SEEMINGLY UNRELATED REGRESSIONS

$$Y_1 = X_1 \beta_1 + \epsilon_1$$
$$Y_2 = X_2 \beta_2 + \epsilon_2$$
$$\vdots$$
$$Y_G = X_G \beta_G + \epsilon_G$$

where Y_i is $t_i \times 1$ vector of obs. on Dep. Var.
X_i is $t_i \times k_i$ matrix of obs. on Regressors
β_i is $k_i \times 1$ vector of parameters
ϵ_i is $t_i \times 1$ vector of disturbances

* Typically $t_i = T \;\; \forall \;\; i = 1, \ldots, G$.

For the derivation of the efficient estimator when sample sizes are unequal, see Schmidt, JEC, 1977.

For example,
G production functions for G different industries

Whether or not they are actually related depends upon the behavior of the ϵ_i ;
ϵ_i & ϵ_j might be correlated.
e.g. a strike in the coal mining industry (a random event) probably affects other industries.

If $\exists$ any such correlation,
then G individual OLS regressions
will <u>not</u> be <u>efficient</u>;
$\exists$ unused information in the data.

How to Model ?
- must make assumptions.

<u>Assumption 1</u>: Each equation <u>separately</u> satisfies the ideal conditions.

Thus, (1) $\epsilon_i \sim N(0, \sigma_{ii} I_T)$

(2) X_i is nonstochastic & of Rank k_i
with $\lim_{T \to \infty} \left(\frac{X_i' X_i}{T} \right)$ finite and nonsingular,
$\forall \; i = 1, \ldots, G$.

<u>Assumption 2</u>: The disturbances are correlated accross equations in the same time period (contemporaneously), but not in different time periods.

i.e. $\text{Cov}(\epsilon_i, \epsilon_j) = E[\epsilon_i \epsilon_j'] = \sigma_{ij} I_T$

$\forall \; i, j = 1, \ldots, G$.

$$\epsilon_1 = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1T} \end{bmatrix} \qquad \epsilon_2 = \begin{bmatrix} \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2T} \end{bmatrix} \qquad \cdots \qquad \epsilon_G = \begin{bmatrix} \epsilon_{G1} \\ \epsilon_{G2} \\ \vdots \\ \epsilon_{GT} \end{bmatrix}$$

$$E[\epsilon_i \epsilon_j'] = E\left\{ \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{iT} \end{bmatrix} \begin{bmatrix} \epsilon_{j1} & \epsilon_{j2} & \cdots & \epsilon_{jT} \end{bmatrix} \right\}$$

$$= E \begin{bmatrix} \epsilon_{i1}\epsilon_{j1} & \epsilon_{i1}\epsilon_{j2} & \cdots & \epsilon_{i1}\epsilon_{jT} \\ \epsilon_{i2}\epsilon_{j1} & \epsilon_{i2}\epsilon_{j2} & \cdots & \epsilon_{i2}\epsilon_{jT} \\ & \vdots & & \\ \epsilon_{iT}\epsilon_{j1} & \epsilon_{iT}\epsilon_{j2} & \cdots & \epsilon_{iT}\epsilon_{jT} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{ij} & 0 & \cdots & 0 \\ 0 & \sigma_{ij} & \cdots & 0 \\ & \vdots & & \\ 0 & 0 & \cdots & \sigma_{ij} \end{bmatrix}$$

$$= \sigma_{ij} I_T$$

Equivalently, $E(\epsilon_{it}\epsilon_{js}) = \begin{cases} \sigma_{ij} & \text{for } t = s \\ 0 & \text{for } t \neq s \end{cases}$.

