Master MIF

Financial Econometrics

Nour Meddahi (Toulouse School of Economics)

Basics on Statistical Inference of Linear Models

I. Examples

• **Example 1: Capital Asset Pricing Model (CAPM).** Denote the arithmetic returns of an asset i, the market index, and the risk-free rate of interest by r_t^i, r_t^M, and r_t^B. Then, the basic CAPM implies

$$E[r_{t+1}^i - r_{t+1}^B] = \beta_i(E[r_{t+1}^M - r_{t+1}^B]),\ \beta_i = \frac{Cov[r_{t+1}^i, r_{t+1}^M]}{Var[r_{t+1}^M]},$$

which means

$$r_{t+1}^i - r_{t+1}^B = \beta_i(r_{t+1}^M - r_{t+1}^B) + \varepsilon_{t+1},\ E[\varepsilon_{t+1}] = 0,\ Cov[r_{t+1}^M, \varepsilon_{t+1}] = 0.$$

• Consequently, one should study the linear regression model

$$r_{t+1}^i - r_{t+1}^B = \alpha_i + \beta_i(r_{t+1}^M - r_{t+1}^B) + \varepsilon_{t+1}.$$

• **Example 2: Forecasting the Market's return.** A common variable used to predict future stock returns is the dividend-price ratio. More precisely, one is interested in the regression

$$r_{t:t+k} = a + b\frac{D_t}{S_t} + \varepsilon_{t+1},\ \text{where } r_{t:t+k} = \frac{S_{t+k}}{S_t} - 1.$$

One could also add other variables in the equation like the short term interest rate or some macroeconomic variables like the "CAY" variable of Lettau and Ludvigson.

II. The Linear Model

We consider a random variable Y_t (called a dependent variable) that we try to explain by other variables X_{2t}, X_{3t}, ..., X_{kt} and the constant (called regressors). We are interested in explaining the mean of Y_t. We postulate the linear regression

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + ... + \beta_k X_{kt} + u_t,\ t = 1, 2, ..., n.$$

The term u is called the *disturbance term*, or the *regression error.* A traditional assumption about u is

$$E[u_t | X_{2t}, X_{3t}, ..., X_{kt}] = 0.$$

In this case,

$$E[Y_t | X_{2t}, X_{3t}, ..., X_{kt}] = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + ... + \beta_k X_{kt}.$$

However, when one considers ad hoc variables, one can not assume $E[u_t | X_{2t}, X_{3t}, ..., X_{kt}] = 0$. However, one can always consider the linear regression model with

$$E[u_t] = 0,\ \text{and}\ Cov[u_t, X_{2t}] = Cov[u_t, X_{3t}] = ... = Cov[u_t, X_{kt}] = 0.$$

In this case, one can adopt the following notation:

$$EL[Y_t | X_{2t}, X_{3t}, ..., X_{kt}] = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + ... + \beta_k X_{kt}.$$

Matrix formulation of the model.

Assuming to have a *sample* of n observations of data, we define the following vectors:

$$\mathbf{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \mathbf{x}_t = \begin{pmatrix} 1 \\ X_{2t} \\ \vdots \\ X_{kt} \end{pmatrix}$$

and

$$\mathbf{X} = (\mathbf{x_1} \;\; \mathbf{x_2} \;\; \cdots \;\; \mathbf{x_n})' \;\; n \times k \text{ matrix.}$$

Observe that

$$Y_t = \mathbf{x}_t' \beta + u_t$$

where

$$\beta = (\beta_1, \beta_2, ..., \beta_k)'.$$

Hence,

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}.$$

III. OLS Estimator

In practice, we do not know the vector β. Likewise, we do not observe the disturbances u_1, u_2,..., u_n. We have too much unknown parameters (k+n) and n equations. Hence, we can not characterize β.

This leads us to introduce a criterion that we will minimize in order to approximate or estimate β. One possible criterion is

$$f(\mathbf{b}) = \sum_{t=1}^{n} (y_t - b_1 - b_2 X_{2t} - ... - b_k X_{kt})^2.$$

An other possible criterion when u_t is symmetric is

$$g(\mathbf{b}) = \sum_{t=1}^{n} |y_t - b_1 - b_2 X_{2t} - ... - b_k X_{kt}|.$$

We will focus on the first criterion.

Observe that me make the difference between the parameter of interest β (called the true parameter) and the vector $\mathbf{b}$ that we use when we define functions. Likewise, we will use the notation

$$\mathbf{u}(\mathbf{b}) = \mathbf{y} - \mathbf{X}\mathbf{b}$$

The function $f(\cdot)$ is called the residuals sum of the squares (RSS):

$$RSS(\mathbf{b}) = \sum_{t=1}^{n}(y_t - b_1 - b_2x_{2t} - ... - b_kx_{kt})^2 = \sum_{t=1}^{n} u_t^2(\mathbf{b}).$$

The **Ordinary Least Squares** (OLS) estimator is defined as the vector that minimizes $RSS(b)$, i.e.,

$$\hat{\beta} = \text{argmin}_{\mathbf{b}\in\mathbf{B}} RSS(b),$$

where $\mathbf{B}$ is some subset of $\mathbf{R}^k$.

Analytical solution.

It turns out that the above minimization problem is easy to be solved, since

$$\begin{aligned} RSS(\mathbf{b}) &= \mathbf{u}'(\mathbf{b})\mathbf{u}(\mathbf{b}) \\ &= (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) = (\mathbf{y}' - \mathbf{b}'\mathbf{X}')(\mathbf{y} - \mathbf{Xb}) \\ &= \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{Xb} + \mathbf{b}'\mathbf{X}'\mathbf{Xb} \\ &= \mathbf{y}'\mathbf{y} - \mathbf{2b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{Xb}. \end{aligned}$$

Differentiating with respect to $\mathbf{b}$, and equating the result to zero, yields the first order conditions (FOC)

$$\frac{\partial RSS(\mathbf{b})}{\partial \mathbf{b}} \mid_{\mathbf{b}=\hat{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} = 0,$$

giving the *normal equations*:

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}.$$

We will assume that the matrix $X'X$ is non-singular, which holds when the matrix $\mathbf{X}$ has a full rank (no perfect linear dependence among 1, X_{2t},..., and X_{kt}). Hence, the OLS estimator is given by

$$\hat{\beta} = [\mathbf{X}'\mathbf{X}]^{-\mathbf{1}}\mathbf{X}'\mathbf{y}. \tag{1}$$

A second differentiation of $RSS(\mathbf{b})$ leads to

$$\frac{\partial^2 RSS(\mathbf{b})}{\partial \mathbf{b} \partial \mathbf{b}'} = 2\mathbf{X}'\mathbf{X},$$

which is a definite positive matrix, implying that $RSS(\mathbf{b})$ is a convex function and that $\hat{\beta}$ is the unique vector that minimizes $RSS(\mathbf{b})$.

