2020 STATA ECONOMETRICS WINTER SCHOOL

Anabela Carneiro[1] João Cerejeira[2] Miguel Portela[2,3] Paulo Guimarães[1,4]

[1]FEP and CEF.UP – U.Porto
[2]NIPE – UMinho
[2]IZA, Bonn
[4]Banco de Portugal
Faculdade de Economia da Universidade do Porto

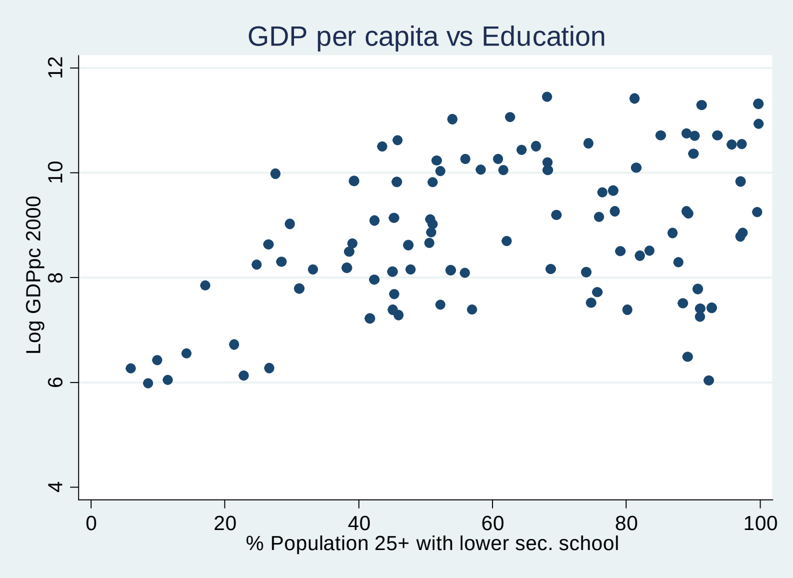
January 20-24, 2020

Outline

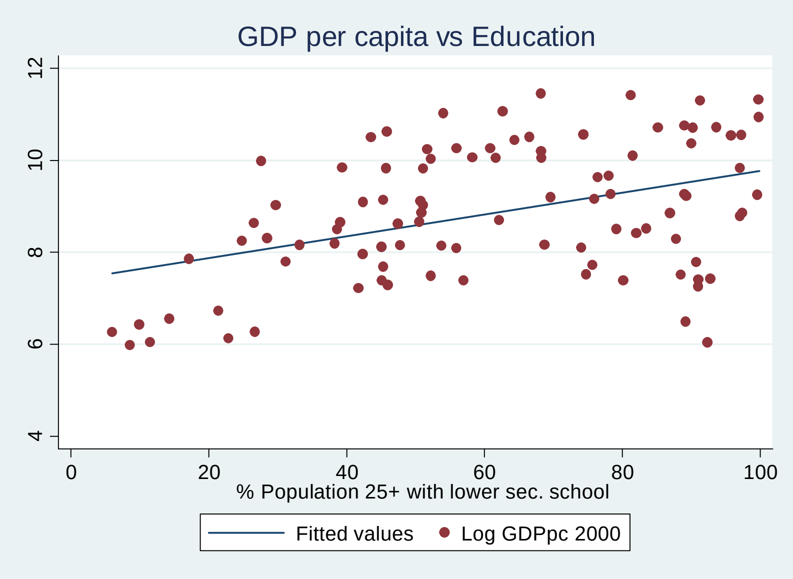
- Computing linear regression estimates
- Regression as a method-of-moments-estimator
- Regression residuals
- Sampling distribution of regression estimates
- Presenting regression estimates
- The ANOVA table
- Hypothesis tests, linear restrictions
- Computing residuals and predicted values
- Specification issues
- Interaction terms and marginal effects
- The generalized linear regression model
- Heteroskedasticity: causes and test
- Types of heteroskedasticity
- Robust estimation
- The GLS and FGLS estimator

The conditional mean of a response variable y as a linear function of k independent variables is:

$$E[y|x_1, x_2, ..., x_k] = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k. \tag{1}$$


GDP per capita vs Education
Log GDPpc 2000
12
10
8
6
4
0
20
40
60
80
100
% Population 25+ with lower sec. school

$$E\left[\text{Gdp per capita}\middle|\text{education}\right] = \beta_1 + \beta_2\left[\text{education}\right]$$


GDP per capita vs Education
4
6
8
10
12
0
20
40
60
80
100
% Population 25+ with lower sec. school
Fitted values
Log GDPpc 2000

But we don't know the population values $\beta_{1,}\beta_2, ..., \beta_k$. We work with a sample of N observations of data from population.
Using this information, we must:
- obtain estimates of the coefficients $\beta_{1,}\beta_2, ..., \beta_k$;
- estimate their variance;
- test coefficients estimate;
- use estimated $\widehat{\beta}_{1,}\widehat{\beta}_2, ..., \widehat{\beta}_k$ to interpret the model.

