$\underset{\sim}{y} = Xb + \underset{\sim}{e}$, $\mathbb{E}\,\underset{\sim}{e} = \underset{\sim}{0}$ now $\text{Cov}\,\underset{\sim}{e} = \sigma^2 I_n$

STAT 714 fa 2023 Lec 03

Gauss-Markov model, Aitken model, generalized least-squares

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Gauss-Markov model

The *Gauss-Markov model* assumes $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, where $\mathbb{E}\mathbf{e} = \mathbf{0}$ and $\operatorname{Cov}\mathbf{e} = \sigma^2\mathbf{I}_n$.

We consider the mean and variance of the LS estimator $\mathbf{c}^T\hat{\mathbf{b}}$ in this model.

We also consider how to estimate the error term variance σ^2.

Result (Must know)

For a random vector $\mathbf{y}$, *vectors* $\mathbf{a}$ *and* $\mathbf{b}$, *and a matrices* $\mathbf{A}$ *and* $\mathbf{B}$, *we have*

1. $\mathbb{E}\mathbf{a}^T\mathbf{y} = \mathbf{a}^T\mathbb{E}\mathbf{y}$.
2. $\operatorname{Var}\mathbf{a}^T\mathbf{y} = \mathbf{a}^T(\operatorname{Cov}\mathbf{y})\,\mathbf{a}$.
3. $\operatorname{Cov}(\mathbf{a}^T\mathbf{y}, \mathbf{b}^T\mathbf{y}) = \mathbf{a}^T(\operatorname{Cov}\mathbf{y})\,\mathbf{b}$.
4. $\operatorname{Cov}\mathbf{A}\mathbf{y} = \mathbf{A}(\operatorname{Cov}\mathbf{y})\mathbf{A}^T$.
5. $\operatorname{Cov}(\mathbf{A}\mathbf{y}, \mathbf{B}\mathbf{y}) = \mathbf{A}(\operatorname{Cov}\mathbf{y})\mathbf{B}^T$.

Result (Variance of estimable contrast)

Let $\mathbf{c}^T\mathbf{b}$ *be an estimable contrast in* $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ *with LS estimator* $\mathbf{c}^T\hat{\mathbf{b}}$. *Then*

$$\operatorname{Var}\mathbf{c}^T\hat{\mathbf{b}} = \sigma^2\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{c}.$$

Exercise: Derive and show invariance to the choice of generalized inverse.

$\underset{\sim}{c}^T \underset{\sim}{b}$ estimable $\iff$ $\underset{\sim}{c} \in Col X^T$. Means $\boxed{\underset{\sim}{c} = X^T \underset{\sim}{a},}$ some $\underset{\sim}{a}$.

$X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{y} \iff X\hat{b} = P_X \underset{\sim}{y}$

$$\underset{\sim}{y} = \overbrace{X\underset{\sim}{b}}^{\text{fixed}} + \underset{\sim}{e}, \quad \mathbb{E}\,\underset{\sim}{e} = \underset{\sim}{0}, \quad Cov(\underset{\sim}{e}) = \sigma^2 I_n.$$

$$Cov(\underset{\sim}{y}) = Cov(\underbrace{X\underset{\sim}{b}}_{\text{const.}} + \underset{\sim}{e}) = Cov(\underset{\sim}{e}) = \sigma^2 I_n.$$

$$
\begin{aligned}
Var(\underset{\sim}{c}^T \hat{\underset{\sim}{b}}) &= Var(\underset{\sim}{a}^T X \hat{\underset{\sim}{b}}) \\
&= Var(\underset{\sim}{a}^T P_X \underset{\sim}{y}) \\
&= \underset{\sim}{a}^T P_X (Cov\, \underset{\sim}{y}) P_X \underset{\sim}{a} \\
&= \underset{\sim}{a}^T P_X (\sigma^2 I_n) P_X \underset{\sim}{a} \\
&= \sigma^2 \underset{\sim}{a}^T P_X P_X \underset{\sim}{a} \\
&= \sigma^2 \underset{\sim}{a}^T P_X \underset{\sim}{a} \\
&= \sigma^2 \boxed{\underset{\sim}{a}^T X} (X^T X)^- \boxed{X^T \underset{\sim}{a}} \\
&= \sigma^2 \underset{\sim}{c}^T (X^T X)^- \underset{\sim}{c}
\end{aligned}
$$

Easier case of full-rank X: We have $\hat{\underset{\sim}{b}} = (X^T X)^{-1} X^T \underset{\sim}{y}$

$$
\begin{aligned}
Var(\underset{\sim}{c}^T \hat{\underset{\sim}{b}}) &= Var(\underset{\sim}{c}^T (X^T X)^{-1} X^T \underset{\sim}{y}) \\
&= \underset{\sim}{c}^T (X^T X)^{-1} X^T [\sigma^2 I] X (X^T X)^{-1} \underset{\sim}{c} \\
&= \sigma^2 \underset{\sim}{c}^T (X^T X)^{-1} X^T X (X^T X)^{-1} \underset{\sim}{c} \\
&= \sigma^2 \underset{\sim}{c}^T (X^T X)^{-1} \underset{\sim}{c}
\end{aligned}
$$

Also: In full-rank case,

$$Cov(\hat{\underset{\sim}{b}}) = Cov((X^T X)^{-1} X^T \underset{\sim}{y}) = (X^T X)^{-1} X^T (\sigma^2 I_n) X (X^T X)^{-1}$$

$$= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1}$$

$$= \sigma^2 (X^T X)^{-1}.$$

Best linear unbiased estimator (BLUE)

A *best linear unbiased estimator* (BLUE) of an estimable contrast $\mathbf{c}^T\mathbf{b}$ is an estimator of the form $a_0 + \mathbf{a}^T\mathbf{y}$ such that $\mathbb{E}[a_0 + \mathbf{a}^T\mathbf{y}] = \mathbf{c}^T\mathbf{b}$ for all $\mathbf{b}$ and such that no other such estimator has smaller variance for any $\mathbf{b}$.

A BLUE achieves the smallest possible variance among linear unbiased estimators.

Theorem (Gauss-Markov Theorem)

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ *with* $\mathbb{E}\mathbf{e} = \mathbf{0}$ *and* $\operatorname{Cov}\mathbf{e} = \sigma^2\mathbf{I}_n$ *and let* $\mathbf{c}^T\mathbf{b}$ *be an estimable contrast. If* $\hat{\mathbf{b}}$ *satisfies* $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$ *then* $\mathbf{c}^T\hat{\mathbf{b}}$ *is the BLUE for* $\mathbf{c}^T\mathbf{b}$.

See Thm 4.1 of Monahan (2008).

Prove the result.

Proof: Let $\underset{\sim}{c}^T \underset{\sim}{b}$ be estimable in $\underset{\sim}{y} = X\underset{\sim}{b} + \underset{\sim}{e}$.

