

Econ 113 – American Economic History

Tuesday 8th February & Thursday 10th February

Richard Halkett – richard@berkeley.edu

Agenda

1. Regression
2. Growth accounting
3. Discussion of the readings

Reading Regression Tables

(developed from a handout by Marit Rehavi)

What is regression?

Essentially, it's correlation in more than two dimensions. The aim is to most-precisely estimate the size of a random variable (which is difficult to measure) based on the values of other measured (and more easily measurable) random variables.

The equation of a regression line is:

$$\hat{y} = mx + b$$

Where \hat{y} is the predicted y-value for a given x-value

Where m is the slope

Where b is the y-intercept

The variable of interest, y , is conventionally called the "dependent variable". The terms "endogenous variable" and "output variable" are also used. The other variables x are called the "independent variables". The terms "exogenous variables" and "input variables" are also used (from *Wikipedia*).

Multiple regression equations take the form:

$$\hat{y} = b + m_1x_1 + m_2x_2 + m_3x_3 + \dots + m_kx_k$$

In these, multiple variables influence the predicted value of the dependent variable. Thus, whilst there remains one y-intercept (the constant, b), there are multiple independent variables ($x_1, x_2, x_3,$

etc) each influencing the *slope* of the line according to the sign and magnitude of their coefficient.

How does it work?

When we look at a regression we are generally interested in three questions:

1. What is the relationship of a particular independent variable to the dependent variable?
2. How strong is that relationship?
3. Using our entire equation, how precisely can we estimate the desired parameter?

We find the answer to the first question in the estimated “coefficient,” the answer to the second question is found in the “T-statistic,” and the answer to the third in reading the coefficient of determination (r^2) statistic.

First, let’s focus on the estimated relationship. Before getting into an example, here are a few points to keep in mind:

- A positive regression coefficient is similar to two things being positively correlated (i.e. they move together).
- Conversely, a negative regression coefficient is similar to two things being negatively correlated (i.e. they move in opposite directions, when one goes up the other goes down).

Example: Say you were given the following table of results (these numbers are all made up) about the weight of tigers:

Dependent Variable: Weight of Tigers

Variable	Coefficient	Standard Error	T-Statistic
Constant	100	50	2
Age	10	2.4	0.5
Calories consumed (per day)	0.1	0.02	5
Length	5	2	2.5
Exercise hours (per day)	-20	15	1.33
Number of Whiskers	0.2	0.05	4

You can always use the information from a regression table to rewrite a regression as an equation. The constant becomes b and the remaining coefficients are applied to the corresponding variable. You must always add an error term, although for statistical purposes, this is equated to being near-zero.

For example, based on the table above you could write:

$$\text{Weight} = 100 + 10 \text{ Age} + 0.1 \text{ Calories} + 5 \text{ Height} - 20 \text{ Exercise} + 0.2 \text{ Whiskers} + \text{error}$$

Now say we have two tigers, Tigger and Tony.

	Tigger:	Tony:
Age	5	5
Calories consumed:	2000	1000
Length	5 feet	5 feet
Exercise Hours	2 hours	2 hours
Whiskers	5 whiskers	5 whiskers

Given our estimates from the table, what would we expect each of their weights to be?

$$\text{Weight Tigger} = 100 + 10*5 + 0.1*2000 + 5*5 - 20*2 + 0.2*5 = 336 \text{ pounds}$$

$$\text{Weight Tony} = 100 + 10*5 + 0.1*1000 + 5*5 - 20*2 + 0.2*5 = 236 \text{ pounds}$$

What is the expected difference in Tigger and Tony's weights? You could estimate each tiger's weight using the formulas above and then take the difference: $336 - 236 = 100$ pounds. However this isn't always convenient and there is a faster way to estimate this. Notice, the only difference between the two equations is the number of calories eaten per day. So you can just take the coefficient of calories (0.1) and multiply it by the difference in calories to get the expected difference in weights (see below).

