Lecture 1: Recap of Linear Model for Cross Section Data and Ordinary Least Squares

Bertel Schjerning

University of Copenhagen, Department of Economics

January 29, 2026

Plan for Lectures on Linear Models

Lecture 1: Recap, Cross section, Least squares (W.4)

Lecture 2: Panel, Fixed effects, First differences (W.10)

Lecture 3: Panel, Random effects, Hausman test (W.10)

Cross Section Data

i	y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$\cdots$	$x^{(K)}$
1	y_1	1	$x_1^{(2)}$	$x_1^{(3)}$	$\cdots$	$x_1^{(K)}$
2	y_2	1	$x_2^{(2)}$	$x_2^{(3)}$	$\cdots$	$x_2^{(K)}$
3	y_3	1	$x_3^{(2)}$	$x_3^{(3)}$	$\cdots$	$x_3^{(K)}$
4	y_4	1	$x_4^{(2)}$	$x_4^{(3)}$	$\cdots$	$x_4^{(K)}$
5	y_5	1	$x_5^{(2)}$	$x_5^{(3)}$	$\cdots$	$x_5^{(K)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
N	y_N	1	$x_N^{(2)}$	$x_N^{(3)}$	$\cdots$	$x_N^{(K)}$

Sampling Scheme for (Micro) Data

In this course **we will assume that:**

- Cross-sectional units (i) are $i$ndependent.
- Observations identically distributed.

Focus: asymptotics as #cross section units grows w/o bound

- Why? Finite-sample results rarely available.
- Limits ("arrows") understood as $N \to \infty$ (all else fixed).
- Implicit assumption for asymptotics to be relevant:

$$N \text{ is (relatively) large.}$$

- I.e., your dataset looks "long and narrow"

Today

- Another look at linear model and OLS in cross section.
- In part, refresher... but also more formal approach.
  - Will argue using asymptotics.
- Useful benchmark for later (panel) lectures.
  - Will essentially transform to cross section.

Outline

Model and Identification

Identification

Equation of interest in error form

$$y = \mathbf{x}\boldsymbol{\beta} + u.$$

Assuming $\mathrm{E}[u \mid \mathbf{x}] = 0$, we get

$$\mathrm{E}[y \mid \mathbf{x}] = \mathbf{x}\boldsymbol{\beta}$$

Identifying assumptions:

OLS.1: $\mathrm{E}[\mathbf{x}'u] = \mathbf{0}$

OLS.2: rank $\mathrm{E}[\mathbf{x}'\mathbf{x}] = K$

Note: $\mathrm{E}[u \mid \mathbf{x}] = 0$ is stronger than OLS.1

...but not necessary for identification of $\boldsymbol{\beta}$.

Identification Argument

Premultiply y equation by $\mathbf{x}'$ and take expectations

$$\mathrm{E}\left[\mathbf{x}'y\right] = \mathrm{E}\left[\mathbf{x}'\mathbf{x}\right]\boldsymbol{\beta} + \mathrm{E}\left[\mathbf{x}'u\right].$$

Under OLS.1. we get

$$\mathrm{E}\left[\mathbf{x}'y\right] = \mathrm{E}\left[\mathbf{x}'\mathbf{x}\right]\boldsymbol{\beta}$$

Under OLS.2, $\mathrm{E}[\mathbf{x}'\mathbf{x}]$ is invertible, so

$$\boldsymbol{\beta} = \left(\mathrm{E}\left[\mathbf{x}'\mathbf{x}\right]\right)^{-1}\mathrm{E}\left[\mathbf{x}'y\right]$$

$\mathrm{E}[\mathbf{x}'\mathbf{x}]$ and $\mathrm{E}[\mathbf{x}'y]$ features of distribution of observables $(y, \mathbf{x})$

$\implies \boldsymbol{\beta}$ is identified

Estimation and Consistency

Convergence in Probability

- $\{W_N\}_{N=1}^{\infty}$ real-valued random variables (r.v.'s)
- $c \in \mathbb{R}$ non-random number (for now)

W_N converges in probability to c if for each $\varepsilon > 0$,

$$\mathrm{P}\left(|W_N - c| > \varepsilon\right) \to 0 \text{ as } N \to \infty.$$

c is then the probability limit of W_N.

- Notation: $W_N \to_p c$ or $c = p\text{-}\lim W_N$.

