# Introductory Econometrics

Introduction to EViews

**Task 1.**

1. Import to EViews the file `wages2.xls` (download from homepage) which contains 3296 observations taken from a US Household Panel. Label the imported data by the following variable list:

   `exper`: experience in years `male`: 1 if male, 0 otherwise `educ`: years of schooling `wage`: wage (in \$) per hour

2. Generate the following series.
   `lnwage`: ln(`wage`)
   `lnexper`: ln(`exper`)
   `lneduc`: ln(`educ`)

3. Create a text object to save and comment your EViews commands.

4. Reproduce the descriptive statistics in the table below.

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| `exper` | 3296 | 8.041869 | 2.290855 | 1 | 18 |
| `male` | 3296 | .5239684 | .499501 | 0 | 1 |
| `educ` | 3296 | 11.63016 | 1.657114 | 3 | 16 |
| `wage` | 3296 | 5.816391 | 4.054694 | .0765556 | 112.7919 |

5. Plot the histogram of `wage` and save the graph by exporting it to your hard disk.

6. Compute sample mean, sample variance, for the variable `wage`. Save them as a scalar object. Compute the 0.01 and the 0.05 empirical quantile of the sample distribution and the sample kurtosis of the variable `wage`. Save the quantiles and the sample kurtosis as series.

7. Estimate the following regression models by OLS and interpret $\beta_2$ (the returns of schooling). How do you call the specification in b) and c)?

   [a)] $\texttt{wage}_i = \beta_1 + \beta_2 \texttt{educ}_i + \beta_3 \texttt{exper}_i + \varepsilon_i$

   [b)] $\texttt{lnwage}_i = \beta_1 + \beta_2 \texttt{educ}_i + \beta_3 \texttt{exper}_i + \varepsilon_i$

   [c)] $\texttt{lnwage}_i = \beta_1 + \beta_2 \texttt{lneduc}_i + \beta_3 \texttt{lnexper}_i + \varepsilon_i$

**<u>Task 2</u>.**
Estimate the Glosten/Harris(1988) model with OLS:

$$\Delta p_t = \beta_1 + \beta_2 \Delta Q_t + \beta_3 Q_t + \beta_4 Q_t V_t + \varepsilon_t$$

Names for the variables in the data set:

- $p_t \mathrel{\hat{=}}$ `p`
- $\Delta p_t \mathrel{\hat{=}}$ `delta_p`
- $\Delta Q_t \mathrel{\hat{=}}$ `delta_q`
- $Q_t \mathrel{\hat{=}}$ `q`
- $Q_t V_t \mathrel{\hat{=}}$ `qv`

1. Import to EViews the file `dcxft_tim_my.xls` (download from homepage).
2. Plot the time series `p` and provide meaningful axis labels and titles for the graph. Export the graphic in your word processor (Word or LaTeX).
3. Take first differences of the time series `p` and label it `d_p`. Compare `d_p` with `delta_p`.
4. Compute the OLS estimator $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ where $\mathbf{y}$ is the vector `delta_p` and $\mathbf{X}$ is the data matrix containing $\Delta Q_t$, $Q_t$ and $Q_t V_t$.
5. Use the same data set and use EXCEL to compute the OLS estimator. First, name the matrices, then determine your output fields where the vector **b** should be written to and use the Excel matrix functions (`mmult`, `mtrans` and `minv`, *read Excel help to learn how those functions are used*) to compute the vector **b**.
6. As a second alternative use the Excel Add-In `Solver` to compute **b**. Therefore, define an empty column vector in your spreadsheet and name it (for convenience $b$). Then, choose a cell where you write down the objective function (sum of squared residuals). Open `Solver` and minimize your objective function. The cells which can be varied by `Solver` to find a minimum are those constituting your pre-specified vector $b$.

1st set assignments

**Task 1.**

Consider the OLS estimator $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Write extensively the matrices $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{y}$ using $\mathbf{x}_i = (x_{i1}, \dots, x_{iK})'$.

**Task 2.**

Show, that in the simple regression model

$$y_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i$$

the OLS estimator $b_2$ can be written as

$$b_2 = \frac{s_{\mathbf{x}_2\mathbf{y}}}{s^2_{\mathbf{x}_2}}$$

where $s_{\mathbf{x}_2\mathbf{y}}$ is the sample covariance of $\mathbf{x}_2$ and $\mathbf{y}$ and $s^2_{\mathbf{x}_2}$ is the sample variance of $\mathbf{x}_2$.

