Kurt Schmidheiny
Unversität Basel

The Multiple Linear Regression Model

1 Introduction

The multiple linear regression model and its estimation using ordinary least squares (OLS) is doubtless the most widely used tool in econometrics. It allows to estimate the relation between a dependent variable and a set of explanatory variables. Prototypical examples in econometrics are:

- Wage of an employee as a function of her education and her work experience (the so-called Mincer equation).
- Price of a house as a function of its number of bedrooms and its age (an example of hedonic price regressions).

The dependent variable is an interval variable, i.e. its values represent a natural order and differences of two values are meaningful. In practice, this means that the variable needs to be observed with some precision and that all observed values are far from ranges which are theoretically excluded. Wages, for example, do strictly speaking not qualify as they cannot take values beyond two digits (cents) and values which are negative. In practice, monthly wages in dollars in a sample of full time workers is perfectly fine with OLS whereas wages measured in three wage categories (low, middle, high) for a sample that includes unemployed (with zero wages) ask for other estimation tools.

2 The Econometric Model

The multiple linear regression model assumes a linear (in parameters) relationship between a dependent variable y_i and a set of explanatory variables $x_i' = (x_{i0}, x_{i1}, ..., x_{iK})$. x_{ik} is also called an independent variable, a covariate or a regressor. The first regressor $x_{i0} = 1$ is a constant unless otherwise specified.

Consider a sample of N observations $i = 1, ..., N$. Every single observation i follows

$$y_i = x_i'\beta + u_i$$

where β is a $(K+1)$-dimensional column vector of parameters, x_i' is a $(K+1)$-dimensional row vector and u_i is a scalar called the error term.

The whole sample of N observations can be expressed in matrix notation,

$$y = X\beta + u$$

where y is a N-dimensional column vector, X is a $N \times (K+1)$ matrix and u is a N-dimensional column vector of error terms, i.e.

$$\underset{N\times 1}{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{bmatrix}} = \underset{N\times(K+1)}{\begin{bmatrix} 1 & x_{11} & \cdots & x_{1K} \\ 1 & x_{21} & \cdots & x_{2K} \\ 1 & x_{31} & \cdots & x_{3K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{NK} \end{bmatrix}} \underset{(K+1)\times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}} + \underset{N\times 1}{\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{bmatrix}}$$

The data generation process (dgp) is fully described by a set of assumptions. Several of the following assumptions are formulated in different alternatives. Different sets of assumptions will lead to different properties of the OLS estimator.

OLS1: Linearity

$y_i = x_i'\beta + u_i$ and $E[u_i] = 0$

OLS1 assumes that the functional relationship between dependent and explanatory variables is linear *in parameters*, that the error term enters additively and that the parameters are constant across individuals i.

OLS2: Independence

$\{x_i, y_i\}_{i=1}^N$ i.i.d. (independent and identically distributed)

OLS2 means that the observations are independently and identically distributed. This assumption is in practice guaranteed by random sampling.

OLS3: Exogeneity

a) $u_i|x_i \sim N(0, \sigma_i^2)$

b) $u_i \perp x_i$ (independent)

c) $E[u_i|x_i] = 0$ (mean independent)

d) $Cov[x_i, u_i] = 0$ (uncorrelated)

OLS3a assumes that the error term is normally distributed conditional on the explanatory variables. *OLS3b* means that the error term is independent of the explanatory variables. *OLS3c* states that the *mean* of the error term is independent of the explanatory variables. *OLS3d* means that the error term and the explanatory variables are uncorrelated. Either *OLS3a* or *OLS3b* imply *OLS3c* and *OLS3d*. *OLS3c* implies *OLS3d*.

OLS4: Error Variance

a) $V[u_i|x_i] = \sigma^2 < \infty$ (homoscedasticity)

b) $V[u_i|x_i] = \sigma_i^2 = g(x_i) < \infty$ (conditional heteroscedasticity)

OLS4a (homoscedasticity) means that the variance of the error term is a constant. *OLS4b* (conditional heteroscedasticity) allows the variance of the error term to depend on the explanatory variables.

