



Computer Lab in Economics Master in International Economics Cross section methods with Stata

Inmaculada Álvarez Ayuso

Office 314 (Módulo I) www.uam.es/inmaculada.alvarez

E-mail: inmaculada.alvarez@uam.es

Ordinary Least Squared (OLS)

OLS is performed with the command **regress**, or **reg**.

Title

[R] **regress** — Linear regression

Syntax

regress *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, *options*]

Where **depvar** is the name of the dependent variable and **indepvars** are regressors.

LINEAR REGRESSION MODEL

$$\ln Y_{it} = \alpha + \text{time}_{it} + \beta_1 \ln L_{it} + \beta_2 \ln K_{it} + u_{it}$$

```
. regress lngdp time lnlabour lncapital
```

Source	SS	df	MS
Model	1043.7426	3	347.914198
Residual	5.15638169	547	.009426658
Total	1048.89898	550	1.90708905

Number of obs = 551
 F(3, 547) = 36907.48
 Prob > F = 0.0000
 R-squared = 0.9951
 Adj R-squared = 0.9951
 Root MSE = .09709

lngdp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
time	.0055023	.0007957	6.91	0.000	.0039392	.0070654
lnlabour	.2859774	.0234426	12.20	0.000	.2399287	.332026
lncapital	.6825611	.0221427	30.83	0.000	.6390661	.7260562
_cons	-6.268584	.06659	-94.14	0.000	-6.399387	-6.13778

LINEAR REGRESSION MODEL

Robust to heteroskedasticity

```
. regress lngdp time lnlabour lncapital, robust
```

Linear regression

Number of obs = 551
 F(3, 547) = 45623.20
 Prob > F = 0.0000
 R-squared = 0.9951
 Root MSE = .09709

lngdp	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
time	.0055023	.0007787	7.07	0.000	.0039726	.007032
lnlabour	.2859774	.0211773	13.50	0.000	.2443785	.3275762
lncapital	.6825611	.0204879	33.32	0.000	.6423166	.7228056
_cons	-6.268584	.0650522	-96.36	0.000	-6.396366	-6.140801

LINEAR REGRESSION MODEL

Post estimation

help regress postestimation

dialogs: predict dfbeta estat
 plot dialogs: acprplot avplots cprplot
 lvr2plot rvfplot rvpplot
 also see: regress
 regress postestimation ts

Title

[R] regress postestimation — Postestimation tools for regress

Description

The following postestimation commands are of special interest after **regress**:

command	description
dfbeta	DFBETA influence statistics
estat hettest	tests for heteroskedasticity
estat imtest	information matrix test
estat ovtest	Ramsey regression specification-error test for omitted variables
estat szroeter	Szroeter's rank test for heteroskedasticity
estat vif	variance inflation factors for the independent variables
acprplot	augmented component-plus-residual plot
avplot	added-variable plot
avplots	all added-variable plots in one image
cprplot	component-plus-residual plot
lvr2plot	leverage-versus-squared-residual plot
rvfplot	residual-versus-fitted plot
rvpplot	residual-versus-predictor plot

LINEAR REGRESSION MODEL

Post estimation

After OLS estimation using the command `regress`, we can use the command `predict` to obtain the linear prediction and the residuals

For example, to forecast the linear prediction:

`predict` «variable name», xb

To obtain the residuals:

`predict` «variable name», r

LINEAR REGRESSION MODEL

Post estimation (command **predict**)

Syntax for predict

```
predict [type] newvar [if] [in] [, statistic]
```

<i>statistic</i>	description
Main	
xb	linear prediction; the default
residuals	residuals
score	score; equivalent to residuals
rstandard	standardized residuals
rstudent	studentized (jackknifed) residuals
cooksd	Cook's distance
leverage hat	leverage (diagonal elements of hat matrix)
pr(<i>a</i>,<i>b</i>)	$\Pr(y \mid a < y < b)$
e(<i>a</i>,<i>b</i>)	$E(y \mid a < y < b)$
ystar(<i>a</i>,<i>b</i>)	$E(y^*)$, $y^* = \max(a, \min(y, b))$
* dfbeta(<i>varname</i>)	DFBETA for <i>varname</i>
stdp	standard error of the linear prediction
stdf	standard error of the forecast
stdr	standard error of the residual
* covratio	COVRATIO
* dfits	DFITS
* welch	welch distance

LINEAR REGRESSION MODEL

Post estimation (command **display**)

To estimate a Linear Regression Model using the command **regress** some values have been saved by default as scalars and matrices.

For example, to obtain the coefficient associated to a variable or the constant, we can use the command

display: **display** _b[variable name]

display _b[_cons]

To obtain the standard error:

display _se[variable name]

To calculate the student t-statistic:

display _b[variable name]/_se[variable name]

LINEAR REGRESSION MODEL

Post estimation (command **ereturn list**)

regress is an e-class command. Therefore, **ereturn list** will summarize that was saved, scalars, macros and functions

scalars:

e(N) = 551	e(df_m) = 3
e(df_r) = 547	e(F) = 45623.19900558261
e(r2) = .9950840054202333	e(rmse) = .0970909756979133
e(mss) = 1043.742595465018	e(rss) = 5.156381686399119
e(r2_a) = .9950570438411852	e(ll) = 505.1630353332038
e(ll_0) = -959.19142247014	e(rank) = 4

macros:

e(cmdline) : "regress lngdp time lnlabour lncapital, robust"	e(title) : "Linear regression"
e(marginsok) : "XB default"	e(vce) : "robust"
e(depvar) : "lngdp"	e(cmd) : "regress"
e(properties) : "b V"	e(predict) : "regres_p"
e(model) : "ols"	e(estat_cmd) : "regress_estat"
e(vcetype) : "Robust"	

matrices:

e(b) : 1 x 4	e(V) : 4 x 4
e(V_modelbased) : 4 x 4	

functions:

e(sample)