Notes on Regression - Approximation of the Conditional Expectation Function

Posted on February 26, 2018

The final installment in my 'Notes on Regression' series! For a review on ways to derive the Ordinary Least Square formula as well as various algebraic and geometric interpretations, check out the previous 5 posts:

- Part 1 - OLS by way of minimising the sum of square errors (/2017/08/16/notes-on-regression-ols/)

- Part 2 - Projection and Orthogonality (/2017/08/23/notes-on-regression-projection/)

- Part 3 - Method of Moments (/2017/08/31/notes-on-regression-method-of-moments/)

- Part 4 - Maximum Likelihood (/2017/09/21/notes-on-regression-maximum-likelihood/)

- Part 5 - Singular Vector Decomposition (/2017/10/21/notes-on-regression-singular-vector-decomposition/)

A common argument against the regression approach is that it is too simple. Real world phenomenons follow non-normal distributions, power laws are everywhere and multivariate relationships possibly more complex. The assumption of linearity in the OLS regression seems way out place of reality. However, if we take into consideration

that the main aim of a statistical model is not to replicate the real world but to yield useful insights, the simplicity of regression may well turn out to be its biggest strength.

In this set of notes I shall discuss the OLS regression as a way of approximating the conditional expectation function (CEF). To be more precise, regression yields the best linear approximation of the CEF. This mathematical property makes regression a favourite tool among social scientist as it places the emphasis on interpretation of an approximation of reality rather than complicated curve fitting. I came across this method from Angrist and Pischke's nearly harmless econometrics (https://press.princeton.edu/titles/8769.html).

What is a Conditional Expectation Function?

Expectation as in the statistics terminology normally refers to the population average of a particular random variable. The conditional expectation as its name suggest is the population average conditional holding certain variables fixed. In the context of regression, the CEF is simply $E[Y_i|X_i]$. Since X_i is random, the CEF is random.[1]

CEF
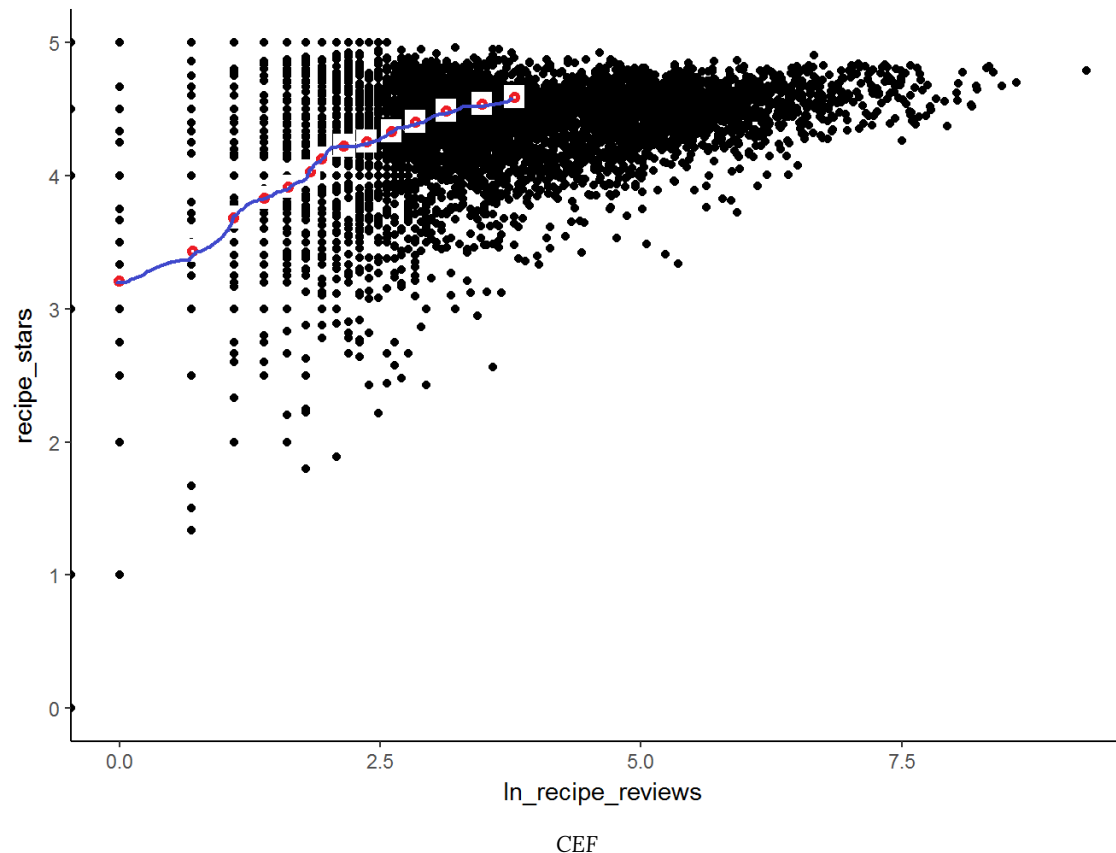
The picture above is an illustrated example of the CEF plotted on a given dataset. Looking at the relationship of the number of stars obtained by a recipe and the log number of reviews, one can calculate the average star rating for a given number of reviews (indicated by the red dots). The CEF function joins all these red dots together (indicated by the blue line).

Nice Properties of the CEF

What could we infer about the relationship between the dependent variable, Y_i and the CEF? Let's split the dependent variable into two components:

$$Y_i = E[Y_i|X_i] + \epsilon_i$$

Using the law of iterated expectation, we can show that $E[\epsilon_i|X_i] = 0$ i.e. mean independence and ϵ_i is uncorrelated with any function of X_i. In other words, we can break the dependent variable into a component that is explained by X_i and another component that is orthogonal to it. Sounds familiar?

