**Constrained Least Squares – Coefficient Calculation and Simulation**

We want to maximize the likelihood equation of β. We know that Y in this context is normally distributed with mean Xβ and variance σ2. This means that the distribution of yi given X is as follows:

So the likelihood of β is therefore:

Taking the natural logarithm gives us the log likelihood:

Here, I have truncated the result to only those involving **β** and, as we are only interested in maximizing with respect to **β**. Note that I have transitioned to a constant, as we are assuming the errors are iid with constant variance . Here, we obtain the same result as minimizing the least squares estimator, so the MLE and the least squares estimators are the same. If we were unconstrained, we would simply take the partial derivative here with respect to **β** to estimate **β** using maximum likelihood.

Since is unconstrained, we can find the maximum likelihood estimator directly and use this to maximize with respect to **β.** To do this, we differentiate the above equation with respect to σ.Doing so yields the normal estimate for :

Where is our new constrained estimates. Plugging this back into the log likelihood equation yields:

However, we want to maximize the above equation with the constraint that: M**β** = d, where d is a known vector and M is an *r x p* matrix of rank *r < p.* This is an equality constraint. We can use the Lagrange multipliers to solve this maximization problem with such a constraint.

We can see that, since the natural logarithm is a monotonically increasing function, and the division within the above logarithm is a constant, and that all other terms do not involve B, maximizing the above equation with respect tois equivalent to minimizing:

Where is the Lagrange multiplier (and so f is the Lagrange function). The Lagrange multiplier is especially useful in this case, as we are trying to minimize the least square error with some constraint on the parameters.

Multiplying the error by the transpose gives:

Now we take derivatives with respect to  **and to obtain**

We can now set both of these equal to zero and solve this system of equations to find the appropriate MLE estimate of **.** However, there is a few simplifying steps which can make the result more useable. The first is to get the expression in terms of the original least squares estimator. If we multiply the first equation above by M(, we obtain:

The first term on the right is the familiar from unconstrained OLS (I will denote this ). Substituting, we get

Solving for lambda:

Plugging this back into the original equation and solving for **,**

This is the desired result, the maximum likelihood estimator given the constraint M.

Now, we return to the first part of the problem.

Is equivalent to the following constraint. **α** where **α** is a vector of zeros in all but the last row, which is a 1. So,

Plugging it into the derived result, we get:

Since in this case, we know that which is a scalar that is the bottom right element of the X’X matrix. So we can simplify this to:

Since is a constant:

So, if we compute the unconstrained and , we can solve directly for the constrained

Now, for the second part of part a), this is equivalent to M being a vector of 1s, lets call it ϕ. Then

ϕ = (1,1,1,1,…,1)’ and d = 1.

Plugging it into the derived result, we get:

The residual sum of squares is defined as:

In this case, **=** X**(**])

So it follows that =

where is the error in the unconstrained case. So

RSSc – RSS =

Each of the errors are normally distributed with mean 0 and variance sigma squared. This implies that:

, iid.

Where the estimator is the estimator from the restrained equation. We can write the sum of squares of the above as:

Whereis the estimator of the unconstrained model. To show this, we remove the identical identity matrices and see that:

Now,

which we know from the above is )’

Rewriting, we have:

Using the fact that **,** we can remove this to obtain.

Thus the original standardized normal variable can be represented as a sum of Q1, and Q2 where:

Q1 =

Q2 =

Q1 is the sum of squared errors for the unconstrained model, and therefore has rank . Q2 has rank *p – r ,* as it is the differences between the two estimators. Thus rank(Q1) + rank (Q2) = *n-r,* which is the rank of the errors of the constrained model, since r<p. Thus by Cochran’s Theorum, the two are independently distributed.

I have also produced simulations to prove that this is the case. I create three normal random variables, two of which are related to the others linearly and randomly. I compute the beta unconstrained, and then use this to compute the beta constrained. I then produce the errors for the constrained beta, and check the correlations in 100 samples of 1000 each. The mean correlation is very close to zero, which is the expected result. This is included in the R file.