Regression analysis: Part I

Daniele Bianchi[1]

whitesphd.com

[1]School of Economics and Finance
Queen Mary, University of London

Summary

This week provides an overview of the linear regression model. While the importance of linear regression models in financial economics diminished, it is still widely employed. More importantly, the theory behind least squares estimate is useful in broader contexts. This week covers model specification, estimation and small-sample inference.

Contents

The linear regression model

Model description

Linear regression expresses a dependent variable as a linear function of independent variables, possibly random, and an error,

$$y_i = X_i\boldsymbol{\beta} + \epsilon_i, \qquad i = 1, \dots, n$$

where y_i is known as the *regressand* or dependent variable, $X_i = (x_{1i}, \dots, x_{ki})$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ the set of *regressors* (or independent variables) and regression coefficients, respectively. ϵ_i is know as innovation or residual.

A more compact notation implies

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_k \end{bmatrix},$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

Example: Cross-sectional asset pricing

The concept of linear regression in financial economics is often explored in the contexts of cross-sectional asset pricing

- Regress returns on a set of factors thought to capture systematic risk
- Examples date back to the CAPM of Markowitz(1959), Sharpe (1964) and Lintner (1965).

The basic model postulates that excess returns are linearly related to systematic risk factors, i.e.,

$$R_i - R_i^f = f_i \boldsymbol{\beta} + \epsilon_i,$$

where $f_i = [F_{1i}, \dots, F_{ki}]$ are the returns on factors such as **market**, **value**, and **size**.[1]

[1] Further details on the construction of portfolios can be found at Fama and French (1993).

Functional forms

A linear relationship is fairly specific and, in some cases, restrictive.

It is important to distinguish specifications that can be examined in the linear regression framework from those that cannot.

Linear regressions require two key features:

- Each term on the right-hand side must have only one coefficient that enters multiplicatively
- The error term must enter additively.

Most specifications that satisfy these two requirements can be treated using the tools of linear regression.

Functional forms

However, some form of "non-linearity" is permitted. That is, any regressor or the regressand can be a non-linear transformation of the original observed data.

Double log (also known as log-log), where both the regressor and the regressand are log transformations of the original (positive) data, are frequently used:

$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + \epsilon_i,$$

Two other transformations of the original data are commonly used:

- A *Dummy variable* that takes value 0 or 1; examples in finance are, calendar effects and group-specific effects.
- Variable *interactions* parameterize non-linearities into a model through products of regressors.

These two variable transformations add significant flexibility to the linear regression model.

Functional forms

Notice that the use of non-linear transformation of the original data changes the interpretation of the coefficients.

If only unmodified regressors are included, i.e., $Y_i = \mathbf{x}_i\beta + \epsilon_i$, then the regression coefficient represents the partial effect $\beta_k = \frac{\partial Y_i}{\partial x_{ki}}$.

Suppose a specification includes both X_i and X_i^2 as regressors, i.e., $Y_i = x_i\beta_1 + x_i^2\beta_2 + \epsilon_i$, then $\frac{\partial Y_i}{\partial X_i} = \beta_1 + \beta_2 x_i$ and the level of the variable enters its partial effect.

Similarly, in a simple log-log model, i.e., $\ln Y_i = \beta_1 \ln X_i + \epsilon_i$, the regression coefficient represents

$$\beta_1 = \frac{\partial Y_i}{\partial X_i} = \frac{\partial Y/Y}{\partial X/X} = \frac{\%\Delta Y}{\%\Delta X},$$

Thus, β_1 corresponds to the *elasticity* of Y_i w.r.t. X_i.

Example: Dummy variables and interactions in asset pricing

The January and December effects are seasonal phenomena that have been widely studied in finance. Simply put, the December effect hypothesizes that returns in December are unusually low due to tax-induced portfolio rebalancing, while the January effect stipulates returns are abnormally high as investors return to the market.

To model such effects one can modify the above factor model as follows

$$R_i - R_i^f = f_i\boldsymbol{\beta} + \gamma_1 I_{1i} + \gamma_2 I_{12i} + \epsilon_i,$$

where I_{1i} if the return was generated in January and $I_{12i} = 1$ in December.

