Classical Linear Regression

Machine Learning & Computational Statistics

Ivailo Dimov

January 29, 2026

Readings

In addition to the lecture notes, the following are required readings:

- Chapter 1.1-1.4, in Hayashi (2000);
- Breiman (2001).

Statistical versus Machine Learning Perspectives? I

Both subfields of statistics, machine learning (ML) and econometrics are closely related with significant overlaps.

However, Breiman (2001) eloquently argues that these fields represent two different cultures in their approaches to statistical data analysis, namely

- statistics focuses on *inference*, whereas
- ML emphasizes *prediction*.

Linear Models I

Without going into the details of their exact specifications, some examples of linear models in finance include:

- The market model

$$r_i = \alpha_i + \beta_i r_{\mathrm{mkt}} + \varepsilon_i \, . \tag{1}$$

- The arbitrage pricing theory (APT) models

$$r_i = \alpha_i + \sum_{k=1}^{K} \beta_{ik} f_k + \varepsilon_i \, . \tag{2}$$

- Various "factor-based" forecasting models

$$r_{i,t+1} = \alpha_i + \sum_{k=1}^{K} \beta_{ik} f_{k,t} + \varepsilon_{i,t+1} \tag{3}$$

Linear Models II

With so many powerful methods such as nonparametric statistics, decision trees, neural networks and Gaussian processes, why do we still care about linear models? There are a number of reasons, including

- When the number of parameters is close to or exceeds the number of observations, particularly if the data matrix is sparse, non-linear and more complex models may be challenging to fit.
- Related to the previous point, when we project our data into high dimensional spaces we may end up having more predictors (that frequently are sparse) than we have observations.
- With linear models we can create "meta-variables" that are used as inputs to other ML techniques.

Linear Models III

- We can use linear models to blend the outputs from other models (this is referred to as *ensemble learning*).
- When we face modeling situations with distributions that are more difficult to work with, such as quantile regression on heavy-tailed errors (i.e., Cauchy, Lévy), linear models are a blessing.

Review of Some Distributions

The Chi-Square Distribution I

Let $x_1, \ldots, x_n \sim \mathcal{N}(0,1)$ be independent random variables and define the random variable

$$y := \sum_{i=1}^{n} x_i^2 \,. \tag{4}$$

Then, y has a chi-square distribution with *n degrees of freedom* (*df*).

We denote this by $y \sim \chi_n^2$.

The Chi-Square Distribution II

The pdf of a chi-square distribution with $k > 0$ dof is:

$$p_{\chi}(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} \tag{5}$$

which is a special case of the pdf of the Gamma distribution from the exponential family:

$$p_G(x) = \frac{1}{\theta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\theta}} \tag{6}$$

The Chi-Square Distribution III

Recall that a matrix $\mathbf{A}$ is *idempotent* if $\mathbf{A}^2 = \mathbf{A}$.

Proposition

Suppose $\mathbf{A}$ is a symmetric idempotent matrix and $\mathbf{x} \sim \mathcal{N}(0, \mathbf{I}_n)$. Then $\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi^2_{\mathrm{trace}(\mathbf{A})}$.

The t-Distribution

Let $x \sim \mathcal{N}(0,1)$ and $y \sim \chi^2_n$, both assumed to be independent.

Then, the random variable

$$t := \frac{x}{\sqrt{y/n}} \tag{7}$$

has a t-distribution with n df.

We denote this by $t \sim t_n$.

The F-Distribution

Let $x_1 \sim \chi^2_{k_1}$ and $x_2 \sim \chi^2_{k_2}$, both assumed to be independent.

Then, the random variable

$$F := \frac{(x_1/k_1)}{(x_2/k_2)} \tag{8}$$

has an F-distribution with (k_1, k_2) df.

We denote this by $F \sim F_{k_1,k_2}$.

A Useful Result from Probability

Proposition

Let $\mathbf{x}$ and $\mathbf{y}$ be two independently distributed random vectors. Then $f(\mathbf{x})$ and $g(\mathbf{y})$ are also independently distributed.

The Classical Linear Regression Model

The Classical Linear Model I

We will now discuss the classical approach to the linear regression problem

$$y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon\,, \tag{9}$$

where $y \in \mathbb{R}$ and $\varepsilon \in \mathbb{R}$ are random variables, $\mathbf{x} \in \mathbb{R}^p$ is a random vector and $\boldsymbol{\beta} \in \mathbb{R}^p$ are parameters.

y is referred to as the dependent variable (or left-hand side variable),

$\mathbf{x}$ are the explanatory variables (right-hand side variables or regressors),

$\boldsymbol{\beta}$ are the regression coefficients, and

ε is referred to as the unobserved error (or residual).