OLS5: Identifiability

$E[x_i x_i'] = Q_{XX}$ is positive definite and finite
$rank(X) = K + 1 < N$

The *OLS5* assumes that the regressors are not perfectly collinear, i.e. no variable is a linear combination of the others. For example, there can only be one constant. Intuitively, *OLS5* means that every explanatory variable adds additional information. *OLS5* also assumes that all regressors (but the constant) have strictly positive variance both in expectations and in the sample and not too many extreme values.

3 Estimation with OLS

Ordinary least squares (OLS) minimizes the squared distances between the observed and the predicted dependent variable y:

$$S(\beta) = \sum_{i=1}^{N} (y_i - x_i'\beta)^2 = (y - X\beta)'(y - X\beta) \to \min_{\beta}$$

The resulting OLS estimator of β is:

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

Given the OLS estimator, we can predict the dependent variable by $\widehat{y}_i = x_i'\widehat{\beta}$ and the error term by $\widehat{u}_i = y_i - x_i'\widehat{\beta}$. $\widehat{u}_i$ is called the *residual*.

4 Goodness-of-fit

The goodness-of-fit of an OLS regression can be measured as

$$R^2 = 1 - \frac{SSR}{SST} = \frac{SSE}{SST}$$

where $SST = \sum_{i=1}^{N}(y_i - \overline{y})^2$ is the total sum of squares and $SSR = \sum_{i=1}^{N} \widehat{u}_i^2$ the residual sum of squares. $SSE = \sum_{i=1}^{N}(\widehat{y}_i - \overline{y})^2$ is called

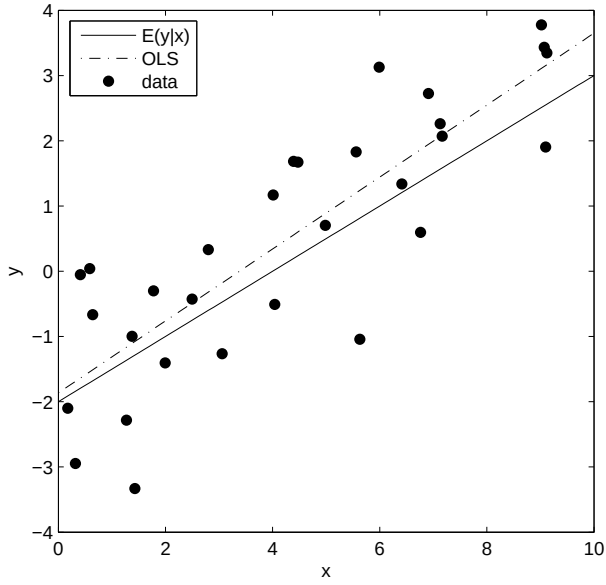


Figure 1: The linear regression model with one regressor. $\beta_0 = -2$, $\beta_1 = 0.5$, $\sigma^2 = 1$, $x \sim \text{uniform}(0, 10)$, $u \sim N(0, \sigma^2)$.

the explained sum of squares if the regression contains a constant and therefore $\overline{y} = \overline{\widehat{y}}$. In this case, R^2 lies by definition between 0 and 1 and reports the fraction of the sample variation in y that is explained by the xs.

Note: R^2 increases by construction with every (also irrelevant) additional regressors and is therefore not a good criterium for the selection of regressors. The *adjusted* R^2 is a modified version that does not necessarily increase with additional regressors:

$$adj.\ R^2 = 1 - \frac{N-1}{N-K-1}\frac{SSR}{SST}.$$

5 Small Sample Properties

Assuming *OLS1*, *OLS2*, *OLS3a*, *OLS4*, and *OLS5*, the following properties can be established for finite, i.e. even small, samples.

- The OLS estimator of β is *unbiased*:

$$E[\widehat{\beta}|X] = \beta$$

- The OLS estimator is (multivariate) normally distributed:

$$\widehat{\beta}|X \sim N\left(\beta, V[\widehat{\beta}|X]\right)$$

  with variance $V[\widehat{\beta}|X] = \sigma^2 (X'X)^{-1}$ under homoscedasticity (*OLS4a*) and $V[\widehat{\beta}|X] = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}$ under known heteroscedasticity (*OLS4b*). Under homoscedasticity (*OLS4a*) the variance V can be *unbiasedly* estimated as

$$\widehat{V}(\widehat{\beta}|X) = \widehat{\sigma^2} (X'X)^{-1}$$

  with

$$\widehat{\sigma^2} = \frac{\widehat{u}'\widehat{u}}{N-K-1}\,.$$

- Gauß-Markov-Theorem: under homoscedasticity (*OLS4a*),

$$\widehat{\beta} \text{ is BLUE (best linear unbiased estimator)}$$

6 Tests in Small Samples

Assume *OLS1*, *OLS2*, *OLS3a*, *OLS4a*, and *OLS5*.