IV. Quality of the regression

Define the OLS residuals $\mathbf{e}$ as

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\beta}.$$

Hence, one gets the FOC

$$\mathbf{X}'\mathbf{e} = \mathbf{0}. \tag{2}$$

Since first column of $\mathbf{X}$ is of ones, then first element of (2) is

$$\sum_{t=1}^{n} e_t = 0, \text{ or } \mathbf{e}'\iota = 0,$$

that is $\bar{e} = \bar{Y} - \hat{\beta}_1 - \hat{\beta}_2\bar{X}_2 - ... - \hat{\beta}_k\bar{X}_k = 0$ (recall that a bar $^{-}$ means the sample mean). The other $k-1$ entries are

$$\mathbf{x}_i'e = 0 \;\; i = 2, ..., k.$$

This also means that the sample covariance between each regressor and the OLS residual is zero, since for $i = 2, ..., k$,

$$\frac{1}{n}\sum_{t=1}^{n}(X_{it} - \bar{X}_i)(e_t - \bar{e}) = \frac{1}{n}\sum_{t=1}^{n} X_{it}(e_t - \bar{e}) = \frac{1}{n}\sum_{t=1}^{n} X_{it}e_t.$$

The following also holds then:

$$\hat{\mathbf{y}}'\mathbf{e} = (\mathbf{X}\hat{\beta})'\mathbf{e} = \hat{\beta}'\mathbf{X}'\mathbf{e} = \mathbf{0}.$$

where $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$ is the vector of *regression values*.

The following decomposition of the sum of the squares then holds:

$$\mathbf{y}'\mathbf{y} = (\hat{\mathbf{y}} + \mathbf{e})'(\hat{\mathbf{y}} + \mathbf{e}) = \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} + \mathbf{e}'\mathbf{e},$$

(no cross terms) or subtracting the squared sample mean of the $\mathbf{y}$ and $\hat{y}$ (which is equal) times sample size

$$\mathbf{y}'\mathbf{y} - n\bar{Y}^2 = (\hat{\mathbf{y}} + \mathbf{e})'(\hat{\mathbf{y}} + \mathbf{e}) - n\bar{Y}^2 = (\hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} - n\bar{Y}^2) + \mathbf{e}'\mathbf{e}.$$

This is also described as $TSS = ESS + RSS$, total sum of the squares equals the sum of explained and residual sum of the squares.

A measure of *goodness of fit* is the *coefficient of multiple correlation* R^2

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \text{ where always } 0 \le R^2 \le 1.$$

It is unity in case of perfect fit and zero in case of worst fit.

But R^2 never decreases if more and more regressors are added, at most it remains constant. Thus *adjusted* R^2:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k}(1 - R^2),$$

which might decrease as new regressors are added if they are non relevant. It can attain negative values though.

V. Properties of the OLS estimator

Mean of the OLS Estimator.

Observe that

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) = \beta + (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{u}.$$

Hence,

$$\mathbf{E}[\hat{\beta} \mid \mathbf{X}] = \beta + (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{E}[\mathbf{u} \mid \mathbf{X}] = \beta \text{ when } \mathbf{E}[\mathbf{u} \mid \mathbf{X}] = \mathbf{0}.$$

Consequently,

$$\mathbf{E}[\hat{\beta}] = \mathbf{E}[\mathbf{E}[\hat{\beta} \mid \mathbf{X}]] = \mathbf{E}[\beta] = \beta,$$

i.e., the OLS estimator is **unbiased**.

Observe that when one assumes $\mathbf{Cov}[\mathbf{u}, \mathbf{X}] = \mathbf{0}$ (and not $\mathbf{E}[\mathbf{u} \mid \mathbf{X}] = \mathbf{0}$), on can show that the ols estimator is asymptotically unbiased, i.e. $\lim_{n \to +\infty} \mathbf{E}[\hat{\beta}] = \beta$.

Variance of the OLS Estimator. We assume again that $\mathbf{E}[\mathbf{u} \mid \mathbf{X}] = \mathbf{0}$.

Given that

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{u},$$

one has

$$\mathbf{Var}[\hat{\beta} \mid \mathbf{X}] = [\mathbf{X}'\mathbf{X}]^{-\mathbf{1}} \ \mathbf{X}' \ \mathbf{Var}[\mathbf{u} \mid \mathbf{X}] \ \mathbf{X} \ [\mathbf{X}'\mathbf{X}]^{-\mathbf{1}}.$$

There are two cases:

1) Homoskedastic case: $\mathbf{Var}[\mathbf{u} \mid \mathbf{X}] = \sigma^{\mathbf{2}}\mathbf{I}$.

Hence,

$$\mathbf{Var}[\hat{\beta} \mid \mathbf{X}] = [\mathbf{X}'\mathbf{X}]^{-\mathbf{1}} \ \mathbf{X}' \ \sigma^{\mathbf{2}}\mathbf{I} \ \mathbf{X} \ [\mathbf{X}'\mathbf{X}]^{-\mathbf{1}} = \sigma^{\mathbf{2}}[\mathbf{X}'\mathbf{X}]^{-\mathbf{1}}.$$

Gauss-Markov Theorem. Under homoskedasticity, The OLS estimator is the *Best Linear Unbiased Estimator* (**BLUE**) of β. Here, linear means an estimator that has the form Gy where G is a matrix. The term best means that for any other linear and unbiased estimator $\tilde{\beta}$, $Var[\tilde{\beta}] - Var[\hat{\beta}]$ is a semi – definite positive matrix.

One can show that the following popular estimator of σ^2 is unbiased:

$$s^2 = \frac{1}{n-k}\mathbf{e}'\mathbf{e}. \tag{3}$$

This leads to the following estimator of $\mathbf{Var}[\hat{\beta} \mid \mathbf{X}]$: $\hat{\mathbf{V}}\mathbf{ar}[\hat{\beta} \mid \mathbf{X}] = \mathbf{s^2}[\mathbf{X}'\mathbf{X}]^{-\mathbf{1}}$.

2) Heteroskedastic case: $\mathbf{Var}[\mathbf{u} \mid \mathbf{X}] = \boldsymbol{\Omega} \neq \sigma^{\mathbf{2}}\mathbf{I}$.

$$\mathbf{Var}[\hat{\beta} \mid \mathbf{X}] = [\mathbf{X}'\mathbf{X}]^{-\mathbf{1}} \ \mathbf{X}' \ \boldsymbol{\Omega} \ \mathbf{X} \ [\mathbf{X}'\mathbf{X}]^{-\mathbf{1}}.$$

The OLS estimator is no longer the best estimator. Next week, we will propose another estimator which is BLUE.

The main problem under heteroskedasticity is that Ω is unknown and therefore it is not easy to estimate $\mathbf{Var}[\hat{\beta} \mid \mathbf{X}]$. There is however a solution provided by Eicker and White.