Show, that $b_1 = \bar{y} - b_2\bar{x}$.

**Task 3.**

Show, that in a semi-log model

$$\ln y_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i$$

$\beta_2$ can be interpreted as a percentage change of $y_i$ if $x_{i2}$ increases one unit.

**Task 4.**

Show the validity of the law of iterated expectations by showing that

$$E_{Z|X}[E_{Y|X,Z}(Y|X,Z)|X] = E_{Y|X}(Y|X)$$

**Task 5.**

Show that $E(\varepsilon_i|1, x_{i1}, x_{i2}) = 0$ implies

$$\begin{aligned} E(\varepsilon_i) &= 0 \\ E(\varepsilon_i x_{i1}) &= 0 \\ E(\varepsilon_i x_{i2}) &= 0 \end{aligned}$$

Show that $E(\varepsilon_i) = 0$ implies that $Cov(\varepsilon_i, x_{i2}) = 0$.

**Task 6.**

Show that for

$$y_i = \beta_1 + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} + \varepsilon_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

the OLS estimate $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ implies that

a) the regression passes through the means of the dependent variable and the regressors

$$\bar{y} = \beta_1 + \beta_2 \bar{x}_{i2} + \ldots + \beta_K \bar{x}_{iK}$$

where $\bar{y} = \frac{1}{n}\sum y_i$ and $\bar{x}_{ik} = \frac{1}{n}\sum x_{ik}$.

b) $\bar{y} = \bar{\hat{y}}$ where $\hat{y}_i = \mathbf{x}_i'\mathbf{b}$ and $\bar{\hat{y}} = \frac{1}{n}\sum \hat{y}_i$

2nd set assignments Introductory Econometrics

**Task 1**

Use the data set `dcx_gh.wf1` to estimate the Glosten/Harris(1988) model. An estimable linear equation can be derived in the following way.

The unknown true value of a stock $m_t$ is given by:

$$m_t = \mu + m_{t-1} + \varepsilon_t + z_t Q_t$$

where $\mu$ is a constant, $\varepsilon_t$ is a random disturbance, $Q_t$ is a trade indicator (taking the value 1 for a buy and $-1$ for a sell) and $z_t = z_0 + z_1 V_t$. With $V_t$ we denote the traded volume. The ask price $p_t^a$ (the price to buy a stock) and the bid price $p_t^b$ (the price to sell a stock) are:

$$\begin{aligned} p_t^a &= m_{t-1} + \mu + \varepsilon_t + z_t + c \\ p_t^b &= m_{t-1} + \mu + \varepsilon_t - z_t - c \end{aligned}$$

Thus, the transaction price in period $t$, $p_t$, and the price in period $t-1$, $p_{t-1}$, can be written as:

$$\begin{aligned} p_t &= \underbrace{m_{t-1} + \mu + \varepsilon_t + z_t Q_t}_{m_t} + cQ_t \\ p_{t-1} &= m_{t-1} + cQ_{t-1} \end{aligned}$$

Subtracting $p_{t-1}$ from $p_t$ and plugging in the above expression for $z_t$ yields:

$$\Delta p_t = \mu + c\Delta Q_t + z_0 Q_t + z_1 Q_t V_t + \varepsilon_t$$

data set description:

- $\Delta p_t \mathrel{\hat{=}}$ `dp`
- $\Delta Q_t \mathrel{\hat{=}}$ `dq`
- $Q_t \mathrel{\hat{=}}$ `q`
- $Q_t V_t \mathrel{\hat{=}}$ `qv`

a) Estimate the OLS regression with EViews.

b) Test if volume $V_t$ has a significant impact on the market makers choice for setting $p_t^a$ and $p_t^b$.

c) Test, if $2c = 0.01$, which would imply that the order processing costs are minimized to the lower bound (1 euro cent) in an electronic order book.