The Classical Linear Model II

We assume that we have collected a dataset of the realizations of dependent and explanatory variables,

$$\mathcal{D} := (\mathbf{X}, \mathbf{y}) = \left\{ (\mathbf{x}_i', y_i) \right\}_{i=1}^{n} \tag{10}$$

with $\mathbf{X} \in \mathbb{R}^{n \times p}$ and $\mathbf{y} \in \mathbb{R}^n$.

Our goal is to determine the parameters $\boldsymbol{\beta}$ that provide the "best fit" for the "model" given our data $\mathcal{D}$, so that

$$\mathbf{y} \approx \mathbf{X}\boldsymbol{\beta}\,. \tag{11}$$

Model Assumptions

Using the notation introduced above, for the classical linear regression model we make the following assumptions

I. *Linearity:*

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,. \tag{12}$$

II. *Strict Exogeneity:*

$$\mathbb{E}\left[\boldsymbol{\varepsilon}|\mathbf{X}\right] = \mathbf{0}\,. \tag{13}$$

III. *No Multicollinearity:*

$$\mathbb{P}\left[\mathrm{rank}(\mathbf{X}) = p\right] = 1\,. \tag{14}$$

IV. *Spherical Errors:*

$$\mathrm{var}\left[\boldsymbol{\varepsilon}|\mathbf{X}\right] = \sigma^2\mathbf{I}_n\,. \tag{15}$$

V. *Normality:*

$$\boldsymbol{\varepsilon}|\mathbf{X} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_n)\,. \tag{16}$$

Strict Exogeneity

The “strict” term refers to conditioning on all of $\mathbf{X}$. It is easy to see that strict exogeneity implies that the unconditional mean of the error term is zero

$$\begin{aligned}\mathbb{E}\left[\varepsilon\right] &= \mathbb{E}\left[\mathbb{E}\left[\varepsilon|\mathbf{X}\right]\right] \\ &= \mathbb{E}\left[\mathbf{0}\right] \\ &= \mathbf{0}\,. \qquad (17)\end{aligned}$$

Spherical Errors

The variance of the errors is given by

$$\mathrm{var}\left[\varepsilon|\mathbf{X}\right] = \sigma^2 \mathbf{I}_n \,.$$

We can break this assumption into two parts:

the *homoscedasticity* assumption

$$\mathbb{E}\left[\varepsilon_i^2|\mathbf{X}\right] = \sigma^2 \tag{18}$$

and the *no autocorrelation* assumption

$$\mathbb{E}\left[\varepsilon_i \varepsilon_j|\mathbf{X}\right] = 0 \ \ , i \neq j \,. \tag{19}$$

Normality

The most restrictive assumption is that the errors follow a multivariate normal distribution

$$\varepsilon|\mathbf{X} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)\,.$$

This assumption is needed to perform inference in the finite-sample case.

Ordinary Least Squares I

Ordinary least squares (OLS) seeks the parameter estimate $\widehat{\boldsymbol{\beta}}$ that minimizes the *sum of squared residuals* (SSR)

$$\mathrm{SSR}(\boldsymbol{\beta}) := \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)^2 . \tag{20}$$

As $\mathrm{SSR}(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2$, we define our estimate via

$$\widehat{\boldsymbol{\beta}} := \mathrm{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 . \tag{21}$$

Some algebra shows

$$\begin{aligned}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= (\mathbf{y}' - \boldsymbol{\beta}'\mathbf{X}')(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} . \end{aligned} \tag{22}$$

Ordinary Least Squares II

In order to find the minimum of the SSR, we need to solve the first order condition (FOC)

$$\nabla_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 = \mathbf{0}\,. \tag{23}$$

First, let us compute the gradient

$$\begin{aligned}\nabla_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 &= \nabla_{\boldsymbol{\beta}} \left(\mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right) && (24)\\ &= 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - 2\mathbf{X}'\mathbf{y}\,. && (25)\end{aligned}$$

We obtain the *normal equations*

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}\,, \tag{26}$$

that have the solution

$$\widehat{\boldsymbol{\beta}} := \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}\,. \tag{27}$$

Solving the OLS Problem I

What numerically is the best way to compute the OLS solution depends on the matrix $\mathbf{X}'\mathbf{X}$ (or $\mathbf{X}$):

(a) how well or poorly conditioned it is,

(b) what the structure of it is (sparse/dense/etc.) and

(c) how large is its dimension.

Yup, *numerical methods* & *numerical analysis* skills come handy in statistics and machine learning!