The linear regression model has the form:

$$y_i = \beta_1 x_1 + \beta_2 x_{2i} + ... + \beta_k x_{ki} + u_i \tag{2}$$

with $i = 1, 2, .., N$.

In matrix form:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u} \tag{3}$$

where $\mathbf{X}$ is an $N \times k$ matrix of sample values.

Regression as a method-of-moments-estimator

The key assumption in the linear regression model is:

$$E[u|\mathbf{x}] = 0. \tag{4}$$

The unobserved factors involved in the regression function are not related systematically to observed factors. For linear relationships, the later assumption implies:

$$E[\mathbf{x}'u] = \mathbf{0}$$

$$E[\mathbf{x}'(\mathbf{y} - \mathbf{x}\beta)] = \mathbf{0}. \tag{5}$$

Substituting calculated moments from our sample into the expression, yields:

$$\begin{aligned} \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\widehat{\beta} &= \mathbf{0} && (6) \\ \widehat{\beta} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}. && (7) \end{aligned}$$

Regression residuals

$$\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widehat{\beta}. \tag{8}$$

The estimator of population variance of the stochastic disturbance is:

$$s^2 = \frac{\sum_{i=1}^{N} \widehat{u}_i^2}{N-k} = \frac{\widehat{\mathbf{u}}'\widehat{\mathbf{u}}}{N-k}. \tag{9}$$

$\sqrt{s^2}$ – standard error of regression or root mean squared error.

Sampling distribution of regression estimates

The OLS estimator $\widehat{\beta}$ is a vector of random variables because it is a function of the random variable y, which in turn is a function of the stochastic disturbance u.

The OLS estimator takes on different values for each sample of N observations drawn from the population.

Assume that u_i are independent draws from a identical distribution (i.i.d).

Large sample theory shows that the sampling distribution of the OLS estimator is approximately normal.

OLS estimator $\widehat{\beta}$ has a large sample normal distribution with mean β and variance $\sigma^2\mathbf{Q}^{-1}$, where $\mathbf{Q}^{-1}$ is the variance-covariance matrix of $\mathbf{X}$ in the population.

Because $\sigma^2\mathbf{Q}^{-1}$ is unknown, a consistent estimator of $\sigma^2\mathbf{Q}^{-1}$ is $s^2(\mathbf{X}'\mathbf{X})^{-1}$.

Example 1

$$
\begin{aligned}
\log(Grow\_GDPper\ capita_c) \quad = \quad & \beta_1 + \\
& +\beta_2[\log GDPpc2000_c] + \\
& +\beta_2[\%\ Educ\_Sec_c] + \\
& +\beta_3[Invest.Grow_c] + \\
& +\beta_4[Trade2000_c] + \\
& +\beta_5[Gov2000_c] + u_c
\end{aligned}
$$

Example 1 – Linear Regression

Source	SS	df	MS			
				Number of obs	=	71
				F(5, 65)	=	18.78
Model	140.653494	5	28.1306987	Prob > F	=	0.0000
Residual	97.3780009	65	1.49812309	R-squared	=	0.5909
				Adj R-squared	=	0.5594
Total	238.031495	70	3.40044992	Root MSE	=	1.224

growthGDPpc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logGDPpc2000	-.7264758	.2168511	-3.35	0.001	-1.159557	-.2933942
educ_sec	.0325741	.006704	4.86	0.000	.0191853	.0459629
invest_growth	.2329561	.0435632	5.35	0.000	.1459543	.3199578
trade2000	.0051979	.0025667	2.03	0.047	.0000719	.0103238
gov2000	-.09683	.3132654	-0.31	0.758	-.7224643	.5288043
_cons	5.557012	1.843849	3.01	0.004	1.874592	9.239432

Presenting regression estimates
The ANOVA table

Model SS=$\widehat{\mathbf{y}}'\widehat{\mathbf{y}} = \sum_{i=1}^{N} \left(\widehat{y}_i - \overline{y}\right)^2$, is the sum of the squares of the deviations of the predicted values of y from the mean value of y.
Residual SS=$\widehat{\mathbf{u}}'\widehat{\mathbf{u}} = \sum_{i=1}^{N} \widehat{u}_i^2 = \sum_{i=1}^{N} \left(y_i - \widehat{y}_i\right)^2$.
Total SS= $\widetilde{\mathbf{y}}'\widetilde{\mathbf{y}} = \sum_{i=1}^{N} \left(y_i - \overline{y}\right)^2$, where $\widetilde{\mathbf{y}} = y - \overline{y}$.