Let $a_0 + \underset{\sim}{a}^T \underset{\sim}{y}$ be an estimator such that

$$\mathbb{E}\left[a_0 + \underset{\sim}{a}^T \underset{\sim}{y}\right] = \underset{\sim}{c}^T \underset{\sim}{b} \quad \forall \; \underset{\sim}{b}.$$

Want to show that $\operatorname{Var}\left(a_0 + \underset{\sim}{a}^T \underset{\sim}{y}\right) \geq \operatorname{Var}\left(\underset{\sim}{c}^T \hat{\underset{\sim}{b}}\right)$.

First, $\mathbb{E}\left[a_0 + \underset{\sim}{a}^T \underset{\sim}{y}\right] = \underset{\sim}{c}^T \underset{\sim}{b} \quad \forall \; \underset{\sim}{b}$, it means

$$a_0 + \underset{\sim}{a}^T \mathbb{E}\underset{\sim}{y} = \underset{\sim}{c}^T \underset{\sim}{b} \quad \forall \; \underset{\sim}{b}$$

$$\Rightarrow a_0 + \underset{\sim}{a}^T X b = \underset{\sim}{c}^T \underset{\sim}{b} \quad \forall \; \underset{\sim}{b}$$

$$\Rightarrow \boxed{a_0 = 0 \quad \text{and} \quad \underset{\sim}{a}^T X = \underset{\sim}{c}.}$$

Now

$$\begin{aligned}
\operatorname{Var}\left(a_0 + \underset{\sim}{a}^T \underset{\sim}{y}\right) &= \operatorname{Var}\left(\underset{\sim}{a}^T \underset{\sim}{y}\right) \\
&= \operatorname{Var}\left(\underset{\sim}{a}^T \underset{\sim}{y} - \underset{\sim}{c}^T \hat{\underset{\sim}{b}} + \underset{\sim}{c}^T \hat{\underset{\sim}{b}}\right) \\
&= \operatorname{Var}\left(\underset{\sim}{a}^T \underset{\sim}{y} - \underset{\sim}{c}^T \hat{\underset{\sim}{b}}\right) + \operatorname{Var}\left(\underset{\sim}{c}^T \hat{\underset{\sim}{b}}\right) \\
&\quad + 2 \operatorname{Cov}\left(\underset{\sim}{a}^T \underset{\sim}{y} - \underset{\sim}{c}^T \hat{\underset{\sim}{b}}, \underset{\sim}{c}^T \hat{\underset{\sim}{b}}\right).
\end{aligned}$$

We have

$$\begin{aligned}
\operatorname{Cov}\left(\underset{\sim}{a}^T \underset{\sim}{y} - \underset{\sim}{c}^T \hat{\underset{\sim}{b}}, \underset{\sim}{c}^T \hat{\underset{\sim}{b}}\right) &= \operatorname{Cov}\left(\underset{\sim}{a}^T \underset{\sim}{y} - \underset{\sim}{a}^T \boxed{X \hat{\underset{\sim}{b}}}, \underset{\sim}{a}^T X \hat{\underset{\sim}{b}}\right) \\
&= \operatorname{Cov}\left(\underset{\sim}{a}^T \underset{\sim}{y} - \underset{\sim}{a}^T P_X \underset{\sim}{y}, \underset{\sim}{a}^T P_X \underset{\sim}{y}\right) \\
&= \operatorname{Cov}\left(\underset{\sim}{a}^T (I - P_X) \underset{\sim}{y}, \underset{\sim}{a}^T P_X \underset{\sim}{y}\right) \\
&= \operatorname{Cov}\left(\underset{\sim}{a}^T (I - P_X) \underset{\sim}{y}, \underset{\sim}{a}^T P_X \underset{\sim}{y}\right)
\end{aligned}$$

3. $\text{Cov}(\mathbf{a}^T\mathbf{y}, \mathbf{b}^T\mathbf{y}) = \mathbf{a}^T(\text{Cov}\,\mathbf{y})\,\mathbf{b}$.

$$= a^T (I - P_X) (\text{Cov}\, y) P_X a$$
$$= \sigma^2 a^T (I - P_X) P_X a$$
$$= \sigma^2 a^T (P_X - P_X) a$$
$$= 0$$
✓

Result (BLUE uncorrelated with unbiased estimators of zero)

The BLUE $\mathbf{c}^T\hat{\mathbf{b}}$ *of an estimable contrast* $\mathbf{c}^T\mathbf{b}$ *is uncorrelated with all unbiased estimators of* 0.

See Res 4.1 of Monahan (2008).

Prove the result.

Result (Expected value of a quadratic form)

Let $\mathbf{z}$ *be a random vector with* $\mathbb{E}\mathbf{z} = \boldsymbol{\mu}$ *and* $\operatorname{Cov}\mathbf{z} = \boldsymbol{\Sigma}$. *Then*

$$\mathbb{E}\mathbf{z}^T\mathbf{A}\mathbf{z} = \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} + \operatorname{tr}(\mathbf{A}\boldsymbol{\Sigma}).$$

See Lem 4.1 of Monahan (2008).

Can be used to prove the following result.

Result (Unbiased estimator of the variance)

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ *with* $\mathbb{E}\mathbf{e} = \mathbf{0}$ *and* $\operatorname{Cov}\mathbf{e} = \sigma^2\mathbf{I}_n$ *and let* $\mathbf{X}$ *have rank* r. *Then*

$n \times 1$

$$\hat{\sigma}^2 = \frac{\|\hat{\mathbf{e}}\|^2}{n - r}, \quad \textit{where} \quad \hat{\mathbf{e}} = (\mathbf{I} - \mathbf{P_X})\mathbf{y},$$

$\hat{y} = X\hat{b} = P_X y$

$\hat{e} = y - \hat{y} = y - P_X y = (I - P_X) y.$

is an unbiased estimator of σ^2.

See Res 4.2 of Monahan (2008).

$\mathbb{E}\hat{\sigma}^2 = \frac{1}{n-r}\mathbb{E}\|\hat{e}\|^2 = \frac{1}{n-r}\mathbb{E}\hat{e}^T\hat{e} = \frac{1}{n-r}\mathbb{E}\, y^T(I - P_X)y$

Prove both results.

Let $\mathbf{z}$ *be a random vector with* $\mathbb{E}\mathbf{z} = \boldsymbol{\mu}$ *and* $\text{Cov}\,\mathbf{z} = \boldsymbol{\Sigma}$. *Then*

$$\mathbb{E}\mathbf{z}^T\mathbf{A}\mathbf{z} = \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} + \text{tr}(\mathbf{A}\boldsymbol{\Sigma}).$$