Solving the OLS Problem II

For example:

- If $\mathbf{X}'\mathbf{X}$ is well-conditioned, then a good choice to solve the normal equations is by Cholesky decomposition.
- If $\mathbf{X}'\mathbf{X}$ is (expected to be) poorly conditioned or even non-invertible (if $\mathbf{X}$ is rank deficient), then it is preferable to solve directly

  $$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

  using (a) *QR decomposition* or (b) *singular value decomposition* (SVD).
- If $\mathbf{X}$ is very large and sparse, then numerical methods that take advantage of the sparsity will frequently be most efficient, such as that of iterative methods.

Additional Concepts

We define the *residuals* as

$$\mathbf{r} := \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\,. \tag{28}$$

Note that

$$\mathrm{SSR} = \mathbf{r}'\mathbf{r} \tag{29}$$

Finally, in practice σ^2 is unknown and needs to be estimated. Later, we will show that

$$s^2 := \frac{\mathbf{r}'\mathbf{r}}{n-p} \tag{30}$$

is an unbiased estimator.

The Geometry of OLS I

We define the *range* of matrix $\mathbf{X}$ as

$$\mathcal{R}(\mathbf{X}) := \{\boldsymbol{\theta} : \boldsymbol{\theta} = \mathbf{X}\mathbf{b},\, \mathbf{b} \in \mathbb{R}^p\} \subset \mathbb{R}^n\,. \tag{31}$$

The range of a matrix is also referred to as its *column space*, as the range is spanned by its column vectors.

Using this definition, we can equivalently formulate the least squares problem as

$$\widehat{\boldsymbol{\theta}} = \text{argmin}_{\boldsymbol{\theta} \in \mathcal{R}(\mathbf{X})} \|\mathbf{y} - \boldsymbol{\theta}\|_2^2\,. \tag{32}$$

Proposition (Geometry of OLS)

The minimum, $\widehat{\boldsymbol{\theta}}$, is attained when $(\mathbf{y} - \widehat{\boldsymbol{\theta}}) \perp \mathcal{R}(\mathbf{X})$. In other words, $(\mathbf{y} - \widehat{\boldsymbol{\theta}})$ is perpendicular to all vectors in $\mathcal{R}(\mathbf{X})$.

The Geometry of OLS II

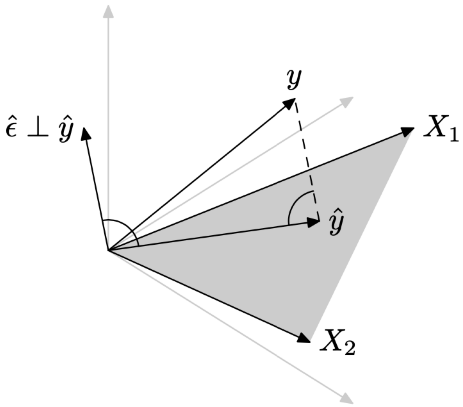

Figure 1: Geometric interpretation of OLS. Here $\mathbf{X} := [\mathbf{X}_1, \mathbf{X}_2]$, $\widehat{\mathbf{y}} := \boldsymbol{\theta}$ and $\widehat{\varepsilon} := \mathbf{y} - \widehat{\mathbf{y}}$. (Source: Rao and Toutenburg (1999)).

Projection and Annihilator Matrices I

This geometric interpretation of OLS, motivates introducing the *projection* and *annihilator* matrices

$$\mathbf{P} := \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \,, \tag{33}$$

$$\mathbf{M} := \mathbf{I}_n - \mathbf{P} \,. \tag{34}$$

Proposition (Properties of $\mathbf{P}$ and $\mathbf{M}$)

1. $\mathbf{P}$ is a symmetric and orthogonal projection from $\mathbb{R}^n$ onto $\mathcal{R}(\mathbf{X})$ (i.e. $\mathbf{P} : \mathbb{R}^n \to \mathcal{R}(\mathbf{X})$, $\mathbf{P} = \mathbf{P}'$ and $\mathbf{P}^2 = \mathbf{P}$).
2. $\mathbf{M}$ is a symmetric and orthogonal projection from $\mathbb{R}^n$ onto $\mathcal{R}(\mathbf{X})^\perp$.