**Task 2**

Analyze the data set `wine.wf1` considering functional form and change of measurement of the variables. Therefore, regress the price and the logarithm of the price, respectively, on the age of the wine. Explain and interpret your results. Further, analyze the effect of using different price measures on your results. Formulate an economic interest rate model in discrete time for wine (take wine as an asset) and interpret the parameters of the linear regression model in this context.

data set description:

- `age`: the age of the wine in years
- `price1`: price for one bottle of wine
- `price12`: price for twelve bottles of wine
- `logprice1`: log-price for one bottle of wine
- `logprice12`: log-price for twelve bottles of wine

3rd set assignments Introductory Econometrics

**Task 1**

Different methods for testing linear restrictions:

Use the data set `dcx_gh.wf1` to estimate the Glosten/Harris(1988) model.

$$\Delta p_t = \mu + c\Delta Q_t + z_0 Q_t + z_1 Q_t V_t + \varepsilon_t$$

Test the joint hypotheses that $z_0 = z_1 = 0$ !

i) Estimate the unrestricted OLS regression with EViews and name the equation (e.g. `UMOD`) object by clicking on `NAME`.

ii) Estimate the restricted OLS regression with EViews and name the equation (e.g. `RMOD`) object by clicking on `NAME`.

iii) Create a scalar object for the sum of squared residuals of each estimated regression by using the following EViews commands in the command line:

```
scalar ussr=UMOD.@ssr;
scalar rssr=RMOD.@ssr;
```

iv) Create a scalar object for the number of observations:

```
scalar n=UMOD.@regobs;
```

v) Compute the F-statistic with help of the three created scalar objects.

**Task 2**

i) Use the Excel spreadsheet `dcxft_tim.xls` to estimate the Glosten/Harris model as in assignment sheet 1. Then, calculate the variance-covariance matrix of the parameter vector $\mathbf{b}$. The VC matrix can be computed as:

$$Var(\mathbf{b}|\mathbf{X}) = \sigma^2 \cdot (\mathbf{X}'\mathbf{X})^{-1}$$

Here, $\sigma^2$ can be replaced by its unbiased estimator $s^2 = \mathbf{e}'\mathbf{e}/(n-K)$, where $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$, $n$ is the number of observations and $K$ is the number of estimated parameters.

ii) Test the joint hypotheses that $2c = 0.01$ and $z_0 = z_1 = 0$! Therefore, create the matrices $\mathbf{R}$ and $\mathbf{r}$ (Hayashi(2000) p.40). Then, compute the $F$-statistic as:

$$F \equiv (\mathbf{Rb} - \mathbf{r})'[\mathbf{R}\widehat{Var(\mathbf{b}|\mathbf{X})}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{r})/\#\mathbf{r}$$

where $\#\mathbf{r}$ is the dimension of $\mathbf{r}$ (number of restrictions).

4th set assignments

**Task 1**

Confidence Intervals:

Suppose you have estimated a parameter vector $\mathbf{b} = (\ 0.55 \quad 0.37 \quad 1.46 \quad 0.01\ )'$ with an estimated variance-covariance matrix

$$\widehat{Var}(\mathbf{b}|\mathbf{X}) = \begin{bmatrix} 0.01 & 0.023 & 0.0017 & 0.0005 \\ 0.023 & 0.0025 & 0.015 & 0.0097 \\ 0.0017 & 0.015 & 0.64 & 0.0006 \\ 0.0005 & 0.0097 & 0.0006 & 0.001 \end{bmatrix}$$

a) Compute the 95% confidence interval each parameter $b_k$.

b) What does the specific confidence interval computed in a) tell you?

c) Why are the bounds of a confidence interval for $\beta_k$ random variables?

d) Another estimation yields an estimated $b_k$ with the corresponding standard error $se(b_k)$. You conclude from computing the t-statistic $t_k = \frac{b_k - \bar{\beta}_k}{se(b_k)}$ that you can reject the null hypothesis $H_0 : b_k = \bar{\beta}_k$ on the $\alpha$% significance level. Now, you compute the $(1-\alpha)$% confidence interval. Will $\bar{\beta}_k$ lie inside or outside the confidence interval?

**Task 2**

More about confidence intervals:

Suppose, computing the lower bound of the 95% confidence interval yields $b_k - t_{\alpha/2}(n-K)se(b_k) = -0.01$. The upper bound is $b_k + t_{\alpha/2}(n-K)se(b_k) = 0.01$ Which of the following statements are correct?