A simple null hypotheses of the form $H_0 : \beta_k = q$ is tested with the t-test. If the null hypotheses is true, the t-statistic

$$t = \frac{\widehat{\beta}_k - q}{\widehat{se}[\widehat{\beta}_k]} \sim t_{N-K-1}$$

follows a t-distribution with $N-K-1$ degrees of freedom. The standard error $\widehat{se}[\widehat{\beta}_k]$ is the square root of the element in the $(k+1)-$th row and $(k+1)-$th column of $\widehat{V}[\widehat{\beta}|X]$. For example, to perform a two-sided test of H_0 against the alternative hypotheses $H_A : \beta_k \neq q$ on the 5% significance level, we calculate the t-statistic and compare its absolute value to the 0.975-quantile of the t-distribution. With $N = 30$ and $K = 2$, H_0 is rejected if $|t| > 2.052$.

A null hypotheses of the form $H_0 : R\beta = q$ with J linear restrictions is jointly tested with the F-test. If the null hypotheses is true, the F-statistic

$$F = \frac{\left(R\widehat{\beta} - q\right)' \left(R\widehat{V}(\widehat{\beta}|X)R'\right)^{-1} \left(R\widehat{\beta} - q\right)}{J} \sim F_{J,N-K-1}$$

follows an F distribution with J numerator degrees of freedom and $N - K - 1$ denominator degrees of freedom. For example, to perform a two-sided test of H_0 against the alternative hypotheses $H_A : R\beta \neq q$ at the 5% significance level, we calculate the F-statistic and compare it to the 0.95-quantile of the F-distribution. With $N = 30$, $K = 2$ and $J = 2$, H_0 is rejected if $F > 3.35$. We cannot perform one-sided F-tests.

Only under homoscedasticity ($OLS4a$), the F-statistic can also be computed as

$$F = \frac{(SSR_{restricted} - SSR)/J}{SSR/(N-K-1)} = \frac{(R^2 - R^2_{restricted})/J}{(1-R^2)/(N-K-1)} \sim F_{J,N-K-1}$$

where $SSR_{restricted}$ and $R^2_{restricted}$ are, respectively, estimated by restriced least squares which minimizes $S(\beta)$ s.t. $R\beta = q$. Exclusionary restrictions of the form $H_0 : \beta_k = 0, \beta_m = 0, ...$ are a special case of $H_0 : R\beta = q$. In this case, restricted least squares is simply estimated as a regression were the explanatory variables $k, m, ...$ are excluded.

7 Confidence Intervals in Small Samples

Assuming *OLS1*, *OLS2*, *OLS3a*, *OLS4a*, and *OLS5*, we can construct confidence intervals for a particular coefficient β_k. The $(1-\alpha)$ confidence interval is given by

$$\left(\widehat{\beta}_k - t_{(1-\alpha/2),(N-K-1)}\widehat{se}[\widehat{\beta}_k]\,,\ \widehat{\beta}_k + t_{(1-\alpha/2),(N-K-1)}\widehat{se}[\widehat{\beta}_k]\right)$$

where $t_{(1-\alpha/2),(N-K-1)}$ is the $(1-\alpha/2)$ quantile of the t-distribution with $N-K-1$ degrees of freedom. For example, the 95 % confidence interval with $N=30$ and $K=2$ is $\left(\widehat{\beta}_k - 2.052\widehat{se}[\widehat{\beta}_k]\,,\ \widehat{\beta}_k + 2.052\widehat{se}[\widehat{\beta}_k]\right)$.

8 Asymptotic Properties of the OLS Estimator

Assuming *OLS1*, *OLS2*, *OLS3d*, *OLS4a* or *OLS4b*, and *OLS5* the following properties can be established for large samples.