Projection and Annihilator Matrices II

The properties of $\mathbf{P}$ and $\mathbf{M}$ are useful. For instance, it follows that we can rewrite the residual $\mathbf{r} := \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$ as

$$\begin{aligned} \mathbf{r} &= \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} \\ &= \mathbf{y} - \mathbf{P}\mathbf{y} \\ &= (\mathbf{I}_n - \mathbf{P})\mathbf{y} \\ &= \mathbf{M}\mathbf{y} \\ &= \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \mathbf{M}\boldsymbol{\varepsilon}\,, \end{aligned} \tag{35}$$

from which we immediately obtain

$$\begin{aligned} \|\mathbf{r}\|_2^2 &= \|\mathbf{M}\boldsymbol{\varepsilon}\|_2^2 \\ &= \boldsymbol{\varepsilon}'\mathbf{M}'\mathbf{M}\boldsymbol{\varepsilon} \\ &= \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}\,. \end{aligned} \tag{36}$$

Finite Sample Properties of OLS

Finite Sample Properties of OLS I

Proposition (Finite-sample properties of OLS)

1. Under assumptions (I)-(III), we have that $\mathbb{E}[\widehat{\boldsymbol{\beta}}|\mathbf{X}] = \boldsymbol{\beta}$.
2. Under assumptions, (I)-(IV), we have that
   (a) $\mathrm{var}[\widehat{\boldsymbol{\beta}}|\mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$,
   (b) $\widehat{\boldsymbol{\beta}}$ is the best linear unbiased estimator (the Gauss-Markov Theorem),
   (c) $\mathrm{cov}[\widehat{\boldsymbol{\beta}}, \mathbf{r}|\mathbf{X}] = 0$, and
   (d) $\mathbb{E}[s^2|\mathbf{X}] = \sigma^2$.
3. If σ^2 is unknown, then an estimate of $\mathrm{var}[\widehat{\boldsymbol{\beta}}|\mathbf{X}]$ is given by

$$\widehat{\mathrm{var}[\widehat{\boldsymbol{\beta}}|\mathbf{X}]} = s^2 \cdot \left(\mathbf{X}'\mathbf{X}\right)^{-1} ,$$

Finite Sample Properties of OLS II

PROOF. We show OLS is unbiased. The error in our estimate can be written as

$$\begin{aligned}\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \boldsymbol{\beta} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) - \boldsymbol{\beta} \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} - \boldsymbol{\beta} \\ &= \mathbf{A}\boldsymbol{\varepsilon}\,, \end{aligned} \tag{37}$$

where $\mathbf{A} := (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Therefore,

$$\begin{aligned}\mathbb{E}[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\mathbf{X}] &= \mathbb{E}[\mathbf{A}\boldsymbol{\varepsilon}|\mathbf{X}] \\ &= \mathbf{A} \cdot \mathbb{E}[\boldsymbol{\varepsilon}|\mathbf{X}] \\ &= \mathbf{0}\,. \end{aligned}$$

□

Hypothesis Testing under Normality

The t-test I

The difference $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ between the OLS estimator, $\widehat{\boldsymbol{\beta}}$, and the true parameter vector, $\boldsymbol{\beta}$ is called the *sampling error*.

From the previous Proposition, we know that the sampling error is distributed as

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})|\mathbf{X} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right) . \tag{38}$$

This is the key observation underlying the t-test.

The *t*-test II

We want to test the null hypothesis

$$H_0 : \beta_j = b_j \,, \tag{39}$$

where $b_j \in \mathbb{R}$ is some known value versus the alternative hypothesis

$$H_a : \beta_j \neq b_j \,. \tag{40}$$

The *t*-test III

From the sampling error distribution we have that under the null

$$(\widehat{\beta}_j - b_j)|\mathbf{X}, H_0 \sim \mathcal{N}\left(0, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}_{jj}\right). \tag{41}$$

If σ^2 is known, then the *z*-ratio (or *z*-statistic)

$$z_j := \frac{\widehat{\beta}_j - b_j}{\sqrt{\sigma^2\left((\mathbf{X}'\mathbf{X})^{-1}_{jj}\right)}}, \tag{42}$$

is distributed (under the null)

$$z_j|\mathbf{X}, H_0 \sim \mathcal{N}(0,1).$$

The t-test IV

Frequently, σ^2 is unknown and needs to be estimated with s^2. Replacing σ^2 with s^2 in the z-ratio , we obtain the t-ratio (or t-statistic)

$$t_j := \frac{\widehat{\beta}_j - b_j}{\sqrt{s^2\left((\mathbf{X}'\mathbf{X})^{-1}_{jj}\right)}} = \frac{\widehat{\beta}_j - b_j}{\mathrm{SE}(\widehat{\boldsymbol{\beta}}_j)},$$

where SE $:= \sqrt{s^2\left((\mathbf{X}'\mathbf{X})^{-1}_{jj}\right)}$ is referred to as the *standard error* of the OLS estimate $\widehat{\beta}_j$.