$$\text{Weight Tigger} = 100 + 10*5 + 0.1*2000 + 5*5 - 20*2 + 0.2*5 = 336 \text{ pounds}$$

$$\text{Weight Tony} = 100 + 10*5 + 0.1*1000 + 5*5 - 20*2 + 0.2*5 = 236 \text{ pounds}$$

$$\text{Tigger} - \text{Tony} = \qquad 0.1 * (2000 - 1000) \qquad = 100 \text{ pounds}$$

Precision and Statistical Significance

While we've talked a lot about the estimates, we haven't looked at how precise they are. How big an error is "too big" depends on what you're estimating. If I tell you that the standard error of my coefficient is 5 you have no idea if I have a precise estimate. It depends on the size of what I'm estimating.

For example, if I'm estimating United States' GDP it's an incredibly precise estimate (5 is tiny compared to \$10 trillion, the size of US GDP). However, if I'm talking about height, 5 feet is huge. If that is the error of your estimate, you have absolutely no idea how tall someone is. By relating the error to the magnitude of the coefficient, T-statistics provide a standardized way for researchers to communicate how precise their estimates are.

$$\text{T-statistic} = \frac{\text{Coefficient}}{\text{Standard Error}}$$

How big is “big enough?”

It depends on what you want to know and how precise you need to be. As a general rule of thumb – and in Econ 113 – we will say if the T-statistic is greater than 2, then the estimate is statistically significant (note if the t-statistic were less than negative 2 it would also be significant).

We use 2 (actually 1.96) as a rule of thumb because you need a t-statistic at least that big to say that 95% of the time the coefficient is **not** just as likely to be 0 as it is to be number we estimated (remember a coefficient of 0 means that there is no relationship between the two variables). In other words, if the t-statistic is greater than or equal to 2 we can be fairly confident that our estimate tells us the sign (positive or negative) of the coefficient.

Back to our example, which coefficient is **not** statistically significant?

Answer: The coefficient on age has a t-statistic of 0.5, therefore it is not statistically significant.

Precision/usefulness of the model

The r^2 statistic is called the coefficient of determination:

$$r^2 = (\text{Explained variation}) / (\text{Total variation})$$

It is the ratio of the explained variation to the total variation. Thus, if the r^2 is 0.90 then 90% of the variation of y can be explained by the model you have developed.

Statistically Significant vs. Economically Significant

Statistical significance **only** tells you how precise your estimate is. It does not tell you anything about whether your estimate is *important*. We say something is statistically significant if it is precisely estimated, but we say it is economically significant if it is important.

- For example, in the table above we were able to very precisely estimate the relationship between whiskers and weight (t-statistic = 4). However, our estimates also tell us that the number of whiskers do not translate into much weight. 5 whiskers are only associated with 1 pound of weight, less than half of a percent of the weight of our smallest tiger, Tony. That’s hardly important. Therefore, we would say that the relationship between tiger weight and whiskers is statistically significant, but it is not economically meaningful.
- Something can be economically significant even if it has a very small coefficient. For example, some theoretical models predict the sign of a coefficient. If we estimate the relationship and find the opposite sign (even if the coefficient is very small), that would be economically significant because it would call a theory into question.

Uses and abuses

1. Confusing correlation and causation – Hannon p1023
 2. Linkages between variables – multicollinearity
 3. Look at the shape of the data!
-

Peter Temin: Labor scarcity and the problem of American Industrial Efficiency in the 1850s

