Econometrics 1, Class Week 2

Back-up Video: Class week 2 video (click here)

Learning Outcomes

(a) Gauss Markov Theorem: Proof.

(b) Orthogonal projectors.

(c) Unbiasedness of the Ordinary Least Squares (OLS) estimator for the variance parameter σ_0^2.

(d) Generalized Least Squares (GLS) estimator of β_0.

Prerequisites

1. Concepts in Linear Algebra:
   - rank of a matrix (p.37 of Auxiliary Math Notes [AMN]);
   - bilinearity of variance-covariance matrices (AMN p.105);
   - positive semi-definiteness of square matrices (AMN p.45);
   - trace of a square matrix (AMN p.43).
2. Concepts in Mathematical Statistics:
   - mean vector, variance-covariance matrix of vector-valued random variables (AMN p.105).

(a) Setting of Gauss Markov Theorem:

T observation collected in outcomes vector $\mathbf{y}$ and matrix of k regressors $\mathbf{X}$, $T > k$, related by *linear* model

$$\underset{T\times 1}{\mathbf{y}} = \underset{T\times k}{\mathbf{X}}\underset{k\times 1}{\beta_0} + \underset{T\times 1}{\mathbf{u}} \tag{1}$$

where $\mathbf{u}$ is a vector or regression errors.

Interpretation: Want to explain the variation in $\mathbf{y}$ through (its co-)variation with the columns of $\mathbf{X}$, assumed to be non-stochastic and exogenous.

This covariation is determined by the true, but unknown population parameter β_0 (a k-vector of population constants) which we seek to estimate.

This parameter enters the relationship (1) linearly.

Gauss - Markov Assumptions

Assumptions about population errors $\mathbf{u}$:

A1 mean zero errors: $\mathbb{E}[\mathbf{u}] = \underset{T\times 1}{\mathbf{0}}$.

A2 homoskedastic errors: $\mathbb{E}\underset{T\times T}{[\mathbf{u}\mathbf{u}']} = \sigma_0^2 \mathbf{I}_T$, where $\mathbf{I}_T$ is the identity matrix of dimension $T \times T$ and $\sigma_0^2 > 0$ is the true, but unknown scalar variance parameter in the population.

I.e. each element u_t of the vector $\mathbf{u}$ has variance $\mathbb{E}[u_t^2] = \sigma_0^2$, and there are no covariances, $\mathbb{E}[u_t u_s] = 0$ whenever $t \neq s$.

A3 Linearity of (1) in β_0.

OLS estimator of β_0

$\hat{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'y}$, *linear* in $\mathbf{y}$ which is premultiplied by $k \times T$ matrix $(\underset{k\times T}{\mathbf{X}'}\underset{T\times k}{\mathbf{X}})^{-1}\underset{k\times T}{\mathbf{X}'}$.

Note: Need additional assumption that

A4 rank of $\mathbf{X}$ is $\text{rk}(\mathbf{X}) = k$; otherwise, inverse does not exist.

Gauss-Markov Theorem: Under the assumptions A1 - A4 about the population, $\hat{\beta}$ is the best linear and unbiased estimator (BLUE) of β_0.

Proof of Linearity: immediate, because $\hat{\beta}$ is linear in $\mathbf{y}$.

Proof of Unbiasedness:

Unbiasedness means: $\mathbb{E}[\hat{\beta}] = \beta_0$.

To verify,

$$
\begin{aligned}
\mathbb{E}[\hat{\beta}] &= \mathbb{E}[(\mathbf{X'X})^{-1}\mathbf{X'y}] \\
&= (\mathbf{X'X})^{-1}\mathbf{X'}\mathbb{E}[\mathbf{y}] \text{ because } \mathbf{X} \text{ is non-stochastic} \\
&= (\mathbf{X'X})^{-1}\mathbf{X'}\mathbb{E}[\mathbf{X}'\beta_0 + \mathbf{u}] \\
&= \beta_0 + (\mathbf{X'X})^{-1}\mathbf{X'}\mathbb{E}[\mathbf{u}] \\
&= \beta_0 \text{ by A1.}
\end{aligned}
$$

Proof of Best (within Class of Linear Unbiased Estimators):

Suppose there exists another linear, unbiased estimator $\tilde{\beta}$ of β_0.

