

4 Gauss-Markov Theorem. Estimation of σ^2 . Distribution of OLS in normal regression

4.1 Gauss-Markov Theorem

Theorem 4.1.1. Consider an $n \times 1$ random vector Y and an $n \times k$ random matrix X .

Assume (no need for iid, large n or normality):

- **GM1** No perfect multicollinearity: $\text{rank}(X) = k$
- **GM2** Strict Exogeneity $E(Y|X) = X\beta$, equivalently $E(\varepsilon|X) = 0$
- **GM3** Homoskedasticity and no serial correlation $\text{Var}(Y|X) = \sigma^2 I$,
equivalently $\text{Var}(\varepsilon|X) = \sigma^2 I$

Then, the OLS estimator $\hat{\beta}_{OLS}$ has the minimum conditional variance in the class of estimators that, conditional on every X , are linear in Y and unbiased. Thus $\hat{\beta}_{OLS}$ is the Best Linear conditionally Unbiased Estimator (BLUE).

A linear estimator of β is any estimator of the form $\tilde{\beta} = A(X)Y$ where $A(X)$ is a $k \times n$ matrix. For OLS $\hat{\beta}_{OLS} = A(X)Y = (X'X)^{-1}X'Y$

Definition 4.1.1

Minimum conditional variance implies:

$$\text{Var}(\tilde{\beta}|X) - \text{Var}(\hat{\beta}_{OLS}|X) \text{ is positive semi-definite } \forall \tilde{\beta}$$

This $k \times k$ matrix A is positive semi-definite iff $z'Az \geq 0$ for all $k \times 1$ vectors z . Thus for any z :

$$z'\text{Var}(\tilde{\beta}|X)z \geq z'\text{Var}(\hat{\beta}_{OLS}|X)z$$

Note this is equivalent to:

$$\text{Var}(z'\tilde{\beta}|X) \geq \text{Var}(z'\hat{\beta}_{OLS}|X)$$

Thus any linear combination of the elements of $\tilde{\beta}$ has a conditional variance that is at least as large as the conditional variance of the corresponding linear combination of the elements of $\hat{\beta}_{OLS}$.

In particular, any component of $\tilde{\beta}$ has a conditional variance that is at least as large as the conditional variance of the corresponding component of $\hat{\beta}_{OLS}$.

4.2 GM PROOF

Lemma 4.2.1. We know

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$$

is conditionally (and unconditionally) unbiased under GM2, and is a linear function of Y.

We also know under GM3 that

$$Var(\hat{\beta}_{OLS}|X) = \sigma^2(X'X)^{-1}$$

Now consider any other linear conditionally unbiased estimator $\tilde{\beta} = A(X)Y$.

$$E(\tilde{\beta}|X) = E(AY|X) = AE(Y|X) = AX\beta \quad \text{under GM3}$$

As we assume conditionally unbiased, for any β

$$AX\beta = \beta \Rightarrow AX = I$$

Lemma 4.2.2.

$$Var(\tilde{\beta}|X) = Var(AY|X) = AVar(Y|X)A' = A\sigma^2I_nA' = \sigma^2AA'$$

Decomposing A:

$$A = A - (X'X)^{-1}X' + (X'X)^{-1}X' = W + (X'X)^{-1}X'$$

Thus:

$$\begin{aligned} Var(\tilde{\beta}|X) &= \sigma^2(W + (X'X)^{-1}X')(W + (X'X)^{-1}X')' \\ &= \sigma^2(W + (X'X)^{-1}X')(W' + X(X'X)^{-1}) \end{aligned}$$

But

$$WX = AX - (X'X)^{-1}X'X = I - I = 0$$

Therefore,

$$\begin{aligned} Var(\tilde{\beta}|X) &= \sigma^2WW' + \sigma^2(X'X)^{-1} \\ &= Var(\hat{\beta}_{OLS}|X) + \sigma^2WW' \end{aligned}$$

Lemma 4.2.3. σ^2WW' is positive semi-definite

For any k-dimensional vector z , denote the k-dimensional vector $W'z$ as $\alpha = (\alpha_1, \dots, \alpha_k)'$

$$z'\sigma^2WW'z = \sigma^2(z'W)(W'z) = \sigma^2\alpha'\alpha = \sigma^2\sum_{i=1}^k \alpha_i^2 \geq 0$$

Thus $Var(\tilde{\beta}|X) - Var(\hat{\beta}_{OLS}|X)$ is psd for any linear conditionally unbiased estimator $\tilde{\beta} = AY$.

□

4.3 Estimation of σ^2

Given the following but how to estimate?:

$$\text{Var}(\hat{\beta}_{OLS}|X) = \sigma^2(X'X)^{-1}$$

We are given X , thus only need to estimate σ^2 .