- British visitors to the 1853 industrial exhibition in New York
 - “on account of the high price of labour the whole energy of the people [in the United States] is directed to improving and inventing labour-saving machinery.”
 - BUT impression largely formed from observation of two industries (woodworking and hardware manufacture)
 - ALSO mainly casual observation
 - AND economic model demonstrating lack of incentive to adopt technology by either side
 - THEREFORE rejects their opinion and suggests that UK and USA were “using essentially the same technology.”
- Erwin Rothbarth stated that inexpensive land in the US made labour more expensive
 - Question: does a difference in technology represent *factor substitution* or an actual technological advance (economically-speaking)?
 - Temin states that the difference in wages between the two countries was “a statistical illusion caused by using the exchange rate to compare wages, rather than the exchange rate plus the tariff.”
 - Summary
 - The availability of land in the US cannot be used as an argument for more labour being used in the UK
 - A lower interest rate would be needed to explain such a phenomenon and this did not exist
 - Therefore increased American use of more capital per worker must have been due to using a different technology
- Overall explanation: asymmetry in the flow of information
 - Americans were more interested in British discoveries than the British were in those made in America
 - Therefore, the two countries’ technologies diverged with America’s developing more quickly
 - Less capital per worker was expended in the US

John R. Nelson Jr – Alexander Hamilton and American Manufacturing: A Reexamination

- Traditional revisionist history
- AH is traditionally presented as the “foremost advocate of American manufacturing”

- Nelson doesn't think that he is – unsurprisingly, our hero is a more nuanced, more confused figure than the normal hagiography would have us believe.
- AH champion of the national bank and establishment of the national debt to strengthen the “infant Government by increasing the number of ligaments between the Government and the interests of Individuals.”
- AH was very pro-Britain and British manufacturing since the import of British MFG produced 80% of the tariffs that kept the government afloat
- On the whole his plan for the US economy “served to deprive manufacturers of capital”
- Also opposed Jefferson's plan to reduce domestic taxation, instead favouring a reduction in import tariffs (which would therefore increase imports and increase competition for local manufacturers)
- Society for Establishing Useful Manufactures (SEUM)
 - Threefold purpose
 - Support price levels of the speculative market by stabilising demands for government bonds
 - Provide a productive outlet in manufacturing for surplus merchant capital
 - Curb the outflow of American securities abroad by requiring their use in subscriptions to SEUM's stock
 - Head of SEUM was William Duer, the “Prince of the Speculators.”
 - OVERALL: not very pro-manufacturing, despite its name
- Report on manufactures
 - AH gave it a low priority
 - Eventually completed one year after a report on banking requested 8 months *after* the one on manufacturing
 - Opens with a strong defence of manufacturing development
 - Suggested three sources of capital for manufacturing
 - Domestic banks – useless because established for short-term loans to merchants
 - Funded debt – useless because most capital found adequate return in land or stock
 - Foreign investment – useless because British investors already had outlets for their capital and were unlikely to finance competitors to their own industry
 - DID NOT suggest direct government loans to the manufacturers
 - DID NOT suggest protective tariffs and prohibitions
 - Depended on “pecuniary bounties” but these only favourable to big business
 - DID NOT protest when report was tabled by the House

Joan Underhill Hannon – Poverty in the Antebellum Northeast: The View from New York State's Poor Relief Rolls

- Rise in the number of able-bodied poor
- Plenty of statistics
- Contemporary politicians & relief officials blame personal character and moral failing
- Social welfare historians stress urbanisation, industrialisation and immigration
- Michael Katz suggests rise of production for market and hence wage labour

- Irregular working
- Low wages leading to small savings
- Break of ties with families and local communities who had previously supported the needy
- BUT lack of direct evidence
- Use of decline of household productivity to indicate move to wage labour
- HOUSEPROD is significantly and negatively correlated with relieved pauperism throughout the antebellum period

Jeffrey G. Williamson – Embodiment, Disembodiment, Learning by Doing and Returns to Scale in Nineteenth-Century Cotton Textiles

- Response to a previous article by Paul David in *The Journal of Economic History*, Sept 1970
- Dispute over where innovation took place – inside or outside the workplace?
- Just under two-thirds of output growth per man-hour can be attributed to learning effects during the 1830s and 1840s
- In 15 years pre-Civil War (1850 onwards) none of the productivity growth can be explained by learning effects
- Capital-augmenting technical change accounts for around one-sixth of productivity growth across the period
- Disputes David's assertion of constant returns to scale across the period
- From 1850, cannot establish whether cost reductions achieved through learning were a result of cumulated output or temporal duration of production experience