Convergence in Probability

Interpretation?

- In large samples, probability mass clusters at c
- In limit, distribution of W_N collapses onto c

How do deal with $\mathbf{W}_N$ and $\mathbf{c}$ vectors (of fixed length K)?

Convergence in probability interpreted coordinatewise:

$\mathbf{W}_N \to_p \mathbf{c} \Leftrightarrow W_{N,j} \to_p c_j$ for all $j \in \{1, \ldots, K\}$.

[Matrices similar. Go through all (j, k) pairs.]

Consistency

- Sample of observations $Z_1, Z_2, \ldots, Z_N$ of size N.
- W_N an estimator of c based on $\{Z_i\}_{i=1}^N$

If $W_N \to_p c$, W_N is called consistent for c.

- In large samples, W_N *tends to be* close to c.
- Consistency is a first-order concern.
- Inconsistent estimators are typically not interesting.

Example

- Sps. access to $\{Z_i\}_{i=1}^{N}$ independently distributed $\mathrm{N}\left(\mu, \sigma^2\right)$.
- Object of interest: μ.

Discussion:

- Estimator of μ?
- Consistent?

Working with Probability Limits

W. Lemma 3.4 (“**Slutsky’s Theorem**”)

- If $\{\mathbf{W}_N\}_{N=1}^{\infty}$ $\mathbb{R}^K$-valued r.v.’s,
- $\mathbf{W}_N \rightarrow_p \mathbf{c} \in \mathbb{R}^K$ non-random, and
- $\mathbf{g} : \mathbb{R}^K \rightarrow \mathbb{R}^J$ is $\underline{\text{continuous at } \mathbf{c}}$, then

$$\mathbf{g}\left(\mathbf{W}_N\right) \xrightarrow{p} \mathbf{g}\left(\mathbf{c}\right).$$

That is,

$$p\text{-}\lim \mathbf{g}\left(\mathbf{W}_N\right) = \mathbf{g}\left(p\text{-}\lim \mathbf{W}_N\right).$$

Probability limits pass through continuous functions

Example (cont.)

- Sps. access to $\{Z_i\}_{i=1}^{N} \sim \mathrm{N}\left(\mu, \sigma^2\right)$ independent

- Know $\overline{Z}_N \to_p \mu$

Discussion: Does...

- $\overline{Z}_N^2 \to_p ?$

- $\cos(\overline{Z}_N) \to_p ?$

- $1/\overline{Z}_N \to_p ?$

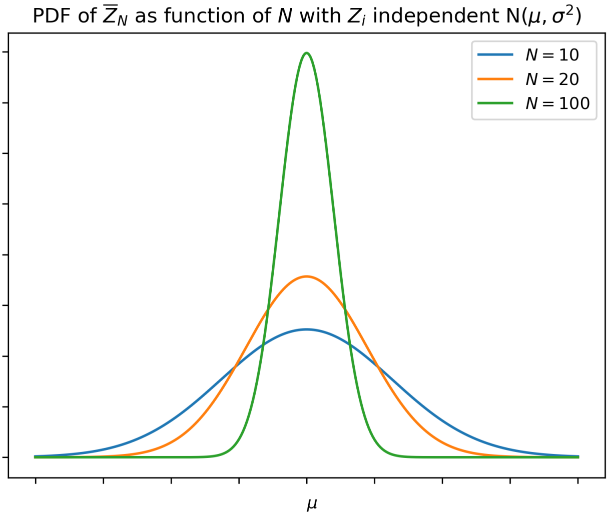

Figure: Convergence in Probability

Estimation in Linear Model

Suppose we have iid data $\{(y_i, \mathbf{x}_i)\}_{i=1}^N$ satisfying

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i, \quad i = 1, 2, \ldots, N.$$

Analogy principle: Replace unknowns with (consistent) estimators.