- The OLS estimator is consistent:

$$\text{plim}\ \widehat{\beta} = \beta$$

- The OLS estimator is asymptotically normally distributed under *OLS4a* as

$$\sqrt{N}(\widehat{\beta} - \beta) \xrightarrow{d} N\left(0, \sigma^2 Q_{XX}^{-1}\right)$$

  and under *OLS4b* as

$$\sqrt{N}(\widehat{\beta} - \beta) \xrightarrow{d} N\left(0, Q_{XX}^{-1} Q_{X\Omega X} Q_{XX}^{-1}\right)$$

  where $Q_{XX} = E[x_i x_i']$ and $Q_{X\Omega X} = E[u_i^2 x_i x_i']$ is assumed positive definite (see handout on "Heteroskedasticity in the Linear Model").

- The OLS estimator is approximately normally distributed

$$\widehat{\beta} \overset{A}{\sim} N\left(\beta, Avar[\widehat{\beta}]\right)$$

where the asymptotic variance $Avar[\widehat{\beta}]$ can be consistently estimated under *OLS4a* (homoscedasticity) as

$$\widehat{Avar}[\widehat{\beta}] = \widehat{\sigma^2}\,(X'X)^{-1}$$

with $\widehat{\sigma^2} = \widehat{u}'\widehat{u}/N$ and under *OLS4b* (heteroscedasticity) as the *robust* or *Eicker-Huber-White* estimator (see handout on "Heteroscedasticity in the linear Model")

$$\widehat{Avar}[\widehat{\beta}] = (X'X)^{-1}\left(\sum_{i=1}^{N}\widehat{u}_i^2 x_i x_i'\right)(X'X)^{-1}\ .$$

Note: In practice we can almost never be sure that the errors are homoscedastic and should therefore always use robust standard errors.

9 Asymptotic Tests

Assume *OLS1*, *OLS2*, *OLS3d*, *OLS4a* or *OLS4b*, and *OLS5*.

A simple null hypotheses of the form $H_0 : \beta_k = q$ is tested with the z-test. If the null hypotheses is true, the z-statistic

$$z = \frac{\widehat{\beta}_k - q}{\widehat{se}[\widehat{\beta}_k]} \overset{A}{\sim} N(0,1)$$

follows approximately the standard normal distribution. The standard error $\widehat{se}[\widehat{\beta}_k]$ is the square root of the element in the $(k+1)-$th row and $(k+1)-$th column of $\widehat{Avar}[\widehat{\beta}]$. For example, to perform a two sided test of H_0 against the alternative hypotheses $H_A : \beta_k \neq q$ on the 5% significance level, we calculate the z-statistic and compare its absolute value to the 0.975-quantile of the standard normal distribution. H_0 is rejected if $|z| > 1.96$.

A null hypotheses of the form $H_0 : R\beta = q$ with J linear restrictions is jointly tested with the Wald test. If the null hypotheses is true, the Wald

statistic

$$W = \left(R\widehat{\beta} - q\right)' \left(R\widehat{Avar}[\widehat{\beta}]R'\right)^{-1} \left(R\widehat{\beta} - q\right) \overset{A}{\sim} \chi^2_J$$

follows approximately an χ^2 distribution with J degrees of freedom. For example, to perform a test of H_0 against the alternative hypotheses $H_A : R\beta \neq q$ on the 5% significance level, we calculate the Wald statistic and compare it to the 0.95-quantile of the χ^2-distribution. With $J = 2$, H_0 is rejected if $W > 5.99$. We cannot perform one-sided Wald tests.

Under *OLS4a* (homoscedasticity) only, the Wald statistic can also be computed as

$$W = \frac{(SSR_{restricted} - SSR)}{SSR/N} = \frac{(R^2 - R^2_{restricted})}{(1 - R^2)/N} \overset{A}{\sim} \chi^2_J$$

where $SSR_{restricted}$ and $R^2_{restricted}$ are, respectively, estimated by restricted least squares which minimizes $S(\beta)$ s.t. $R\beta = q$. Exclusionary restrictions of the form $H_0 : \beta_k = 0, \beta_m = 0, ...$ are a special case of $H_0 : R\beta = q$. In this case, restricted least squares is simply estimated as a regression were the explanatory variables $k, m, ...$ are excluded.

