Observation Theory

Script V42B – Best Linear Unbiased Estimation

To understand the background and implications of best linear unbiased estimation, I would like to go a bit deeper into its origins

Here you see the final appearance of the best linear unbiased estimator x-hat.

It is clear that, apart from the observations in vector y, it is dependent on two components:

The design matrix A.

And the covariance matrix Qy.

(or actually, the inverse of this covariance matrix)

The other estimators, yhat and ehat remain unchanged from the way we introduced them earlier.

In this video, I would like to look a bit deeper into the elements of BLUE.

Blue stands for Best Linear Unbiased Estimation.

We already know what 'estimation' means: we want to find a value for a parameter which is as close as possible to its true, but unknown value.

But what about the other three words?

Let's start with 'Linear'.

When we talk about a linear function, we typically mean that there is a linear relationship between two variables, as in this example.

Here we only have two scalars, which are linearly related.

If we write this in a general form, it would look like this

Note that the underlines imply that both the observations as well as the residuals are stochastic.

If we want to write this for an arbitrary number of observations and unknowns, we rely on matrix equations, as you see here.

So, we can confirm that such an expression is a linear function.

This would also hold for the estimator of x itself:

… which is also a linear function if we lump all matrix multiplications together.

This means that all these estimators, for x-hat, but also for for y-hat and for e-hat are linear estimators.

The third element in the acronym BLUE represents Unbiasedness as a property.

When an estimator is unbiased, this means the we expect that the value of the estimator is equal to the true (but unknown) value of the parameter.

In mathematics, we use the 'Expectation Operator' for this property.

You may remember that we used this linear model of observation equations.

It means that we expect that our observations are well represented by the linear model Ax.

This is an assumption.

Following this assumption, we can now verify whether the parameters we are estimating are 'unbiased'.

Let's take x-hat as an example.

If we now want to verify whether xhat is an unbiased estimator, we want to check whether the expectation of xhat is equal to the true (but unknown) value of x.

Thus, we use the expectation operator at both sides of the equation.

But since the expectation operator is only meaningful for stochastic parameters, we can move all non-stochastic parts of the equation out.

By inserting our starting assumption of a correct model, we find this expression:

If we reorder this a bit,

we see that this actually

is a multiplication of a matrix with its inverse matrix, which yields the unit matrix, and finally we see that the expectation of the estimator of x (which is xhat) is dentical to the unknown value of x itself.

In other words, the estimator is unbiased.

Now, the first letter of the acronym BLUE is actually the most important one.

It tells us that the estimator should be 'Best'.

The term 'Best' always needs to be defined specifically.

In the context of estimation theory, Best means that out of all possible estimators that one could dream of, we want the one for which the precision of the estimator has the lowest value.

Lets spend one more minute to grasp this.

In this generic expression of the weighted least squares estimator xhat, there is only one thing we can influence to change the value of xhat.

This is the weight matrix.

Therefore, if we introduce one million different weight matrices, we will get one million different values for xhat.

Now, the term 'Best' implies that out of these one million values, I would like to find the value that has the best precision.

If x would be a scalar, this means that the best precision is equal to the lowest variance of x, or the lowest standard deviation.

Let’s work this out in some more detail.

If this is again the Generic Weighted Least Squares Estimator of x, than we are interested in the covariance matrix of xhat.

This follows from error propagation, (which will be elaborated on in another module.)

As you can see here:

And if we fill in all the terms, we get a long expression, which we can simplify a bit to this expression

Note the symmetry of the expression.

Now, without going into too much detail, I would like to show that this equation can collapse to a much more elegant form.

For this to happen we should replace the weight matrix W with the inverse of the covariance matrix Qy

Let’s clean up what we have obtained now.

The center part of this equation can be cleaned up further, by removing the product of the covariance matrix with its inverse.

This yields a product of three identical square matrices, of which two are inverted.

We can therefore eliminate another product of a matrix with its inverse, and we find a short and elegant expression for Q-xhat.

Let's look at what we've got now.

We see that if we would use a specific weight matrix; the inverse of the covariance matrix of the observations; than the covariance matrix of xhat has an elegant expression.

Moreover, this expression is now only dependent of the A-matrix, and of the covariance matrix of the observations.

Even if we do not go further in depth on these results, it is intuitively clear that now there is no more need to find alternative weight matrices.

We simply use our knowledge on the quality of the observations to find the estimator of the parameters.

And there is no more need to information from outside.

It can be proven that this expression for the covariance matrix of the parameters is 'Best'.

Of all possible weight matrices, this particular one gives us the optimal estimator of xhat.

To finish this video, I would like to return to the easy sea-level rise example, and recap what the equations will look like if we use BLUE as our estimation procedure.

We formulate our y=Ax problem, using the expectation operator, and the corresponding stochastic model, expressing the dispersion of the observations

by its covariance matrix.

We then find the least-squares estimate x-hat.

The quality of the observations is captured in the covariance matrix of y, and it is indicated conceptually in the figure by the errorbars.

Essentially, we minimize the differences between our actual measurements (the red dots) and the linear model which we assumed to be applicable (which is the green dashed line).

We do this by giving some observations more weight than others, which follows from the inverse of the covariance matrix.

In the plot, the observations with a bigger errorbar represent observations with less precision.

The differences (or 'residuals') are indicated by the little vertical blue lines, and we call these the "least squares estimates of the residuals", or e-hat.

Mathematically, we obtain this solution by minimizing the "weighted sum of the squared residuals", which then yields the "Best Linear Unbiased estimate".

In this video, we elaborated on the meaning of the three components of BLUE: the 'linearity', the unbiasedness and the 'best' properties of the estimator.