R-squared=Model SS / Total SS=1-(Residual SS / Total SS)=R^2
Adj R-squared=$1 - \left(1 - R^2\right)\frac{N-1}{N-k}$

The other measures to compare competing regression models are the Akaike information criterion (AIC) and Bayesian information criterion (BIC, or Schwarz criterion).
These measures account for both the goodness of fit and its parsimony by rewarding improvements in the goodness of fit and penalizing the additional degrees of freedom.
The preferred model is the one with the minimum AIC or BIC value.The AIC penalizes the number of parameters less strongly than does the Bayesian information criterion.

The `estat ic` command will calculate the AIC and BIC after estimation.

The ANOVA table

Source	SS	df	MS			
Model	140.653494	5	28.1306987	Number of obs	=	71
Residual	97.3780009	65	1.49812309	F(5, 65)	=	18.78
				Prob > F	=	0.0000
				R-squared	=	0.5909
				Adj R-squared	=	0.5594
Total	238.031495	70	3.40044992	Root MSE	=	1.224

R-squared = 0.5909 = 140.653494 / 238.031495

F statistic=Model MS / Residual MS

The ANOVA table

Source	SS	df	MS			
				Number of obs	=	71
				F(5, 65)	=	18.78
Model	140.653494	5	28.1306987	Prob > F	=	0.0000
Residual	97.3780009	65	1.49812309	R-squared	=	0.5909
				Adj R-squared	=	0.5594
Total	238.031495	70	3.40044992	Root MSE	=	1.224

```
R-squared    = 0.5909 = 140.653494 / 238.031495
F(5, 65)     = 18.78  = 28.1306987 / 1.49812309
Root MSE     = 1.224   = sqrt(1.49812309)
```

$$\text{Root MSE}=\sqrt{\text{Residual MS}}$$

The ANOVA table

```
      Source |       SS           df       MS      Number of obs   =        71
-------------+----------------------------------   F(5, 65)        =     18.78
       Model |  140.653494         5  28.1306987   Prob > F        =    0.0000
    Residual |  97.3780009        65  1.49812309   R-squared       =    0.5909
-------------+----------------------------------   Adj R-squared   =    0.5594
       Total |  238.031495        70  3.40044992   Root MSE        =     1.224
```

```
R-squared   = 0.5909 = 140.653494 / 238.031495
F(5, 65)    = 18.78  = 28.1306987 / 1.49812309
Root MSE    = 1.224   = sqrt(1.49812309)
```

Saved results can be recovered with `ereturn` command.

The `e(sample)` function returns 1 if an observation was included in the estimation sample and 0 otherwise.

With `e(sample)` we can retain the estimation sample for later use:

- example: `generate reg1sample=e(sample)`

The `estat` and `matrix list` commands displays several items after any estimation command:

`estat summarize`

`matrix list e(b)`

`estat vce` - estimated variance covariance (VCE) matrix $s^2(\mathbf{X}'\mathbf{X})^{-1}$.

`estat ic` - Akaike's information criterion and Bayesian information criterion.

Beta estimates

Source	SS	df	MS			
				Number of obs	=	71
				F(5, 65)	=	18.78
Model	140.653494	5	28.1306987	Prob > F	=	0.0000
Residual	97.3780009	65	1.49812309	R-squared	=	0.5909
				Adj R-squared	=	0.5594
Total	238.031495	70	3.40044992	Root MSE	=	1.224

growthGDPpc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logGDPpc2000	-.7264758	.2168511	-3.35	0.001	-1.159557	-.2933942
educ_sec	.0325741	.006704	4.86	0.000	.0191853	.0459629
invest_growth	.2329561	.0435632	5.35	0.000	.1459543	.3199578
trade2000	.0051979	.0025667	2.03	0.047	.0000719	.0103238
gov2000	-.09683	.3132654	-0.31	0.758	-.7224643	.5288043
_cons	5.557012	1.843849	3.01	0.004	1.874592	9.239432

VCE estimates

e(V)	logGD~2000	educ_sec	invest_g~h	trade2000	gov2000	_cons
logGDPpc2000	.0470244					
educ_sec	-.00035488	.00004494				
invest_gro~h	.0015976	.00002345	.00189775			
trade2000	-5.557e-06	7.068e-07	-7.114e-06	6.588e-06		
gov2000	-.05055677	-.00010745	.0015163	-.00018711	.09813524	
_cons	-.38381007	.00028083	-.02381767	-.00046136	.42696198	3.3997779

Detecting Collinearity in Regression

When Stata determines that $(\mathbf{X}'\mathbf{X})$ is numerically singular, it drops variables until the resulting regressor matrix is invertible, marking their coefficients with (`dropped`) in place of a value.

