Bedrocks of Quantitative Finance: The Linear Regression
Riley Dunnaway (03/28/24)

Abstract

An exploration of linear regression and various methods of derivation. Exploration of ordinary least squares, method of moments, and the method of maximum likelihood. These estimation methods represent staples in the tool belt of all quantitative analysts and warrant deep mathematical understanding.

1 Motivation for Linear Regression

Suppose we have two data sets corresponding to variables x and y we suspect are linearly related. The general equation for a line is given by

$$y = \alpha + \beta x \tag{1}$$

The goal of linear regression is to fit a line of form (1) to our data. In the real world this line will never fit our data perfectly, so we must introduce an error variable, ϵ into equation (1):

$$y = \alpha + \beta x + \epsilon \tag{2}$$

The methods we will discuss focus on minimizing the error term, and yielding the most accurate linear representation of our data.

2 Ordinary Least Squares

Note that the error, ϵ_i for each observation x_i is equal to the difference between the model output $\hat{y_i}$ and the actual data value y_i. One method involves minimizing the sum of all $\hat{\epsilon}_i^2$. Notice we square the error so all error is positive and negative/positive errors don't cancel out.

The *residual sum of squares* is given by:

$$\sum_{i=1}^{N} \hat{\epsilon}_i^2 = \sum_{i=1}^{N} (y_i - \hat{y_i})^2 \tag{3}$$

Where N is equal to the number of data observations.

$$\sum_{i=1}^{N} \hat{\epsilon}_i^2 = \sum_{i=1}^{N} (y_i - \hat{y_i})^2 = \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 \tag{4}$$

To minimize this residual sum of squares, we take the first derivatives with respect to $\hat{\alpha}$ and $\hat{\beta}$ and set them equal to 0 to find maxes and mins.

$$\frac{\partial}{\partial \hat{\alpha}} \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 = -2 \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0 \tag{5}$$

$$\frac{\partial}{\partial \hat{\beta}} \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 = -2 \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)(x_i) = 0 \tag{6}$$

From equation (6):

$$-2 \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i) = \sum_{i=1}^{N} y_i - N\hat{\alpha} - \hat{\beta} \sum_{i=1}^{N} x_i = N\bar{y} - N\hat{\alpha} - N\hat{\beta}\bar{x} = \bar{y} - \hat{\alpha} - \hat{\beta}\bar{x} = 0 \tag{7}$$

Where the bar notation indicates the mean of all y or x data values. Since equation (8) shows $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$, we substitute this into equation (7):

$$\begin{aligned} -2 \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)(x_i) = \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)(x_i) = \sum_{i=1}^{N} (y_i - \bar{y} + \hat{\beta}\bar{x} - \hat{\beta} x_i)(x_i) \\ = \sum_{i=1}^{N} x_i y_i - \bar{y} \sum_{i=1}^{N} x_i + \hat{\beta}\bar{x} \sum_{i=1}^{N} x_i - \hat{\beta} \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i y_i - N\bar{y}\bar{x} + \hat{\beta} N \bar{x}^2 - \hat{\beta} \sum_{i=1}^{N} x_i^2 = 0 \end{aligned} \tag{8}$$

Solving equation (9) for $\hat{\beta}$ gives:

$$\hat{\beta} = \frac{\sum_{t_1}^{N} x_i y_i - N\bar{x}\bar{y}}{\sum_{i=1}^{N} x_i^2 - N\bar{x}^2} \tag{9}$$

And there we have it! equations (8) and (10) give us the values of $\hat{\alpha}$ and $\hat{\beta}$, or the equation of the line which minimizes error, in terms of the mean values of the x and y data sets.

3 Method of Moments

This same result can be achieved by using moments with one additional condition. In statistics a moment is a quantitative measure on our dataset, namely mean, variance, skewness, and kurtosis. In this case, we will be looking at the mean.