Also, if we were to try to find a function of X, $m(X)$ that minimises the squared mean error i.e. $min\ E[(Y_i - m(X_i))^2]$, we would find that the optimum choice of $m(X)$ is exactly the CEF! To see that expand the squared error term:

$$\begin{aligned}(Y_i - m(X_i))^2 &= ((Y_i - E[Y_i|X_i]) + (E[Y_i|X_i] - m(X_i)))^2 \\ &= (Y_i - E[Y_i|X_i])^2 + 2(Y_i - E[Y_i|X_i])(E[Y_i|X_i] - m(X_i)) + (E[Y_i|X_i] - m(X_i))^2\end{aligned}$$

The first term on the right does not factor in the arg min problem. $(Y_i - E[Y_i|X_i])$ in the second term is simply ϵ_i and a function of X multiplied with ϵ_i would still give an expectation of zero. Hence, the problem can be simplified to minimising the last term which is only minimised when $m(X_i)$ = CEF.

Regression and the CEF

Now let's link the regression back to the discussion on the CEF. Recall the example of the number of stars a recipe has and the number of reviews submitted. Log reviews is a continuous variable and there are lots of points to take into consideration. Regression offers a way of approximating the CEF linearly i.e.

$$\beta = \arg \min_b E[E[Y_i|X_i] - X_i'b]$$

To get this result, one can show that minimising $(Y_i - X_i'b)^2$ is equivalent to minimising the above equation.[2] Thus, even if the CEF is non-linear as in the recipe and star rating example, the regression line would provide the best linear approximation to it (drawn in green below).
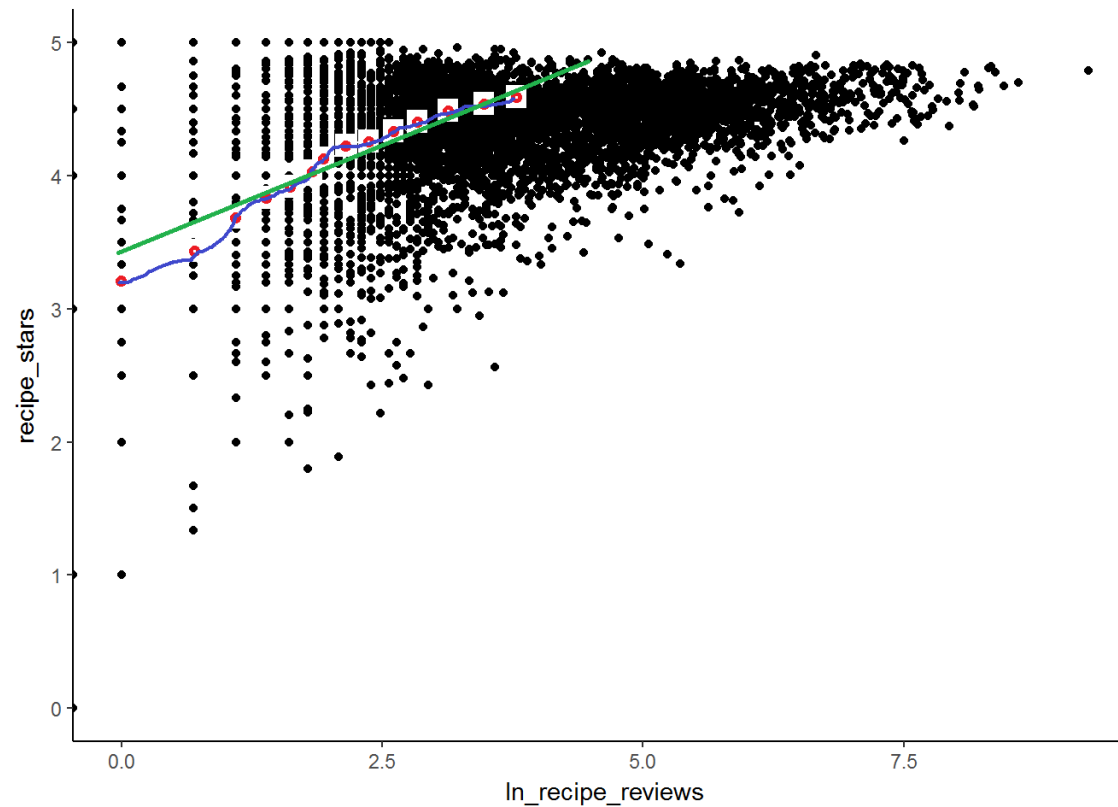
Regression as the best linear approximation to the CEF

1. In practice, one obtains a sample of the population data and uses the sample to make an approximation of the population CEF.↵

2. just add and subtract $E[Y_i|X_i]$ and manipulate the terms in a similar way to the previous proof using $m(X)$.↵

(mailto:timothy.lin@alumni.ubc.ca)　(https://github.com/timlrx)

(https://linkedin.com/in/timothy-lin-0600ba141)　(https://www.timlrx.com/index.xml)

Timothy Lin　•　2019　•　Quasilinear Musings (https://www.timlrx.com/)

Hugo v0.19 (http://gohugo.io) powered　•　Theme by Beautiful Jekyll (http://deanattali.com/beautiful-jekyll/)
adapted to Beautiful Hugo (https://github.com/halogenica/beautifulhugo)