Prove:

$$\text{tr}(AB) = \text{tr}(BA)$$

$$\text{Cov}\, z = \mathbb{E}\, z z^T - (\mathbb{E} z)(\mathbb{E} z)^T$$

$$Y \in \mathbb{R} \qquad \text{Var}\, Y = \mathbb{E} Y^2 - (\mathbb{E} Y)^2$$

$$\mathbb{E}\, \underset{1\times 1}{z^T A z} = \mathbb{E}\, \text{tr}\left(z^T A z\right)$$

$$= \mathbb{E}\left[\text{tr}\left(z z^T A\right)\right]$$

$$= \text{tr}\left[\mathbb{E}\left(z z^T A\right)\right]$$

$$= \text{tr}\left(\mathbb{E}(z z^T)\, A\right)$$

$$= \text{tr}\left(\left((\mathbb{E} z)(\mathbb{E} z)^T + \text{Cov}\, z\right) A\right)$$

$$= \text{tr}\left(\left(\mu \mu^T + \Sigma\right) A\right)$$

$$= \text{tr}\left(\mu \mu^T A + \Sigma A\right)$$

$$= \text{tr}\left(\mu \mu^T A\right) + \text{tr}(\Sigma A)$$

$$= \text{tr}\left(\mu^T A \mu\right) + \text{tr}(A \Sigma)$$

$$= \mu^T A \mu + \text{tr}(A \Sigma).$$

Now:

$$\mathbb{E}\,\hat{\sigma}^2 = \mathbb{E}\left[\frac{\|\hat{e}\|^2}{n-r}\right] = \frac{1}{n-r}\,\mathbb{E}\left[\hat{e}^T \hat{e}\right]$$

$$= \frac{1}{n-r}\,\mathbb{E}\left[y^T (I - P_X)\, y\right]$$

$\mathbb{E}\,\underset{\sim}{y} = X\underset{\sim}{b}$

$\text{Cov}\,\underset{\sim}{y} = \sigma^2 I_n$

$P_X X\underset{\sim}{b} = X\underset{\sim}{b}$

$(I - P_X) X\underset{\sim}{b} = X\underset{\sim}{b} - X\underset{\sim}{b} = \underset{\sim}{0}$

$$= \frac{1}{n-r}\Big[\underbrace{(X\underset{\sim}{b})^T (I - P_X)\, X\underset{\sim}{b}}_{\underset{\sim}{0}} + \text{tr}\big[(I - P_X)\,\sigma^2 I_n\big]\Big]$$

$$= \frac{\sigma^2}{n-r}\,\underbrace{\text{tr}\,(I - P_X)}_{= n-r \text{ because}}$$

(i) $\text{tr}\,(I - P_X) = \text{tr}\, I_n - \text{tr}\, P_X$

$= n - \text{tr}\, P_X$.

(ii) P_X idempotent, so eigenvals are 0 or 1.

(iii) P_X symmetric, so rank = # nonzero eigenvalues

(iv) $\text{tr}\, P_X$ = sum of eigenvalues
= # nonzero eigenvalues
= rank X
= r

$$= \sigma^2.$$

Let $\mathbf{z}$ *be a random vector with* $\mathbb{E}\mathbf{z} = \boldsymbol{\mu}$ *and* $\text{Cov}\,\mathbf{z} = \boldsymbol{\Sigma}$. *Then*

$$\mathbb{E}\mathbf{z}^T\mathbf{A}\mathbf{z} = \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} + \text{tr}(\mathbf{A}\boldsymbol{\Sigma}).$$

A setup for "underfitting" or omitting important covariates

Suppose $\mathbf{y}$ is truly given by

" eta " is fixed

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\eta} + \mathbf{e}, \quad \mathbb{E}\mathbf{e} = 0, \quad \operatorname{Cov}\mathbf{e} = \sigma^2 \mathbf{I}_n,$$

but one assumes the model $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ and finds $\hat{\mathbf{b}}$ satisfying $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$.

So $\boldsymbol{\eta}$ represents the effects of omitted covariates.

Exercise: In the above setup find the expected value of

1. $\mathbf{c}^T\hat{\mathbf{b}}$, where $\mathbf{c} \in \operatorname{Col}\mathbf{X}^T$ — $c = X^T a$ for some a.
2. $\hat{\sigma}^2 = \|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\|^2/(N - r)$, where $r = \operatorname{rank}\mathbf{X}$. ←

In each case give a condition on $\boldsymbol{\eta}$ under which the estimator will be unbiased.

$\underset{\sim}{y} = \boxed{X\underset{\sim}{b} + \underset{\sim}{\eta}} + \underset{\sim}{e}$, but $\hat{\underset{\sim}{b}}$ satisfies $X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{y}$

(1)

$$\mathbb{E}\, \underset{\sim}{c}^T \hat{\underset{\sim}{b}} = \mathbb{E}\, \underset{\sim}{a}^T X \hat{\underset{\sim}{b}}$$

$$= \mathbb{E}\, \underset{\sim}{a}^T P_X \underset{\sim}{y}$$

$$= \underset{\sim}{a}^T P_X \mathbb{E}\, \underset{\sim}{y}$$

$$= \underset{\sim}{a}^T P_X [X\underset{\sim}{b} + \underset{\sim}{\eta}]$$

$$= \underset{\sim}{a}^T P_X X \underset{\sim}{b} + \underset{\sim}{a}^T P_X \underset{\sim}{\eta}$$

$$= \underset{\sim}{a}^T \underbrace{X (X^T X)^- X^T X}\, \underset{\sim}{b} + \underset{\sim}{a}^T P_X \underset{\sim}{\eta}$$

$$= \underset{\sim}{a}^T X \underset{\sim}{b} + \underset{\sim}{a}^T P_X \underset{\sim}{\eta}$$

$$= \underset{\sim}{c}^T \underset{\sim}{b} + \underbrace{\underset{\sim}{a} P_X \underset{\sim}{\eta}}_{\text{bias term}}$$

$P_X \underset{\sim}{\eta} = 0$ if $\underset{\sim}{\eta} \in (\text{Col } X)^\perp = \text{Nul } X^T$

Because if $X^T \underset{\sim}{\eta} = 0$,

then $P_X \underset{\sim}{\eta} = X (X^T X)^- X^T \underset{\sim}{\eta} = \underset{\sim}{0}$

If $\underset{\sim}{\eta}$ is orthogonal to all columns of X, $\underset{\sim}{c}^T \hat{\underset{\sim}{b}}$ is still unbiased.

Don't omit predictors which are correlated with your predictors in the model.

A setup for "overfitting" or including unimportant covariates

Suppose $\mathbf{y}$ is truly given by

$$\mathbf{y} = \mathbf{X}_1\mathbf{b}_1 + \mathbf{e}, \quad \mathbb{E}\mathbf{e} = 0, \quad \operatorname{Cov}\mathbf{e} = \sigma^2\mathbf{I}_n,$$

but one assumes $\mathbf{y} = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2 + \mathbf{e}$ and finds $\hat{\mathbf{b}} = [\mathbf{b}_1^T \ \mathbf{b}_2^T]^T$ satisfying

$$\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}, \quad \text{where} \quad \mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2].$$

In truth $\mathbf{b}_2 = \mathbf{0}$, so $\mathbf{X}_2$ contains unimportant (extra) covariate information.