We assume that the individual observations i and j are independent for any i and j, which implies that the matrix Ω is diagonal. Hence, one has

$$Var[u_t \mid \mathbf{X}_t] = \sigma_t^2.$$

Consequently,

$$\mathbf{Var}[\hat{\beta} \mid \mathbf{X}] = [\mathbf{X}'\mathbf{X}]^{-\mathbf{1}} \left(\sum_{\mathbf{t=1}}^{\mathbf{n}} \mathbf{x_t x_t'} \sigma_{\mathbf{t}}^{\mathbf{2}} \right) [\mathbf{X}'\mathbf{X}]^{-\mathbf{1}}.$$

The Eicker-White is defined as

$$\hat{\mathbf{V}}\mathbf{ar}[\hat{\beta} \mid \mathbf{X}] = [\mathbf{X}'\mathbf{X}]^{-\mathbf{1}} \left(\sum_{\mathbf{t=1}}^{\mathbf{n}} \mathbf{x_t x_t' e_t^2} \right) [\mathbf{X}'\mathbf{X}]^{-\mathbf{1}}. \tag{4}$$

One can show that

$$\mathbf{Plim}\ \mathbf{n}[\hat{\mathbf{V}}\mathbf{ar}[\hat{\beta} \mid \mathbf{X}]\ - \mathbf{Var}[\hat{\beta} \mid \mathbf{X}]] = \mathbf{0}.$$

Here, n shows up because $\mathbf{Plim}\ [\hat{\mathbf{V}}\mathbf{ar}[\hat{\beta} \mid \mathbf{X}] = \mathbf{Plim}\mathbf{Var}[\hat{\beta} \mid \mathbf{X}] = \mathbf{0}$, i.e., we are taking the next order.

Financial data often present heteroskedasticity at daily and weekly frequencies. Therefore, in practice, it is suggested to use the Eicker-White estimator. Note that one can test the homoskedasticity assumption.

VI. Distribution of the OLS estimator under normality and homoskedasticity

Here, we assume that

$$u_t \mid \mathbf{x_t} \sim \mathcal{N}(0, \sigma^2),$$

i.e., the disturbances are normal and homoskedastic. Note that

$$y_t \mid \mathbf{x_t} \sim \mathcal{N}(\mathbf{x_t'}\beta, \sigma^2).$$

Therefore,

$$\hat{\beta} \mid \mathbf{X} \sim \mathcal{N}(\beta, \sigma^2[\mathbf{X'X}]^{-1}).$$

In what follows, $c_{i,j}$ denotes the $i \times j$ element of $[\mathbf{X'X}]^{-1}$. Then, one can show for $i = 1, ...k$,

$$\frac{(\hat{\beta}_i - \beta_i)}{s \ \sqrt{c_{i,i}}} \sim \mathcal{T}(n-k).$$

We have a pivotal statistic and therefore we can build confidence intervals for any component of β.

We can also build a confidence set (ellipsoid) of β given that one can show

$$\frac{(\hat{\beta} - \beta)'[\mathbf{X'X}](\hat{\beta} - \beta)}{k \ s^2} \sim F(k, n-k).$$

VII. Asymptotic Distribution of the OLS Estimator

• In practice, the normality and homoskedasticity assumption is quite stringent and is unlikely to hold. We would like to be able to make inference without imposing distributional assumptions on $\mathbf{u}$. A way out is to derive what is the behaviour of $\hat{\beta}$ as $n \to \infty$. Then, we will use such result as an *approximation* for any finite (possible large) n.

• We now assume that:

$$E(u_t \mid \mathbf{x}_{t-s}) = 0 \text{ for any } s > 0.$$

$$\frac{\mathbf{X}'\mathbf{X}}{n} \to_p \mathbf{Q} \text{ where } \mathbf{Q} \text{ is non-random positive definite.}$$

$$\frac{1}{\sqrt{n}}\mathbf{X}'\mathbf{u} \to_d N(\mathbf{0}, \sigma^2\mathbf{Q}).$$

• One can show:

$$\hat{\beta} \to_p \beta, \quad \sqrt{n}(\hat{\beta} - \beta) \to_d N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1}).$$

• A consistent estimate of $\mathbf{Q}$ is $\frac{1}{n}(\mathbf{X}'\mathbf{X}) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{x_t}\mathbf{x_t'}$ which is definite positive for n sufficiently large.

• Confidence intervals: same method as the one in the normal and homoskedastic case. Use the normal distribution $\mathcal{N}(0,1)$ instead of the $\mathcal{T}(n-k)$.

VIII. Non-linear Transforms of the Parameters

Assume that $\beta \in \mathbb{R}^k$ is a vector and that

$$\sqrt{n}(\hat{\beta}_n - \beta) \rightarrow_d N(0, \Sigma).$$

Consider a function $g(\cdot) : \mathbb{R}^k \rightarrow \mathbb{R}^q$ which is C^1, i.e., it is differentiable and its derivative is a continuous function. Then one has

$$\sqrt{n}(g(\hat{\beta}_n) - g(\beta)) \rightarrow_d N\left(0, \frac{\partial g}{\partial \beta'}(\beta)\ \Sigma\ \frac{\partial g'}{\partial \beta}(\beta)\right).$$

Useful result when one is interested in non-linear transforms of parameters.

IX. Testing Hypotheses

We wish to test whether the null hypothesis is rejected or not:

$$H_0 : \ \ \mathbf{R}\beta = r,$$

where $\mathbf{R}$ is $q \times k$ constant matrix and $\mathbf{r}$ is q constant vector. Examples of this are:

$$H_0 : \ \ \beta_i = 0.$$
$$H_0 : \ \ \beta_i = \beta_{i0}, \text{ for some given value } \beta_{i0}.$$
$$H_0 : \ \ \beta_2 + \beta_3 = 1.$$
$$H_0 : \ \ \beta_3 = \beta_4.$$

The idea is compute sample equivalent $\mathbf{R}\hat{\beta} - r$ and see whether it is large or small (in a statistical sense). We assume that $Rank(\mathbf{R}) = q$.

We have two cases: 1) The normal and homoskedastic case and 2) the non-normal or heteroskedastic case. In the first case, we will use the finite sample distribution of the OLS while we will use the asymptotic theory in the second case.

