GLS Explanation

Caleb Leedy

31 January 2024

Introduction

This is a write up explaining the dual form of generalized least squares (GLS). We think of GLS as an extension of ordinary least squares (OLS), so I will discuss this first and then comment on how these are connected and the dual formulation for both.

The Gauss-Markov Theorem

It all comes back to the Gauss-Markov Theorem, which states that for a linear model in the form,

$$Y = X\beta + \varepsilon$$

with $E[\varepsilon] = 0$ and $\text{Var}(\varepsilon) = \sigma^2 I_n$, the BLUE (best linear unbiased estimator) of $C\beta$ for $C = AX$ for some matrix A is (assuming $X'X$ has an inverse),

$$C\hat{\beta} = C(X'X)^{-1}X'Y = AX(X'X)^{-1}X'Y.$$

GLS Extension

To extend this to the GLS setting consider the case where $\text{Var}(\varepsilon) = \sigma^2 V$ for some known positive-definite matrix V. In this case, we can rewrite $Y = X\beta + \varepsilon$ as

$$Z = W\beta + \tilde{\varepsilon}$$

where $Z = V^{-1/2}Y$, $W = V^{-1/2}X$ and $\tilde{\varepsilon} = V^{-1/2}\varepsilon$. But this formulation implies that $E[\tilde{\varepsilon}] = 0$ and $\text{Var}(\tilde{\varepsilon}) = \sigma^2 I_n$ and so hence by the Gauss-Markov Theorem, the BLUE is

$$(W'W)^{-1}W'Z = (X'V^{-1}V)^{-1}X'V^{-1}Y.$$

Dual Form

All of this comes back to the fact that we want to estimate the BLUE. However, we could do this directly. By definition, the BLUE is the minimum variance linear unbiased estimator. This means that we consider a class of estimators $CW'Y$ such that $E[CW'Y] = C\beta$ that minimizes the variance $CW'\text{Var}(Y)WC' \propto CW'WC'$. Since we assume the model form of $Y = X\beta + \varepsilon$, the unbiasedness condition is equivalent to,

$$E[CW'Y] = CW'E[Y] = CW'X\beta = C\beta.$$

Since this is true for all β, it must be the case that

$$W'X = I_p \text{ for } p = \dim(\beta).$$

Hence, we have a Lagrangian

$$L(W) \propto CW'WC' - (W'X - I_p)\lambda.$$

Thus, the first-order conditions are $W - X\lambda = 0$ and $W'X - I_p = 0$. This yields a solution of $W = X(X'X)^{-1}$ which is exactly what we get from OLS because the final result is $CW'Y = C(X'X)^{-1}X'Y$. To get the GLS form, we can keep the same constraint except minimize the variance $W'VW$.

Estimating the Variance

When estimating GLS, we assume that the covariance matrix V is fully known. What do we do if we have to estimate V? In this case, we need to use (finite-dimensional) semiparametric theory to estimate the optimal W such that W belongs to the orthogonal nuisance tangent space of $\hat{V}$. This is something that we could explore further if we want to.