Strategy: Will show its variance-covariance matrix is larger (in a positive semi-definite sense) than that of $\hat{\beta}$.

Linearity of $\tilde{\beta}$: $\underset{k\times 1}{\tilde{\beta}} = \underset{k\times T}{\mathbf{A}}\underset{T\times 1}{\mathbf{y}}$, for some $k \times T$ matrix $\mathbf{A}$.

Unbiasedness of $\tilde{\beta}$: $\mathbb{E}[\tilde{\beta}] = \beta_0$, i.e.

$$\begin{aligned}
\beta_0 &= \mathbb{E}[\tilde{\beta}] \\
&= \mathbb{E}[\mathbf{A}\mathbf{y}] \\
&= \mathbb{E}[\mathbf{A}(\mathbf{X}\beta_0 + \mathbf{u})] \\
&= \mathbf{A}\mathbf{X}\beta_0 \text{ by A1,}
\end{aligned}$$

for *any* conceivable β_0. Hence,

$$\mathbf{A}\mathbf{X} = \mathbf{I}_k. \tag{2}$$

So linearity and unbiasedness place this restriction on $\mathbf{A}$ in definition of $\tilde{\beta}$.

Proof of Best (continued)

$$
\begin{aligned}
\underset{k\times k}{\text{var}(\tilde{\beta})} &= \mathbb{E}[\underset{k\times 1}{(\tilde{\beta}-\beta_0)}\underset{1\times k}{(\tilde{\beta}-\beta_0)'}] \\
&= \text{var}((\tilde{\beta}-\hat{\beta})+\hat{\beta}) \\
&= \text{var}(\hat{\beta})+\text{var}(\tilde{\beta}-\hat{\beta})+\cdots \\
&\quad \cdots \text{cov}(\hat{\beta},\tilde{\beta}-\hat{\beta})+\text{cov}(\tilde{\beta}-\hat{\beta},\hat{\beta}) \\
\text{cov}(\tilde{\beta}-\hat{\beta},\hat{\beta}) &= \text{cov}(\mathbf{Ay}-(\mathbf{X'X})^{-1}\mathbf{X'y},(\mathbf{X'X})^{-1}\mathbf{X'y}) \\
&= (\mathbf{A}-(\mathbf{X'X})^{-1}\mathbf{X}')\text{cov}(\mathbf{y},\mathbf{y})\mathbf{X}(\mathbf{X'X})^{-1} \\
&\quad \text{(see p. 105 of auxiliary math notes)} \\
&= (\mathbf{A}-(\mathbf{X'X})^{-1}\mathbf{X}')\sigma_0^2\mathbf{I}_T\mathbf{X}(\mathbf{X'X})^{-1} \text{ by A2} \\
&= \sigma_0^2(\mathbf{A}-(\mathbf{X'X})^{-1}\mathbf{X}')\mathbf{X}(\mathbf{X'X})^{-1} \\
&= \sigma_0^2(\mathbf{AX}-\mathbf{I}_T)(\mathbf{X'X})^{-1} \\
&= \mathbf{0}.
\end{aligned}
$$

where the last equality follows from (2).

Therefore,

$$
\begin{aligned}
\text{var}(\tilde{\beta}) &= \text{var}(\hat{\beta})+\text{var}(\tilde{\beta}-\hat{\beta}) \\
\text{var}(\tilde{\beta}) &\geq \text{var}(\hat{\beta}),
\end{aligned}
$$

because variance-covariance matrices are positive semi-definite.

This completes the proof of the Gauss-Markov Theorem.

If $\mathbf{X}$ were stochastic, all conclusions hold *conditional on* $\mathbf{X}$. This is the typical setting in econometrics.

(b) Orthogonal Projectors $P_X = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X}'$ and $M = \mathbf{I}_T - P_X$.