Note: the Wald statistic can also be calculated as

$$W = J \cdot F \overset{A}{\sim} \chi^2_J$$

where F is the small sample F-statistic. This formulation differs by a factor $(N - K - 1)/N$ but has the same asymptotic distribution.

10 Confidence Intervals in Large Samples

Assuming *OLS1*, *OLS2*, *OLS3d*, *OLS5*, and *OLS4a* or *OLS4b*, we can construct confidence intervals for a particular coefficient β_k. The $(1 - \alpha)$ confidence interval is given by

$$\left(\widehat{\beta}_k - z_{(1-\alpha/2)}\widehat{se}[\widehat{\beta}_k]\,,\ \widehat{\beta}_k + z_{(1-\alpha/2)}\widehat{se}[\widehat{\beta}_k]\right)$$

where $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. For example, the 95 % confidence interval is $\left(\widehat{\beta}_k - 1.96\widehat{se}[\widehat{\beta}_k]\,,\ \widehat{\beta}_k + 1.96\widehat{se}[\widehat{\beta}_k]\right)$.

11 Small Sample vs. Asymptotic Properties

The t-test, F-test and confidence interval for small samples depend on the normality assumption *OLS3a* (see Table 1). This assumption is strong and unlikely to be satisfied. The asymptotic z-test, Wald test and the confidence interval for large samples rely on much weaker assumptions. Although most statistical software packages report the small sample results by default, we would typically prefer the large sample approximations. In practice, small sample and asymptotic tests and confidence intervals are very similar already for relatively small samples, i.e. for $(N-K) > 30$. Large sample tests also have the advantage that they can be based on heteroscedasticity robust standard errors.

12 More Known Issues

Non-linear functional form: The true relationship between the dependent variable and the explanatory variables is often not linear and thus in violation of assumption *OLS1*. The multiple linear regression model allows for many forms of non-linear relationships by transforming both dependent and explanatory variables. See the handout on "Functional Form in the Linear Model" for details.

Aggregate regressors: Some explanatory variables may be constant within groups (clusters) of individual observations. For example, wages of individual workers are regressed on state-level unemployment rates. This is a violation of the independence across individual observations (*OLS2*). In this case, the usual standard errors will be too small and t-statistics too large by a factor of up to $\sqrt{M}$, where M is the average number of individual observations per group (cluster). For example, the average number of

workers per state. Cluster-robust standard errors will provide asymptotically consistent standard errors for the usual OLS point estimates. See the handout on "Clustering in the Linear Model" for more details and generalizations.

Omitted variables: Omitting explanatory variables in the regression generally violates the exogeneity assumption (*OLS3*) and leads to biased and inconsistent estimates of the coefficients for the included variables. This omitted-variable bias does not occur if the omitted variables are uncorrelated with all included explanatory variables.

Irrelevant regressors: Including irrelevant explanatory variables, i.e. variables which do not have an effect on the dependent variable, does not lead to biased or inconsistent estimates of the coefficients for the other included variables. However, including too many irrelevant regressors may lead to very imprecise estimates, i.e. very large standard errors, in small datasets.

Reverse causality: A reverse causal effect of the dependent variable on one or several explanatory variables is a violation of the exogeneity assumption (*OLS3*) and leads to biased and inconsistent estimates. See the handout on "Instrumental Variables" for a potential solution.

Measurement error: Imprecise measurement of the explanatory variables is a violation of *OLS3* and leads to biased and inconsistent estimates. See the handout on "Instrumental Variables" for a potential solution.

Multicollinearity: Perfectly correlated explanatory variables violate the identifiability assumption (*OLS5*) and their effects cannot be estimated separately. The effects of highly but not perfectly correlated variables can in principle be separately estimated. However, the estimated coefficients will be very imprecise, i.e. the standard errors will be very large. If variables are (almost) perfectly correlated in all conceivable states of the world, there is no theoretical meaning of separate effects. If multicollinearity is only a feature of a specific sample, collecting more data may provide the necessary variation to estimate separate effects.