1) The normal and homoskedastic case: We have

$$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$$

Therefore, under H_0,

$$\mathbf{R}\hat{\beta} \sim N(\mathbf{r}, \sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{R}'),$$

given that $E(\mathbf{R}\hat{\beta}) = \mathbf{R}\beta$ and $\text{var}(\mathbf{R}\hat{\beta}) = \mathbf{R}\text{var}(\hat{\beta})\mathbf{R}'$. Hence,

$$(\mathbf{R}\hat{\beta} - \mathbf{r})'[\sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta} - \mathbf{r}) \sim \chi^2_q.$$

The previous statistic is not pivotal given that we do not know σ^2. It can be shown that $\dfrac{\mathbf{e}'\mathbf{e}}{\sigma^2} \sim \chi^2_{n-k}$ and that the latter is independently distributed of $\hat{\beta}$. In addition, one has

$$F \equiv \frac{(\mathbf{R}\hat{\beta} - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta} - \mathbf{r})/q}{\mathbf{e}'\mathbf{e}/(n-k)} \sim F(q, n-k),$$

which is a pivotal statistic.

When $q = 1$ and $H_0: \;\beta_i = \beta_{i0}$ then $F = \dfrac{(\hat{\beta}_i - \beta_{i0})^2}{s^2 c_{ii}} \sim F(1, n-k)$, where c_{ii} is (i,i)th element on diagonal of $(X'X)^{-1}$. However, since now $q = 1$, we often use

$$T \equiv \frac{\hat{\beta}_i - \beta_{i0}}{s\sqrt{c_{ii}}} \sim \mathcal{T}(n-k),$$

• At the other end we can test whether *all* the regression coefficients (other than β_1) are zero and the F-statistic becomes

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F(k-1, n-k).$$

Softwares often provide the previous statistic.

• Another important example is when null hypothesis says that only *subset* of coefficients are zero. Partition $\beta = (\beta_a' \, \beta_b')'$, where β_b is of dimension $k_b \times 1$, with $1 \leq k_b < k$, and let

$$H_0 : \quad \beta_b = 0.$$

Then

$$F = \frac{(\mathbf{e}^* \mathbf{e}^* - \mathbf{e}' \mathbf{e})/k_b}{\mathbf{e}' \mathbf{e}/(n-k)} \sim F(k_b, n-k).$$

$\mathbf{e}^* \mathbf{e}^*$ is the RSS from the restricted regression, that is regressing Y_t on $(X_1, ..., X_{k_a})$. This contrast with the (regular) regression, which yields $\hat{\beta}$, called then the *unrestricted regression* .

2) Non-normal or heteroskedastic case: One should use the same statistics. The difference is their distribution. Instead of using a $\mathcal{T}(n-k)$, one should use a $\mathcal{N}(0,1)$. Likewise, instead of using a $F(q, n-k)$, one should use a χ^2_q.

X. Generalized Least Square Estimator

What happens when

$$E\mathbf{uu}' \neq \sigma^2\mathbf{I} \ ?$$

This arises when u_t are either *heteroskedastic* (variance not constant) or *autocorrelated* or both. Say in general that

$$E\mathbf{uu}' = \sigma^2\mathbf{\Omega},$$

where $\mathbf{\Omega}$ is a $T \times T$ matrix, either diagonal (heteroskedaticity only) or full but certainly non singular. Now σ^2 is just a scaling factor.

The OLS estimator is no longer BLUE. One can derive a BLUE estimator, called the generalized least square (GLS) estimator, as follows: By using linear algebra, one cab find a non-singular matrix $\mathbf{P}$ such that

$$\mathbf{\Omega}^{-1} = P'P.$$

Then if we multiply the regression model by $\mathbf{P}$

$$\mathbf{P}y = PX\beta + Pu \text{ or equivalently } \mathbf{y}^* = X^*\beta + u^*$$

defining $\mathbf{y}^* \equiv Py$, $X^* \equiv PX, u^* \equiv Pu$. Then

$$E\mathbf{u}^*u^{*\prime} = \sigma^2\mathbf{P}\Omega P' = \sigma^2 PP^{-1}P^{-1\prime}P' = \sigma^2 I$$

Therefore the linear model with the starred ($*$) variables does satisfy the Gauss-Markov conditions and for it OLS is the best thing we could do. Let's call it $\hat{\beta}_{GLS}$:

$$\hat{\beta}_{GLS} \equiv (X^{*\prime}X^{*})^{-1}X^{*\prime}y^{*} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$

The problem is that Ω is unknown. One should specify a parametric model for Ω, estimate it to get $\hat{\Omega}$, and then define the feasible GLS (FGLS) estimator defined by

$$\hat{\beta}_{FGLS} \equiv (X'\hat{\Omega}^{-1}X)^{-1}X\hat{\Omega}^{-1}y.$$

There still a problem. What is the parametric model of Ω? The financial theory in the CAPM does not specify it. Likewise, in reduced form models, one could be wrong in specifying a model. Finally, one needs an assumption (strong exogeneity) to show the consistence of the GLS that is often too restrictive and unrealistic in financial model. In practice, one does not use the GLS method in financial econometrics.

XI. Serial Correlation in the Disturbances: The HAC Estimator

While one does not use the GLS estimator when Ω is unknown, one has to estimate consistently $Var[\hat{\beta}^{OLS}]$. Under heteroskedasticity, one should use the Eicker-White estimator. However, the Eicker-White estimator is not consistent when the disturbances u_t are serially correlated. There are two leading examples:

1) Multi-horizon forecasting: $r_{t+1:t+k} = x_t'\beta + \varepsilon_{t+k}$. Due to the overlapping of periods, the disturbances ε_{t+k} are correlated. The OLS estimator is still consistent, biased in finite sample and not asymptotically. We need to estimate $Var[\hat{\beta}]$.

2) We want to estimate the mean of the short term interest rate r_t, $\bar{r}$, and a variance of $\bar{r}$. The problem is that the short term interest rate is highly correlated with unknown correlation (if we do not specify a model).

Let us focus on the second example.

$$Var[\bar{r}] = Var\left[\frac{1}{n}\sum_{t=1}^{n} r_t\right] = \frac{1}{n^2}\sum_{1\leq i,j\leq n} Cov[r_i, r_j] = \frac{1}{n}Var[r_t] + \frac{2}{n}\sum_{l=1}^{n-1}\left(1 - \frac{l}{n}\right) Cov[r_t, r_{t+l}],$$

under the assumption $E[r_i] = E[r_{i+h}]$ and $Cov[r_i, r_j] = Cov[r_{i+h}, r_{j+h}]$ for any i, j, h. In this case, we will say that the process r_t is a second order stationary process.

One can show that $lim_{n\to\infty} nVar[\bar{r}] = Var[r_t] + 2\sum_{l=1}^{\infty} Cov[r_t, r_{t+l}]$.

A potential estimator of $Var[\sqrt{n}\bar{r}]$ is $\hat{V}ar[r_t] + 2\sum_{l=1}^{\infty} \hat{C}ov[r_t, r_{t+l}]$, where $\hat{C}ov[r_t, r_{t+l}] = \frac{1}{n-k}\sum_{t=1}^{n-l}(r_t - \bar{r})(r_{t+l} - \bar{r})$. There are three problems. First, we have finite sample, se will not be able to estimate an infinite number of parameters. Second, we should estimate a small number of parameters, otherwise the quality of the estimators is poor. Finally, we have to be sure that the estimator is positive (univariate case) or positive definite (regression case).