Equivalently, the vectors, $\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_G \end{bmatrix}$, are iid $N[0, \Sigma]$

where $$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1G} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2G} \\ \vdots & & & \\ \sigma_{G1} & \sigma_{G2} & \cdots & \sigma_{GG} \end{bmatrix}$$

Saying that the vectors, $\begin{bmatrix}\epsilon_1\\ \epsilon_2\\ \vdots\\ \epsilon_G\end{bmatrix}$, are iid $N[0,\Sigma]$

is like saying that the stacked vector,

$$\begin{bmatrix}\epsilon_1\\ \epsilon_2\\ \vdots\\ \epsilon_G\end{bmatrix} \sim N\left\{\begin{bmatrix}0\\ 0\\ \vdots\\ 0\end{bmatrix}, \begin{bmatrix}\sigma_{11}I_T & \sigma_{12}I_T & \cdots & \sigma_{1G}I_T\\ \sigma_{21}I_T & \sigma_{22}I_T & \cdots & \sigma_{2G}I_T\\ \vdots & & & \\ \sigma_{G1}I_T & \sigma_{G2}I_T & \cdots & \sigma_{GG}I_T\end{bmatrix}\right\}$$

$GT \times 1$ $\qquad$ $GT \times 1$ $\qquad$ $GT \times GT$

or $\sim N[0, \Sigma \otimes I_T]$

or $\sim N[0, \Omega]$.

Hence the stacked vector of disturbances, ϵ^*, has the covariance matrix,

$$Cov(\epsilon^*) = E[\epsilon^* \epsilon^{*\prime}] = \Sigma \otimes I_T = \Omega .$$

This represents a very specific assumption about the behavior of the disturbances.

DIGRESSION on Kronecker Products

let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$, and $B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ \vdots & & & \\ b_{p1} & b_{p2} & \cdots & b_{pq} \end{bmatrix}$

$(m \times n)$ $(p \times q)$

Then the Kronecker Product of A times B is

"A Kronecker B";

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ \vdots & & & \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}$$

$$(mp \times nq)$$

Facts:

(1) $(A \otimes B)' = A' \otimes B'$

(2) $[A \otimes B][C \otimes D] = AC \otimes BD$

(3) If A & B are nonsingular,
$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

Note: $\quad \mathrm{Cov}(\epsilon^*) = \Sigma \otimes I_T = \Omega$

$$\Longrightarrow \quad \Omega^{-1} = \Sigma^{-1} \otimes I_T \qquad \text{by Kronecker Fact (3).}$$

later

The Joint Equation

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_G \end{bmatrix} = \begin{bmatrix} X_1 & & & 0 \\ & X_2 & & \\ & & \ddots & \\ 0 & & & X_G \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_G \end{bmatrix}$$

$$(GT \times 1) \qquad \Big(GT \times \sum_{i=1}^{G} k_i\Big) \quad \Big(\sum_{i=1}^{G} k_i \times 1\Big) \qquad (GT \times 1)$$

or: $\quad Y^* = \underset{\Uparrow}{X^*} \beta^* + \epsilon^*$

(Block Diagonal)

Proposition: OLS applied to the joint equation is the same as OLS applied to each equation separately.

Proof: $\hat{\beta}^* = (X^{*\prime} X^*)^{-1} X^{*\prime} Y^*$

$$= \left\{ \begin{bmatrix} X_1' & & & 0 \\ & X_2' & & \\ & & \ddots & \\ 0 & & & X_G' \end{bmatrix} \begin{bmatrix} X_1 & & & 0 \\ & X_2 & & \\ & & \ddots & \\ 0 & & & X_G \end{bmatrix} \right\}^{-1} \begin{bmatrix} X_1' & & & 0 \\ & X_2' & & \\ & & \ddots & \\ 0 & & & X_G' \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_G \end{bmatrix}$$

$$= \begin{bmatrix} X_1'X_1 & & & 0 \\ & X_2'X_2 & & \\ & & \ddots & \\ 0 & & & X_G'X_G \end{bmatrix}^{-1} \begin{bmatrix} X_1'Y_1 \\ X_2'Y_2 \\ \vdots \\ X_G'Y_G \end{bmatrix}$$

