Reading Regression Results

A Quick Guide

Econometrics is the principal tool to provide evidence. It consists mainly of regressions. Below is a guide to reading and understanding regression results.

Equation

We start with an equation, which has a dependent variable (DV) on the left side and (IDV) independent variables on the right side. Basically we are trying to determine if they are correlated. When the IDV variable is higher, is the DV higher, when IDV is lower, is DV lower? We are not saying the one causes the other but we are sort of implying that it does. That is, does IDV cause DV?

We write:

$DV = IDV_1 + IDV_2 + IDV_3$

Or we can use Y and Xs.

$Y = X_1 + X_2 + X_3$

Or we can use other letters (see below, L for lung capacity, etc.).

Basic example

We might think that people who cycle 3 hours per week and swim 2 times per week and eat a bag of applies have better health. How do we measure better health? We can use lung capacity.

Lung capacity = Swim + Cycling + Apples

$L = S + C + A$

So we get data on many people on their lung capacity, weekly exercise and carrot consumption. People who swim more should have better lung capacity, etc. but we want to test this with a regression to see if it is true. We want to test which variables on the right side are statistically significantly related to lung capacity. It is likely that swim and cycling are significant but that apples is not. Apples might improve health but not lung capacity. Apples will be insignificant.

Presentation

In a journal paper, the equation is written horizontally (as they are above). However, the results of the regressions (the numbers) are normally presented vertically (i.e. in columns). You need to read down the column.

Look at the print out from Balart, which is partly reproduced at the end of this note. There is an equation just above the text. It starts G_c = ... That is the equation presented horizontally. The S variable is for 'starting performance' and the PD variable is for 'performance decline'. Do you see them? Now look at Table 5, these variables are listed, but vertically.

There is also normally more than one column. These are slightly different versions of the same regression or equation. One regression (column) might include an extra variable, one might drop a variable, one might use a slightly different regression technique, etc.

The DV (left side) is not shown. But it should say somewhere in the table (at the top, or in the notes) what this variable is. It normally would say, 'Dependent variable is the growth rate of GDP per capita' or 'Dependent variable = growth of GDP p.c'. Always determine first what the DV is so you know what the regression is about.

The IDVs are written in words or letters in the first column and the results for each regression are presented (in numbers) in the subsequent columns.

What you want to know are two things:

What variables are in each regression (swimming, years of education, etc.)?
Is the variable significant?

Most results will put asterisks (*) beside the variables to indicate if they are significant. Normally,

*** is very significant (at the 1% level).
** is medium significance (at the 5% level)
* is weak significance (at the 10% level).
if there is no * the variable is not significant.

Reading by Balart et al.

To show how this works, see the reading by Balart et al. on cognitive and noncognitive skills.

Table 4

The DV is the growth rate of GDP per capita (a measure of economic growth). It says this in the notes.

Panel A has 9 regressions and Panel B also has 9. Why there are two panels is not that important but I will explain it. The difference is that the education quality variable used in A is the HW Index whereas B uses PISA. HW is made up of international student test results for a number of different tests. PISA is a single test and is used alone in B. We will focus on Panel A.

Column 1 includes only a variable for years of schooling. The space for the HW Index is blank, which means that variable is not included in this regression. Years of schooling is significantly correlated with the growth rate of GDP per capita. We know this because of the * (i.e. 0.187*). What is means is that if the people in a country have a higher number of years of schooling, the economic growth rate will be higher.

Column 2, the HW Index is included but not the years of schooling. The HW is significant as indicated by 2.256***.

In column 3, both variables are included and here the years of schooling is no longer significant (no *) but the HW Index is. This is true for all the regressions in Panel A and in Panel B.

Interpretation

What does this mean? Both the HW Index and years of schooling are both measures of human capital. The first is a measure of schooling *quality* (how well students do on the tests and therefore how much they learn in school). The second is a measure of schooling *quantity* (number of years they were in school).

The results show that each of the two human capital variables is significant when included in the regression by itself. However, when they are included together, schooling quality is significant and schooling quantity is insignificant. The interpretation is: school quality is important for economic growth and cancels out the impact of the years of schooling.

t statistics

The asterisks indicate significance. Below the coefficients is often a number in parentheses (   ) that is the t-statistic. It also indicates significance. As a general rule: take off the negative sign (if there is one) and if the number is greater than 2 it is significant and if it less than about 2 it is not significant. This is a rough rule. It may be the case that if the t is a bit below 2 it might still be significant. It depends on the nature of the data. As I said, it is a rough guide. Certainly if the t stat is 4 or 5 you can be sure it is significant and if it is 0.5 it is not significant.

If asterisks are provided, you don't really need to read the t statistic because you already know whether it is significant or not.

End

Table 4
Growth regressions with the HW-index and PISA scores using the PISA sample.

