New York University
Wilf Family Department of Politics

Quantitative Research in Political Science I

Professor Patrick Egan

A Proof that OLS is the Best Linear Unbiased Estimator (BLUE) for β

In our work with the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, we required four assumptions to establish that the OLS estimator $\widehat{\boldsymbol{\beta}} \equiv \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$ is unbiased for $\boldsymbol{\beta}$:

1. The DGP can accurately be written as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where $\mathbf{X}$ is an $N \times K$ matrix;
2. No perfect multicollinearity among the x's, i.e. that $rank\left(\mathbf{X}\right) = K$;
3. We have a random sample, making our observations i.i.d.;
4. $E\left(u|\mathbf{X}\right) = 0$ [no omitted variables correlated with both $\mathbf{X}$ and y]; implies fixed $\mathbf{X}$.

Together, these assumptions established that $E\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$. Unbiasedness is of course a very nice property for an estimator to have. But another question that arises when we consider the desirability of an estimator is its precision. In particular, we would like an estimator that is as precise as possible. In fact, when examining estimators that are otherwise similar, the estimator whose sampling distribution has the smallest variance is considered, intuitively, to be the "best" of these estimators.

Recall that we invoked an additional assumption in order to fully specify the sampling distribution of $\widehat{\boldsymbol{\beta}}$ and thus establish that $VAR\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$:

5. This assumption has two parts, together which are known as "sphericality of errors":
   (a) No autocorrelation: $cov\left(u_i, u_j\right) = 0 \forall i \neq j$;
   (b) Homoskedasticity: $var\left(u_1\right) = var\left(u_2\right) = ... = var\left(u_N\right) = \sigma^2$.

With Assumptions 1 through 5 in place, it can be shown that the OLS estimator $\widehat{\boldsymbol{\beta}} \equiv \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$ is not only an unbiased linear estimator of $\widehat{\boldsymbol{\beta}}$; it's the *best* linear unbiased estimator (BLUE). Together, Assumptions 1-5 are known as the *Gauss-Markov Assumptions*. The **Gauss-Markov Theorem** demonstrates that OLS is (the) BLUE. A proof follows:

Consider some alternate linear unbiased estimator for $\boldsymbol{\beta}$, which we will write as $\widetilde{\boldsymbol{\beta}}$. By "unbiased," we of course mean $E\left(\widetilde{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$. By "linear," we mean that $\widetilde{\boldsymbol{\beta}}$ can be written as a linear function of $\mathbf{y}$ and some set of weights, $\mathbf{C}$:

$$\widetilde{\boldsymbol{\beta}} \equiv \mathbf{C}\mathbf{y}.$$

Specifically, let's rewrite $\mathbf{C}$ as our OLS estimator plus some $K \times N$ matrix $\mathbf{D}$:

$$\begin{aligned} \mathbf{C} &\equiv \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' + \mathbf{D} \text{ and so} \\ \widetilde{\boldsymbol{\beta}} &= \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' + \mathbf{D}\right]\mathbf{y} \\ \widetilde{\boldsymbol{\beta}} &= \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' + \mathbf{D}\right]\mathbf{X}\boldsymbol{\beta} + \mathbf{u}. \end{aligned}$$

Goal is to show that no $\widetilde{\boldsymbol{\beta}}$ exists such that $var\left(\widetilde{\beta}_k\right) < var\left(\widehat{\beta}_k\right)$ for any k. First let's consider $E\left(\widetilde{\boldsymbol{\beta}}\right)$:

$$\begin{aligned} E\left(\widetilde{\boldsymbol{\beta}}\right) &= E\left\{\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' + \mathbf{D}\right]\mathbf{X}\boldsymbol{\beta} + \mathbf{u}\right\} \\ &= E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right] + E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}\right] + E\left[\mathbf{D}\mathbf{X}\boldsymbol{\beta}\right] + E\left[\mathbf{D}\mathbf{u}\right] \\ &= E\left(\boldsymbol{\beta}\right) + \mathbf{0} + E\left[\mathbf{D}\mathbf{X}\boldsymbol{\beta}\right] + \mathbf{D}E\left[\mathbf{u}\right] \\ &= \boldsymbol{\beta} + \mathbf{D}\mathbf{X}\boldsymbol{\beta} \text{ (Treating } \mathbf{D}, \mathbf{X} \text{ as fixed; } E\left[\mathbf{u}\right] = \mathbf{0}) \\ &= \left(\mathbf{I} + \mathbf{D}\mathbf{X}\right)\boldsymbol{\beta}. \end{aligned}$$