1. With probability of 5% the true parameter $\beta_k$ lies in the interval -0.01 and 0.01.
2. The null hypothesis $H_0 : \beta_k = \bar{\beta}_k$ cannot be rejected for values $(-0.01 \leq \bar{\beta}_k \leq 0.01)$ on the 5% significance level.
3. The null hypothesis $H_0 : \beta_k = 1$ can be rejected on the 5% significance level.
4. The true parameter $\beta_k$ is with probability $1 - \alpha = 0.95$ greater than -0.01 and smaller than 0.01.
5. The stochastic bounds of the $1 - \alpha$ confidence interval overlap the true parameter with probability $1 - \alpha$.
6. If the hypothesized parameter value $\bar{\beta}_k$ falls within the range of the $1 - \alpha$ confidence interval computed from the estimates $b_k$ and $se(b_k)$ then we do not reject $H_0 : \beta_k = \bar{\beta}_k$ at the significance level of 5%.

**Task 3**

Goodness of fit:

a) Show that if the regression includes a constant:

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i$$

then the variance of the dependent variable can be written as:

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2 + \frac{1}{N}\sum_{i=1}^{N} e_i^2$$

Hint: $\bar{y} = \bar{\hat{y}}$

b) Take your result from a) and formulate an expression for the coefficient of determination $R^2$.

c) Suppose, you estimated a regression with an $R^2 = 0.63$. Interpret this value.

d) Suppose, you estimate the same model as in c) without a constant. You know that you cannot compute a meaningful centered $R^2$. Therefore, you compute the uncentered $R^2_{uc}$:

$$R^2_{uc} = \frac{\hat{\mathbf{y}}'\hat{\mathbf{y}}}{\mathbf{y}'\mathbf{y}} = 0.84$$

Compare the two goodness of fit measures in c) and d). Would you conclude that the constant can be excluded because $R^2_{uc} > R^2$?

**Task 4**

Regression with EViews:

In a hedonic price model the price of an asset is explained with its characteristics. In the following we assume that housing pricing can be explained by its size $sqrft$ (measured as *square feet*), the number of bedrooms $bdrms$ and the size of the lot $lotsize$ (also measured as *square feet*. Therefore, we estimate the following equation with OLS:

$$log(price) = \beta_0 + \beta_1 log(sqrft) + \beta_2 bdrms + \beta_3 \log(lotsize)$$

Results of the estimation can be found in the following table:

Dep. Variable: LPRICE
Incl. observations: 88

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -1.29704 | 0.65128 | -1.99152 | 0.0497 |
| LSQRFT | 0.70023 | — | 7.54031 | 0.0000 |
| BDRMS | 0.03696 | 0.02753 | — | — |
| LLOTSIZE | 0.16797 | 0.03828 | 4.38771 | 0.0000 |
| R-squared | 0.64297 | Mean dependent var | | 5.6332 |
| Adjusted R-squared | 0.63021 | S.D. dependent var | | 0.3036 |
| S.E. of regression | 0.18460 | Akaike info criterion | | -0.4968 |
| Sum squared resid | 2.86256 | Schwarz criterion | | -0.3842 |
| Log likelihood | 25.86066 | F-statistic | | — |
| Durbin-Watson stat | 2.08900 | Prob(F-statistic) | | 0.0000 |

(a) Interpret the estimated coefficients $\hat{\beta}_1$ und $\hat{\beta}_2$.

(b) Compute the missing values for *Std. Error* and *t-Statistic* in the table and comment on the statistical significance of the estimated coefficients ($H_0$: $\beta_j = 0$ vs. $H_1$: $\beta_j \neq 0$, $j = 0, 1, 2, 3$).

(c) Test the null hypothesis $H_0$: $\beta_1 = 1$ vs. $H_1$: $\beta_1 < 0$.

(d) Estimate the p-value for $\hat{\beta}_2$ as close as possible and interpret.

(e) What is the null hypothesis of this specific $F - Statistic$? Compute the missing value and interpret the result.

(f) Interpret the value of *R-squared*.

(g) An alternative specification of the model that excludes the lot size as an explanatory variable provides you with values for the Akaike information criterion of -0.313 and a Schwartz criterion of -0.229. Which specification would you prefer?