With near-collinearity, small changes in the data matrix may cause large changes in the parameter estimates since they are nearly unidentified. The coefficients may also have large standard errors. Remember that the k^{th} diagonal element of the VCE can be written as $\frac{S^2}{(1-R_k^2)S_{kk}}$, where R_k^2 is the the R^2 from a regression of variable k on all other variables.

The VIF_k (variance inflation factor), $(1 - R_k^2)^{-1}$, measures the degree to which the variance has been inflated because regressor k is not orthogonal to the other regressors. After fitting a model, type:

```
estat vif
```

The mean VIF should be smaller than unity and the largest VIF smaller than 10.

Hypothesis tests, linear restrictions

Three tests are commonly used in econometrics: Wald tests, Lagrange multiplier (LM) tests and likelihood-ratio (LR) tests.
Here I present the Wald tests.
Given the population regression equation:

$$y = x\beta + u$$

any set of linear restrictions on the coefficient vector may be expressed as

$$\mathbf{R}\beta = \mathbf{r}.$$

Example: Wald test

Given:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

we want to test $H_0 : \beta_2 = 0$. The restriction is:

$$\mathbf{R} = \{0 \quad 1 \quad 0\}$$

$$\mathbf{r} = (0)$$

Given the hypothesis $H_0 = \mathbf{R}\beta = 0$, the Wald statistic is:

$$W = (\mathbf{R}\widehat{\beta} - \mathbf{r})'\{\mathbf{R}(\widehat{\mathbf{VCE}})\mathbf{R}'\}^{-1}(\mathbf{R}\widehat{\beta} - \mathbf{r}) \tag{10}$$

w has a large-sample χ^2 distribution when H_0 is true. In small samples w/q is better approximated by an F distribution with q (the number of restrictions) and $(N - k)$ degress of freedom. If $q = 1$, $\sqrt{w}$ can be approximated by a Student t distribution with $(N - k)$ d.f.

Since we know the distribution of w when H_0 is true, the standard hypothesis test is:

$$\Pr(\text{Reject } H_0 \mid H_0 \text{ is true}) = \alpha \tag{11}$$

where α is the significance level of the test.
Stata presents $p-$values, which measure the evidence against H_0 - the largest significance level at which a test can be conducted without rejecting H_0.

Example:

Coefficient of `education` `.0325741`, with s.d. `.006704`.
t statistic of the the null

$$H_0 : \beta_{[Educ\_Sec]} = 0$$

is

$$= \widehat{\beta}_{[Educ\_Sec]} / s.d. \left(\widehat{\beta}_{[Educ\_Sec]} \right) = .0325741 / .006704 = 4.86$$

Confidence intervals:

$$\widehat{\beta}_{[Educ\_Sec]} - s.d.(\widehat{\beta}_{[Educ\_Sec]}) \times t_{crit,\ 5\%} \leq \beta_{[Educ\_Sec.]} \leq \widehat{\beta}_{[Educ\_Sec]} + s.d.(\widehat{\beta}_{[Ed}$$

$$.03257 - .00670 \times 1.99714 \leq \beta_{[Educ\_Sec]} \leq ..03257 + .00670 \times 1.99714$$

$$.01919 \leq \beta_{[Educ\_Sec]} \leq .04597$$

You can get the critical value $t_{crit,5\%}$ using `di invttail(65,.025)` function.
To get the p-values, use the function `ttail(N-k, t)`, e.g.:
$\texttt{di } 2 * \texttt{ttail}(65, (.0325741/.006704)) = 0.000$ - gives the p-value for $H_0 : \beta_{[Educ\_Sec]} = 0$.

Wald tests with command test

Test the significance of a coefficient:

`test education`

note 1: $(t_{N-k})^2 = F^1_{N-k}$

note 2: `ttest` performs t tests on the equality of means while `test` test linear hypotheses after estimation.