The t-test V

Proposition

Under the classical linear regression assumptions (I)-(V) the t-ratio follows a t-distribution with $(n - p)$ degrees of freedom under the null hypothesis, that is

$$t_j | \mathbf{X}, H_0 \sim t_{n-p} \, .$$

The t-test VI

PROOF. Observe that we can rewrite the t-ratio as follows

$$\begin{aligned} t_j &= \frac{\widehat{\beta}_j - b}{\sqrt{\sigma^2\left((\mathbf{X}'\mathbf{X})^{-1}_{jj}\right)}} \cdot \sqrt{\frac{\sigma^2}{s^2}} = \frac{z_j}{\sqrt{s^2/\sigma^2}} \\ &= \frac{z_j}{\sqrt{\frac{\mathbf{r}'\mathbf{r}/(n-p)}{\sigma^2}}} \\ &= \frac{z_j}{\sqrt{q/(n-p)}} \end{aligned} \tag{43}$$

where $q := \mathbf{r}'\mathbf{r}/\sigma^2$ and z_j is the z-ratio.

We need to show that
(1) $q|\mathbf{X} \sim \chi^2_{n-p}$ and
(2) z_j and q are independent conditional on $\mathbf{X}$.

The t-test VII

(1) Using that $\mathbf{r}'\mathbf{r} = \varepsilon' \mathbf{M} \varepsilon$, we have

$$q = \frac{\mathbf{r}'\mathbf{r}}{\sigma^2} \tag{44}$$

$$= \frac{\varepsilon'}{\sigma} \mathbf{M} \frac{\varepsilon}{\sigma} . \tag{45}$$

From the normality assumption (V), we have that

$$\frac{\varepsilon}{\sigma} \sim \mathcal{N}(0, \mathbf{I}_n) , \tag{46}$$

and hence $q \sim \chi^2_{\mathrm{trace}(\mathbf{M})}$. It is not hard to see that $\mathrm{trace}(\mathbf{M}) = n - p$, which finishes the first part.

The t-test VIII

(2) As both random variables $\widehat{\beta}_j$ and $\mathbf{r}$ are linear combinations of ε it follows that they are jointly normally distributed. Therefore, to show that they are independent, it is enough to show that they are uncorrelated.

That they are uncorrelated is one of the finite-sample properties of OLS.

Finally, that z_j and q (as functions of $\widehat{\beta}_j$ and $\mathbf{r}$) are independent follows from an earlier Proposition. □

The t-test IX

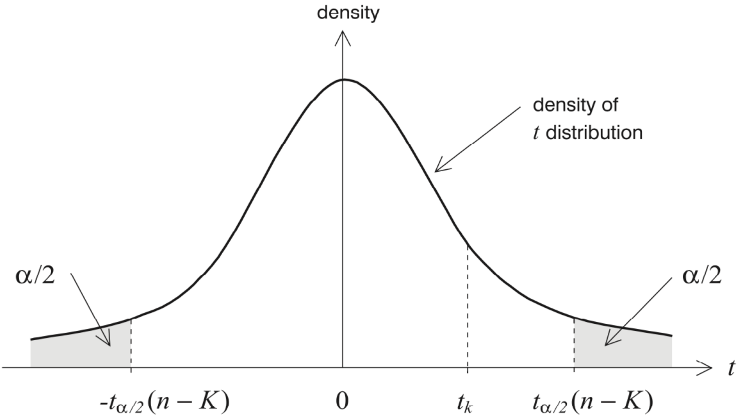

Figure 2: Illustration of how to perform a t-test at significance level α. (Source: Hayashi (2000)).

The F-test

We want to test the null hypothesis

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{d}\,, \tag{47}$$

where $\mathbf{R} \in \mathbb{R}^{d \times p}$ with rank d, and $\mathbf{d} \in \mathbb{R}^d$ are known.

We define the F-statistic

$$F := \frac{(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{d})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{d})/d}{s^2}\,. \tag{48}$$

Proposition

Under the classical linear regression assumptions (I)-(V) the F-statistic follows an F-distribution with $(d, n-p)$ degrees of freedom under the null hypothesis, that is

$$F|\mathbf{X}, H_0 \sim F_{(d,n-p)}\,.$$

References

- Breiman, Leo (2001). “Statistical Modeling: The Two Cultures”. In: *Statistical Science* 16.3, pp. 199–231.
- Hayashi, Fumio (2000). *Econometrics*. Princeton University Press.
- Rao, Radhakrishna C. and Helge Toutenburg (1999). *Linear Models: Least Squares and Alternatives*. 2nd ed. Springer.