Note: $\sigma^2 = E(\varepsilon_i^2|X)$ and since trivially $\sigma^2 = E(\sigma^2)$ we have:

$$\sigma^2 = E(E(\varepsilon_i^2|X)) = E(\varepsilon_i^2)$$

This suggests the MOM estimator:

$$\hat{\sigma}_{MOM}^2 = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2$$

except we don't know ε_i as β is unknown.

But with $\hat{\beta} = \hat{\beta}_{OLS}$ we can use the sample analogue $\hat{\varepsilon}_i = y_i - x_i' \hat{\beta}_{OLS}$ and thus:

Theorem 4.3.1. The biased (ML) estimator of σ^2 is:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 \left(= \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i' \hat{\varepsilon}_i \right)$$

4.3.1 Unbiased Estimator of σ^2

We first compute the bias of $E(\hat{\sigma}^2|X)$ to then correct for it:

Lemma 4.3.1.

$$\hat{\varepsilon} = M_X Y = M_X (X\beta + \varepsilon) = M_X \varepsilon$$

Then

$$\begin{aligned} E(\hat{\sigma}^2|X) &= E(\hat{\varepsilon}' \hat{\varepsilon}|X) \\ &= E(\varepsilon' M_X' M_X \varepsilon|X) = E(\varepsilon' M_X \varepsilon|X) = E(\text{tr}(\varepsilon' M_X \varepsilon)|X), \quad \text{since argument is a scalar} \\ &\quad \because \text{trace multiplications are commutative if conformations exists} \Rightarrow \\ &= E(\text{tr}(M_X \varepsilon \varepsilon')|X) = \text{tr}(E(\varepsilon' M_X \varepsilon|X)) = \text{tr}(M_X E(\varepsilon \varepsilon'|X)) = \text{tr}(M_X \sigma^2 I_n) = \sigma^2 \text{tr}(M_X) \end{aligned}$$

Lemma 4.3.2.

$$\text{tr}(M_X) = n - k$$

Proof.

$$\begin{aligned} M_X &= (I - X(X'X)^{-1}X') \\ \text{tr}(M_X) &= \text{tr}(I_n - X(X'X)^{-1}X') = \text{tr}(I_n) - \text{tr}((X'X)^{-1}X'X) \\ &= \text{tr}(I_n) - \text{tr}(I_k) \\ &= n - k \end{aligned}$$

□

Thus

$$E(\hat{\sigma}^2|X) = \frac{n - k}{n} \sigma^2$$

Theorem 4.3.2.

$$\hat{\sigma}_u^2 = \frac{n}{n-k} \hat{\sigma}_{ML}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k}$$

$\hat{\sigma}_u^2$ is an unbiased estimator of σ^2 $\hat{\sigma}_{ML}^2$ is an unbiased estimator of σ^2

4.4 Estimation of standard errors

By default STATA computes s.e. of component $\hat{\beta}_j$ of $\hat{\beta}_{OLS}$ as the square roots from the i-th diagonal element of $\hat{\sigma}^2(X'X)^{-1}$ or more explicitly with the partitioned regression formulae:

$$se(\hat{\beta}_i) = \frac{\hat{\sigma}_u}{\sqrt{X_i' M_{-i} X_i}}$$

where X_i is the i-th column of X (i-th regressor) and M_{-i} is the residual maker matrix in the regression on all the other explanatory variables but X_i .

4.5 Distribution of OLS

Knowing the mean and variance of OLS is not sufficient to test hypotheses about β . We need to also know the distribution. In small samples this is easy to derive with the following assumption:

Normal Regression : $\varepsilon|X \sim N(0, \sigma^2 I_n)$

This subsumes GM2 and GM3 and adds normality.

Claim 4.5.1. There are several properties of the multivariate Gaussian which become useful in derivations.

- If $\mathbf{Z} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ and \mathbf{A} be any deterministic $r \times n$ matrix, then $\mathbf{AZ} \sim N(\mathbf{0}, \sigma^2 \mathbf{AA}')$. In particular any linear combinations of normals is normal.
- The Normal distribution $N(0, \sigma^2 I_n)$ is invariant to rotations/orthogonal transformations. If $\mathbf{Z} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ and \mathbf{Q} is any $n \times n$ orthogonal matrix, then $\mathbf{QZ} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, i.e. \mathbf{QZ} has the same distribution as \mathbf{Z} .

Theorem 4.5.1. Using the first property and the normal regression assumption, we obtain:

$$\hat{\beta}_{OLS}|X = \beta + (X'X)^{-1}X'\varepsilon|X \sim N(\beta, \sigma^2(X'X)^{-1})$$