Return to the problem set up in section one:

$$y = \alpha + \beta x + \epsilon \tag{10}$$

Our assumption for the method of moments will be that the error is normally distributed with a mean of 0, i.e.:

$$\epsilon \sim N(0, \sigma^2) \tag{11}$$

Given this assumption, one can easily see that the expected value of ϵ_i would be 0. So,

$$\mathbb{E}(\epsilon_i) = 0 \tag{12}$$

Rearranging equation (11), we see that $y - \alpha - \beta x = \epsilon$, so by taking the expected value and substituting equation (13) we see

$$\mathbb{E}(y_i - \alpha - \beta x_i) = \mathbb{E}(\epsilon_i) = 0 \tag{13}$$

Since ϵ is normally distributed, we also see

$$\mathbb{E}(\epsilon_i x_i) = 0 \tag{14}$$

Then plugging in equation (14) into (15),

$$\mathbb{E}((y_i - \alpha - \beta x_i)x_i) = 0 \tag{15}$$

Lastly, notice that the variance of our error, ϵ^2 should have the expected value of our standard deviation squared by definition.

$$\mathbb{E}(\epsilon_i^2) = \sigma^2 \tag{16}$$

Now by rewriting the expected values of equations (14), (16), and (17) as the mean of our data, we get the following. Notice the introduction of hat notation to differentiate coefficients of the sample means from those of population moments.

$$\begin{gathered} \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \\ \frac{1}{N}\sum_{i=1}^{N}x_i(y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \\ \frac{1}{N}\sum_{i=1}^{N}\hat{\epsilon}_i^2 = \hat{\sigma}^2 \end{gathered} \tag{17}$$

Notice the similarity between (18) and equations (6) and (7) derived in the least squares method. By following the results of the previous section, we see that the special case of normally distributed error yields the same coefficients as ordinary least squares but is a biased estimator.

4 Method of Maximum Likelihood

One final method of for determining linear regression coefficients uses the likelihood function, or joint-density function, from statistics. Given the same set of assumptions used in Section 3, we get the likelihood function

$$\begin{aligned} L &= \prod_{i=1}^{N}\frac{1}{(2\pi)^{\frac{1}{2}}\sigma}e^{-\frac{1}{2\sigma^2}\epsilon_i^2} \\ &= \frac{1}{(2\pi)^{N/2}\sigma^N}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\epsilon_i^2} \\ &= \frac{1}{(2\pi)^{N/2}\sigma^N}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i-\alpha-\beta x_i)^2} \end{aligned} \tag{18}$$

To maximize this likelihood function, we want the exponential term to be minimized. To do this we take the partial derivatives of the exponent with respect to α and β and set them equal to 0.

$$\begin{aligned} \frac{\partial}{\partial \alpha} \sum_{i=1}^{N} (y_i - \alpha - \beta x_i)^2 = 0 \\ \frac{\partial}{\partial \beta} \sum_{i=1}^{N} (y_i - \alpha - \beta x_i)^2 = 0 \end{aligned} \tag{19}$$

Notice again the similarity to the least squares equations (7) and (8). By solving (19), we get our linear regression coefficients.

5 Comparisons and Conclusion

Given the similarities that appear in the above derivations, one may wonder about the differences between the three methods. Firstly, it is important to notice that the Method of Moments and Method of Maximum Likelihood are only equivalent to the Least Squares estimation given the normality assumption. When working with non-normal distributions of error, one can not assume equivalence between the three derivations.

So when is one method more proper than another? Least squares is viewed as accessible and easily applied in most cases, especially with normal distributions of error. However, the method of moments and maximum likelihood may be more desirable with alternative error distributions. While the method of moments is also fairly accessible, the method of maximum likelihood is generally viewed as the most desirable when computation allows for it. When comparing the method of maximum likelihood to least squares, one must consider the end goal before determining regression method. One method returns the line with the least amount of error while the other returns the most statistically likely line. Both can be useful, but one may be more appropriate based on the distributions of data.