Test the joint significance of a set of coefficients

`testparm *2000` provides a useful alternative to test that permits *varlist* rather than a list of coefficients allowing the use of standard Stata notation, including '-' and '*'

Tests involving linear combinations of parameters:

`test educ_sec=trade2000` is equivalent to test

$H_0 : \beta_{[Educ\_Sec]} = \beta_{[Trade2000]} \Leftrightarrow \beta_{[Educ\_Sec]} - \beta_{[Trade2000]} = 0$

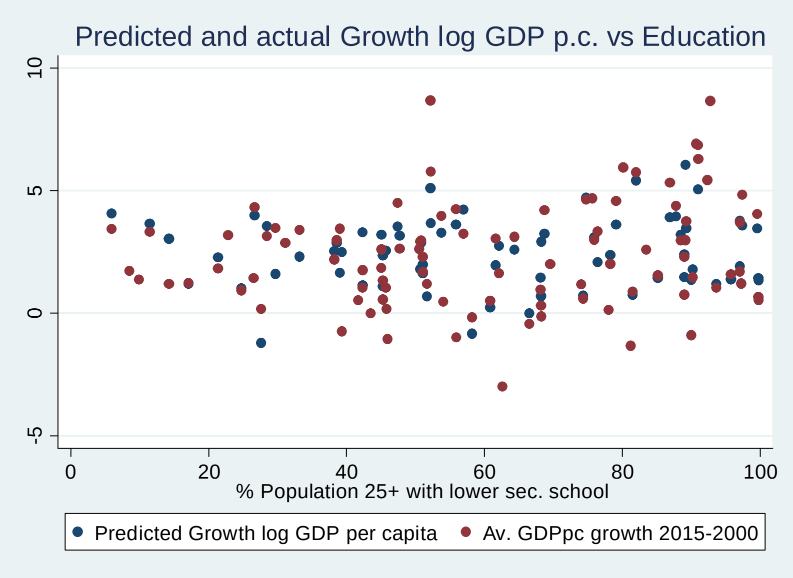
Tests involving nonlinear combinations of parameters

testnl _b[$Educ\_Sec$]/_b[$Trade2000$]=10

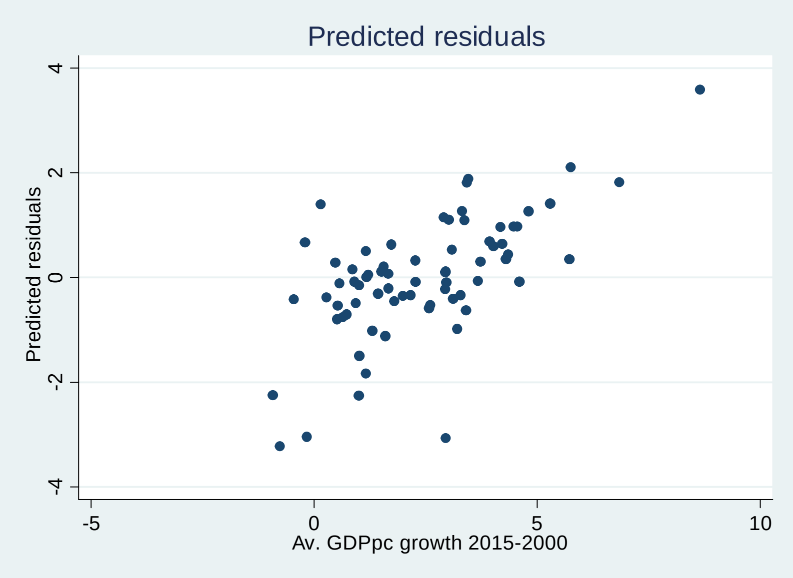
Computing residuals and predicted values

After fitting a linear regression model with `regress`, we can compute the predicted values or regression residuals:

```
predict growthGDPpc_hat, xb
predict residuals_hat, residual
```