5th set assignments

**Task 1**

Practical example: Nerlove(1963): Returns to Scale in Electricity Supply

A complete discussion of the example can be found in Hayashi(2000), pp.60. To summarize the problem: We want to estimate a microeconomic cost function for the electricity supply sector. Therefore, consider the output of firm $i$ to be determined by a Cobb-Douglas production function:

$$Q_i = A_i x_{i1}^{\alpha_1} x_{i2}^{\alpha_2} x_{i3}^{\alpha_3}$$

where $Q_i$ is firm $i$'s output, $x_{i1}$ is labor input for firm $i$, $x_{i2}$ is capital input and $x_{i3}$ is fuel. $A_i$ captures unobservable differences in production efficiency (or firm heterogeneity). The $\alpha_j$ are the output elasticities with respect to factor $j$. The sum $\alpha_1 + \alpha_2 + \alpha_3 \equiv r$ is the degree of returns to scale (note, that $r = 1$ would imply constant returns to scale). The cost function associated with such a production function is:

$$TC_i = r \cdot (A_i \alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \alpha_3^{\alpha_3})^{-1/r} Q_i^{1/r} p_{i1}^{\alpha_1/r} p_{i2}^{\alpha_2/r} p_{i3}^{\alpha_3/r}$$

where $TC_i$ are the total costs for firm $i$. Linearizing by taking logs yields:

$$\log(TC_i) = \mu_i + \frac{1}{r}\log(Q_i) + \frac{\alpha_1}{r}\log(p_{i1}) + \frac{\alpha_2}{r}\log(p_{i2}) + \frac{\alpha_3}{r}\log(p_{i3})$$

where $\mu_i = \log[r \cdot (A_i \alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \alpha_3^{\alpha_3})^{-1/r}]$. Now, let $\mu = E(\mu_i)$ and define $\varepsilon_i \equiv \mu_i - \mu$ so that $E(\varepsilon_i) = 0$. This $\varepsilon_i$ represents the industry's average efficiency relative to the firm's efficiency. We can write for the total cost equation:

$$\log(TC_i) = \beta_1 + \beta_2 \log(Q_i) + \beta_3 \log(p_{i1}) + \beta_4 \log(p_{i2}) + \beta_5 \log(p_{i3}) + \varepsilon_i$$

where

$$\beta_1 = \mu,\ \beta_2 = \frac{1}{r},\ \beta_3 = \frac{\alpha_1}{r}\ \beta_4 = \frac{\alpha_2}{r}\ \beta_5 = \frac{\alpha_3}{r}$$

- Download the Excel spreadsheet `nerlove.xls` and import the data in EViews.
- Create the necessary variables to estimate the cost equation with OLS.
- Estimate the cost equation and interpret the parameters.
- Discuss, whether the assumptions of the classic linear regression model are satisfied.
- Which restriction is implied by the notion that the degree of returns to scale is constant ($\alpha_1 + \alpha_2 + \alpha_3 \equiv r$)?
- Test with two alternative procedures if the above restriction can be rejected.
  Hint: The restricted model reads as:

$$\log\left(\frac{TC_i}{p_{i3}}\right) = \beta_1 + \beta_2 \log(Q_i) + \beta_3 \log\left(\frac{p_{i1}}{p_{i3}}\right) + \beta_4 \log\left(\frac{p_{i2}}{p_{i3}}\right) + \varepsilon_i$$

- Reformulate the model to get the log of the average costs as the dependent variable. Then, create the new independent variable and estimate this model. What happens to parameter $\beta_2$ in this new setting? Discuss the difference in the two values for $R^2$ obtained in the total cost and the average cost equation.
- Test the hypothesis that there exist constant returns to scale $(H_0 : r = 1)$
- Eventually, make a residual series from your restricted model estimated previously. Plot the residuals against $\log(Q_i)$. Are you disturbed after looking at the plot?