Identification result: Under OLS.1–2,

$$\boldsymbol{\beta} = \left(\mathrm{E}\left[\mathbf{x}'\mathbf{x}\right]\right)^{-1} \mathrm{E}\left[\mathbf{x}'y\right].$$

Law of large numbers (LLN) suggests

$$\widehat{\boldsymbol{\beta}} := \left(\frac{1}{N}\sum_{i=1}^N \mathbf{x}_i'\mathbf{x}_i\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^N \mathbf{x}_i' y_i\right) \tag{OLS}$$

OLS Consistency

Inserting model,

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i' u_i \right).$$

By random sampling + LLN,

$$\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i' u_i \xrightarrow{p} \mathrm{E}\left[\mathbf{x}' u\right] \overset{\text{OLS.1}}{=} \mathbf{0}.$$

OLS Consistency

Similarly

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i \xrightarrow{p} \mathrm{E}\left[\mathbf{x}'\mathbf{x}\right],$$

which is invertible (OLS.2).

So Slutsky's theorem yields

$$p\text{-lim}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right)^{-1} \overset{\text{OLS.2}}{=} \left(\mathrm{E}\left[\mathbf{x}'\mathbf{x}\right]\right)^{-1}$$

- Invertible = matrix version of "no division by zero"

OLS Consistency

Put together: Slutsky's theorem (w/ product mapping)

$$\begin{aligned}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'u_i\right) &\xrightarrow{p} \left(\mathrm{E}\left[\mathbf{x}'\mathbf{x}\right]\right)^{-1}\mathrm{E}\left[\mathbf{x}'u\right] \\ &= \left(\mathrm{E}\left[\mathbf{x}'\mathbf{x}\right]\right)^{-1}\mathbf{0} \\ &= \mathbf{0}.\end{aligned}$$

Conclude: $\widehat{\boldsymbol{\beta}} \to_p \boldsymbol{\beta}$.

OLS.1–2 imply consistency of OLS.

Inference?

Asymptotic Normality

Convergence in Distribution

- $\{W_N\}_{N=1}^{\infty}$ real-valued r.v.'s (for now)
- Let F_N be cumulative distribution function (CDF) of W_N,

$$F_N(w) := \mathrm{P}(W_N \leqslant w), \quad w \in \mathbb{R}.$$

$\{W_N\}_{N=1}^{\infty}$ converges in distribution to r.v. W_∞ with CDF F_∞ if

$$F_N(w) \to F_\infty(w) \text{ as } N \to \infty$$

at every point w for which F_∞ is continuous.

- Notation: $W_N \to_d W_\infty$ or $W_\infty = d\text{-}\lim W_N$

Convergence in Distribution

Interpretation?

- In large samples, W_N distributed approximately as W_∞.
- That is, the graphs of F_N and F_∞ are close.

If $\mathbf{W}_N$ and $\mathbf{W}_\infty$ (K-)vectors, understand $\leqslant$ coordinatewise, e.g.

$$F_N(\mathbf{w}) = \mathrm{P}\left(W_{N,1} \leqslant w_1, \ldots, W_{N,K} \leqslant w_K\right).$$

Otherwise same definition.

Example (cont.)

- Sps. again $\{Z_i\}_{i=1}^{N} \sim \mathrm{N}\left(\mu, \sigma^2\right)$ independent
- By properties of normals,

$$\frac{\overline{Z}_N - \mu}{\sigma/\sqrt{N}} \overset{d}{=} \mathrm{N}\left(0, 1\right)$$

$$\Leftrightarrow \mathrm{P}\left(\frac{\overline{Z}_N - \mu}{\sigma/\sqrt{N}} \leqslant w\right) = \Phi\left(w\right) \text{ for any } w \in \mathbb{R}.$$

- RHS independent of N, so certainly

$$\frac{\overline{Z}_N - \mu}{\sigma/\sqrt{N}} \xrightarrow{d} \mathrm{N}\left(0, 1\right).$$

- Exact normality unrealistic.
- Latter true even if only $\{Z_i\}_{i=1}^{N} \sim_{\text{indep.}} ?\left(\mu, \sigma^2\right)$

Central Limit Theorem

- We'll get $\rightarrow_d$ by means of a central limit theorem (CLT).

Theorem (**CLT, W. Theorem 3.2**)

If $\{\mathbf{Z}_i\}_{i=1}^{\infty}$ *iid* $\mathbb{R}^K$*-valued r.v.'s with zero mean, then*

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{Z}_i \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \mathrm{E}\left[\mathbf{Z}\mathbf{Z}'\right]\right) \text{ as } N \to \infty.$$

- Hence, even if distribution of $\mathbf{Z}_i$ unknown:

  Scaled averages of mean-zero r.v.'s are approximately normal in large samples.