Exercise: In the above setup, assume (for simplicity) that $\mathbf{X}$ is full-rank. Then:

$$\hat{b} = (X^TX)^{-1}X^Ty = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix}$$

1. Find $\mathbb{E}\mathbf{c}_1^T\hat{\mathbf{b}}_1$ and $\operatorname{Var}\mathbf{c}_1^T\hat{\mathbf{b}}_1$
2. In what situation does adding "extra" covariates have no effect on $\operatorname{Var}\mathbf{c}_1^T\hat{\mathbf{b}}_1$?
3. Find $\mathbb{E}\|(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{y}\|^2/(N - \operatorname{rank}\mathbf{X})$ and $\mathbb{E}\|(\mathbf{I} - \mathbf{P}_{\mathbf{X}_1})\mathbf{y}\|^2/(N - \operatorname{rank}\mathbf{X}_1)$.

①

$$\mathbb{E}\,\hat{\underset{\sim}{b}} = \mathbb{E}\,(X^TX)^{-1}X^T\underset{\sim}{y} = (X^TX)^{-1}X^T\,\mathbb{E}\,\underset{\sim}{y} = (X^TX)^{-1}X^TX\underset{\sim}{b} = \underset{\sim}{b} = \begin{bmatrix}\underset{\sim}{b}_1\\ \underset{\sim}{b}_2\end{bmatrix}$$

$$\mathbb{E}\,\underset{\sim}{c}_1^T\hat{\underset{\sim}{b}}_1 = \underset{\sim}{c}_1^T\underset{\sim}{b}_1.$$ Unbiased.

$$\underline{\mathrm{Var}\ \underset{\sim}{c}_1^T\hat{\underset{\sim}{b}}_1} = \underset{\sim}{c}_1^T\left(\mathrm{Cov}\ \hat{\underset{\sim}{b}}_1\right)\underset{\sim}{c}_1$$

$$\mathrm{Cov}\ \hat{\underset{\sim}{b}} = \mathrm{Cov}\left((X^TX)^{-1}X^T\underset{\sim}{y}\right)$$

$$\mathrm{Cov}(A\underset{\sim}{y}) = A(\mathrm{Cov}\,\underset{\sim}{y})A^T$$

$$= (X^TX)^{-1}X^T[\sigma^2 I_n]X(X^TX)^{-1}$$

$$= \sigma^2(X^TX)^{-1}.$$

$X = [X_1\ X_2]$

X^TX

block inverse

$$\mathrm{Cov}\begin{bmatrix}\hat{\underset{\sim}{b}}_1\\ \hat{\underset{\sim}{b}}_2\end{bmatrix} = \sigma^2\begin{bmatrix} X_1^TX_1 & X_1^TX_2\\ X_2^TX_1 & X_2^TX_2\end{bmatrix}^{-1}$$

pg 268 of Monahan

$$= \sigma^2\begin{bmatrix}\left[X_1^TX_1 - X_1^TX_2(X_2^TX_2)^{-1}X_2^TX_1\right]^{-1} & * \\ * & *\end{bmatrix}$$

$$\mathrm{Cov}\ \hat{\underset{\sim}{b}}_1 = \sigma^2\left[X_1^TX_1 - X_1^T\underline{X_2(X_2^TX_2)^{-1}X_2^T}X_1\right]^{-1}$$

$$= \sigma^2\left(X_1^T[I - P_{X_2}]X_1\right)^{-1}$$

$$\mathrm{Var}\ \underset{\sim}{c}_1^T\hat{\underset{\sim}{b}}_1 = \sigma^2\ \underset{\sim}{c}_1^T\left(X_1^T[I - P_{X_2}]X_1\right)^{-1}\underset{\sim}{c}_1$$

If the columns of X_2 are uncorrelated with those of X_1, there is no effect

If we did NOT include extra covariates X_2.

We would have $\text{Var}\, c_1^T \hat{b}_1 = \sigma^2 c_1^T (X_1^T X_1)^{-1} c_1$

of including extra covariates on $\text{Var}\, c_1^T \hat{b}_1$.

If $X^T X \hat{b} = X^T y$, then $c^T \hat{b}$ might not be BLUE.

Aitken model

The *Aitken model* assumes $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, where $\mathbb{E}\mathbf{e} = \mathbf{0}$ and $\operatorname{Cov}\mathbf{e} = \sigma^2\mathbf{V}$, $\mathbf{V}$ pd.

The ordinary least-squares estimator $\mathbf{c}^T\mathbf{b}$ may not be the BLUE anymore...

Generalized least-squares estimator of an estimable contrast

Under the Aitken model, the *GLS estimator* of a contrast is $\mathbf{c}^T\hat{\mathbf{b}}_{\text{gls}}$, where $\hat{\mathbf{b}}_{\text{gls}}$ is any vector such that $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})\hat{\mathbf{b}}_{\text{gls}} = \mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$.

$Q(b) = \|y - Xb\|^2 = (y - Xb)^T (y - Xb)$

Exercise: Show that $\hat{\mathbf{b}}_{\text{gls}}$ minimizes $Q_{\mathbf{V}}(\mathbf{b}) = (\mathbf{y} - \mathbf{X}\mathbf{b})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b})$.

$$\frac{\partial}{\partial b} Q_V(b) = \frac{\partial}{\partial b}\left[y^T V^{-1} y - 2 y^T V^{-1} X b + b^T X^T V^{-1} X b \right]$$

$$= -2X^T V^{-1} \underset{\sim}{y} + 2X^T V^{-1} X \underset{\sim}{b} \overset{set}{=} 0$$

$$\Rightarrow \boxed{X^T V^{-1} X \underset{\sim}{b} = X^T V^{-1} \underset{\sim}{y}}$$

We will need these results for establishing some properties of the GLS estimator.

Result ("Generalized cool result" and another gen. inverse of $\mathbf{X}$)

1. *If* $\mathbf{V}$ *is positive definite there exists a* [nonsingular] *matrix* $\mathbf{R}$ *such that* $\mathbf{RR} = \mathbf{V}^{-1}$.
2. $\operatorname{Nul} \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} = \operatorname{Nul} \mathbf{X}$.
3. $\mathbf{X}^T\mathbf{V}^{-1}\mathbf{XA} = \mathbf{X}^T\mathbf{V}^{-1}\mathbf{XB} \iff \mathbf{XA} = \mathbf{XB}$ [generalized "cool result"]
4. $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{V}^{-1}$ *is a generalized inverse of* $\mathbf{X}$.

Prove the results.

(1) V pd. $\Rightarrow$ V nonsingular / invertible.

$V = PDP^T$ (spectral), $P^TP = I$ $\Rightarrow P^{-1} = P^T$ (left inverse is right inverse) $\Rightarrow PP^T = I$.

$V^{-1} = PD^{-1}P^T$ since $PD^{-1}P^TPDP^T = PP^T = I$.

Then set $R = P D^{-1/2} P^T$, $D^{-1/2} = \text{diag}\left(\frac{1}{\sqrt{\lambda_1}}, \dots, \frac{1}{\sqrt{\lambda_p}} \right)$

Then

$$RR = P D^{-1/2} \underbrace{P^T P}_{I} D^{-1/2} P^T$$
$$= P D^{-1} P^T$$
$$= V^{-1}.$$

So R is an invertible symmetric square root of V^{-1}.