Fact (3)
$[A \otimes I]^{-1} = A^{-1} \otimes I$

$$= \begin{bmatrix} (X_1'X_1)^{-1} & & & 0 \\ & (X_2'X_2)^{-1} & & \\ & & \ddots & \\ 0 & & & (X_G'X_G)^{-1} \end{bmatrix} \begin{bmatrix} X_1'Y_1 \\ X_2'Y_2 \\ \vdots \\ X_G'Y_G \end{bmatrix}$$

$$= \begin{bmatrix} (X_1'X_1)^{-1} X_1'Y_1 \\ (X_2'X_2)^{-1} X_2'Y_2 \\ \vdots \\ (X_G'X_G)^{-1} X_G'Y_G \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_G \end{bmatrix}$$

QED

Viewed as a system, OLS estimates are unbiased, consistent, ~~and appropriate tests work~~
BUT, viewed as a system, $Cov(\epsilon^*) \neq \sigma^2 I_{GT}$;
the ideal conditions are not satisfied;
Should use GLS to obtain more efficient estimate

<u>Theorem:</u> The efficient estimator of β^* is

$$\tilde{\beta}^* = (X^{*\prime} \Omega^{-1} X^*)^{-1} X^{*\prime} \Omega^{-1} Y^*$$

Recall; if $\epsilon \sim N(0, \Omega)$ with $\Omega \neq \sigma^2 I$,
then $\hat{\beta}$ unbiased & consistent, but inefficient
and s^2 biased and inconsistent.
GLS is the cure. <u>Proved already!</u>

Problem: We must invert Ω, which is $GT \times GT$.

Solution: Given assumptions,

$$\Omega = \Sigma \otimes I_T ;$$

$$\Rightarrow \quad \Omega^{-1} = \Sigma^{-1} \otimes I_T \qquad \text{by Kronecker Fact (3)}$$

$$= \begin{bmatrix} \sigma^{11} I & \sigma^{12} I & \cdots & \sigma^{1G} I \\ \vdots & & & \\ \sigma^{G1} I & \sigma^{G2} I & \cdots & \sigma^{GG} I \end{bmatrix}$$

[Notation; σ^{11} = top left element of Σ^{-1}
(superscripts instead of sub-).]

Hence, to invert Ω $(GT \times GT)$,
we need only invert Σ $(G \times G)$!

$\Rightarrow$ [usefulness of assumptions!]

What does $\tilde{\beta}^* = (X^{*\prime}\Omega^{-1}X^*)^{-1}(X^{*\prime}\Omega^{-1}Y^*)$ look like?

Consider $(X^{*\prime}\Omega^{-1}X^*)$.

$$= \begin{bmatrix} X_1' & & 0 \\ & X_2' & \\ 0 & \ddots & X_G' \end{bmatrix} \begin{bmatrix} \sigma^{11}I & \cdots & \sigma^{1G}I \\ \vdots & & \\ \sigma^{G1}I & \cdots & \sigma^{GG}I \end{bmatrix} \begin{bmatrix} X_1 & & 0 \\ & X_2 & \\ 0 & \ddots & X_G \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^{11}X_1' & \sigma^{12}X_1' & \cdots & \sigma^{1G}X_1' \\ \sigma^{21}X_2' & \sigma^{22}X_2' & \cdots & \sigma^{2G}X_2' \\ \vdots & & & \\ \sigma^{G1}X_G' & \sigma^{G2}X_G' & \cdots & \sigma^{GG}X_G' \end{bmatrix} \begin{bmatrix} X_1 & & 0 \\ & X_2 & \\ 0 & \ddots & X_G \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^{11}X_1'X_1 & \sigma^{12}X_1'X_2 & \cdots & \sigma^{1G}X_1'X_G \\ \sigma^{21}X_2'X_1 & \sigma^{22}X_2'X_2 & \cdots & \sigma^{2G}X_2'X_G \\ \vdots & & & \\ \sigma^{G1}X_G'X_1 & \sigma^{G2}X_G'X_2 & \cdots & \sigma^{GG}X_G'X_G \end{bmatrix}$$

Consider $(X^{*\prime}\Omega^{-1}Y^*)$.