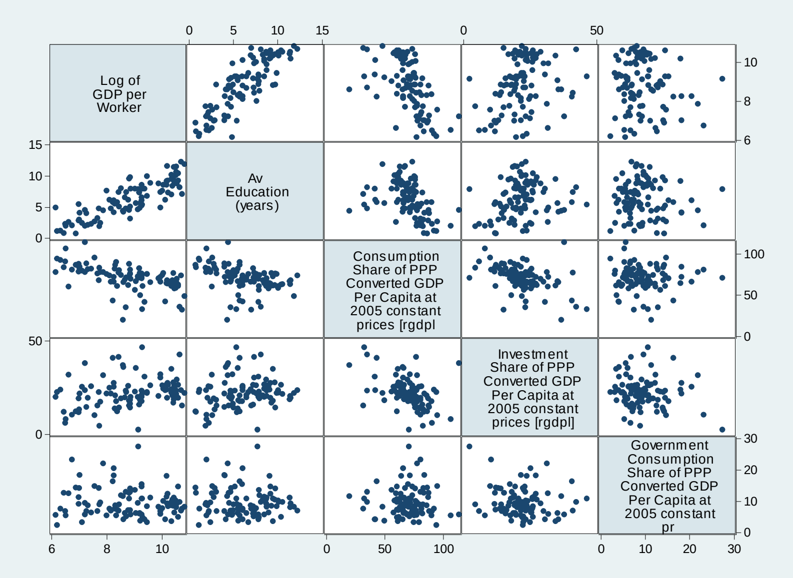
Predicted and actual Growth log GDP p.c. vs Education
-5
0
5
10
0
20
40
60
80
100
% Population 25+ with lower sec. school
Predicted Growth log GDP per capita
Av. GDPpc growth 2015-2000


Predicted residuals
Predicted residuals
4
2
0
-2
-4
-5
0
5
10
Av. GDPpc growth 2015-2000

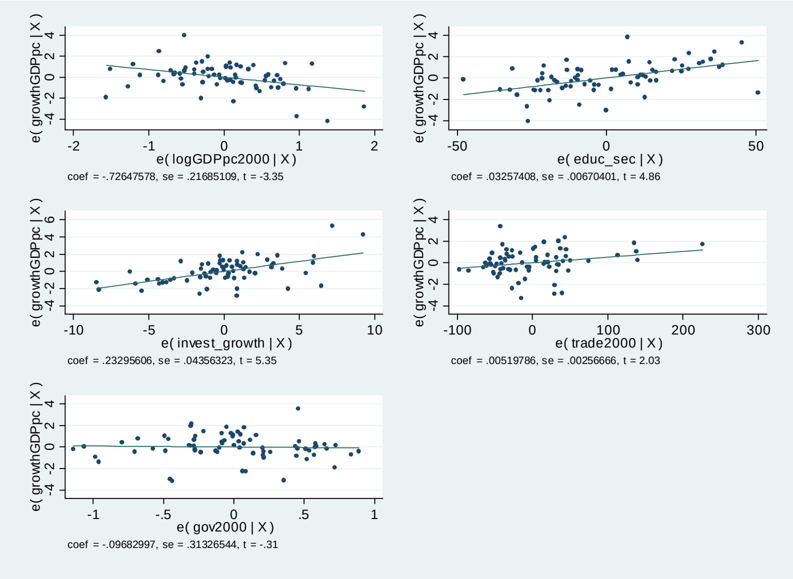
Specification issues: graphically analyzing regression data

`graph matrix` - generates a set of plots illustrating the bivariate relationships underlying the regression model.
`corr` - correlations (covariances) of variables.
`avplots` - added-variable plot decomposes the multivariate relationship into a set of two-dimensional plots.

```
graph matrix growthGDPpc logGDPpc2000 educ_sec
invest_growth trade2000 gov2000, msize(small)
```



```
avplots, msize(small) col(2)
```



Detecting Outliers

An outlier is a data point with an unusual value (observed or residual). Evidence that the model's coefficients are strongly influenced by a few data points casts doubt on the fitted model's worth in a broder context. A data point has a high degree of leverage on the estimates if including it in the sample alters considerably the estimated coefficients

Use the following commands to detect outliers:

```
lvr2plot, mlabel(countrycode)
predict lev if e(sample), leverage
```

The leverage values are computed from the diagonal elements of the matrix $h_i = x_i(\mathbf{X}'\mathbf{X})^{-1}x_i'$.

Alternatively

`predict dfits if e(sample), dfits` (combines levarage values with magnitude of residuals)

Interaction terms and marginal effects

Suppose that education is a complement of capital. How to test it? First: include the interaction term $Educ\_Sec \times Invest.Grow$ in the regression model:

$$
\begin{aligned}
\log(Grow\_GDPper\ capita_c) \quad = \quad & \beta_1 + \\
& +\beta_2[\log GDPpc2000_c] + \\
& +\beta_2[\%\ Educ\_Sec_c] + \\
& +\beta_3[Invest.Grow_c] + \\
& +\beta_4[Trade2000_c] + \\
& +\beta_5[Gov2000_c] + \\
& +\beta_6[\%\ Educ\_Sec_c \times Invest.Grow_c] \\
& +u_c
\end{aligned}
$$