	(1)	(2)	(3)	(4)[a]	(5)[b]	(6)[c]	(7)[d]	(8)[e]	(9)[f]
Panel A: HW-index as a measure of human capital with restricted sample									
HW-index		2.256***	2.260***	2.310***	2.186***	1.144**	1.399***	1.378***	2.213***
		(9.15)	(9.22)	(9.14)	(9.59)	(2.71)	(3.74)	(3.74)	(9.92)
Years of schooling	0.187*		−0.00375	−0.0320	−0.0661	0.0420	0.0582	0.0115	−0.0250
	(1.80)		(−0.05)	(−0.49)	(−1.02)	(0.51)	(0.85)	(0.15)	(−0.30)
N	37	37	37	37	37	37	36	36	37
Adj. R^2	0.208	0.730	0.722	0.723	0.809	0.784	0.770	0.799	0.717
Panel B: PISA 2006 as a measure of human capital with restricted sample									
PISA 2006		2.282***	2.245***	2.235***	2.223***	1.181***	1.265***	1.220**	2.299***
		(7.98)	(7.59)	(6.84)	(7.16)	(3.53)	(2.83)	(2.74)	(9.73)
Years of schooling	0.187*		0.0426	0.0316	−0.0241	0.0654	0.104	0.0526	0.0305
	(1.80)		(0.53)	(0.38)	(−0.28)	(0.84)	(1.46)	(0.63)	(0.35)
N	37	37	37	37	37	37	36	36	37
Adj. R^2	0.208	0.700	0.694	0.692	0.691	0.803	0.754	0.781	0.728

Notes: t statistics in parentheses, heteroskedasticity robust standard errors * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: average annual growth rate in GDP per capita, 1960–2000 Regressions include a constant and GDP per capita in 1960.

[a] Measure of years of schooling refers to the average over the period 1960–2000.
[b] Controlling for outliers by using rreg command in Stata.
[c] Includes indicators for the eight world regions.
[d] Controlled for openness of economy and protection against expropriation.
[e] Controls in d plus fertility and tropical location.
[f] GDP per capita 1960 measured in logs.

P. Balart et al. Economics of Education Review 63 (2018) 134–153

Table 5
Regressions of economic growth on the starting performance and performance decline.

	(1)	(2)	(3)	(4)[a]	(5)[b]	(6)[c]	(7)[d]	(8)[e]	(9)[f]
Starting performance	2.143***		1.330***	1.345**	1.049***	0.766*	0.701	0.249	1.934***
	(4.84)		(2.76)	(2.69)	(3.05)	(2.00)	(1.38)	(0.49)	(4.45)
Performance decline		1.872***	1.300***	1.257***	1.467***	0.680*	0.917***	1.120***	0.701**
		(6.36)	(4.44)	(4.17)	(5.14)	(1.83)	(3.09)	(4.95)	(2.64)
Years of schooling	0.0415	0.173*	0.0871	0.0651	0.0483	0.0793	0.131*	0.0839	0.0344
	(0.53)	(1.86)	(1.23)	(0.96)	(0.64)	(1.00)	(1.93)	(1.14)	(0.44)
N	37	37	37	37	37	37	36	36	37
Adj. R^2	0.579	0.623	0.726	0.721	0.746	0.787	0.754	0.798	0.716

Notes: t statistics in parentheses, heteroskedasticity robust standard errors * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Dependent variable: average annual growth rate in GDP per capita, 1960–2000; Regressions include a constant and GDP per capita in 1960 [a–f] See the description of Table 4.

$$G_c = \beta_0 + \beta_1 S_c + \beta_2 PD_c + \beta_3 GDP_{0c} + \sum_n \delta_n Z_{nc} + \epsilon_c \tag{3}$$

When estimating Eq. (3), we try to stay as close as possible to HW (2012). We use the same data on economic growth, the same growth period (1960–2000), identical covariates, estimate the same model specifications, and start with the same sample of countries. However, the decomposition method that we apply in this paper exploits a specific feature of the PISA test, namely the random allocation of the PISA booklets (see Section 3). Hence, we can apply the decomposition method only to one of the tests included in the HW-index. This has two implications for the estimations. First, the sample of countries that participated in the PISA test differs from the sample used in HW (2012). As a first step in our analysis, we check whether the reduction in sample size from 50 to 37 countries, due to our reliance on the PISA, changes the results obtained in HW (2012). The second implication is that we use the PISA test only for measuring skills, and not the complete set of tests used for the HW-index. However, the PISA scores are highly correlated with the HW-index ($r = 0.91$). Further we will show below that re-estimating the main models from HW (2012) with PISA scores instead of the HW-index delivers very similar results. This suggests that PISA scores are a good proxy for the HW-index and, therefore, we use the PISA scores for estimating Eq. (1). Next, we decompose these PISA scores into the two components and we use these two components for