In addition, Dummy interactions can be used to produce models that have both different intercepts and slopes in January and December, i.e.,

$$R_i - R_i^f = f_i\boldsymbol{\beta} + \gamma_1 I_{1i} + \gamma_2 I_{12i} + I_{1i} f_i \boldsymbol{\delta} + I_{12i} f_i \boldsymbol{\mu} + \epsilon_i,$$

Estimation

Estimation

Linear regression is also known as Ordinary Least Squares (OLS) or simply least squares.

The least-squares estimator minimizes the squared distance between the fit line and the regressand, i.e.,

$$\min_{\beta} \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)' \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right) = \min_{\beta} \sum_{i=1}^{n} \left(Y_i - \mathbf{x}_i\boldsymbol{\beta}\right)^2 ,$$

First-order conditions of this optimization problem are

$$-2\mathbf{X}' \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right) = -2 \left(\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right) = -2 \sum_{i=1}^{n} \mathbf{x}_i \left(Y_i - \mathbf{x}_i\boldsymbol{\beta}\right) = \mathbf{0},$$

Definition (OLS Estimator)

The ordinary least-squares estimator $\hat{\beta}$ is defined by

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{y},$$

Estimation

Definition (OLS Variance Estimator)

The OLS residual variance estimator, denoted $\hat{\sigma}^2$, is defined

$$\hat{\sigma}^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{n-k},$$

where $\hat{\epsilon} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ the estimated residuals.

Definition (OLS Standard errors)

The standard error of the regression is defined as

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2},$$

Estimation

Notice that the OLS estimator is reasonable only if $\mathbf{X}'\mathbf{X}$ is invertible, which is equivalent to $\operatorname{rank}(\mathbf{X}) = k$.

This requirement states that:

- No column of $\mathbf{X}$ can be exactly expressed as a combination of the $k-1$ remaining columns.
- The number of observations is at least as large as the number of regressors $(n \geq k)$.

Dummy variables create one further issue worth of attention. Suppose dummy variables corresponding to the quarters of the year $I_{1i}, \ldots, I_{4i}$, are constructed from a quarterly data set of returns. Then consider the model

$$R_i = \beta_0 + \beta_1 I_{1i} + \beta_2 I_{2i} + \beta_3 I_{3i} + \beta_4 I_{4i} + \epsilon_i,$$

It is not possible to estimate this model because the constant is a perfect linear combination of the dummy variables.

Example: Cross-sectional asset pricing

Let us consider the standard Fama-French factors:

- The market factor (MKT): Returns on a value-weighted portfolio of all NYSE, AMEX, and NASDAQ stocks in excess of the risk-free rate.
- The "size" factor (SMB): Returns on the Small-minus-Big factor, a zero-cost portfolio that is long on small market cap firms and short big caps.
- The "value" factor (HML): Returns on the High-minus-Low factor, a zero-cost portfolio that is long high BE/ME firms and short low BE/ME firms.

and consider the following regression model

$$R_i^e = \beta_0 + \beta_1 MKT_i + \beta_2 SMB_i + \beta_3 HML_i + \epsilon_i,$$

with R_i^e the return on a given portfolio in excess of the risk-free rate.

Example: Cross-sectional asset pricing

Figure 1: Returns and risk factors

Variable	Description
VWM	Returns on a value-weighted portfolio of all NYSE, AMEX and NASDAQ stocks
SMB	Returns on the Small minus Big factor, a zero investment portfolio that is long small market capitalization firms and short big caps.
HML	Returns on the High minus Low factor, a zero investment portfolio that is long high BE/ME firms and short low BE/ME firms.
MOM	Returns on a portfolio that is long winners and short losers as defined by their performance over the past 12 months, excluding the last month. Includes the large and small cap stocks but excludes mid-cap stocks.
SL	Returns on a portfolio of small cap and low BE/ME firms.
SM	Returns on a portfolio of small cap and medium BE/ME firms.
SH	Returns on a portfolio of small cap and high BE/ME firms.
BL	Returns on a portfolio of big cap and low BE/ME firms.
BM	Returns on a portfolio of big cap and medium BE/ME firms.
BH	Returns on a portfolio of big cap and high BE/ME firms.
RF	Risk free rate (Rate on a 3 month T-bill).
DATE	Date in format YYYYMM.