(2) $\text{Nul } X^T V^{-1} X = \text{Nul } X$ [$\text{Nul } X^T X = \text{Nul } X$]

(i) $\text{Nul } X \subset \text{Nul } X^T V^{-1} X$

Let $\underset{\sim}{x} \in \text{Nul } X$. Then $X\underset{\sim}{x} = \underset{\sim}{0}$. So $X^T V^{-1} X \underset{\sim}{x} = 0$.
So $\underset{\sim}{x} \in \text{Nul } X^T V^{-1} X$.

(ii) $\text{Nul } X^T V^{-1} X \subset \text{Nul } X$

Let $\underset{\sim}{x} \in \text{Nul } X^T V^{-1} X$. Then $X^T V^{-1} X \underset{\sim}{x} = \underset{\sim}{0}$

Therefore $\underset{\sim}{x}^T X^T V^{-1} X \underset{\sim}{x} = \underset{\sim}{0}$

$$\Rightarrow \underset{\sim}{x}^T X^T R^T R X \underset{\sim}{x} = 0$$

$$\| R X \underset{\sim}{x} \|^2 = 0$$

$$\Rightarrow R X \underset{\sim}{x} = \underset{\sim}{0}$$

$$\Rightarrow X\underset{\sim}{x} \in \text{Nul } R = \{\underset{\sim}{0}\} \quad (R \text{ has full-column rank}).$$

□

(3) $X^T V^{-1} X A = X^T V^{-1} X B \iff XA = XB$.

"<=" Let $XA = XB$. Then $X^T V^{-1} X A = X^T V^{-1} X B$

"=>" Let $X^T V^{-1} X A = X^T V^{-1} X B$

Then $X^T V^{-1} X (A - B) = 0$.

So columns of $A-B$ are in $\text{Nul}\, X^T V^{-1} X = \text{Nul}\, X$.

So $X(A-B) = 0$

$\Rightarrow XA = XB$.

(4) $(X^T V^{-1} X)^- X^T V^{-1}$ is a gen. inv. of X.

We have

$$X^T V^{-1} X \underbrace{\left[(X^T V^{-1} X)^- X^T V^{-1} X \right]}_{A} = X^T V^{-1} X \, \boxed{I}$$

$$\Rightarrow \quad X \underbrace{(X^T V^{-1} X)^- X^T V^{-1}}_{\text{gen. inv. of } X} X = X$$

Note: Estimability of $\mathbf{c}^T\mathbf{b}$ is still equivalent to $\mathbf{c} \in \operatorname{Col} \mathbf{X}^T$.

Result (Properties of the GLS estimator of an estimable contrast)

Let $\mathbf{c}^T\mathbf{b}$ be an estimable contrast. Then the GLS estimator $\mathbf{c}^T\hat{\mathbf{b}}_{\text{gls}}$

1. *is invariant to the choice of $\hat{\mathbf{b}}_{\text{gls}}$ which satisfies $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})\hat{\mathbf{b}}_{\text{gls}} = \mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$.*
2. *has expected value equal to $\mathbf{c}^T\mathbf{b}$ for all $\mathbf{b}$.*

Prove the results.

① Let $\hat{b}_1$ and $\hat{b}_2$ solve $(X^TV^{-1}X)\,b = X^TV^{-1}y$.

$\Rightarrow (X^TV^{-1}X)\hat{b}_1 = X^TV^{-1}y$ and $(X^TV^{-1}X)\hat{b}_2 = X^TV^{-1}y$.

$\Rightarrow (X^TV^{-1}X)\hat{b}_1 = (X^TV^{-1}X)\hat{b}_2$

$$\Rightarrow \quad (X^T V^{-1} X)(\hat{\underset{\sim}{b}}_1 - \hat{\underset{\sim}{b}}_2) = \underset{\sim}{0}$$

$$\Rightarrow \quad \hat{\underset{\sim}{b}}_1 - \hat{\underset{\sim}{b}}_2 \in \text{Nul } X^T V^{-1} X = \text{Nul } X$$

$$\Rightarrow \quad X \hat{\underset{\sim}{b}}_1 = X \hat{\underset{\sim}{b}}_2 .$$

Now, if $\underset{\sim}{c}^T \underset{\sim}{b}$ is estimable, $\underset{\sim}{c} \in \text{Col } X^T$.

So $\underset{\sim}{c} = X^T \underset{\sim}{a}$ for some $\underset{\sim}{a}$.

That means $\underset{\sim}{c}^T \hat{\underset{\sim}{b}}_1 = \underset{\sim}{a}^T X \hat{\underset{\sim}{b}}_1 = \underset{\sim}{a}^T X \hat{\underset{\sim}{b}}_2 = \underset{\sim}{c}^T \hat{b}_2 .$

② for some $\underset{\sim}{a}$ because $\underset{\sim}{c} \in \text{Col } X^T$.

$$\mathbb{E}\, \underset{\sim}{c}^T \hat{\underset{\sim}{b}}_{gls} = \mathbb{E}\, \underset{\sim}{a}^T X \hat{b}_{gls}$$

$$= \mathbb{E}\, \underset{\sim}{a}^T X (X)^- X \hat{b}_{gls}$$

$$= \mathbb{E}\, \underset{\sim}{a}^T X (X^T V^{-1} X)^- X^T V^{-1} X \hat{\underset{\sim}{b}}_{gls}$$

$$= \mathbb{E}\, \underset{\sim}{a}^T X (X^T V^{-1} X)^- X^T V^{-1} \underset{\sim}{y}$$

$$= \underset{\sim}{a}^T X \underbrace{(X^T V^{-1} X)^- X^T V^{-1}}_{\text{g. inv of } X} X \underset{\sim}{b}$$

$$= \underset{\sim}{a}^T X \underset{\sim}{b}$$

$$= \underset{\sim}{c}^T \underset{\sim}{b}$$

Theorem (Aitken's Theorem)

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, *where* $\mathbb{E}\mathbf{e} = \mathbf{0}$ *and* $\operatorname{Cov}\mathbf{e} = \sigma^2\mathbf{V}$, $\mathbf{V}$ *pd, and let* $\mathbf{c}^T\mathbf{b}$ *be an estimable contrast. Then* $\mathbf{c}^T\hat{\mathbf{b}}_{\text{gls}}$ *is the BLUE for* $\mathbf{c}^T\mathbf{b}$.

See Thm 4.2 of Monahan (2008).

Prove the result.

Let $a_0 + a^T y$ be unbiased for $c^T b$ $\forall$ b.

$\Rightarrow \mathbb{E}[a_0 + a^T y] = a_0 + a^T X b = c^T b$ $\forall$ b $\Rightarrow$ $a_0 = 0$, $a^T X = c$.

Just consider $a^T y$.

Let $\mathbb{E}[\underset{\sim}{a}^T \underset{\sim}{y}] = \underset{\sim}{c}^T \underset{\sim}{b} \quad \forall \ \underset{\sim}{b}$.