$$= \begin{bmatrix} \sigma^{11}X_1' & \sigma^{12}X_1' & \cdots & \sigma^{1G}X_1' \\ \sigma^{21}X_2' & \sigma^{22}X_2' & \cdots & \sigma^{2G}X_2' \\ \vdots & & & \\ \sigma^{G1}X_G' & \sigma^{G2}X_G' & \cdots & \sigma^{GG}X_G' \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_G \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^{11}X_1'Y_1 + \sigma^{12}X_1'Y_2 + \cdots + \sigma^{1G}X_1'Y_G \\ \sigma^{21}X_2'Y_1 + \sigma^{22}X_2'Y_2 + \cdots + \sigma^{2G}X_2'Y_G \\ \vdots \\ \sigma^{G1}X_G'Y_1 + \sigma^{G2}X_G'Y_2 + \cdots + \sigma^{GG}X_G'Y_G \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{G} \sigma^{1j}X_1'Y_j \\ \sum_{j=1}^{G} \sigma^{2j}X_2'Y_j \\ \vdots \\ \sum_{j=1}^{G} \sigma^{Gj}X_G'Y_j \end{bmatrix}$$

(On bd)

Finally,

$$\tilde{\beta}^* = \begin{bmatrix} \sigma^{11} X_1'X_1 & \sigma^{12} X_1'X_2 & \cdots & \sigma^{1G} X_1'X_G \\ \sigma^{21} X_2'X_1 & \sigma^{22} X_2'X_2 & \cdots & \sigma^{2G} X_2'X_G \\ \vdots & & & \\ \sigma^{G1} X_G'X_1 & \sigma^{G2} X_G'X_2 & \cdots & \sigma^{GG} X_G'X_G \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{G} \sigma^{1j} X_1' Y_j \\ \sum_{j=1}^{G} \sigma^{2j} X_2' Y_j \\ \vdots \\ \sum_{j=1}^{G} \sigma^{Gj} X_G' Y_j \end{bmatrix}$$

And

$$\tilde{\beta}^* \longrightarrow N\left[\beta^*, \ (X^{*\prime}\Omega^{-1}X^*)^{-1}\right].$$

Fact: $\tilde{\beta}^*$ is efficient relative to $\hat{\beta}^*$, except in 2 cases (again, proved in Spring) (in which $\tilde{\beta}^*$ and $\hat{\beta}^*$ are equivalent).

Case 1) Σ is diagonal. (then Σ^{-1} is also diagonal.)

ie. $\sigma_{ij} = 0$ if $i \neq j$;
the disturbances accross equations are uncorrelated even in the same time period.

Then

$$\tilde{\beta}^* = \begin{bmatrix} \sigma^{11} X_1'X_1 & & & 0 \\ & \sigma^{22} X_2'X_2 & & \\ & & \ddots & \\ 0 & & & \sigma^{GG} X_G'X_G \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{11} X_1'Y_1 \\ \sigma^{22} X_2'Y_2 \\ \vdots \\ \sigma^{GG} X_G'Y_G \end{bmatrix} = \hat{\beta}^*$$

The σ^{ij} are all zero,
and the σ^{ii} cancel out. $\leftarrow \boxed{\left(\frac{1}{\sigma^{ii}}\right)\sigma^{ii} = 1}$

That $\tilde{\beta}^* = \hat{\beta}^*$ when Σ is diagonal makes sense!

When disturbances are not correlated accross equations, $\nexists$ no reason why the estimation of any particular regression should be improved by incorporating the information from other regressions.

Note, however, that in this case, the joint disturbance term, ϵ^*, does not satisfy the ideal conditions since σ_{ii} need not $= \sigma_{jj}$ (heterosked.).

Nevertheless, OLS is efficient in this case. (because X^* is block-diagonal)

<u>Case 2)</u> Same regressors in each equation.

i.e. $X_1 = X_2 = \cdots = X_G = X$;

$X^* = I_G \otimes X$.