Implementation in Stata 14

The multiple linear regression model is estimated by OLS with the `regress` command. For example,

```
webuse auto.dta
regress mpg weight displacement
```

regresses the mileage of a car (`mpg`) on `weight` and `displacement` (see annotated output next page). A constant is automatically added if not suppressed by the option `noconst`

```
regress mpg weight displacement, noconst
```

Estimation based on a subsample is performed as

```
regress mpg weight displacement if weight>3000
```

where only cars heavier than 3000 lb are considered. Transformations of variables are included with new variables

```
generate logmpg = log(mpg)
generate weight2 = weight^2
regress logmpg weight weight2 displacement
```

The Eicker-Huber-White covariance is reported with the option `robust`

```
regress mpg weight displacement, vce(robust)
```

F-tests for one or more restrictions are calculated with the post-estimation command `test`. For example

```
test weight
```

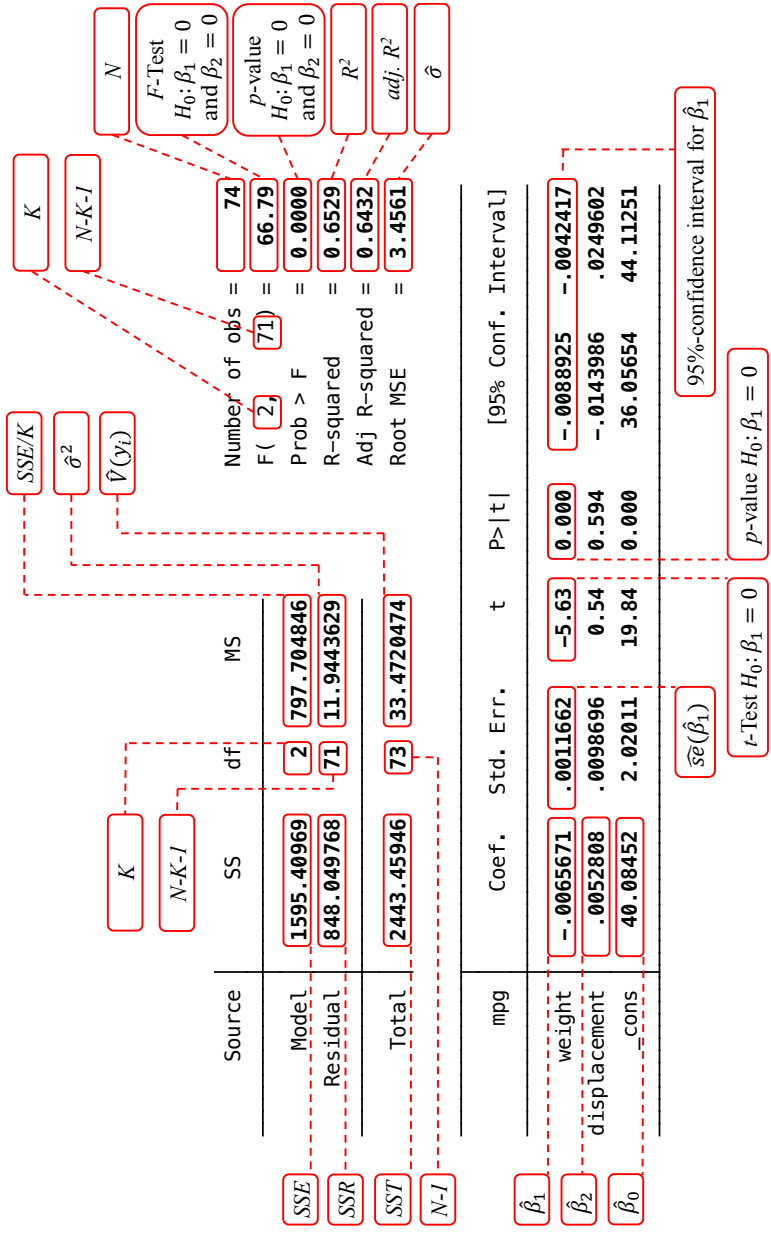
tests $H_0 : \beta_1 = 0$ against $H_A : \beta_1 \neq 0$, and

```
test weight displacement
```

tests $H_0 : \beta_1 = 0$ and $\beta_2 = 0$ against $H_A : \beta_1 \neq 0$ or $\beta_2 \neq 0$

