"secures" unbiasedness

An unbiased estimator of σ^2 under the GMM is given by

SSE

$$\widehat{\sigma}^2 \equiv \frac{\boldsymbol{y}^\top(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y}}{n-r}, \text{ where } r = rank(\boldsymbol{X}).$$

idempotent symmetry

Because $\boldsymbol{I}-\boldsymbol{P_X} \overset{\text{idempotent}}{=} (\boldsymbol{I}-\boldsymbol{P_X})(\boldsymbol{I}-\boldsymbol{P_X}) \overset{\text{symmetry}}{=} (\boldsymbol{I}-\boldsymbol{P_X})^\top(\boldsymbol{I}-\boldsymbol{P_X})$,

$(I - P_x)$

$$\begin{aligned}
\boldsymbol{y}^\top(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y} &= \boldsymbol{y}^\top\boxed{(\boldsymbol{I}-\boldsymbol{P_X})^\top}(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y} \\
&= \{(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y}\}^\top\{(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y}\} \\
&= ||(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y}||^2 = ||\boldsymbol{y}-\boldsymbol{P_X}\boldsymbol{y}||^2 \\
&= ||\boldsymbol{y}-\widehat{\boldsymbol{y}}||^2 \\
&= \text{“Sum of Squared Errors” (SSE).}
\end{aligned}$$

Gauss-Markov Model with Normal Errors (GMMNE)

Suppose

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where

- $\boldsymbol{y} \in \mathbb{R}^n$ is the response vector,
- $\boldsymbol{X}$ is an $n \times p$ matrix of known constants,
- $\boldsymbol{\beta} \in \mathbb{R}^p$ is an unknown parameter vector, and
- $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$ for some unknown variance parameter $\sigma^2 \in \mathbb{R}^+$.

Why? linear transformations of normally distr. RV are also normally distributed

The GMMNE is a special case of the GMM.

We have added the assumption $\boldsymbol{\epsilon}$ is multivariate normal.

The GMMNE implies $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$.

The GMMNE is useful for drawing statistical inferences regarding estimable $\boldsymbol{C\beta}$.

Throughout the remainder of these slides we will assume

- the GMMNE model holds,
- $\boldsymbol{C}$ is a $q \times p$ matrix such that $\boldsymbol{C\beta}$ is estimable,
- $\boxed{rank(\boldsymbol{C}) = q,}$ and
- $\boldsymbol{d}$ is a known $q \times 1$ vector.

Specifies what we assume under H_0 about $C\beta$

These assumptions imply $\boxed{H_0 : \boldsymbol{C\beta} = \boldsymbol{d} \text{ is a \textit{testable hypothesis}.}}$

The Distribution of $C\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$

In the GMMNE model, it can be shown that $C\widehat{\boldsymbol{\beta}}$ follows a Normal distribution with mean and variance given as follows:

$X\hat{\beta} \sim N(X\beta, \sigma^2 I)$

Distribution of $C\widehat{\boldsymbol{\beta}}$

$$\boldsymbol{C}\widehat{\boldsymbol{\beta}} \sim \mathcal{N}\left(\boldsymbol{C}\boldsymbol{\beta},\ \underline{\sigma^2 \boldsymbol{C}(\boldsymbol{X}^\top \boldsymbol{X})^- \boldsymbol{C}^\top}\right)$$

The distribution of $\widehat{\sigma}^2$ is a scaled χ^2_{n-r} distribution:

scaled χ^2 random variable

Distribution of $\widehat{\sigma}^2$

$$\frac{(n-r)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-r} \Longleftrightarrow \frac{\widehat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2_{n-r}}{n-r} \Longleftrightarrow \widehat{\sigma}^2 \sim \boxed{\frac{\sigma^2}{n-r}\chi^2_{n-r}}$$

Note that $\boxed{C\widehat{\boldsymbol{\beta}} \text{ and } \widehat{\sigma}^2 \text{ are independent.}}$

slide 25

To test $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$, we can use the following statistic

$$\begin{aligned} F &\equiv (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})^\top [\widehat{\text{Var}}(\boldsymbol{C}\widehat{\boldsymbol{\beta}})]^{-1} (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})/q \\ &= (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})^\top [\widehat{\sigma}^2 \boldsymbol{C}(\boldsymbol{X}^\top \boldsymbol{X})^- \boldsymbol{C}^\top]^{-1} (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})/q \\ &= \frac{(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})^\top [\boldsymbol{C}(\boldsymbol{X}^\top \boldsymbol{X})^- \boldsymbol{C}^\top]^{-1} (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})/q}{\widehat{\sigma}^2}. \end{aligned}$$

F has a non-central F–distribution with non-centrality parameter θ

$$\theta = \frac{(\boldsymbol{C\beta} - \boldsymbol{d})^\top [\boldsymbol{C}(\boldsymbol{X}^\top \boldsymbol{X})^- \boldsymbol{C}^\top]^{-1} (\boldsymbol{C\beta} - \boldsymbol{d})}{2\sigma^2}$$

and df q and $n - r$. — df associated with SSE

rank(C)

The F–Test continued ($H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$)

The non-negative non-centrality parameter

$$\frac{(\boldsymbol{C\beta} - \boldsymbol{d})^\top [\boldsymbol{C}(\boldsymbol{X}^\top \boldsymbol{X})^- \boldsymbol{C}^\top]^{-1} (\boldsymbol{C\beta} - \boldsymbol{d})}{2\sigma^2}$$

is equal to zero if and only if $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$ is true.