Source: Financial Econometrics Notes by Kevin Sheppard.

Example: Cross-sectional asset pricing

Figure 2: Asset pricing model estimates

	Constant	VWM^e	SMB	HML	MOM	$\hat{\sigma}$
SL^e	-0.15	1.09	1.02	-0.26	-0.03	0.99
SM^e	0.08	0.96	0.82	0.35	-0.00	0.77
SH^e	0.05	1.00	0.87	0.69	-0.00	0.56
BL^e	0.12	0.99	-0.15	-0.28	-0.00	0.69
BM^e	-0.05	0.98	-0.13	0.31	-0.00	1.15
BH^e	-0.09	1.08	0.00	0.76	-0.04	1.06

Table 3.3: Estimated regression coefficients from the model $R_i^{p_i} = \beta_1 + \beta_2 VWM_i^e + \beta_3 SMB_i + \beta_4 HML_i + \beta_5 MOM_i + \varepsilon_i$, where $R_i^{p_i}$ is the excess return on one of the six size and value sorted portfolios. The final column contains the standard error of the regression.

Source: Financial Econometrics Notes by Kevin Sheppard.

Assessing fit

Once the parameters have been estimated, the next step is to determine whether the model fits the data.

The minimised sum of squared errors, the optimization's objective, is an obvious choice to assess fit.

However, there is an important drawback to using the sum of squared errors: changes in the scale of Y_i alter the sum without changing the fit. It is necessary to distinguish between portions of $\mathbf{y}$ explained by $\mathbf{X}$ from those that are not.

Definition (R^2)

The centered R^2 is defined as

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{SSE}{TSS},$$

where

- Total Sum of Squares (TSS) $\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2$
- Regression Sum of Squares (RSS) $\sum_{i=1}^{n} \left(\mathbf{x}_i\hat{\boldsymbol{\beta}} - \overline{\mathbf{x}}\hat{\boldsymbol{\beta}}\right)^2$
- Sum of Squared Errors (SSE) $\sum_{i=1}^{n} \left(Y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)^2$

and where $\overline{\mathbf{x}} = n^{-1} \sum_{i=1}^{n} \mathbf{x}_i$.

Assessing fit

The R^2 does have some caveats:

- First, adding an additional regressor will always (weakly) increase the R^2 since the sum of squared errors cannot increase by the inclusion of an additional regressor.
- This makes the R^2 useless in discriminating two nested models. One solution is to use the degrees-of-freedom adjusted R^2 defined as

$$\overline{R}^2 = 1 - \frac{SSE\; n-1}{TSS\; n-k},$$

- The $\overline{R}^2$ will increase if the reduction in the SSE is large enough to compensate for a loss of one degree of freedom, captured by the $n-k$ term.
- A second caveat is that the R^2 is not invariant to the regressand. That is, you cannot compare two models with different dependent variables based on the R^2.

Example: Cross-sectional asset pricing

Figure 3: R^2 vs adj R^2

Regressand	Regressors	R^2_U	$\bar{R}^2_U$	R^2_C	$\bar{R}^2_C$
BH^e	$1, VME^e$	0.7620	0.7616	–	–
BH^e	$1, VME^e, SMB$	0.7644	0.7637	–	–
BH^e	$1, VME^e, SMB, HML$	0.9535	0.9533	–	–
BH^e	$1, VME^e, SMB, HML, MOM$	0.9543	0.9541	–	–
BH^e	VWM^e	–	–	0.7656	0.7653
$10 + BH^e$	$1, VME^e$	0.7620	0.7616	–	–
$10 + BH^e$	VME^e	–	–	0.2275	0.2264
$10 \times BH^e$	$1, VME^e$	0.7620	0.7616	–	–
$10 \times BH^e$	VME^e	–	–	0.7656	0.7653
$BH^e - VME^e$	$1, VME^e$	0.0024	0.0009	–	–
$\sum_Y BH^e$	$1, \sum_Y VME^e$	0.6800	0.6743	–	–

Table 3.4: Centered and uncentered R^2 and $\bar{R}^2$ from models with regressor or regressand changes. Only the correct version of the R^2 is shown – centered for models that contain a constant as indicated by 1 in the regressor list, or uncentered for models that do not. The top rows demonstrate how R^2 and its adjusted version change as additional variables are added. The bottom two rows demonstrate how changes in the regressand – the left-hand-side variable – affect the R^2.