Then

$$\tilde{\beta}^* = (X^{*\prime}\Omega^{-1}X^*)^{-1}(X^{*\prime}\Omega^{-1}Y^*)$$

$$= \{(I\otimes X')(\Sigma^{-1}\otimes I)(I\otimes X)\}^{-1}(I\otimes X')(\Sigma^{-1}\otimes I)Y^*$$

by Kronecker Fact (1)

$$= \{(\Sigma^{-1}\otimes X')(I\otimes X)\}^{-1}(\Sigma^{-1}\otimes X')Y^*$$

by Kronecker Fact (2).

$$= (\Sigma^{-1} \otimes X'X)^{-1} (\Sigma^{-1} \otimes X') Y^* \qquad \text{Kronecker Fact(}$$

$$= \left[\Sigma \otimes (X'X)^{-1}\right] (\Sigma^{-1} \otimes X') Y^* \qquad \text{Kronecker Fact(}$$

$$= \left[I_G \otimes (X'X)^{-1} X'\right] Y^* \qquad \text{Kronecker Fact(}$$

$$= \begin{bmatrix} (X'X)^{-1}X' & & & 0 \\ & (X'X)^{-1}X' & & \\ 0 & & \ddots & \\ & & & (X'X)^{-1}X' \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_G \end{bmatrix}$$

$$= \begin{bmatrix} (X'X)^{-1}X'Y_1 \\ (X'X)^{-1}X'Y_2 \\ \vdots \\ (X'X)^{-1}X'Y_G \end{bmatrix} = \hat{\beta}^*$$

This is true even though there may be correlation accross equations ($\sigma_{ij} \neq 0$); the σ_{ij} are cancelled out in the Kronecker Product multiplication, as $\Sigma\Sigma^{-1} = I_G$.

Comment: Rewrite the joint equation in the Case (2) situation.

$$\underset{T\times G}{\left[y_{t1}\ y_{t2} \cdots y_{tG}\right]} = \underbrace{\underset{T\times K}{\left[x_{t1}\ x_{t2} \cdots x_{tK}\right]}}_{X} \underset{K\times G}{\left[\beta_1\ \beta_2 \cdots \beta_G\right]} + \underset{T\times G}{\left[\epsilon_1\ \epsilon_2 \cdots \epsilon_G\right]}$$

This is G equations, each with the same regressors, at time t.

To statisticians, this is simply a multivariate regression model. If the regressors are the same (or if not, include appropriate zeroes in β), the Maximum Likelihood Estimator will turn out to be the OLS estimator.

$$\underset{T\times G}{\left[Y_1\ Y_2 \cdots Y_G\right]} = \underset{T\times K}{X} \underset{K\times G}{\left[\beta_1\ \beta_2 \cdots \beta_G\right]} + \underset{T\times G}{\left[\epsilon_1\ \epsilon_2 \cdots \epsilon_G\right]}$$

<u>Problem:</u> Don't know Σ.
— so estimate it!

<u>Proposition:</u> Let $\hat{\Omega} = \hat{\Sigma} \otimes I$,
where $\hat{\Sigma}$ is <u>any</u> consistent estimate of Σ

Define $\hat{\beta}^* = (X^{*\prime}\hat{\Omega}^{-1}X^*)^{-1} X^{*\prime}\hat{\Omega}^{-1}Y^*$.

This is consistent, and has the same asymptotic distribution as $\tilde{\beta}^*$.

Proved <u>already</u> wrt GLS.
— See Schmidt.

The usual estimate is S, defined by

$$\hat{\sigma}_{ij} = S_{ij} = \frac{1}{T}(Y_i - X_i\hat{\beta}_i)'(Y_j - X_j\hat{\beta}_j)$$

$$= \frac{1}{T} e_i' e_j$$

$$= \frac{1}{T}\sum_{t=1}^{T} e_{it}\, e_{jt}$$

where the $\hat{\beta}$'s are OLS estimates.

Then $S = [\hat{\sigma}_{ij}]$; $i, j = 1, \dots, G$.

← [This is the obvious estimator of Σ!]

3 - Step Procedure :

(1) Run OLS on every equation, $i = 1, \dots, G$;
obtain residuals, $e_i = Y_i - X_i \hat{\beta}_i$.

(2) Let S be the matrix composed of
the $S_{ij} = \hat{\sigma}_{ij} = \frac{1}{T} e_i' e_j$ for $i, j = 1, \dots, G$.

