Deriving the equation

The weighted least squares estimation method was introduced earlier in an intuitive approach. There, we postulated the equation for the least squares estimate without further derivation. In this video, I would like to show how we arrive at these equations.

Here we see the equations for the least squares estimate, as we introduced them earlier. The top equation shows the unweighted form of the estimate x-hat, where we see that is actually a linear function of the vector of observations y.

The second equation is exactly identical to the first one, but we introduced a unit-matrix at two positions

This unit matrix consists of the values one on the diagonal, and zeros at all other positions.

We can consider this unit matrix to be a weight matrix, but it is a very special one, because it assigns the same weight to every observation.

Therefore it is identical to the 'unweighted' least squares estimate.

The third equation replaces this unit matrix with an arbitrary weight matrix W.

The weights can have specific values, depending on the problem we are facing.

Now, let's have a look how we can arrive at these equations. What are the basic ideas behind it?

If you remember the problem of fitting a model to the observations, as shown here, it is clear that the model we prefer is a model where the residuals are minimal. These are the blue lines in the figure.

Since there are many observations, and therefore many residuals, we would like to capture them all in a single number. And for this we use the sum of all the squared residuals. The summation will result in one single number, and squaring the residuals ensures that we don't need to consider positive or negative values; via the squares they are all positive.

'Least' squares refers to this criterion. Now let's see how this results in the weighted least squares equation.

We need to find the value e-hat, which minimizes the sum of the squared residuals. However, the e-hat is the difference between the original observations, y, and the y-hat. Y-hat refers to the value the observations would have if they would fit perfectly to the model.

So, the problem shifts from finding the value y-hat, that minimizes the sum of the squared residuals.

Subsequently, y-hat is a function of x-hat, which follows from our standard 'y=Ax' model.

So, the problem shifts from finding the value x-hat, that minimizes the sum of the squared residuals.

This is an equation we can work out further by discriminating the individual components. Let's work this out.

Using the standard rules for multiplication, we rewrite the equation into the individual elements. Remember that our goal still is to minimize x-hat.

If we now look at the dimensions of the elements in this summation, we see that each individual product has the dimension 1-times-1.

In other words, these are all scalars, or single 'numbers'.

Therefore we can write the equation even easier, and lump a couple of elements together.

Remember, we are still on our way to minimize x-hat.

The minimization of a function means that the first derivative of that function is equal to zero.

So here, the minimization of this function means that the first derivative of that function (with x-hat as free parameter) is equal to zero.

This can be worked out: all elements that do not contain an x-hat drop out, and for the other elements we take these first derivatives. We need to take into account that we are dealing with matrices.

Simplifying this leads to the so called 'Normal equations', which contain a linear function of the estimated parameters on the left hand side, and a linear function of the observations on the right hand side.

The product "A-transpose A" is called the "Normal Matrix".

This way we find the expression for x-hat, and subsequently also for y-hat and e-hat.

This derivation still focused on the unweighted least-squares estimate.

Now let's repeat the same equation for the case with weighted observations.

Here we start with the weighted sum of the squared residuals, where the minimization problem starts with minimizing e-hat. By replacing e-hat by y minus y-hat, and later by 'y minus A x-hat', we find the final expression that should be minimized.

Using standard matrix multiplication, we can write this into subsequent terms.

Again, the minimization problem is decomposed in the separate elements, which all have the dimension 1-by-1, which enables us to simplify the equation considerably.

From this point we need to find the minimum again, which means that the first derivative needs to be equal to zero.

We therefore take the derivative of each element that contains the free parameter x-hat, and set the sum of these elements to zero.

Working out these steps leads to the 'normal equations', but now for the weighted case, and consequently to an expression of x-hat, the estimate of the unknown parameter vector x.

In other words, doing some reverse engineering: if we use this expression of x-hat, we find the minimum value of this function.

Since "y-hat is A times x-hat", and "e-hat is y minus y-hat", this also means that this expression for x-hat 'minimizes the sum of the squared residuals', which was our starting point.

For this reason, x-hat is our Least Squares estimate.

In the end, this gives us the expressions of the weighted least squares estimators for x, y and e.

In this video, we elaborated on how we could derive the set of equations for Weighted Least Squares estimation. When the weight matrix is given, this now enables us to estimate the unknowns.

In the following videos, we will find out how we can find the optimal weight matrix for each problem.