Source: Financial Econometrics Notes by Kevin Sheppard.

Assumptions and small-sample properties

Assumptions

It is necessary to make some assumptions about the innovations ϵ and the regressors $\mathbf{x}$ before providing a statistical interpretation to $\boldsymbol{\beta}$.

Two broad sets of assumptions can be used to analyze the behavior of $\hat{\boldsymbol{\beta}}$:

- The classical framework (also known as small-sample or finite-sample).
- Asymptotic analysis (also known as the large-sample framework).

Neither of these methods are ideal:

- The small-sample framework is precise but comes with some inapplicable (in financial markets) restrictions.
- The large-sample framework requires few restrictive assumptions and is broadly applicable to financial data, although results are only exact if the number of observations is infinite.

Small-sample assumptions

Assumption (Linearity)

$Y_i = \mathbf{x}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i.$

Assumption (Conditional mean)

$E\left[\boldsymbol{\epsilon}_i|\mathbf{X}\right] = 0, \qquad i = 1, 2, \ldots, n$

- This assumption states that the mean of each $\boldsymbol{\epsilon}_i$ is zero given any x_{ki}, any function of any x_{ki} or combination of these.
- This assumption is typically violated in time-series data.

Assumption (Rank)

The rank of $\mathbf{X}$ is k with probability 1.

This assumption is needed to ensure that $\hat{\boldsymbol{\beta}}$ is identified and can be estimated. In practice, it requires that no regressor is perfectly co-linear with the others and that the number of observations is $n \geq k$.

Small-sample assumptions

Assumption (Conditional homoskedasticity)

$V[\boldsymbol{\epsilon}_i|\mathbf{X}] = \sigma^2$, *i.e., residuals have the same variance.*

This assumption is required to establish the optimality of OLS.

Assumption (Conditional correlation)

$E[\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_j|\mathbf{X}] = 0, \qquad i = 1, 2, \ldots, n, j = i+1, \ldots, n$

This is convenient when coupled with the homoskedasticity assumption since the residual covariance is $\sigma^2\mathbf{I}_n$.

Assumption (Conditional normality)

$\boldsymbol{\epsilon}|\mathbf{X} \sim N(0, \Sigma)$

This is very restrictive but allows for precise statements about the finite-sample distribution of $\hat{\boldsymbol{\beta}}$ and test statistics.

Small-sample properties of OLS

Using these assumptions, many useful properties of $\hat{\boldsymbol{\beta}}$ can be derived. Recall that $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

Theorem (Bias of $\hat{\beta}$)

Under the first three assumptions $E\left[\hat{\beta}|\mathbf{X}\right] = \beta$

Although unbiasedness is desirable, it is not particularly meaningful without increasing precision as the sample size increases.

Theorem (Variance of $\hat{\beta}$)

Under the first five assumptions $V\left[\hat{\beta}|\mathbf{X}\right] = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$

Small-sample properties of OLS

The OLS estimate $\hat{\beta}$ has an even stronger property under the first five assumptions: it is the Best Linear Unbiased Estimator (BLUE).

Theorem (Gauss-Markov theorem)

Under assumptions 1-5 $\hat{\beta}$ is the mininum variance estimator among all linear unbiased estimators. That is $V\left[\hat{\beta}|\mathbf{X}\right] < V\left[\tilde{\beta}|\mathbf{X}\right]$ where $\tilde{\beta}$ is any other linear, unbiased estimator of β.

Finally, making use of the normality assumption, it is possible to determine the conditional distribution of $\hat{\beta}$.

Theorem (Distribution of $\hat{\beta}$)

Under all assumptions $\hat{\beta}|\mathbf{X} \sim N\left(\beta, \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$.

Small-sample hypothesis testing

Hypothesis specification

Hypothesis testing is the mechanism used to determine whether data and theory are congruent.