New variables with residuals and fitted values are generated by

```
predict uhat if e(sample), resid
predict pricehat if e(sample)
```


Source
SS
df
MS
Model
1595.40969
2
797.704846
Residual
848.049768
71
11.9443629
Total
2443.45946
73
33.4720474
mpg
Coef.
Std. Err.
t
P>|t|
[95% Conf. Interval]
weight
-.0065671
.0011662
-5.63
0.000
-.0088925
-.0042417
displacement
.0052808
.0098696
0.54
0.594
-.0143986
.0249602
_cons
40.08452
2.02011
19.84
0.000
36.05654
44.11251
Number of obs = 74
F(2, 71) = 66.79
Prob > F = 0.0000
R-squared = 0.6529
Adj R-squared = 0.6432
Root MSE = 3.4561
K
N-K-1
SSE/K
$\hat{\sigma}^2$
$\hat{V}(y_i)$
N
F-Test $H_0: \beta_1 = 0$ and $\beta_2 = 0$
p-value $H_0: \beta_1 = 0$ and $\beta_2 = 0$
R^2
adj. R^2
$\hat{\sigma}$
SSE
SSR
SST
N-1
$\hat{\beta}_1$
$\hat{\beta}_2$
$\hat{\beta}_0$
$\widehat{se}(\hat{\beta}_1)$
t-Test $H_0: \beta_1 = 0$
p-value $H_0: \beta_1 = 0$
95%-confidence interval for $\hat{\beta}_1$

Implementation in R

The multiple linear regression model is estimated by OLS with the `lm` function. For example,

```
> library(foreign)
> auto <- read.dta("http://www.stata-press.com/data/r11/auto.dta")
> fm <- lm(mpg~weight+displacement, data=auto)
> summary(fm)
```

regresses the mileage of a car (`mpg`) on `weight` and `displacement`. A constant is automatically added if not suppressed by `-1`

```
> lm(mpg~weight+displacement-1, data=auto)
```

Estimation based on a subsample is performed as

```
> lm(mpg~weight+displacement, subset=(weight>3000), data=auto)
```

where only cars heavier than 3000 lb are considered. Tranformations of variables are directly included with the `I()` function

```
> lm(I(log(mpg))~weight+I(weight^2)+ displacement, data=auto)
```

The Eicker-Huber-White covariance is reported after estimation with

```
> library(sandwich)
> library(lmtest)
> coeftest(fm, vcov=sandwich)
```

F-tests for one or more restrictions are calculated with the command `waldtest` which also uses the two packages `sandwich` and `lmtest`

```
> waldtest(fm, "weight", vcov=sandwich)
```

tests $H_0 : \beta_1 = 0$ against $H_A : \beta_1 \neq 0$ with Eicker-Huber-White, and

```
> waldtest(fm, .~.-weight-displacement, vcov=sandwich)
```

tests $H_0 : \beta_1 = 0$ and $\beta_2 = 0$ against $H_A : \beta_1 \neq 0$ or $\beta_2 \neq 0$.

New variables with residuals and fitted values are generated by

```
> auto$uhat <- resid(fm)
> auto$mpghat <- fitted(fm)
```

References

Introductory textbooks

Stock, James H. and Mark W. Watson (2020), Introduction to Econometrics, 4th Global ed., Pearson. Chapters 4 - 9.

Wooldridge, Jeffrey M. (2009), Introductory Econometrics: A Modern Approach, 4th ed., Cengage Learning. Chapters 2 - 8.

Advanced textbooks

Cameron, A. Colin and Pravin K. Trivedi (2005), Microeconometrics: Methods and Applications, Cambridge University Press. Sections 4.1-4.4.

Wooldridge, Jeffrey M. (2002), Econometric Analysis of Cross Section and Panel Data, MIT Press. Chapters 4.1 - 4.23.

Companion textbooks

Angrist, Joshua D. and Jörn-Steffen Pischke (2009), Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press. Chapter 3.

Kennedy, Peter (2008), A Guide to Econometrics, 6th ed., Blackwell Publishing. Chapters 3 - 11, 14.