Then we want to show $Var(\underset{\sim}{a}^T \underset{\sim}{y}) \geq Var(\underset{\sim}{c}^T \hat{\underset{\sim}{b}}_{gls})$.

$$Var(\underset{\sim}{a}^T \underset{\sim}{y}) = Var\left((\underset{\sim}{a}^T \underset{\sim}{y} - \underset{\sim}{c}^T \hat{b}_{gls}) + \underset{\sim}{c}^T \hat{\underset{\sim}{b}}_{gls} \right)$$

$$= Var(\underset{\sim}{a}^T \underset{\sim}{y} - \underset{\sim}{c}^T \hat{\underset{\sim}{b}}_{gls}) + Var(\underset{\sim}{c}^T \hat{\underset{\sim}{b}}_{gls})$$

$$+ \ 2\, Cov(\underset{\sim}{a}^T \underset{\sim}{y} - \underset{\sim}{c}^T \hat{b}_{gls}, \ \underset{\sim}{c}^T \hat{b}_{gls}) .$$

From here we show that the covariance is zero.

Can write $\underset{\sim}{c}^T \hat{b}_{gls} = \underset{\sim}{a}^T X (X^T V^{-1} X)^- X^T V^{-1} \underset{\sim}{y}$

$$Cov(\underset{\sim}{a}^T \underset{\sim}{y}, \underset{\sim}{b}^T \underset{\sim}{y}) = \underset{\sim}{a}^T (Cov\, \underset{\sim}{y})\, \underset{\sim}{b}$$

$$Cov(\underset{\sim}{a}^T \underset{\sim}{y} - \underset{\sim}{c}^T \hat{b}_{gls}, \ \underset{\sim}{c}^T \hat{b}_{gls})$$

$$= Cov\left(\underset{\sim}{a}^T \underset{\sim}{y} - \underset{\sim}{a}^T X (X^T V^{-1} X)^- X^T V^{-1} \underset{\sim}{y}, \ \underset{\sim}{a}^T X (X^T V^{-1} X)^- X^T V^{-1} \underset{\sim}{y} \right)$$

$$= Cov\left(\underset{\sim}{a}^T (I - X (X^T V^{-1} X)^- X^T V^{-1}) \underset{\sim}{y}, \ \underset{\sim}{a}^T X (X^T V^{-1} X)^- X^T V^{-1} \underset{\sim}{y} \right)$$

Cov $\underset{\sim}{y}$

$$= \underset{\sim}{a}^T (I - X (X^T V^{-1} X)^- X^T V^{-1}) [\sigma^2 V] V^{-1} X (X^T V^{-1} X)^- X^T \underset{\sim}{a}$$

X

$$= \sigma^2 \underset{\sim}{a}^T \left(X (X^T V^{-1} X)^- X^T - X (X^T V^{-1} X)^- X^T V^{-1} X (X^T V^{-1} X)^- X^T \right) \underset{\sim}{a}$$

g. inv of X

$$= 0.$$

Is $\underbrace{X}\underbrace{(X^TV^{-1}X)^{-}X^TV^{-1}}$ a proj matrix?

$\underbrace{X(X^TV^{-1}X)^{-}X^TV^{-1}X}_{X}(X^TV^{-1}X)^{-}X^TV^{-1} = X(X^TV^{-1}X)^{-}X^TV^{-1}$. yes.

Result (Unbiased estimator of the variance in the Aitken model)

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ *with* $\mathbb{E}\mathbf{e} = \mathbf{0}$ *and* $\operatorname{Cov}\mathbf{e} = \sigma^2\mathbf{V}$ *and let* $\mathbf{X}$ *have rank* r. *Then*

$$\hat{\sigma}^2_{\text{gls}} = \frac{1}{n-r}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}_{\text{gls}})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}_{\text{gls}})$$

is an unbiased estimator of σ^2.

$\mathbb{E}\, z^TAz = (\mathbb{E}z)^TA\,\mathbb{E}z + \operatorname{tr}(A \operatorname{Cov} z)$

Prove the result.

In G-M model $y = Xb + e$, $\mathbb{E}e = 0$, $\operatorname{Cov} e = \sigma^2 I_n$, rank $X = r$.

$\hat{\sigma}^2 = \frac{\|\hat{e}\|^2}{n-r} = \frac{\|(I - P_X)y\|^2}{n-r}$. $\mathbb{E}\hat{\sigma}^2 = \sigma^2$

$\mathbb{E}\hat{\sigma}^2 = \frac{1}{n-r}\mathbb{E}\, y^T(I-P_X)y = \frac{1}{n-r}\left[(\mathbb{E}y)^T(I-P_X)\underbrace{\mathbb{E}y}_{Xb} + \operatorname{tr}\left((I-P_X)\underbrace{(\operatorname{Cov} y)}_{\sigma^2 I_n}\right)\right]$

$= \frac{1}{n-r}\sigma^2 \underbrace{\operatorname{tr}(I - P_X)}_{n-r} = \sigma^2$

$$\mathbb{E}\,\hat\sigma^2_{gls} = \mathbb{E}\,(\tilde y - X\hat{\tilde b}_{gls})^T V^{-1} (\tilde y - X\hat{\tilde b}_{gls}) \frac{1}{n-r}$$

$RR = V^{-1}$,
R is invertible, symmetric

$$(\tilde y - X\hat{\tilde b}_{gls})^T V^{-1} (\tilde y - X\hat{\tilde b}_{gls}) = (\tilde y - X\hat{\tilde b}_{gls})^T R^T R (\tilde y - X\hat{\tilde b}_{gls})$$

$$= (R\tilde y - RX\hat b_{gls})^T (R\tilde y - RX\hat{\tilde b}_{gls})$$

$\tilde z = R\tilde y$

$U = RX$

$$= (\tilde z - U\hat{\tilde b}_{gls})^T (\tilde z - U\hat{\tilde b}_{gls})$$

We have $\hat{\tilde b}_{gls}$ satisfies

$$X^T V^{-1} X \hat{\tilde b}_{gls} = X^T V^{-1} \tilde y$$

$$\iff X^T R R X \hat b_{gls} = X^T R R \tilde y$$

$$\iff X^T R^T R X \hat{\tilde b}_{gls} = X^T R^T R \tilde y$$

$$\iff U^T U \hat{\tilde b}_{gls} = U^T \tilde z$$

$$\iff U\hat b_{gls} = P_U \tilde z$$

$RR = V^{-1}$

$(RR)^{-1} = V$

$$= (\tilde z - P_U \tilde z)^T (\tilde z - P_U \tilde z)$$

$$= \tilde z^T (I - P_U)\tilde z$$

$$\mathrm{Cov}\,\tilde z = \mathrm{Cov}(R\tilde y) = R(\mathrm{Cov}\,\tilde y)R^T$$

$$= R\sigma^2 V R = \sigma^2 R (RR)^{-1} R$$

$$= \sigma^2 I_n$$

$$\mathbb{E}\,(\tilde y - X\hat{\tilde b}_{gls})^T V^{-1} (\tilde y - X\hat{\tilde b}_{gls}) = \mathbb{E}\,\tilde z^T (I - P_U)\tilde z$$

$$= (\mathbb{E}\,\tilde z)^T (I - P_U)(\mathbb{E}\,\tilde z) + \mathrm{tr}\big((I - P_U)(\mathrm{Cov}\,\tilde z)\big)$$