For $\widetilde{\boldsymbol{\beta}}$ to be unbiased, it must be that $E\left(\widetilde{\boldsymbol{\beta}}\right) = \left(\mathbf{I} + \mathbf{D}\mathbf{X}\right)\boldsymbol{\beta} = \boldsymbol{\beta}$. Thus $\mathbf{D}\mathbf{X}$ must be equal to zero. Let's use this to rewrite $\widetilde{\boldsymbol{\beta}}$. From above we have

$$\begin{aligned} \widetilde{\boldsymbol{\beta}} &= \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' + \mathbf{D}\right]\mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ &= \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u} + \mathbf{D}\mathbf{u}. \end{aligned}$$

Now let's consider $var\left(\widetilde{\boldsymbol{\beta}}\right)$:

$$\begin{aligned} var\left(\widetilde{\boldsymbol{\beta}}\right) &\equiv E\left\{\left[\widetilde{\boldsymbol{\beta}} - E\left(\widetilde{\boldsymbol{\beta}}\right)\right]\left[\widetilde{\boldsymbol{\beta}} - E\left(\widetilde{\boldsymbol{\beta}}\right)\right]'\right\} \\ &= E\left[\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\right] \text{ (since } \widetilde{\boldsymbol{\beta}} \text{ is unbiased for } \boldsymbol{\beta}) \\ &= E\left\{\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u} + \mathbf{D}\mathbf{u}\right]\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u} + \mathbf{D}\mathbf{u}\right]'\right\} \text{ (the } \boldsymbol{\beta}\text{'s drop out)} \\ &= E\left\{\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' + \mathbf{D}\right]\mathbf{u}\mathbf{u}'\left[\mathbf{D}' + \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right]\right\} \text{ (factor out } \mathbf{u}\text{'s, take the transpose)} \end{aligned}$$

Relying on the assumption of sphericality, this simplifies to

$$\begin{aligned} var\left(\widetilde{\boldsymbol{\beta}}\right) &= \sigma^2\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' + \mathbf{D}\right]\left[\mathbf{D}' + \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right] \\ &= \sigma^2\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{D}' + \mathbf{D}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} + \left(\mathbf{X}'\mathbf{X}\right)^{-1} + \mathbf{D}\mathbf{D}'\right] \\ &= \sigma^2\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} + \mathbf{D}\mathbf{D}'\right] \quad (\mathbf{D}\mathbf{X} = \mathbf{0} = \mathbf{X}'\mathbf{D}') \\ &= \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} + \sigma^2\mathbf{D}\mathbf{D}' \end{aligned}$$

The first term in this expression for $var\left(\widetilde{\boldsymbol{\beta}}\right)$ should look familiar. It's $var\left(\widehat{\boldsymbol{\beta}}\right)$! So for $var\left(\widetilde{\boldsymbol{\beta}}\right) < var\left(\widehat{\boldsymbol{\beta}}\right)$, the diagonal elements of $\mathbf{D}\mathbf{D}'$ must be negative. But multiplying a matrix by its transpose yields a matrix with sums of squares on its diagonal, which must be non-negative. Thus it must be that $var\left(\widetilde{\boldsymbol{\beta}}\right) \geq var\left(\widehat{\boldsymbol{\beta}}\right)$, and so no unbiased linear estimator for $\boldsymbol{\beta}$ exists that is more precise (a.k.a. is "better") than $\widehat{\boldsymbol{\beta}}$. So $\widehat{\boldsymbol{\beta}}$ is (the) BLUE for $\boldsymbol{\beta}$. $\square$.