Data set description:

- `TC` $\hat{=}$ total costs
- `Q` $\hat{=}$ output
- `PL` $\hat{=}$ wage rate (factor price for labor)
- `PF` $\hat{=}$ factor price for fuel
- `PF` $\hat{=}$ rental price for capital

6th set assignments

**Task 1**

Consider the following assumptions:

1. linearity
2. rank condition: $K \times K$ matrix $E(\mathbf{x}_i\mathbf{x}_i') = \mathbf{\Sigma_{xx}}$ is nonsingular
3. predetermined regressors: $E(\mathbf{g}_i) = 0$ where $\mathbf{g}_i = \mathbf{x}_i \cdot \varepsilon_i$
4. $\mathbf{g}_i$ is a martingale difference sequence with finite second moments

i) Show, that under those assumptions, the OLS estimator is distributed asymptotically normal:

$$\sqrt{n}(\mathbf{b} - \boldsymbol{\beta}) \underset{d}{\rightarrow} N(0, \mathbf{\Sigma_{xx}}^{-1} E(\varepsilon_i^2 \mathbf{x}_i\mathbf{x}_i')\mathbf{\Sigma_{xx}}^{-1})$$

ii) Further, show that assumption 4 implies that the $\varepsilon_i$ are serially uncorrelated or $E(\varepsilon_i\varepsilon_{i-j}) = 0$.

**Task 2**

Show, that the test statistic

$$t_k \equiv \frac{\sqrt{n}(b_k - \beta_k)}{\sqrt{[\widehat{Avar(b_k)}]}} \underset{d}{\rightarrow} N(0,1)$$

converges in distribution to a standard normal distribution. Note, that $b_k$ is the k-th element of $\mathbf{b}$ and $Avar(b_k)$ is the (k,k) element of the $K \times K$ matrix $Avar(\mathbf{b})$. Use the facts, that $\sqrt{n}(b_k - \beta_k) \underset{d}{\rightarrow} N(0, Avar(b_k))$ and $\widehat{Avar(b)} \underset{p}{\rightarrow} Avar(b)$. Use Lemma 2.4(c) for argumentation.

**Task 3**

Show, that the test statistic

$$W \equiv (\mathbf{R}\sqrt{n}\mathbf{b} - \mathbf{r})'[\mathbf{R}\widehat{Avar(\mathbf{b})}\mathbf{R}']^{-1}(\mathbf{R}\sqrt{n}\mathbf{b} - \mathbf{r}) \underset{d}{\rightarrow} \chi^2(\#\mathbf{r})$$

converges in distribution to a Chi-square with $\#\mathbf{r}$ degrees of freedom. As a hint, rewrite the equation above as $W \equiv \mathbf{c}_n'\mathbf{Q}_n^{-1}\mathbf{c}_n$. Use Lemma 2.4(d) and the footnote on page 41 for argumentation.

7th set assignments

**Task 1**

Practical example: Testing the Efficient Market Hypothesis

Literature: Hayashi(2000), pp.150-159, Mishkin(1992), Fama(1975)

A complete discussion of the example can be found in Hayashi(2000), pp.150. To summarize the problem: We want to test the efficient market hypothesis of Fama(1976). Following the arguments on p.152 in Hayashi there are two testable implications of the efficient market hypothesis when applied to Treasury bills in a rational expectation framework:

1. The ex-post real interest rate $r$ has a constant mean and is serially uncorrelated
2. The nominal interest rate $R_t$ is the best inflation forecast on the basis of all available information. Since the nominal interest rate is determined by the price of a Treasury bill this implication is equivalent to the notion that asset prices contain all available information. Formally, we can write $E(\pi_{t+1}|I_t) = -r + R_t$ where $I_t$ is the full information set available at time $t$.

Data set description:

- `Year`
- `Month`
- `PAI1` $\hat{=}$ one-month inflation rate (in percent, annual rate)
- `PAI3` $\hat{=}$ three-month inflation rate (in percent, annual rate)
- `TB1` $\hat{=}$ one-month T-Bill rate (in percent, annual rate)
- `TB3` $\hat{=}$ three-month T-Bill rate (in percent, annual rate)
- `CPI` $\hat{=}$ consumer price index

- Download the Excel spreadsheet **`mishkin.xls`** and import the data in EViews.
- Create another inflation rate series calculated from the given CPI data as:

$$\pi_{t+1} = \left[ \left( \frac{P_{t+1}}{P_t} \right)^{12} \right] \times 100$$

- Explore your data graphically and provide descriptive statistics.
- To test for stationarity, conduct Dickey/Fuller tests for all series in your data set and interpret the results. Therefore, open the respective data series and use `VIEW` $\rightarrow$ `UNIT ROOT TEST`

- **Test implication 1:** Compute the ex-post real rate $r$ by subtracting the newly constructed inflation rate from the one-month T-Bill rate and test the real rate series for serial correlation. Use `VIEW` $\rightarrow$ `CORRELOGRAM`. First, look at the serial correlation for the period 1953:1 to 1971:7 (period used by Fama(1975)). Then, look at the post-Fama period. How do you conclude?