A solution has been proposed by Newey and West. They show that the following estimator is positive and consistent (under some assumptions)

$$\hat{V}ar[\sqrt{n}\bar{r}] = \hat{V}ar[r_t] + 2\sum_{l=1}^{L}\left(1 - \frac{l}{L}\right)\hat{C}ov[r_t, r_{t+l}].$$

Such estimator is called a Heteroskedasticity and Autocorrelation Consistent (HAC) estimator of the standard errors. The parameter L is called the truncation parameter of the HAC estimator. L must be chosen such that it is large in large samples, although still much less than n. A good guideline is $L = 0.75n^{1/3}$.

In the regression case, form the formula $\hat{\beta} = \beta + [\mathbf{X'X}]^{-\mathbf{1}}\mathbf{X'u}$, one gets

$$
\begin{aligned}
Var[\hat{\beta} \mid \mathbf{X}] &= [\mathbf{X'X}]^{-\mathbf{1}} Var[\mathbf{X'u}][\mathbf{X'X}]^{-\mathbf{1}} \\
&= \left[\sum_{t=1}^{n} \mathbf{x_t x_t'}\right]^{-1} Var\left[\sum_{t=1}^{n} x_t u_t\right] \left[\sum_{t=1}^{n} \mathbf{x_t x_t'}\right]^{-1} \\
&= \frac{1}{n}\left[\frac{1}{n}\sum_{t=1}^{n} \mathbf{x_t x_t'}\right]^{-1} Var\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t u_t\right] \left[\frac{1}{n}\sum_{t=1}^{n} \mathbf{x_t x_t'}\right]^{-1}.
\end{aligned}
$$

The Newey and West estimator of $Var\left[\sqrt{n}\sum_{t=1}^{n} x_t u_t\right]$ is given by

$$
\hat{\Sigma}_{xu} = \hat{V}ar[x_t u_t] + \sum_{l=1}^{L}\left(1 - \frac{l}{L}\right)\left(\hat{C}ov[x_t u_t, x_{t+l}u_{t+l}] + \hat{C}ov[x_t u_t, x_{t+l}u_{t+l}]'\right),
$$

where

$$
\hat{C}ov[x_t u_t, x_{t+l}u_{t+l}] = \frac{1}{n-k}\sum_{t=1}^{n-l}(x_t u_t - \bar{xu})(x_{t+l}u_{t+l} - \bar{xu})'.
$$

Then, a positive definite estimator of the variance of $\hat{\beta}$ is

$$
Var[\hat{\beta}] = \frac{1}{n}\left[\frac{1}{n}\sum_{t=1}^{n} \mathbf{x_t x_t'}\right]^{-1} \hat{\Sigma}_{xu} \left[\frac{1}{n}\sum_{t=1}^{n} \mathbf{x_t x_t'}\right]^{-1}.
$$

Observe that the Eicker-White estimator is a special case of the HAC estimators.

XII. Instrumental Variable Estimation

Assume that we want to estimate the simple model $y = \beta\tilde{x} + u$, $E[u\tilde{x}] = 0$, $E[u] = 0$. However, the variable $\tilde{x}$ is measured with an error, $x = \tilde{x} + v$ with v independent with $\tilde{x}$ and with u, and $E[v] = 0$. For simplicity, we assume $E[\tilde{x}] = 0$. What are the properties of the OLS estimator $\hat{\beta}$?

$$\hat{\beta} = ArgMin_b \sum_{t=1}^{n}(y_t - bx_t)^2 = \frac{\sum_{t=1}^n x_t y_t}{\sum_{t=1}^n x_t^2} = \frac{\sum_{t=1}^n x_t(\beta\tilde{x}_t + u_t)}{\sum_{t=1}^n x_t^2} = \frac{\sum_{t=1}^n x_t(\beta(x_t - v_t) + u_t)}{\sum_{t=1}^n x_t^2}$$

$$= \beta - \beta\frac{\sum_{t=1}^n(\tilde{x}_t + v_t)v_t}{\sum_{t=1}^n x_t^2} + \frac{\sum_{t=1}^n x_t u_t}{\sum_{t=1}^n x_t^2}$$

$$= \beta - \beta\frac{\frac{1}{n}\sum_{t=1}^n \tilde{x}_t v_t}{\frac{1}{n}\sum_{t=1}^n x_t^2} - \beta\frac{\frac{1}{n}\sum_{t=1}^n v_t^2}{\frac{1}{n}\sum_{t=1}^n x_t^2} + \frac{\frac{1}{n}\sum_{t=1}^n x_t u_t}{\frac{1}{n}\sum_{t=1}^n x_t^2}.$$

Hence,

$$\text{Plim}\hat{\beta} = \beta - \beta\frac{E[\tilde{x}_t v_t]}{E[x_t^2]} - \beta\frac{E[v_t^2]}{E[x_t^2]} + \frac{E[x_t u_t]}{E[x_t^2]}$$

$$= \beta - \beta\frac{Var[v_t]}{Var[\tilde{x}_t] + Var[v_t]} = \beta\frac{Var[\tilde{x}_t]}{Var[\tilde{x}_t] + Var[v_t]}.$$

Hence, the OLS estimator is biased and inconsistent when $Cov[x_t, u_t] \neq 0$. The same inconsistent problem happens when the variable $\tilde{x}$ is endogenous.

Solution: Instrumental variable.

Assume that we have a variable z_t such that $Cov[z_t, \tilde{x}_t] \neq 0$, $Cov[z_t, v_t] = 0$ and $Cov[z_t, u_t] = 0$. Define $\hat{x}_t$ as $\hat{x}_t = \gamma z_t$ where $x_t = \gamma z_t + \eta_t$ with $Cov[\hat{z}_t, \eta_t] = 0$ (regression of x_t on z_t). Then,

$$\begin{aligned}
\hat{\beta}_{IV} = ArgMin_b \sum_{t=1}^{n} (y_t - b\hat{x}_t)^2 &= \frac{\sum_{t=1}^n \hat{x}_t y_t}{\sum_{t=1}^n \hat{x}_t^2} = \frac{\sum_{t=1}^n \hat{x}_t(\beta\tilde{x}_t + u_t)}{\sum_{t=1}^n \hat{x}_t^2} = \frac{\sum_{t=1}^n \hat{x}_t(\beta(x_t - v_t) + u_t)}{\sum_{t=1}^n \hat{x}_t^2} \\
&= \frac{\sum_{t=1}^n \hat{x}_t(\beta(\hat{x}_t + \eta_t - v_t) + u_t)}{\sum_{t=1}^n \hat{x}_t^2} \\
&= \beta + \beta\frac{\frac{1}{n}\sum_{t=1}^n \hat{x}_t\eta_t}{\frac{1}{n}\sum_{t=1}^n \hat{x}_t^2} - \beta\frac{\frac{1}{n}\sum_{t=1}^n \hat{x}_t v_t}{\frac{1}{n}\sum_{t=1}^n \hat{x}_t^2} + \frac{\frac{1}{n}\sum_{t=1}^n \hat{x}_t u_t}{\frac{1}{n}\sum_{t=1}^n \hat{x}_t^2}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\text{Plim}\hat{\beta}_{IV} &= \beta + \beta\frac{E[\hat{x}_t\eta_t]}{E[\hat{x}_t^2]} - \beta\frac{E[\hat{x}_t v_t]}{E[\hat{x}_t^2]} + \frac{E[\hat{x}_t u_t]}{E[\hat{x}_t^2]} \\
&= \beta.
\end{aligned}$$

