federal Funds Rate
>>> r_Taylor_rule1=[0.0335, 0.0335, 0.0333, 0.0304, 0.0302, 0.0219, 0.0145, 0.0103, 0.0100, 0.0109, 0.0134, 0.0180, 0.0113,
                    0.0158, 0.0155, 0.0147, 0.0223, 0.0222, 0.0280, 0.0280, 0.0272, 0.0352, 0.0347, 0.0376, 0.0393, 0.0345,
                    0.0256, 0.0297, 0.0328, 0.0314, 0.0431, 0.0411, 0.0434, 0.0485, 0.0147, -0.0033, -0.0114, -0.0169, 0.0066,
                    0.0130, 0.0118, 0.0095, 0.0086, 0.0174, 0.0260, 0.0287, 0.0276, 0.0249, 0.0190, 0.0182, 0.0182, 0.0172,
                    0.0150, 0.0180, 0.0169, 0.0146, 0.0217, 0.0213]
#Federal Funds Rate calculated considering Taylor rule 1. In this case the output gap is estimated using the potential GDP
provided by HP filter
>>> r_Taylor_rule2=[-0.1391, -0.1405, -0.1388, -0.1436, -0.1406, -0.1475, -0.1491, -0.1451, -0.1444, -0.1437, -0.1462,
                    -0.1438, -0.1384, -0.1255, -0.1169, -0.1098, -0.1023, -0.0941, -0.0864, -0.0739, -0.0680, -0.0587,
                    -0.0521, -0.0400, -0.0358, -0.0335, -0.0274, -0.0228, -0.0169, -0.0129, -0.0129, -0.0189, -0.0152,
                    -0.0189, -0.0365, -0.0484, -0.0536, -0.0534, -0.0488, -0.0458, -0.0358, -0.0282, -0.0210, -0.0245,
                    -0.0164, -0.0128, -0.0047, 0.0019, 0.0069, 0.0130, 0.0126, 0.0183, 0.0212, 0.0307, 0.0381, 0.0327,
                    0.0428, 0.0508]
#Federal Funds Rate calculated considering Taylor Rule 2. In this case the output gap is estimated using the potential GDP
from FRED Database
>>> plt.subplot(311)
>>> plt.plot(r, label="Federal Funds Rate from U.S.")
>>> plt.grid(True)
>>> plt.legend()
>>> plt.subplot(312)
>>> plt.plot(r_Taylor_rule1, label="Federal Funds Rate from U.S. according to Taylor rule1")
>>> plt.grid(True)
>>> plt.legend()
>>> plt.subplot(313)
>>> plt.plot(r_Taylor_rule2, label="Federal Funds Rate from U.S. according to Taylor rule2")
>>> plt.grid(True)
>>> plt.legend()
```



References

Amemiya, T. (1977). The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model, *Econometrica*, 45, issue 4, p. 955-68, https://EconPapers.repec.org/RePEc:ecm:emetrp:v:45:y:1977:i:4:p:955-68.

Daroczi G. et al. (2013). *Introduction to R for Quantitative Finance*. Packt Publishing.

Greene, W. (2000). *Econometric Analysis, Prentice-Hall*, NY.

Gujarati, D. (2004). *Basic Econometrics*, McGraw-Hill.

Horrace, W. C. and Oaxaca, R. L., 2006. "Results on the bias and inconsistency of ordinary least squares for the linear probability model," *Economics Letters*, Elsevier, vol. 90(3), pages 321-327.

Jeet P. and Vats P. (2017). *Learning Quantitative Finance. with R*. Packt Publishing.

Scott, M. et al. (2013). *Financial Risk Modelling and Portfolio Optimization with R*. Wiley.
people.ku.edu/~p112m883/pdf/Econ526/Ch11Part1_slides.docx

3.3 Collaborative Review Task

During each module students will undertake a short case-study assignment, which is then marked by their peers according to a grading rubric. The question is provided below.

In the following table, the results of 6 models attempt to explain a dependent variable of interest, y. You may assume that there is sufficient theoretical reason to consider any or all of the explanatory variables x_1, x_2, x_3 and x_1 in a model for y, but it is unknown whether all of them are necessary to effectively model the data generating process of y.

	Model 1		**Model 2**		**Model 3**		**Model 4**		**Model 5**		**Model 6**	
	Coefficient	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**	**Coefficient**	**p-value**
Constant	0.06906	*0.001575*	0.07629	*0.000336*	0.06969	*0.00148*	0.07697	*0.000312*	0.07374	*0.000307*	0.07472	*0.000264*
x_1	-0.08039	*0.000135*	~	~	-0.08118	*0.000124*	~	~	-0.0813	*<0.0001*	-0.0825	*<0.0001*
x_2	0.23038	*<0.0001*	0.3267	*<0.0001*	0.23188	*<0.0001*	0.32874	*<0.0001*	0.33752	*<0.0001*	0.34071	*<0.0001*
x_3	~	~	0.23239	*<0.0001*	~	~	0.23374	*<0.0001*	0.23387	*<0.0001*	0.2359	*<0.0001*
x_4	~	~	~	~	0.01192	*0.577191*	0.01208	*0.559985*	~	~	0.01796	*0.367086*
R^2	0.3775	~	0.4148	~	0.3785	~	0.4158	~	0.4638	~	0.466	~
F-statistic	42.55	*<0.0001*	69.81	*<0.0001*	39.79	*<0.0001*	46.5	*<0.0001*	56.51	*<0.0001*	42.55	*<0.0001*

Provide a thorough, rigorous analysis of which of the models is the preferred model for the explanatory variable of interest. Your analysis should include features of each coefficient, each model, and each of the diagnostic statistics. Do NOT analyse them one-by-

one, but by theme as identified in Module 2 of Econometrics. For the preferred model, give an analysis of the likely correlation among the explanatory variables.

Your answer will be evaluated on overall coverage, logical progression, and style of presentation.