$\mathbb{E} R\tilde y = RX\tilde b = Ub$; $\sigma^2 I_n$

$$= \underbrace{(U\underset{\sim}{b})^T (I - P_U)(U\underset{\sim}{b})}_{0} + \sigma^2 \, \underbrace{tr(I - P_U)}$$

$$= \sigma^2 (n - tr\, P_U)$$

$$tr\, P_U = \text{rank}\, U = \text{rank}(RX) = \text{rank}\, X = r$$

$$\left.\begin{array}{lll} U = RX & \Rightarrow & \text{Row}\, U \subset \text{Row}\, X \\ X = R^{-1} U & \Rightarrow & \text{Row}\, X \subset \text{Row}\, U \end{array}\right) \Rightarrow \text{Row}\, X = \text{Row}\, U$$

$$\Rightarrow \dim \text{Row}\, X = \dim \text{Row}\, U$$

$$\Rightarrow \dim \text{Col}\, X = \dim \text{Col}\, U$$

$$\Rightarrow \text{rank}\, X = \text{rank}\, U.$$

$$= \sigma^2 (n - r).$$

$\square$.

$$\text{Cov}\, \underset{\sim}{e} = \sigma^2 \overbrace{\begin{bmatrix} x_1^2 & & \\ & \ddots & \\ & & x_n^2 \end{bmatrix}}^{V}$$

$\underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ $\underset{\sim}{e} = \begin{bmatrix} \varepsilon_1 |x_1| \\ \vdots \\ \varepsilon_n |x_n| \end{bmatrix}$, $\mathbb{E}\,\underset{\sim}{e} = \underset{\sim}{0}$, $\text{Cov}(\underset{\sim}{e}) = \begin{bmatrix} \sigma^2 x_1^2 & & \\ & \ddots & \\ & & \sigma^2 x_n^2 \end{bmatrix}$

Exercise: Let $Y_i = x_i\beta + \varepsilon_i|x_i|$, $i = 1, \dots, n$, where $\varepsilon_1, \dots, \varepsilon_n \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma^2)$.

1. Write the model in matrix form as an Aitken model.
2. Give $\hat{\beta}_{\text{gls}}$ as well as $\hat{\beta}_{\text{ols}}$. ✓
3. Give $\hat{\sigma}^2_{\text{gls}}$. ✓
4. Give $\mathbb{E}\hat{\sigma}^2_{\text{gls}}$ (unbiased) as well as $\mathbb{E}\hat{\sigma}^2_{\text{ols}}$. ← biased
5. Give $\mathbb{E}\hat{\beta}_{\text{gls}}$ as well as $\mathbb{E}\hat{\beta}_{\text{ols}}$. both unbiased
6. Give $\text{Var}\,\hat{\beta}_{\text{gls}}$ as well as $\text{Var}\,\hat{\beta}_{\text{ols}}$.
7. Show that $\text{Var}\,\hat{\beta}_{\text{ols}} \geq \text{Var}\,\hat{\beta}_{\text{gls}}$ using the Cauchy-Schwarz inequality.

(2) $\hat{\beta}_{\text{ols}} = (X^TX)^{-1}X^T\underset{\sim}{Y} = \dfrac{\sum_{i=1}^n x_iY_i}{\sum_{i=1}^n x_i^2}$ $\quad X^TX\,\hat{b}_{\text{ols}} = X^T\underset{\sim}{Y}$

$X^TV^{-1}X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}^T \begin{pmatrix} x_1^2 & & \\ & \ddots & \\ & & x_n^2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ $\quad \underbrace{X^TV^{-1}X}_{n}\,\hat{b}_{\text{gls}} = X^TV^{-1}\underset{\sim}{Y}$

$$= \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}^T \begin{pmatrix} x_1^{-2} & & \\ & \ddots & \\ & & x_n^{-2} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$\underset{\sim}{x}^T A \underset{\sim}{x} = \sum_i \sum_j x_i A_{ij} x_j$$

$$= \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}^T \begin{pmatrix} x_1^{-1} & & \\ & \ddots & \\ & & x_n^{-1} \end{pmatrix} \begin{pmatrix} x_1^{-1} & & \\ & \ddots & \\ & & x_n^{-1} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$= \underset{\sim}{1}_n^T \underset{\sim}{1}_n$$

$$= n.$$

$$\hat{\beta}_{gls} = X^T V^{-1} \underset{\sim}{y} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}^T \begin{pmatrix} x_1^{-2} & & \\ & \ddots & \\ & & x_n^{-2} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \frac{1}{n} \sum_{i=1}^{2} \frac{x_i y_i}{x_i^2} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{x_i}$$

$$\hat{\sigma}^2_{ols} = \frac{\|\hat{e}\|^2}{n-1} = \frac{\|(I - P_x)\underset{\sim}{y}\|^2}{n-1} \Rightarrow$$

$$\mathbb{E}\,\hat{\sigma}^2_{ols} = \frac{1}{n-1} \mathbb{E}\,\|(I - P_x)\underset{\sim}{y}\|^2 = \frac{1}{n-1}\left[0 + \operatorname{tr}\left((I - P_x) \operatorname{Cov} \underset{\sim}{y}\right)\right]$$

$$= \frac{1}{n-1} \operatorname{tr}\left((I - P_x)\, \sigma^2 \begin{pmatrix} x_1^2 & & \\ & \ddots & \\ & & x_n^2 \end{pmatrix}\right)$$

$$P_x = x(\underset{\sim}{x^T x})^{-1} x^T$$

$$P_x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \frac{1}{\sum_{i=1}^{n} x_i^2} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}^T$$

$$= \frac{\sigma^2}{n-1}\left(\sum x_i^2 - \operatorname{tr}\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \frac{1}{\sum x_i^2} \begin{pmatrix} x_1^3 & & \\ & \ddots & \\ & & x_n^3 \end{pmatrix} \right)\right)$$

$$P_x \begin{pmatrix} x_1^2 & & \\ & \ddots & \\ & & x_n^2 \end{pmatrix}$$

$$= \frac{\sigma^2}{n-1}\left(\sum x_i^2 - \operatorname{tr}\left(\begin{pmatrix} x_1^4 & & \\ & \ddots & \\ & & x_n^4 \end{pmatrix} \frac{1}{\sum x_i^2} \right)\right)$$

$$= \frac{\sigma^2}{n-1}\left(\sum x_i^2 - \frac{\sum x_i^4}{\sum x_i^2} \right) \leftarrow \text{check}$$

$$\mathbb{E}\,\hat{\sigma}^2_{gls} = \sigma^2. \quad (\text{proved in general}).$$

$$\mathbb{E}\,\hat{\beta}_{ols} = \mathbb{E}\left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}\right) = \mathbb{E}\,\frac{\sum_{i=1}^{n} x_i\,(x_i\beta + \varepsilon_i|x_i|)}{\sum_{i=1}^{n} x_i^2} = \beta. \quad \text{Unbiased!}$$