Hence, the IV estimator is a consistent estimator of β. In practice, be sure that $Cov[z_t, x_t] \neq 0$, otherwise one faces the problem of weak instruments.

General approach: A consistent estimator of the general model $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$, when $Cov[\mathbf{X}, \mathbf{U}] \neq 0$, can be obtained though if we could find a matrix $\mathbf{Z}$ of order $n \times l$, with $l \geq k$ (more instruments than variables) such that: 1) the variables in $\mathbf{Z}$ correlated with those in $\mathbf{X}$ and plim $\mathbf{Z}'X/n = \Sigma_{\mathbf{Z}X}$ finite and full rank. 2) plim $\mathbf{Z}'\mathbf{u}/n = 0$.

Pre-multiplying the regression model by $\mathbf{Z}'$ yields

$$\mathbf{Z}'\mathbf{y} = \mathbf{Z}'\mathbf{X}\beta + \mathbf{Z}'\mathbf{u},\ \mathrm{var}(\mathbf{Z}'\mathbf{u}) = \sigma^2(\mathbf{Z}'\mathbf{Z}).$$

This suggests using GLS yielding the so-called instrumental variable estimator:

$$\hat{\beta}_{IV} = (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}'\mathbf{y} = (\mathbf{X}'\mathbf{P}_{\mathbf{Z}}\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{P}_{\mathbf{Z}}\mathbf{y},$$

setting $\mathbf{P}_{\mathbf{Z}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}'$. The covariance matrix is

$$\mathrm{var}(\hat{\beta}_{IV}) = \sigma^2(\mathbf{X}'\mathbf{P}_{\mathbf{Z}}\mathbf{X})^{-1},$$

and disturbance variance may be estimated by

$$\hat{\sigma}^2_{IV} = (\mathbf{y} - \mathbf{X}\hat{\beta}_{\mathbf{IV}})'(\mathbf{y} - X\hat{\beta}_{IV})/n.$$

Special case when $l = k$. Then $\mathbf{Z}'\mathbf{X}$ non-singular yielding

$$\hat{\beta}_{IV} = (\mathbf{Z}'\mathbf{X})^{-\mathbf{1}}\mathbf{Z}'\mathbf{y} \text{ with } \mathrm{var}(\hat{\beta}_{IV}) = \sigma^2(\mathbf{Z}'\mathbf{X})^{-\mathbf{1}}(\mathbf{Z}'\mathbf{Z})(\mathbf{X}'\mathbf{Z})^{-\mathbf{1}},$$

Master MIF

Financial Econometrics

Nour Meddahi (Toulouse School of Economics)

Basics on Estimation and Testing

I. Basics on Estimation

Main goal: Often, we want to characterize the relationship between two variables (e.g. the market's and Vodafone's returns) or to understand the behavior of a random variable (e.g. the return of an asset) in order to forecast it, etc.... For this purpose, we need to know some parameters of interest. For instance, the conditional distribution of Vodafone's return given the market's return will depend on some parameters θ that are **unknown**. Therefore, we have to use the **data** to extract the information about θ. More precisely, we will propose a function of the data, called an **estimator** of θ and denoted $\hat{\theta}$, that **approximates** θ. We call this method **estimation**. We will study the properties of $\hat{\theta}$.

Definition: Random Sample. A sample of n observations on one or more variables, denoted $x_1, x_2, ..., x_n$ is a **random sample** if the n observations are drawn independently from the same population, or probability distribution, $F(x; \theta)$. The random sample is said to be **independent, identically distributed** and denoted **i.i.d.**

When one studies time series, one considers **consecutive** observation of the same variable (like daily returns of an stock). In that case, the sample is not i.i.d.

Definition: Statistic. Any function of the data $x_1, x_2, ..., x_n$ is called a **statistic**.

Definition: Empirical Distribution. Let $(x_1, ..., x_n)$ be a random sample. The empirical distribution is the distribution that assigns probability $1/n$ to each x_i, i.e.

$$P_n[X = x_i] = \frac{1}{n}.$$

Definition: Sample Moments. The moments of the empirical distributions are called the sample moments: for any function $h(X)$, its sample moments is

$$E_n[h(X)] \equiv \frac{1}{n} \sum_{i=1}^{n} h(x_i).$$

Definition: Sample Frequency. A sample frequency is the observed frequency of an event $X \in A$.

Note that $P(X \in A) = E[\mathbf{1}_{X \in A}]$ where $\mathbf{1}_B$ is the indicator function, i.e. $\mathbf{1}_B = 1$ when the event B holds and $\mathbf{1}_B = 0$ otherwise. Hence,

$$P_n[X \in A] = E_n[\mathbf{1}_{X \in A}] = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{x_i \in A}.$$

Application: Histogram, i.e., the empirical density function.

Properties of estimators.

• An estimator is a random variable given that it is a function of the data which are random. Therefore, we will try to characterize its density function, expectation, variance, etc...

• In general, it is very hard to characterize the distribution of an estimator for a given sample size n. We will need the use of the asymptotic theory (or large sample theory).

• Consequently, we will focus for the moment on the expectation and the variance of an estimator.

• A desirable property of any estimator is to be unbiased, i.e., if you repeat the experiment, on average, you will obtain the unknown parameter θ.

Definition: Unbiased Estimator. $\hat{\theta}$ is unbiased when

$$E[\hat{\theta}] = \theta.$$

Definition: Bias of an Estimator. The bias of an estimator $\hat{\theta}$ is defined as

$$\text{Biais}[\hat{\theta}] = E[\hat{\theta}] - \theta.$$

Examples: Assume that $E[X] = \mu$ and $Var[X] = \sigma^2$.

- The sample average is an unbiased estimator of $\mu = E[X]$.

$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n}\sum_{i=1}^{n} E[X] = E[X] = \mu.$$

- The sample variance is biased. One can show that

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2\right] = \frac{n-1}{n}\sigma^2.$$

This leads to the introduction of a new unbiased estimator of σ^2 (used in practice and in softwares) given by

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2.$$

• Another desirable property of an unbiased estimator is relate to its variance. We would prefer a small variance.