(i) Idempotency of M:

$$\begin{aligned} MM &= (\mathbf{I}_T + P_X)(P_X - 2P_X) \\ &= \mathbf{I}_T + P_X - 2P_X \text{ by idempotency of } P_X \\ &= \mathbf{I}_T - P_X = M. \end{aligned}$$

(ii) Orthogonality of M and P_X:

$$\begin{aligned} MP_X &= (\mathbf{I}_T - P_X)P_X \\ &= P_X - P_X P_X \\ &= P_X - P_X \text{ by idempotency of } P_X \\ &= \mathbf{0}. \end{aligned}$$

(iii) Regression residuals $\hat{\mathbf{u}}$:

$$\begin{aligned} \hat{\mathbf{u}} &= \mathbf{y} - \mathbf{X}\hat{\beta} \\ &= (\mathbf{I}_T - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X}')\mathbf{y} \\ &= M(\mathbf{X}\beta_0 + \mathbf{u}) \\ &= M\mathbf{u} \quad \text{because } M\mathbf{X} = \mathbf{X} - \mathbf{X} = \mathbf{0}. \end{aligned}$$

(c) Unbiasedness of OLS estimator $s^2 = \frac{1}{T-k}\hat{\mathbf{u}}'\hat{\mathbf{u}}$ of σ_0^2

$$
\begin{aligned}
\mathbb{E}[s^2] &= \frac{1}{T-k}\mathbb{E}[\hat{\mathbf{u}}'\hat{\mathbf{u}}] \\
&= \frac{1}{T-k}\mathbb{E}[\mathbf{u}'M\mathbf{u}] \text{ by (i) and (iii)} \\
&= \frac{1}{T-k}\mathrm{tr}[\mathbb{E}[\mathbf{u}'M\mathbf{u}]] \text{ b/c trace of scalar is scalar} \\
&= \frac{1}{T-k}\mathbb{E}[\mathrm{tr}[\mathbf{u}'M\mathbf{u}]] \text{ b/c lin. operators commute} \\
&= \frac{1}{T-k}\mathbb{E}[\mathrm{tr}[\mathbf{u}\mathbf{u}'M]] \text{ by properties of trace} \\
&= \frac{1}{T-k}\mathrm{tr}[\mathbb{E}[\mathbf{u}\mathbf{u}']M] \text{ b/c lin. operators commute} \\
&\quad \text{and } M = \mathbf{I}_T - P_X \text{ is non-stochastic} \\
&= \frac{1}{T-k}\mathrm{tr}[\sigma_0^2 M] \text{ by A2} \\
&= \frac{\sigma_0^2}{T-k}(\mathrm{tr}[\mathbf{I}_T] - \mathrm{tr}[P_X]) \\
&= \frac{\sigma_0^2}{T-k}(\mathrm{tr}[\mathbf{I}_T] - \mathrm{tr}[(\mathbf{X'X})^{-1}\mathbf{X'X}]) \\
&= \frac{\sigma_0^2}{T-k}(\mathrm{tr}[\mathbf{I}_T] - \mathrm{tr}[\mathbf{I}_k]) \\
&= \sigma_0^2.
\end{aligned}
$$

(d) Generalized Least Squares (GLS)

Change A2 to $\mathbb{E}[\mathbf{uu}'] = \sigma_0^2\Omega$ (i.e. errors are no longer homoskedastic).

What is variance of OLS estimator?

$$\begin{aligned}
\mathrm{var}(\hat{\beta}) &= \mathrm{var}(\beta_0 + (\mathbf{X'X})^{-1}\mathbf{X'u}) \\
&= (\mathbf{X'X})^{-1}\mathbf{X'}\mathrm{var}(\mathbf{u})\mathbf{X}(\mathbf{X'X})^{-1} \\
&= \sigma_0^2(\mathbf{X'X})^{-1}\mathbf{X'}\Omega\mathbf{X}(\mathbf{X'X})^{-1}.
\end{aligned}$$

GLS estimator $\tilde{\beta}$ solves

$$\mathbf{0} = \mathbf{X}'\Omega^{-1}(\mathbf{y} - \mathbf{X}\tilde{\beta}), \tag{3}$$

so

$$\tilde{\beta} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{y}. \tag{4}$$