⑥

$$\text{Var}\,\hat{\beta}_{ols} = \text{Var}\left((X^T X)^{-1} X^T \underset{\sim}{y}\right)$$

$$= (X^T X)^{-1} X^T \left[\sigma^2 V\right] X (X^T X)^{-1}$$

$$= \sigma^2 \frac{1}{\sum x_i^2} \underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}^T \begin{pmatrix} x_1^2 & & \\ & \ddots & \\ & & x_n^2 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}} \frac{1}{\sum x_i^2}$$

$$= \sigma^2 \frac{\sum x_i^4}{\left(\sum x_i^2\right)^2}$$

$$\sum a_i b_i \le \sqrt{\sum a_i^2}\sqrt{\sum b_i^2}$$

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} \underset{a_i}{1} \cdot \underset{b_i}{x_i^2} \le \sqrt{\sum_{i=1}^{n} 1^2}\sqrt{\sum_{i=1}^{n} x_i^4} = \sqrt{n}\sqrt{\sum_{i=1}^{n} x_i^4}$$

$$\Rightarrow \left(\sum_{i=1}^{n} x_i^2\right)^2 \le n \sum_{i=1}^{n} x_i^4$$

$$\text{Var}\,\hat{\beta}_{gls} = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} \frac{y_i}{x_i}\right)$$

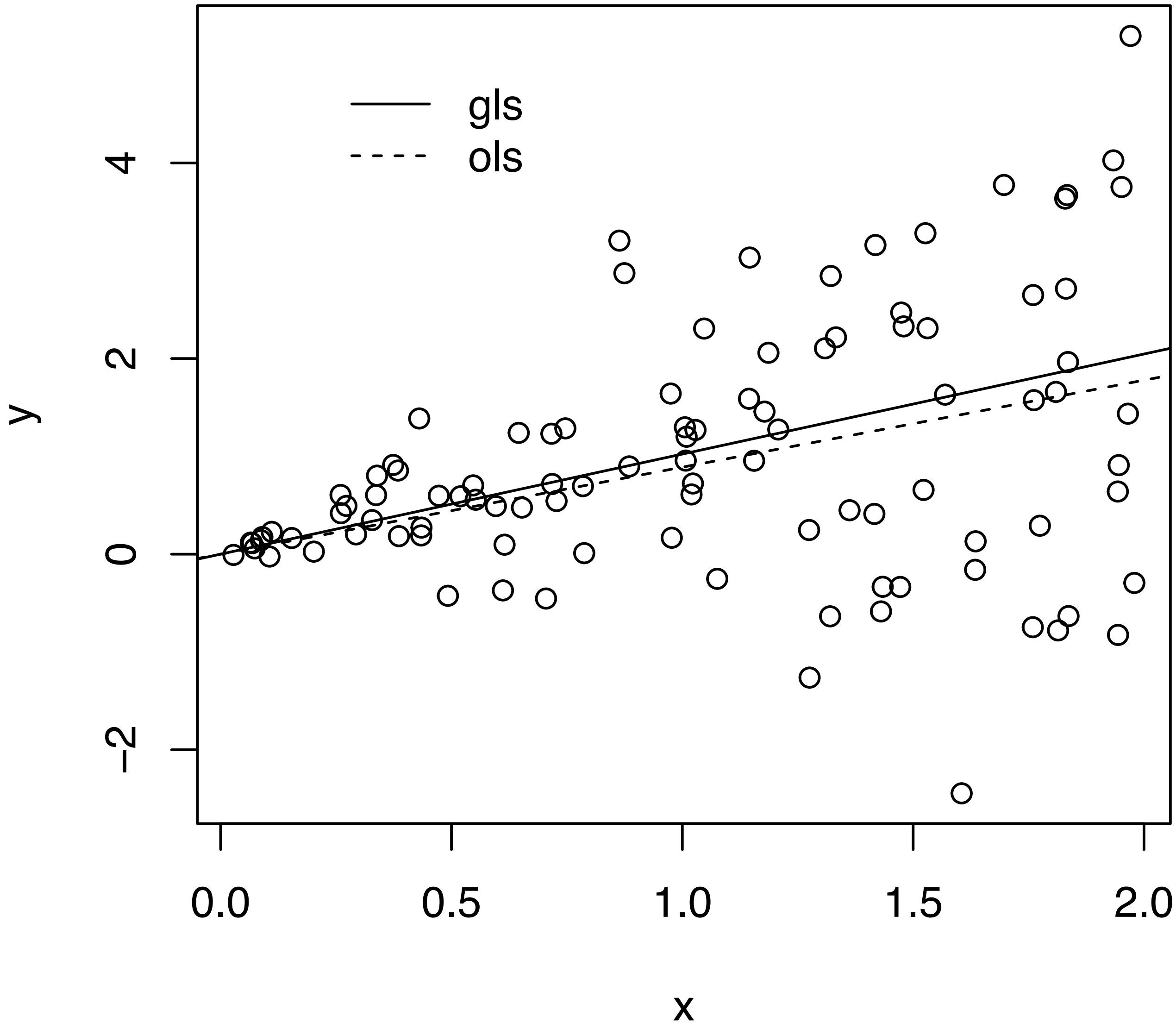
$$= \frac{1}{n^2}\sum_{i=1}^{n} \frac{\text{Var}\,y_i}{x_i^2}$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \frac{\sigma^2 x_i^2}{x_i^2}$$

$$= \frac{\sigma^2}{n}.$$

$$\frac{\sum x_i^4}{\left(\sum x_i^2\right)^2} \ge \frac{1}{n}$$

$$y_i = \beta x_i + \varepsilon|x_i|$$



$\varepsilon_i \mid x_i \mid$

Go back to the Gauss-Markov model, impose $\mathbf{P}^T\mathbf{b} = \delta$, and consider BLUE...

Theorem (Best linear unbiased estimator in the restricted G-M model)

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, where $\mathbb{E}\mathbf{e} = \mathbf{0}$ and $\operatorname{Cov}\mathbf{e} = \sigma^2\mathbf{I}_n$, with the restriction $\mathbf{P}^T\mathbf{b} = \delta$.

Let $\hat{\mathbf{b}}_H$ and $\hat{\mathbf{u}}$ satisfy $\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}}_H \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \delta \end{bmatrix}$ and let $\mathbf{c}^T\mathbf{b}$ be estimable.

Then $\mathbf{c}^T\hat{\mathbf{b}}_H$ is the BLUE for $\mathbf{c}^T\mathbf{b}$ in the restricted model.

See Res 4.5 of Monahan (2008).

Prove the above in the steps:

1. Show consistency of the equations $\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \hat{\mathbf{v}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{c} \\ \mathbf{0} \end{bmatrix}$. Lemma 4.2 of M.
2. Show $\mathbb{E}\mathbf{c}^T\hat{\mathbf{b}}_H = \mathbf{c}^T\mathbf{b}$ for all $\mathbf{b}$. Lemma 4.3 of M.
3. Show that any unbiased estimator $a_0 + \mathbf{a}^T\mathbf{y}$ has variance $\geq \operatorname{Var}\mathbf{c}^T\hat{\mathbf{b}}_H$.

Monahan, J. F. (2008). *A primer on linear models*. CRC Press.