Formalized in terms of β, the null hypothesis is formulated as

$$H_0 : \mathbf{R}\left(\boldsymbol{\beta}\right) - \mathbf{r} = 0,$$

where $\mathbf{R}\left(\cdot\right)$ is a function from $\mathbb{R}^k$ to $\mathbb{R}^m$, $m \leq k$ and $\mathbf{r}$ is an $m \times 1$ vector. When testing linear equality hypotheses the null is formulated as

$$H_0 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r} = 0,$$

and $\mathbf{R}$ is an $m \times k$-matrix.

Hypothesis specification

Take as an example the unrestricted linear model

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \beta_4 X_{4,i} + \epsilon_i,$$

Linear equality constraints can be used to test parameter restrictions such as $\beta_1 = 0$, $3\beta_1 + \beta_2 = 1$, $\sum_{j=1}^{k} \beta_j = 0$ and $\beta_1 = \beta_2 = \beta_3 = 0$.

These hypotheses can be described in terms of $\mathbf{R}$ and $\mathbf{r}$ as

H_0	$\mathbf{R}$	$\mathbf{r}$
$\beta_1 = 0$	$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \end{bmatrix}$	0
$3\beta_1 + \beta_2 = 1$	$\begin{bmatrix} 0 & 3 & 1 & 0 & 0 \end{bmatrix}$	1
$\sum_{j=1}^{k} \beta_j = 0$	$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \end{bmatrix}$	0
$\beta_1 = \beta_2 = \beta_3 = 0$	$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

Hypothesis testing

When using linear equality constraints, alternative hypotheses are specified as $H_1 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r} \neq 0$.

Once both the null and the alternative hypotheses have been postulated, it is necessary to discern whether the data are consistent with the null hypothesis.

Three types of statistics will be described here:

- t-tests
- Wald
- Likelihood Ratio

t-tests

T-tests can be used to test a single hypothesis on one or more coefficients,

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

where $\mathbf{R}$ is a 1 by k vector and r is a scalar. Recall from the theorem above that $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim N\left(0, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$.

Under the null, and applying the properties of normal random variables

$$\mathbf{R}\hat{\boldsymbol{\beta}} - r \sim N\left(0, \sigma^2 \mathbf{R} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{R}'\right),$$

A simple test can be constructed

$$z = \frac{\mathbf{R}\hat{\boldsymbol{\beta}} - r}{\sqrt{\sigma^2 \mathbf{R} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{R}'}} \sim N(0,1),$$

To perform a test with size α, z should be compared to the critical values of the standard normal and rejected if $|z| > C_\alpha$ where C_α is the $1-\alpha$ quantile of the standard normal.

t-tests

However z is an infeasible statistic since it depends on an unknown quantity σ^2.

The natural solution is to replace the unknown parameter σ^2 with the estimate $\hat{\sigma}^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{n-k}$.

Theorem (t-test)

Under the above assumptions

$$t = \frac{\mathbf{R}\hat{\boldsymbol{\beta}} - r}{\sqrt{\hat{\sigma}^2 \mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'}} \sim t_{n-k},$$

As $\nu \to \infty$, the Student's t distribution converges to a standard normal.

t-stat

While any single linear restriction can be tested with a t-test, the expression t-stat has become synonymous with a specific null hypothesis.

Definition (t-stat)

The t-stat of a coefficient β_k is the t-test value of a test of the null $H_0 : \beta_k = 0$ against the alternative $H_1 : \beta_k \neq 0$, and is computed

$$\frac{\hat{\beta}_k}{\sqrt{\hat{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}_{[kk]}}},$$

where $\left(\mathbf{X}'\mathbf{X}\right)^{-1}_{[kk]}$ is the kth diagonal element of $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$.

The T-test is also unique among existing test statistics in that it can easily be applied against both one-sided alternatives and two-sided alternatives.

However, there is often a good argument to test a one-sided alternative.

- For instance, in tests of the market premium, theory indicates that it must be positive to induce investment.