Asymptotic Normality of OLS

$$\text{Recall}: \quad \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i' u_i \right)$$

Rearrange and scale to get

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{x}_i' u_i \right).$$

OLS.2 implies $(N^{-1} \sum_{i=1}^{N} \mathbf{x}_i' \mathbf{x}_i)^{-1} \to_p (\mathrm{E}[\mathbf{x}'\mathbf{x}])^{-1}$.
By random sampling + CLT + OLS.1,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{x}_i' u_i \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \mathrm{E}\left[u^2 \mathbf{x}'\mathbf{x}\right]\right).$$

Combining Modes of Convergence

Product Rule: If

1. $\mathbf{Y}_N \to_p \mathbf{C}$ constant (number/vector/matrix), and
2. $\mathbf{Z}_N \to_d \mathbf{Z}$,

then $\mathbf{Y}_N \mathbf{Z}_N \to_d \mathbf{CZ}$.

Will use this product rule again and again...

Warning: Constancy of $\mathbf{C}$ cannot be disposed of.

Asymptotic Normality of OLS

It follows that

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i\right)^{-1} \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \mathbf{x}_i' u_i\right)$$

$$\xrightarrow{d} \left(\mathrm{E}\left[\mathbf{x}'\mathbf{x}\right]\right)^{-1} \mathrm{N}\left(\mathbf{0}, \mathrm{E}\left[u^2\mathbf{x}'\mathbf{x}\right]\right). \quad \text{(Product Rule)}$$

Normal property:

1. If $\mathbf{C} \in \mathbb{R}^{L\times K}$ is constant (conformable matrix)
2. and $\mathbf{Z} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$ in $\mathbb{R}^K$,

then $\mathbf{CZ} \sim \mathrm{N}\left(\mathbf{0}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}'\right)$ in $\mathbb{R}^L$.

Normal family closed under linear transformations.

Asymptotic Normality of OLS

Conclude:

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right),$$
$$\text{where} \quad \mathbf{A} := \mathrm{E}\left[\mathbf{x}'\mathbf{x}\right],$$
$$\mathbf{B} := \mathrm{E}\left[u^2\mathbf{x}'\mathbf{x}\right],$$

Under OLS.1–2, OLS is ($\sqrt{N}$-)asymptotically normal.

Variance Estimation

Asymptotic Variance Estimation

Treating $\rightarrow_d$ as $=_d$ (or $\sim$), asymptotic distribution suggests

$$\widehat{\boldsymbol{\beta}} \overset{d}{\approx} \mathrm{N}\left(\boldsymbol{\beta}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}/N\right).$$

Interpretation:

- In large samples, OLS approximately ($\approx_d$) normal.
- (W. also uses $\sim_a$, but not great notation.)

Variance parts $\mathbf{A}$ and $\mathbf{B}$ typically unknown. To be estimated...

Asymptotic Variance Estimation

Potential source of confusion:

Limit theory implies

$$\mathrm{var}\left[\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right] \to \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1} =: \mathbf{V}.$$

So $\mathbf{V}$ is the limit variance of $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$.

Call $\mathrm{Avar}(\widehat{\boldsymbol{\beta}}) := \mathbf{V}/N$ the asymptotic variance of $\widehat{\boldsymbol{\beta}}$.

"$\widehat{\mathbf{V}}/N$ consistently estimates $\mathrm{Avar}(\widehat{\boldsymbol{\beta}})$" means "$\widehat{\mathbf{V}} \to_p \mathbf{V}$."

Convenient, but imprecise, shorthand.

Asymptotic Variance Estimation

A consistent estimator $\widehat{\mathrm{Avar}(\widehat{\beta})}$ of $\mathrm{Avar}(\widehat{\beta})$ is

$$\left(\sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^{N} \widehat{u}_i^2 \mathbf{x}_i'\mathbf{x}_i\right) \left(\sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i\right)^{-1}, \quad \widehat{u}_i := y_i - \mathbf{x}_i\widehat{\beta}.$$

Just

$$\widehat{\mathbf{V}}/N := \widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1}/N$$

having applied the analogy principle to get

$$\widehat{\mathbf{A}} := \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i \quad \text{and} \quad \widehat{\mathbf{B}} := \frac{1}{N}\sum_{i=1}^{N} \widehat{u}_i^2 \mathbf{x}_i'\mathbf{x}_i.$$

Robust to cond'l heteroskedasticity ($\mathrm{E}\left[u^2 \mid \mathbf{x}\right]$ can vary w/ $\mathbf{x}$).