In Stata:

```
regress growthGDPpc logGDPpc2000
c.educ_sec##c.invest_growth gov2000 trade2000
```

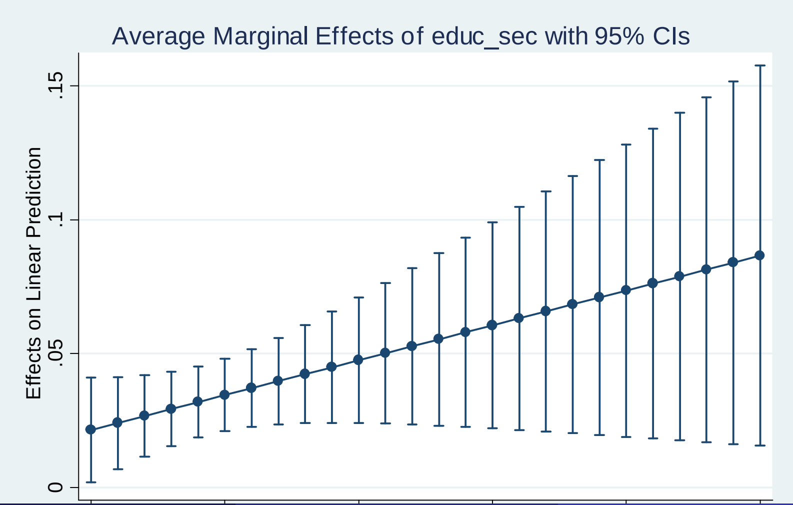
Second: test the significance of $\widehat{\beta}_6$.

Stata has the new command `margins` to compute marginal means, predictive margins, and marginal effects.
Typing `margins, dydx(*)` we get the marginal effects evaluated at the mean of each variable.

$$\frac{\partial \log(Grow\_GDPper\ capita_c)}{\partial \%\ Educ\_Sec_c} = \widehat{\beta}_2 + \widehat{\beta}_6 Invest.Grow$$

$$\frac{\partial \log\left(Grow\_GDPpercapita_c\right)}{\partial ki} = \widehat{\beta}_3 + \widehat{\beta}_6 \%\ Educ\_Sec_c$$

With `margins, dydx(educ_sec) at(invest_growth=(0 (1) 25))` we get marginal effects of education evaluated in different values of `invest_growth`.
After `margins`, we can plot the marginal effects with `marginsplot` command.



The generalized linear regression model

Suppose that $\Sigma_u \neq \sigma^2 I_N$. The OLS estimator is unbiased, consistent, but is no longer efficient as demonstrated by:

$$\begin{aligned}
\widehat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}'\beta + \mathbf{u}) \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\
& E[\widehat{\beta} - \beta] = 0. \qquad (12)
\end{aligned}$$

$$\begin{aligned}
Var[\widehat{\beta}|\mathbf{X}] &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\
&= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\Sigma_u\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}. \qquad (13)
\end{aligned}$$

The VCE computed by regress is $s_u^2(\mathbf{X}'\mathbf{X})^{-1}$. When $\Sigma_u \neq \sigma^2 I_N$ this estimator of the VCE is not consistent and the usual inference procedures are inappropriate.

Heteroskedasticity: causes and test

Potential causes of heteroskedasticity:

- disturbances are often related to some measure of scale (e.g. income);
- disturbances are homoskedastic within groups but heteroskedastic between groups;
- grouped data, in which each observation is the average of microdata.

Stata heteroskedasticity test

`estat hettest, iid`

The `hettest` is the Breusch-Pagan test which test the null hypothesis that the variance of the residuals is homogenous. Therefore, if the p-value is very small, we would have to reject the hypothesis and accept the alternative hypothesis that the variance is not homogenous. The test is very sensitive to model assumptions, such as the assumption of normality. We can use the option `iid` that causes `estat hettest` to compute the $N * R^2$ version of the score test that drops the normality assumption.