However, the rejection rule differs if a one-sided or a two-sided test:

- In the case of a two-sided test, reject the null hypothesis if $|t| > F_{t_\nu}(1-\alpha/2)$ where $F_{t_\nu}(\cdot)$ is the CDF of a t_ν-distributed random variable.
- In the case of a one-sided upper-tail test, reject if $t > F_{t_\nu}(1-\alpha)$
- In the case of a one-sided lower-tail test, reject if $t < F_{t_\nu}(\alpha)$

t-stat

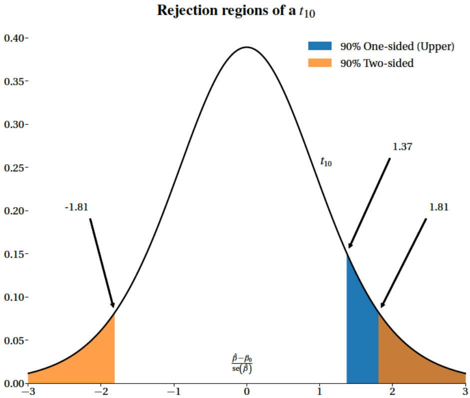

Figure 3.1: Rejection region for a t-test of the nulls $H_0 : \beta = \beta^0$ (two-sided) and $H_0 : \beta \leq \beta^0$. The two-sided rejection region is indicated by dark gray while the one-sided (upper) rejection region includes both the light and dark gray areas in the right tail.

Source: Financial Econometrics Notes by Kevin Sheppard.

Wald test

Wald test directly examines the distance between $\mathbf{R}\boldsymbol{\beta}$ and $\mathbf{r}$.

Intuitively, if the null hypothesis is true, then $\mathbf{R}\boldsymbol{\beta} - r \approx 0$.

In the small-sample framework, the distribution of $\mathbf{R}\boldsymbol{\beta} - \mathbf{r}$ follows directly from the properties of normal random variables. Specifically,

$$\mathbf{R}\boldsymbol{\beta} - \mathbf{r} \sim N\left(\mathbf{0}, \sigma^2 \mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right),$$

Thus, to test the null $H_0 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r} = 0$ against the alternative $H_1 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r} \neq 0$ the Wald statistics take the form

$$W = \frac{\left(\mathbf{R}\boldsymbol{\beta} - \mathbf{r}\right)'\left[\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)\mathbf{R}'\right]^{-1}\left(\mathbf{R}\boldsymbol{\beta} - \mathbf{r}\right)/m}{\hat{\sigma}^2}, \sim F_{m,n-k}$$

Wald test

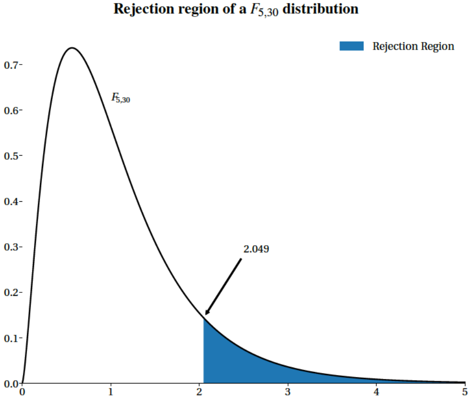

Figure 3.3: Rejection region for a $F_{5,30}$ distribution when using a test with a size of 10%. If the null hypothesis is true, the test statistic should be relatively small (would be 0 if exactly true). Large test statistics lead to rejection of the null hypothesis. In this example, a test statistic with a value greater than 2.049 would lead to a rejection of the null at the 10% level.

Source: Financial Econometrics Notes by Kevin Sheppard.

Likelihood ratio test

Likelihood ratio (LR) test are based on the relative probability of observing the data if the null is valid to the probability of observing the data under the alternative. The test statistic is defined

$$LR = -2\ln\left(\frac{\max_{\beta,\sigma^2} f\left(\mathbf{y}|\mathbf{X};\boldsymbol{\beta},\sigma^2\right) \quad \text{subject to} \quad \mathbf{R}\beta = \mathbf{r}}{\max_{\beta,\sigma^2} f\left(\mathbf{y}|\mathbf{X};\boldsymbol{\beta},\sigma^2\right)}\right),$$