Efficiency

Relative Efficiency

Suppose:

1. Both $\widehat{\boldsymbol{\theta}}^{(1)}$ and $\widehat{\boldsymbol{\theta}}^{(2)}$ $\sqrt{N}$-asymptotically normal ($\sqrt{N}$-a.n.)
2. $\mathbf{V}^{(j)} = \lim_{N\to\infty} \text{var}[\sqrt{N}(\widehat{\boldsymbol{\theta}}^{(j)} - \boldsymbol{\theta})], j = 1, 2.$

Makes sense to choose most precise procedure

... i.e. one w/ "smallest" (limit) variance.

$\widehat{\boldsymbol{\theta}}^{(1)}$ asymptotically efficient relative to $\widehat{\boldsymbol{\theta}}^{(2)}$ if $\mathbf{V}^{(2)} - \mathbf{V}^{(1)}$ positive semidefinite.

Note: A <u>second</u>-order concern.

- Don't care about precision of inconsistent estimator.

Efficiency

OLS.3: $\mathrm{E}\left[u^2 \mathbf{x}'\mathbf{x}\right] = \sigma^2 \mathrm{E}\left[\mathbf{x}'\mathbf{x}\right]$

- Implies OLS also asymptotically efficient
  - Relative to *all* $\sqrt{N}$-a.n. estimators linear in $\mathbf{y}$.
- Simpler ("non-robust") variance estimator

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\beta}})} = \widehat{\sigma}^2 \left(\sum_{i=1}^{N} \mathbf{x}_i' \mathbf{x}_i \right)^{-1}, \quad \widehat{\sigma}^2 := \frac{1}{N-K} \sum_{i=1}^{N} \widehat{u}_i^2 .$$

If OLS 3. is violated?

$\widehat{\boldsymbol{\beta}}$ need not have smallest variance.

Inconsistent $\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\beta}})} \Rightarrow$ rely on robust version for inference.

Testing Linear Hypotheses

Testing Linear Hypotheses

Interest in $\mathrm{H}_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$.

- $\mathbf{R}$: $Q \times K$ with $\operatorname{rank} \mathbf{R} = Q \leqslant K$
- $\mathbf{r}$: $Q \times 1$

Wald statistic:

$$W := \left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left[\mathbf{R}\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\beta}})}\mathbf{R}'\right]^{-1} \left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}\right).$$

Under H_0, $W \to_d \chi^2_Q$.

Wald test:

$$\text{Reject } \mathrm{H}_0 \text{ at level } \alpha \Leftrightarrow W > (1-\alpha)\text{-quantile of } \chi^2_Q.$$

On Wald Statistic Form

Fact: If $\mathbf{Z} \sim \mathrm{N}\left(\mathbf{0}_{G\times 1}, \mathbf{\Sigma}\right)$ then $\mathbf{Z}'\mathbf{\Sigma}^{-1}\mathbf{Z} \sim \chi^2_G$.

- Multivariate version of $Z \sim \mathrm{N}\left(0, \sigma^2\right) \Rightarrow \left(Z/\sigma\right)^2 \sim \chi^2_1$.

OLS.1–2 imply $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \approx_d \mathrm{N}\left(\mathbf{0}_{K\times 1}, \mathbf{V}/N\right)$.

Linearly transform: $\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta} \approx_d \mathrm{N}\left(\mathbf{0}_{Q\times 1}, \mathbf{R}\left(\mathbf{V}/N\right)\mathbf{R}'\right)$

Conclude:

$$(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})'\left[\mathbf{R}\left(\mathbf{V}/N\right)\mathbf{R}'\right]^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}) \overset{d}{\approx} \chi^2_Q.$$

Under H_0, $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$.

Wald arises from consistent $\mathrm{Avar}(\widehat{\boldsymbol{\beta}})$ estimator.

Example Hypotheses ($K = 3$)

1. "$\beta_1 = \beta_2$" corresponds to

2. "$\beta_1 + \beta_2 = 1$" corresponds to

3. "$\beta_1 = \beta_2 = \beta_3$" corresponds to