$\Longleftrightarrow$

$\theta = 0$

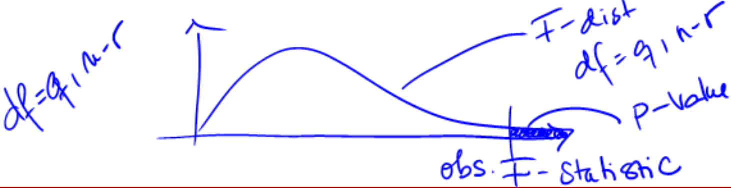
If $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$ is true, the statistic F has a central F–distribution with q and $n - r$ degrees of freedom ($F_{q,n-r}$).

use p-values to assess the strength evidence against H_0 and in favor of the alternative H_a.

Thus, to test $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$, we compute the test statistic F and compare the observed value of F to the $F_{q,n-r}$-distribution.

If F is so large that it seems unlikely to have been a draw from the $F_{q,n-r}$-distribution, we reject H_0 and conclude $\boldsymbol{C\beta} \neq \boldsymbol{d}$.

The p-value of the test is the probability that a random variable with distribution $F_{q,n-r}$ matches or exceeds the observed value of the test statistic F.



The t–Test ($H_0 : \boldsymbol{c}^\top \boldsymbol{\beta} = d$) for estimable $\boldsymbol{c}^\top \boldsymbol{\beta}$

$q = 1$

$c^\top$ is a row vector,

d is a scalar

The test statistic

$$t \equiv \frac{\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} - d}{\sqrt{\widehat{\mathrm{Var}}(\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}})}} = \frac{\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} - d}{\sqrt{\widehat{\sigma}^2 \boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^- \boldsymbol{c}}}.$$

t has a non-central t-distribution with non-centrality parameter

$$\frac{\boldsymbol{c}^\top \boldsymbol{\beta} - d}{\sqrt{\sigma^2 \boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^- \boldsymbol{c}}}$$

and df$= n - r$.

The t–Test (continued)

The non-centrality parameter

$$\frac{\boldsymbol{c}^\top \boldsymbol{\beta} - d}{\sqrt{\sigma^2 \boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^- \boldsymbol{c}}}$$

is equal to zero if and only if $H_0 : \boldsymbol{c}^\top \boldsymbol{\beta} = d$ is true.

If $H_0 : \boldsymbol{c}^\top \boldsymbol{\beta} = d$ is true, the statistic t has a central t-distribution with $n - r$ degrees of freedom (t_{n-r}).

The t–Test (continued)

large values of t compared to what we should see under H_0 are evidence

Thus, to test $H_0 : \boldsymbol{c}^\top \boldsymbol{\beta} = d$, we compute the test statistic t and compare the observed value of t to the t_{n-r}-distribution. against H_0 and in favor of H_a

If t is so far from zero that it seems unlikely to have been a draw from the t_{n-r}-distribution, we reject H_0 and conclude $\boldsymbol{c}^\top \boldsymbol{\beta} \neq d$.

The p-value of the test is the probability that a random variable with distribution t_{n-r} would be as far or farther from 0 than the observed value of the t test statistic.

A $100(1-\alpha)\%$ Confidence Interval for Estimable $\boldsymbol{c}^\top\boldsymbol{\beta}$

allows us to judge the practical value of a statistically significant difference, keep in

A $100(1-\alpha)\%$ confidence interval for estimable $\boldsymbol{c}^\top\boldsymbol{\beta}$ is given as

mind as

$$\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}} \pm t_{n-r,1-\alpha/2}\sqrt{\widehat{\sigma}^2\boldsymbol{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^-\boldsymbol{c}}$$

long as we have a sufficiently large sample

estimate $\pm$ (distribution quantile) $\times$ (estimated standard error)

size, the smallest difference between $c^\top\beta - d$ can be made statistically significant!

Form of the t Statistic for Testing $H_0 : \boldsymbol{c}^\top \boldsymbol{\beta} = d$

$$t \quad = \quad \frac{\text{estimate} - d}{\text{estimated standard error}} = \frac{\text{estimate} - d}{\sqrt{\widehat{\text{Var}}(\text{estimator})}}$$

$$\begin{aligned} t^2 \quad &= \quad \frac{(\text{estimate} - d)^2}{\widehat{\text{Var}}(\text{estimator})} \\ &= \quad (\text{estimate} - d)\left[\widehat{\text{Var}}(\text{estimator})\right]^{-1}(\text{estimate} - d)/1 \end{aligned}$$

Revisiting the F Statistic for Testing $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$

$$
\begin{aligned}
F &= (\textbf{estimate} - \boldsymbol{d})^\top \left[\widehat{\mathrm{Var}}(\textbf{estimator})\right]^{-1} (\textbf{estimate} - \boldsymbol{d})/q \\
&= (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})^\top [\widehat{\mathrm{Var}}(\boldsymbol{C}\widehat{\boldsymbol{\beta}})]^{-1} (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})/q \\
&= (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})^\top [\widehat{\sigma}^2 \boldsymbol{C}(\boldsymbol{X}^\top \boldsymbol{X})^- \boldsymbol{C}^\top]^{-1} (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})/q \\
&= \frac{(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})^\top [\boldsymbol{C}(\boldsymbol{X}^\top \boldsymbol{X})^- \boldsymbol{C}^\top]^{-1} (\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})/q}{\widehat{\sigma}^2}
\end{aligned}
$$

rank(C)

end lecture 3

01-27-25