Types of heteroskedasticity

In the identically distributed assumption:

$$\Sigma_u = \begin{pmatrix} \sigma^2 & 0 & ... & 0 \\ 0 & \sigma^2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix} = \sigma^2 I_N. \tag{14}$$

If the diagonal elements differ:

$$\Sigma_u = \begin{pmatrix} \sigma_1^2 & 0 & ... & 0 \\ 0 & \sigma_2^2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & 0 & \sigma_N^2 \end{pmatrix} \tag{15}$$

If errors are correlated within clusters (m clusters) of observations, we have:

$$\Sigma_u = \begin{pmatrix} \Sigma_1 & 0 & \dots & 0 \\ 0 & \Sigma_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \Sigma_M \end{pmatrix} \tag{16}$$

Serial correlation in time-series regression models:

$$\Sigma_u = \sigma_u^2 \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{N-1} \\ \rho_1 & 1 & \dots & \rho_{2N-3} \\ \dots & \dots & \dots & \dots \\ \rho_{N-1} & \rho_{2N-3} & 0 & 1 \end{pmatrix} \tag{17}$$

The robust estimator of the VCE

The term we must estimate $\{\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}\} = \{\mathbf{X}'E[\mathbf{u}\mathbf{u}'|\mathbf{X}]\mathbf{X}\}$ is sandwiched between the $(\mathbf{X}'\mathbf{X})^{-1}$ terms. Huber (1967) and White (1980) showed that:

$$\widehat{S}_0 = \frac{1}{N}\sum_{i=1}^{N}\widehat{u}_i^2\mathbf{x}_i'\mathbf{x}_i, \tag{18}$$

consistently estimates $\{\mathbf{X}'E[\mathbf{u}\mathbf{u}'|\mathbf{X}]\mathbf{X}\}$ when the u_i are conditionally heteroskedastic. The robust estimator of the VCE is:

$$Var[\widehat{\beta}|\mathbf{X}] = \frac{N}{N-k}(\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{i=1}^{N}\widehat{u}_i^2\mathbf{x}_i'\mathbf{x}_i\right)(\mathbf{X}'\mathbf{X})^{-1}. \tag{19}$$

The `robust` option available with `regression` command implements the estimator described above.

After `regress y x, robust` Wald tests produced by `test` will be robust to conditional heteroskedasticity of unknown form.

The cluster estimator of the VCE

If errors are correlated within clusters of observations but uncorrelated between different clusters, we can use an estimator of the VCE referred to as the cluster-robust-VCE estimator:

$$Var[\widehat{\beta}|\mathbf{X}] = \frac{N-1}{N-k}\frac{M}{M-1}(\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{j=1}^{M}\widetilde{u}_j'\widetilde{u}_j\right)(\mathbf{X}'\mathbf{X})^{-1}. \qquad (20)$$

Like `robust` option, application of the `cluster()` option does not affect the point estimates but only modifies the estimated VCE:

```
reg growthGDPpc logGDPpc2000 educ_sec invest_growth gov2000
trade2000 , robust
reg growthGDPpc logGDPpc2000 educ_sec invest_growth gov2000
trade2000 , cluster(open)
```

In the presence pf heteroskedasticity and autocorrelation we can use the Newey-West (HAC) estimator of the VCE, with the command `newey`.

The GLS and FGLS estimator

With a known Σ_u matrix, we can premultiply the model by $\mathbf{P}'=\Sigma_u^{-1}$:

$$\begin{aligned} \mathbf{P}'\mathbf{y} &= \mathbf{P}'\mathbf{X}\beta+\mathbf{P}'\mathbf{u} \quad (21) \\ \mathbf{y}^* &= \mathbf{X}^*\beta+\mathbf{u}^* \quad (22) \end{aligned}$$

where

$$Var[\mathbf{u}^*] = E[\mathbf{u}^*\mathbf{u}^{*\prime}] = \mathbf{P}'\Sigma_u\mathbf{P}' = \mathbf{I}_N. \quad (23)$$

Then,

$$\widehat{\beta}_{GLS} = (\mathbf{X}^{*'}\mathbf{X}^{*})^{-1}(\mathbf{X}^{*\prime}\mathbf{y}^{*}) \tag{24}$$

and

$$Var[\widehat{\beta}_{GLS}|\mathbf{X}] = (\mathbf{X}'\Sigma_u^{-1}\mathbf{X})^{-1}. \tag{25}$$

The FGLS estimator is applied when Σ_u is not known and if we have a consistent estimator of Σ_u, denoted $\widehat{\Sigma}_u$, replacing $\mathbf{P}'$ with $\widehat{\mathbf{P}}'$.
In grouped data we can estimate FGLS models multiplying original data with proper weights. The "analytical weights" (`aw`) are the inverse of the observations variances, and the original data are multiplied by them.

References

Baum, Christopher (2006), An Introduction to Modern Econometrics Using Stata, Revised Edition, Stata Press.
Cameron, Colin and Pravin Trivedi (2010), Microeconometrics Using Stata, Stata Press.
Wooldridge, J. (2013) Introductory Econometrics: A Modern Approach, 5th Ed., South-Western College Publishing.