Linear regression model (LRM)

- This is a REVIEW of concepts you should familiar with (from the previous course). We will focus on applications

- Multiple regression model and assumptions
- Hypothesis testing
- Functional forms
- Reporting and interpreting results
- Logit and probit models (depending on time, in review session)

Note: If you don't understand or remember something after this review, you can use the Wooldridge – you can ask me for recommended chapters and pages

References for this part

- Wooldridge, J.M. (2007) Introducción a la econometría: un enfoque moderno. Thomson-Paraninfo.
- Gujarati, D. N. (2011). Econometrics by example. Hampshire, UK: Palgrave Macmillan.
- Same examples and definitions from: http://www.ats.ucla.edu/
- Studenmund, A. H. (2016). Using Econometrics: A Practical Guide, Global Edition. Pearson Education Limited.
- Wheelan, C. (2013). Naked statistics: stripping the dread from the data. WW Norton & Company.
- Bertrand, M., Goldin, C., & Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. American Economic Journal: Applied Economics, 228-255.

The power of linear regression models

Wine quality = 12.145 + 0.00117 * Winter Rainfall + 0.0614 average growing season temperature – 0.00386 harvest rainfall

What We Can Learn About Investing From the Business of French Wine

Algorithms are better than humans when it comes to predicting future events—no matter wine prices or equities in the financial markets.

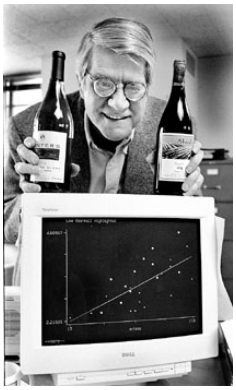
Originally published: November 20, 2013

KEY TAKEAWAYS

- Human memory tends to be biased, and we are inconsistent at summarizing complex information.
- In fields where predictability is difficult, like markets or wine prices, experts are often inferior to algorithms.

To gauge the price for a rare vintage Bordeaux, you could travel around the French countryside asking grape growers to recall weather from decades ago.

They would tell you that the most valuable vintages follow a similar pattern. Hot and dry summers yield grapes with higher sugar concentrations, compared to wetter conditions that result in grapes with more diluted flavors.



Prof. Ashenfelter

https://www.betterment.com/resources/investment-strategy/behavioral-finance-investing-strategy/human-vs-algorithm-investing-a-lesson-from-wine-country/

Why do we care about a linear regression model?

- Example: Is there gender discrimination in the workplace for young professionals after a MBA? (Bertrand, Goldin & Katz, 2010)
  - "After 10 years in the workplace, however, a huge gap has opened up; women on average are earning a striking 45 percent less than their male class mates" ("Naked Statistics" by C. Wheelan, p. 203)
  - What if we control for other factors? (e.g. perhaps women tend to choose part-time jobs)
- LRM allows to account for associations (and hopefully causal ones) between variables we care about while controlling for other factors

What do you learn?

TABLE 3—WAGE REGRESSIONS FOR POOLED SAMPLE

	Dependent variable: Log (annual earnings)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	−0.287 [0.035]***	−0.190 [0.033]***	−0.146 [0.032]***	−0.173 [0.030]***	−0.094 [0.029]***	−0.064 [0.029]**	−0.054 [0.028]	−0.038 [0.025]
MBA GPA		0.429 [0.054]***	0.406 [0.053]***		0.369 [0.051]***	0.351 [0.051]***	0.367 [0.049]***	0.347 [0.043]***
Fraction finance classes		1.833 [0.211]***	1.807 [0.206]***		1.758 [0.199]***	1.737 [0.194]***	1.65 [0.193]***	0.430 [0.180]**
Actual post-MBA exp			0.046 [0.075]			0.085 [0.071]	0.056 [0.068]	0.029 [0.064]
Actual post-MBA exp^2			0.010 [0.004]***			0.005 [0.004]	0.008 [0.003]**	0.007 [0.003]**
Any no work spell			−0.290 [0.067]***			−0.228 [0.062]***	−0.218 [0.061]***	−0.173 [0.054]***
Dummy variables:								
Weekly hours worked	No	No	No	Yes	Yes	Yes	Yes	Yes
Pre-MBA characteristics	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Reason for choosing job	No	No	No	No	No	No	Yes	Yes
Job function	No	No	No	No	No	No	No	Yes
Employer type	No	No	No	No	No	No	No	Yes
Cohort × year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	12.156 [0.018]***	9.493 [0.585]***	8.809 [0.667]***	10.385 [0.151]***	8.08 [0.603]***	7.525 [0.694]***	8.229 [0.733]***	8.324 [0.547]***
Observations	18,272	18,272	18,272	18,272	18,272	18,272	18,272	18,272
R^2	0.15	0.31	0.34	0.26	0.40	0.41	0.43	0.54

Linear regression model

- The general form of the LRM model is (population model):

$$Y_i = B_0 + B_1 X_{1i} + B_2 X_{2i} + \ldots + B_k X_{ki} + u_i$$

- Or, as written in short (matrix) form:

$$Y = BX + u$$

- What's the meaning of this?
- Y is the regressand or dependent variable, X is a vector of regressors, predictors, covariates or independent/explanatory variables, and u is the error term.
- Regression coefficients: B_0 is the intercept, B_1 to B_k are the slope coefficients.

This review is based on "Econometrics by Example" (Gujarati) and "Introductory Econometrics" (Wooldridge)