(3) $\hat{\beta}^* = (X^{*\prime} \hat{\Omega}^{-1} X^*)^{-1} X^{*\prime} \hat{\Omega}^{-1} Y^*$

where $\hat{\Omega} = \hat{\Sigma} \otimes I_T$.
$\doteq S \otimes I_T$

Note: In SAS,

```
PROC SYSREG OUT[illegible];
  EQU1: MODEL Y1 = X1 X2 ... ;
  EQU2: MODEL Y2 =           ;
   :
  EQUG: MODEL YG =           ;
  SYSTEM EQU1 EQU2 ... EQUG;
```

Note: This 3-Step Procedure can be iterated.
Then $\hat{\Sigma}$ will converge to the efficient (MLE)
estimator of Σ.

* Should <u>really</u> $\div (T-K)$ ← (= the degrees of freedom)
in each element in S, instead of by T.

—— dividing by T → biased but consistent $\hat{\sigma}_{ij}$.

Small Sample Properties of $\hat{\beta}^*$.

Proposition: With $\hat{\Omega} = S \otimes I$, $\hat{\beta}^*$ is unbiased.

Proof: $\hat{\beta}^* = (X^{*\prime}(S^{-1} \otimes I) X^*)^{-1} X^{*\prime}(S^{-1} \otimes I)(X^* \beta^* + \epsilon^*)$

$$(\hat{\beta}^* - \beta^*) = \underbrace{(X^{*\prime}(S^{-1} \otimes I) X^*)^{-1} X^{*\prime}(S^{-1} \otimes I)}_{g(\epsilon^*)} \epsilon^*$$

why?

$$= g(\epsilon^*) \cdot \epsilon^*$$

Fact; $g(-\epsilon^*) = g(\epsilon^*)$.

This is true because the only place ϵ^* appears in $g(\epsilon^*)$ is in the terms in S, where they are multiplied times themselves in a Sum of Squares. Hence, any sign Δ is cancelled out.

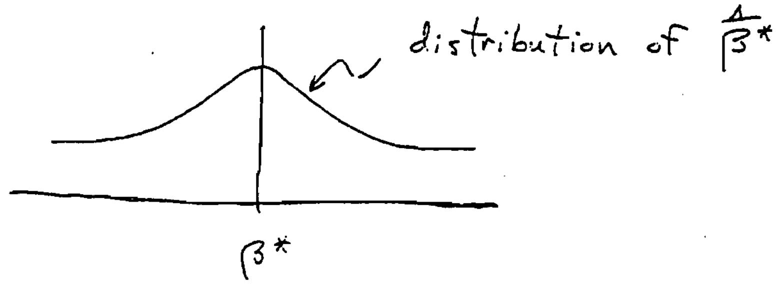
This implies that changing the sign of ϵ^* just changes the sign of the deviation of $\hat{\beta}^*$ from β^*:

$$\begin{aligned} \text{i.e.} \quad g(-\epsilon^*) \cdot (-\epsilon^*) &= g(\epsilon^*) \cdot (-\epsilon^*) \\ &= -(\hat{\beta}^* - \beta^*) \\ &= \beta^* - \hat{\beta}^* . \end{aligned}$$

i.e. a deviation in one direction is just as likely as the same deviation in the other direction.

Hence, $(\hat{\beta}^* - \beta^*)$ and $(\beta^* - \hat{\beta}^*)$ have the same probability densities.

$\rightarrow$ $\hat{\beta}^*$ is distributed symmetrically about β^*.



$\rightarrow$ If the mean of $\hat{\beta}^*$ exists, it is β^*.
$\rightarrow$ $\tilde{\beta}^*$ unbiased.

QED.

Comment: Monte Carlo experiments re: the distribution of these estimators have shown that a sample size of about 20 is large enough for the desirable asymptotic properties to hold.
(See Guilkey & Schmidt.)

Extensions of this Topic:

(1) Autocorrelation in ϵ^* – Parks, JASA, 1967.
– Guilkey & Schmidt, JASA, Sept, 7

(2) Unequal #'s of observations – Schmidt, JEC, 1977.