Letting $\hat{\beta}_R$ denote the constrained estimate of β, this test statistic can be reformulated as

$$\begin{aligned} LR &= -2\ln\left(\frac{f\left(\mathbf{y}|\mathbf{X};\hat{\boldsymbol{\beta}}_R,\hat{\sigma}^2_R\right)}{f\left(\mathbf{y}|\mathbf{X};\hat{\boldsymbol{\beta}},\hat{\sigma}^2\right)}\right), \\ &= -2\left[l\left(\hat{\beta}_R,\hat{\sigma}^2_R;\mathbf{y}|\mathbf{X}\right) - l\left(\hat{\beta},\hat{\sigma}^2;\mathbf{y}|\mathbf{X}\right)\right], \end{aligned}$$

Likelihood ratio test

The distribution of the LR statistic can be determined by noting that

$$LR = n \ln\left(\frac{SSE_R}{SSE_U}\right) = N \ln\left(\frac{\hat{\sigma}_R^2}{\hat{\sigma}_U^2}\right),$$

and that

$$\frac{n-k}{m}\left[\exp\left(\frac{LR}{n}\right) - 1\right] = W,$$

The transformation between W and LR is monotonic so the transformed statistic has the same distribution as W, a $F_{m,n-k}$.

After having calculated W from LR, the test statistic is compared to the critical value C_α and the null is rejected if $W > C_\alpha$.

Example: Cross-sectional asset pricing

t-Tests

	$\hat{\beta}$	s.e. $\left(\hat{\beta}\right)$	t-stat	p-value
Constant	-0.086	0.042	-2.04	0.042
VWM^e	1.080	0.010	108.7	0.000
SMB	0.002	0.014	0.13	0.893
HML	0.764	0.015	50.8	0.000
MOM	-0.035	0.010	-3.50	0.000

Wald Tests

Null	Alternative	W	M	p-value
$\beta_j = 0,\ j = 2, \ldots, 5$	$\beta_j \neq 0,\ j = 2, \ldots, 5$	3558.8	4	0.000
$\beta_j = 0,\ j = 3, 4, 5$	$\beta_j \neq 0,\ j = 3, 4, 5$	956.5	3	0.000
$\beta_j = 0,\ j = 1, 5$	$\beta_j \neq 0,\ j = 1, 5$	10.1	2	0.000
$\beta_j = 0,\ j = 1, 3$	$\beta_j \neq 0,\ j = 1, 3$	2.08	2	0.126
$\beta_5 = 0$	$\beta_5 \neq 0$	12.3	1	0.000

Table 3.5: The upper panel contains t-stats and p-values for the regression of Big-High excess returns on the four factors and a constant. The lower panel contains test statistics and p-values for Wald tests of the reported null hypothesis. Both sets of tests were computed using the small-sample assumptions and may be misleading since the residuals are both non-normal and heteroskedastic.

Exercises

Exercises

Problem 3.1: Once the assumption that the innovations are conditionally normal has been made, the maximum likelihood is an obvious method to estimate the unknown parameters $\left(\beta, \sigma^2\right)$. Conditioning on $\mathbf{X}$, and assuming the innovations are normal, homoskedastic, and conditionally uncorrelated, the log-likelihood is given by

$$l\left(\beta, \sigma^2; \mathbf{y}|\mathbf{X}\right) = -\frac{n}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\left(\sigma^2\right) - \frac{\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)}{2\sigma^2},$$

derive the maximum likelihood estimates of β and σ.

Problem 3.2: Derive the OLS estimator for the model $Y_i = \alpha + \epsilon_i$

Problem 3.3: Derive the OLS estimator for the model $Y_i = \beta X_i + \epsilon_i$

Problem 3.4: Show that the variance of the OLS estimate is $\sigma^2\left(\mathbf{X}'\mathbf{X}\right)$.

Exercises

Problem 3.5: Imagine you have been given the task of evaluating the relationship between the return on a mutual fund and the number of years its manager has been a professional. Consider the regression

$$r_{it} = \alpha + \beta \text{exper}_{it} + \epsilon_{it},$$

where r_{it} is the return on fund i in year t and $exper_{it}$ is the number of years the fund manager has held her job in year t. Questions:

- What test statistic would you use to determine whether experience has a positive effect?
- What are the null and alternative hypothesis for the above test?

Problem 3.6: Show that the OLS estimate is unbiased.

Problem 3.7: Show that the distribution of the OLS estimate $\hat{\beta} \sim N\left(\beta, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$.

Problem 3.8: Show that the variance of the OLS depends on the sample size.