Example: variance of the sample mean.

$$Var[\bar{X}] = \frac{\sigma^2}{n}$$

Definition: Efficiency. Consider two unbiased estimator $\hat{\theta}_1$ and $\hat{\theta}_2$ of θ. $\hat{\theta}_1$ will be called more efficient $\hat{\theta}_2$ when

$$Var[\hat{\theta}_1] < Var[\hat{\theta}_2].$$

When θ is vector, the previous definition means that the matric $Var[\hat{\theta}_2] - Var[\hat{\theta}_1]$ is a positive definite matrix.

• It is also of interest to compare estimators that are biased. In this case, we need a criterion that combines the bias and the variance of the estimator.

Definition: Mean-squared error of an estimator. It is defined as

$$\begin{aligned} MSE[\hat{\theta}] &\equiv E[(\hat{\theta}-\theta)^2] = Var[\hat{\theta}] + (Bias[\hat{\theta}])^2 \text{ when } \theta \text{ is a scalar} \\ &\equiv E[(\hat{\theta}-\theta)(\hat{\theta}-\theta)'] = Var[\hat{\theta}] + (Bias[\hat{\theta}])(Bias[\hat{\theta}])' \text{ when } \theta \text{ is a vector.} \end{aligned}$$

We always prefer the estimator that has the smallest MSE (when it is possible to have one).

Example: When $X \sim \mathcal{N}(\mu, \sigma^2)$, one can show that

$$Var[\hat{V}ar[X]] = \frac{2(n-1)\sigma^4}{n^2} \; < \; Var[s^2] = \frac{2\sigma^4}{n-1}$$

while

$$MSE[\hat{V}ar[X]] - MSE[s^2] = \sigma^4 \left[\frac{2n-1}{n^2} - \frac{2}{n-1} \right] > 0.$$

In other words, we prefer the estimator s^2.

• There are other properties of estimators that are important, in particular **consistency**. These properties are defined when the sample size $n \to +\infty$.

Concepts of Convergence.

Convergence in probability . The sequence of random variables Z_n converges in probability to a random variable (or a constant) Z if, for any $\epsilon > 0$,

$$Prob(\mid Z_n - Z \mid > \epsilon) \to 0 \text{ as } n \to \infty.$$

We also write $Z_n \to_p x$ as $n \to \infty$ or $\text{Plim}_{n\to\infty} Z_n = Z$. In many cases of interest, Z will be a non-stochastic constant.

Consistent estimator: An estimator θ_n of θ is called consistent if Plim $\theta_n = \theta$.

Convergence in Mean-Square or L^2. The sequence of random variables Z_n converges in mean-square to a random variable (or a constant) Z if

$$E[(Z_n - Z)^2] \to 0 \text{ as } n \to \infty \text{ (scalar case)}$$

$$E[||Z_n - Z||^2] \to 0 \text{ as } n \to \infty \text{ (multivariate and matricx cases)}$$

$$\textbf{Convergence in mean} - \textbf{square} \Longrightarrow \textbf{Convergence in probability}.$$

Application: sample mean.

$$E[(\bar{X}_n - \mu)^2] = \frac{\sigma^2}{n} \to 0 \text{ as } n \to +\infty.$$

i.e., $\bar{X}_n$ converges in mean-square to μ. Consequently, $\bar{X}_n$ converges probability to μ, i.e., $\bar{X}_n$ is a consistent estimator of μ.

We will now study the asymptotic distribution of the estimators. We need to introduce the concept of of **convergence in distribution**.

Convergence in distribution. Let Z_n be a sequence of random variables, each with a cdf $F_n(z)$. Then Z_n converges in distribution to the random variable Z with cdf $F(z)$ if for any z

$$F_n(z) \to F(z) \text{ as } n \to \infty.$$

We shall indicate this as $Z_n \to_d Z$ as $n \to \infty$. We shall call $F(z)$ the **asymptotic distribution** of Z_n.

Central Limit Theorem: By Central Limit Theorem (CLT) we mean the set of results which state the conditions under which a sequence of rv, suitably normalized, converges to a standard normal. A simple example is the **Lindberg-Levy** CLT: if x_i are $i.i.d.$ with finite mean μ and finite variance σ^2 then

$$\sqrt{n}(\bar{X}_n - \mu) \to_d N(0, \sigma^2).$$

Application:

$$\sqrt{n}\frac{(\bar{X}_n - \mu)}{\hat{\sigma}} \to_d N(0,1) \quad \text{and} \quad \sqrt{n}\frac{(\bar{X}_n - \mu)}{s} \to_d N(0,1).$$

II. Basics on Finite Sample Distribution of estimators.

• Ideally, we would like to know the finite sample distribution (i.e. the exact distribution) of the estimator $\hat{\theta}$ to fully know the properties of the estimator. Unfortunately, there are few examples where one can obtain the exact distribution. Instead of the exact distribution, one will use the asymptotic theory to approximate the distribution of the estimator.

• An important example where one can derive the exact distribution of some estimators is the normal case.

The normal case: Assume that x_1, x_2, ..., x_n, are i.i.d. and follow a $\mathcal{N}(\mu, \sigma^2)$. Then one can show:

1. $$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$$

2. $$\frac{s^2}{\sigma^2} \sim \frac{\chi^2(n-1)}{n-1}$$

3. The random variables $\bar{X}$ and s^2 are independent

4. $$T = \sqrt{n}\,\frac{(\bar{X} - \mu)}{s} \sim \mathcal{T}(n-1).$$

Point Estimation and Interval Estimation.

• So far, we were interested in estimating θ, i.e., we focused on **point estimation**.

• A different approach is the **interval estimation**: We want to characterize an interval (or a set) such that the probability that the unknown parameter θ is in the interval equals given number, say 95%, and denoted $1 - \alpha$.

• For this purpose, we need a **pivotal statistic**, i.e., a statistic that its distribution is known and does not depend on any parameter.

Example:

1. $\sqrt{n}(\bar{X} - \mu)$ is not a pivotal statistic because its distribution is $\mathcal{N}(0, \sigma^2)$ where σ^2 is unknown.
2. The "T" statistic $\sqrt{n}\,\dfrac{(\bar{X} - \mu)}{s}$ is pivotal because it distribution is $\mathcal{T}(n-1)$.
3. $\dfrac{s^2}{\sigma^2}$ is a pivotal statistic because it distribution $\dfrac{\chi^2(n-1)}{n-1}$ is known.

- We have

$$P\left[\left|\sqrt{n}\ \frac{(\bar{X}-\mu)}{s}\right| \leq \mathcal{T}(n-1)_{\alpha/2}\right] = 1-\alpha$$

where $\mathcal{T}(n-1)_\alpha$ is the $1-\alpha$ quantile of the $\mathcal{T}(n-1)$ distribution $(0 < \alpha < 1)$.