- **Test implication 2:** Let $y_t \equiv \pi_{t+1}$, $\mathbf{x}_t \equiv (1, R_t)'$, $\varepsilon_t \equiv \eta_{t+1}$ (the inflation forecast error) which is a martingale difference sequence with $E(\eta_{t+1}|I_t) = 0$.
  Estimate the following regression for the two time periods (Fama sample and post Fama sample):
  $$\pi_{t+1} = \beta_1 + \beta_2 R_t + \eta_{t+1} \;\; or \;\; y_t = \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t$$
  In the estimation window you can choose under `Options` the standard errors being reported. Use classic standard errors as well as heteroskedasticity-robust standard errors and compare the results.

- Test the hypothesis coming from theory that the parameter $\beta_2$ is equal to one. Conclusion? How can you interpret the intercept? For which values $\bar{r}$ can you not reject $H_0 : r = \bar{r}$? Hint for computing confidence bounds :

  ```
  scalar up=eq01.@coefs(i)+2*eq01.@stderrs(i)
  scalar low=eq01.@coefs(i)-2*eq01.@stderrs(i)
  ```

  where `eq01` is the name of the equation object and `i` denotes the i$^{th}$ element in the estimated coefficient vector.

- It is argued that the CPI series exhibits a strong seasonal pattern. In order to account for seasonality you are asked to replace the constant in the estimation with twelve dummy variables for each month. This can be easily done by just adding `@expand(month)` in the estimation window where you specify your regression configuration. Why do you have to remove the constant first if twelve dummies for each month are added? What happens if you do not remove the constant from the model?

**Task 2**

Read pp. 150 and reconsider why the efficient market hypothesis implies the important parts of the assumptions justifying the t-test in the regression above, i.e. the OLS estimator is asymptotically normal.

8th set assignments

Literature: Hayashi(2000), pp.187-191

**Task 1**

Deriving the endogeneity bias

The structural model of a simple model of supply and demand looks like:

$$\begin{aligned} q_i^d &= \alpha_0 + \alpha_1 p_i + u_i \\ q_i^s &= \beta_0 + \beta_1 p_i + v_i \\ q_i^d &= q_i^s \end{aligned}$$

- Solve for the market clearing price $p_i$ and the market clearing quantity $q_i$ to derive the reduced form.
- Use the reduced form to derive expressions for $Cov(p_i, u_i)$ and $Cov(p_i, v_i)$.
- Considering the demand equation, we know that the OLS estimator for $\alpha_1$ converges in probability to:

$$\hat{\alpha}_{1,OLS} \underset{p}{\rightarrow} \frac{Cov(p_i, q_i)}{Var(p_i)}$$

  Show, using the demand equation in the structural form, that

$$Cov(p_i, q_i) = \alpha_1 Var(p_i) + Cov(p_i, u_i)$$

  and, hence

$$\hat{\alpha}_{1,OLS} \underset{p}{\rightarrow} \alpha_1 + \frac{Cov(p_i, u_i)}{Var(p_i)}$$

**Task 2**

Instrumental variables

Suppose, you have extracted an observable factor $x_i$ from the supply shifter $v_i$.
You receive the following structural form:

$$\begin{aligned} q_i^d &= \alpha_0 + \alpha_1 p_i + u_i \\ q_i^s &= \beta_0 + \beta_1 p_i + \beta_2 x_i + \zeta_i \\ q_i^d &= q_i^s \end{aligned}$$

- Which properties should $x_i$ have to be a valid instrument?
- Solve for the market clearing price $p_i$ and the market clearing quantity $q_i$ to derive the reduced form as in task 1.
- Use the reduced form to derive expressions for $Cov(x_i, p_i)$ and $Cov(x_i, q_i)$.
- What would be a consistent estimator for the slope parameter in the demand equation $\alpha_1$?