- This probability statement is equivalent to

$$P\left[\mu \in \left[\bar{X} - \mathcal{T}(n-1)_{\alpha/2}\frac{s}{\sqrt{n}}\ ,\ \bar{X} + \mathcal{T}(n-1)_{\alpha/2}\frac{s}{\sqrt{n}}\right]\right] = 1-\alpha.$$

In other words, the interval $\left[\bar{X} - \mathcal{T}(n-1)_{\alpha/2}\frac{s}{\sqrt{n}}\ ,\ \bar{X} + \mathcal{T}(n-1)_{\alpha/2}\frac{s}{\sqrt{n}}\right]$ contains μ with probability $1-\alpha$. It is called a **confidence interval**.

- One can do the same analysis for σ^2 by using the pivotal statistic $\dfrac{s^2}{\sigma^2}$.

- **Example**: $n = 25$, $\bar{X} = 1.63$ and $s = 0.51$. One has

$$P(-2.064 \leq \frac{5(\bar{X}-\mu)}{s} \leq 2.064) = 0.95.$$

Hence, the 95% confidence interval of μ is $[1.4195, 1.8405]$.

III. Basics on Testing Hypotheses

Main goal: Often, financial theory and economic theory imply restrictions that the data should follow. Likewise, when one wants to forecast a variable y, one wants to know whether it is useful to include another variable x. In both case, one has to test the corresponding hypothesis.

• The classical testing procedures are based on constructing a statistic from a random sample that will enable us to decide with reasonable confidence whether or not the data in the sample would have been generated by a hypothesized population. This procedure involves a statement of hypothesis, usually in terms of **null** or maintained hypothesis and **an alternative**, conventionally denoted by H_0 and H_1, respectively.

• Classical (or Neyman-Person) methodology involves splitting sample space into two regions: if the test statistic falls in **rejection region** (also known as critical region), then null hypothesis H_0 is rejected; if the test statistic falls in **acceptance region** , then H_0 is not rejected.

• In order to do a tests, one has to find a pivotal statistic.

• Since the sample is random, the test statistic is also random. The same test procedure can lead to different conclusions in different samples. There are two ways such a procedure can be error:

1. Type I error. The procedure may lead to rejection of the null hypothesis when it is true.
2. Type II error. The procedure may fail to reject the null hypothesis when it is false.

Definition 1 *The probability of type I error is the* **size of the test**. *This is conventionally denoted by* α *and is also called significance level. In other words,*

$$\alpha = P[\text{ Reject the null} \mid \text{the null is true}].$$

The type I error could be eliminated by making the rejection region very small. By eliminating the probability of type I error, that is making it unlikely that the hypothesis is rejected, we must increase the probability of a type II error.

Definition 2 *The* **power of a test** *is the probability that it will correctly lead to rejection of a false null hypothesis:*

$$\text{power} = 1 - \beta = 1 - \mathrm{P}(\text{type II error})$$

with

$$\beta = \mathrm{P}(\text{type II error}) = \mathrm{P}[\text{ Does not Reject the null} \mid \text{the null does not hold}].$$

Definition 3 *A test is* **most powerful** *if it has greater power than any other test of the same size.*

Definition 4 **A test is consistent** *if its power goes to one as the sample size grows to infinity.*

Rejection region and acceptance region.

• These regions are characterized by using a pivotal statistic.

• One should characterize the rejection region, which depends on the alternative. We will use an example to show how one should proceed.

The normal case: Assume that x_1, x_2, ..., x_n, are i.i.d. and follow a $\mathcal{N}(\mu, \sigma^2)$. Then one has the pivotal statistic

$$T = \sqrt{n}\,\frac{(\bar{X} - \mu)}{s} \sim \mathcal{T}(n-1).$$

One should rewrite the pivotal statistic **under the null**. In our example, the null is $H_0 : \mu = \mu^0$, so the pivotal statistic becomes

$$T = \sqrt{n}\,\frac{(\bar{X} - \mu^0)}{s} \sim \mathcal{T}(n-1).$$

First case: A two sided test. $H_0 : \mu = \mu^0,\ H_a : \mu \neq \mu^0$. One has to consider the pivotal statistic written **under the null** and find out the area where the statistic is unlikely to follow the desirable distribution. Here, if the data are generated under the alternative, the absolute value of T should be large, meaning that $\bar{X}$ is far from μ^0. In other words, the rejection region is

$$\left| T = \sqrt{n}\ \frac{(\bar{X} - \mu^0)}{s} \right| > \mathcal{T}(n-1)_{\alpha/2}.$$

Second case: A one sided test. $H_0 : \mu = \mu^0,\ H_a : \mu > \mu^0$. One has to consider the pivotal statistic written **under the null** and find out the area where the statistic is unlikely to follow the desirable distribution. Here, if the data are generated under the alternative, the statistic T should be **positive and large**. Hence, the rejection region is

$$T = \sqrt{n}\ \frac{(\bar{X} - \mu^0)}{s} > \mathcal{T}(n-1)_{\alpha}.$$

• **Example**: $n = 25$, $\bar{X} = 1.63$ and $s = 0.51$.

- $H_0 : \mu = 1.5,\ H_a : \mu \neq 1.5$. One gets $T = 1.27$.

- $H_0 : \mu = 1.5,\ H_a : \mu > 1.5$. One gets $T = 1.27$.

- $H_0 : \mu = 1.3,\ H_a : \mu < 1.3$. One gets $T = 3.23$

• **P-value**. Softwares provide the p-value, which corresponds to the probability of the rejection region when the T statistic is its boundary. One should reject the null when the p-value is smaller than α.

Connection between interval estimation and hypothesis testing.

There is a link between interval estimation and the hypothesis test. The confidence interval give a range of plausible values for the parameter. Therefore, it stands to reason that if a hypothesized value of the parameter does not fall in this range of plausible values, the data are not consistent with the hypothesis and it should be rejected. Consider testing

$$H_0 : \theta = \theta_0 \quad \text{against} \quad H_1 : \theta \neq \theta_0.$$

We know from the previous lecture how to build a confidence interval of θ. If θ^0 does not belong to the confidence set, then we reject the null. Otherwise, we do not reject it.

• **Example**: $n = 25$, $\bar{X} = 1.63$ and $s = 0.51$. The 95% confidence interval of μ is $[1.4195, 1.8405]$.

a) $H_0 : \mu = 1.5$, $H_a : \mu \neq 1.5$. At the 5 percent significance test, we do not reject the null given that 1.5 is in the confidence interval $[1.4195, 1.8405]$.

b) $H_0 : \mu = 2$, $H_a : \mu \neq 2$. At the 5 percent significance test, we reject the null given that 2 is not in the confidence interval $[1.4195, 1.8405]$.