UNIVERSITÄT
BIELEFELD
Faculty of Business Administration
and Economics

Faster and more robust maximum likelihood estimation for random utility models

Lennart Oelschläger Dietmar Bauer

Bielefeld University, Empirical Methods Department, Econometrics Group

ICMC April 1-3 2024 in Puerto Varas, Chile

Faster and more robust MLE for RUMs

Faster and more robust MLE for RUMs

$$\rightarrow \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \boldsymbol{U}$$

UNIVERSITÄT
BIELEFELD
Faculty of Business Administration
and Economics

Faster and more robust MLE for RUMs

$$\rightarrow \quad \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\beta + \gamma) + \varepsilon$$

Faster and more robust MLE for RUMs

$$\rightarrow \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \boldsymbol{U} = \mu + \boldsymbol{X}(\beta + \gamma) + \varepsilon$$

- μ are alternative-specific constants

Faster and more robust MLE for RUMs

$$\rightarrow \quad \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\beta + \gamma) + \varepsilon$$

- $\boldsymbol{\mu}$ are alternative-specific constants
- $\boldsymbol{X}$ is a matrix of decider and / or alternative attributes

Faster and more robust MLE for RUMs

$$\rightarrow \quad \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\boldsymbol{\beta} + \boldsymbol{\gamma}) + \boldsymbol{\varepsilon}$$

- $\boldsymbol{\mu}$ are alternative-specific constants
- $\boldsymbol{X}$ is a matrix of decider and / or alternative attributes
- $\boldsymbol{\beta}$ encodes how $\boldsymbol{X}$ influences $\boldsymbol{U}$

Faster and more robust MLE for RUMs

$$\rightarrow \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\boldsymbol{\beta} + \boldsymbol{\gamma}) + \boldsymbol{\varepsilon}$$

- $\boldsymbol{\mu}$ are alternative-specific constants
- $\boldsymbol{X}$ is a matrix of decider and / or alternative attributes
- $\boldsymbol{\beta}$ encodes how $\boldsymbol{X}$ influences $\boldsymbol{U}$
- $\boldsymbol{\gamma}$ are random effects, typically $\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega})$

Faster and more robust MLE for RUMs

$$\rightarrow \quad \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\boldsymbol{\beta} + \boldsymbol{\gamma}) + \boldsymbol{\varepsilon}$$

- $\boldsymbol{\mu}$ are alternative-specific constants
- $\boldsymbol{X}$ is a matrix of decider and / or alternative attributes
- $\boldsymbol{\beta}$ encodes how $\boldsymbol{X}$ influences $\boldsymbol{U}$
- $\boldsymbol{\gamma}$ are random effects, typically $\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$
- $\boldsymbol{\varepsilon}$ is a random error, typically $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ (probit) or $\boldsymbol{\varepsilon} \sim$ iid EV (logit)

Faster and more robust MLE for RUMs

$$\rightarrow \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\beta + \gamma) + \varepsilon \rightarrow \text{choice } y = \arg\max \boldsymbol{U}$$

- $\boldsymbol{\mu}$ are alternative-specific constants
- $\boldsymbol{X}$ is a matrix of decider and / or alternative attributes
- β encodes how $\boldsymbol{X}$ influences $\boldsymbol{U}$
- γ are random effects, typically $\gamma \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega})$
- ε is a random error, typically $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ (probit) or $\varepsilon \sim$ iid EV (logit)

Faster and more robust MLE for RUMs

$$\rightarrow \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\beta + \gamma) + \varepsilon \rightarrow \text{choice } y = \arg\max \boldsymbol{U}$$

Faster and more robust MLE for RUMs

$$\rightarrow \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\beta + \gamma) + \varepsilon \rightarrow \text{choice } y = \arg\max \boldsymbol{U}$$

Let $\boldsymbol{\theta}$ denote the set of model parameters and assume a distribution for ε is fixed.

Faster and more robust MLE for RUMs

$$\rightarrow \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\beta + \gamma) + \varepsilon \quad \rightarrow \quad \text{choice } y = \arg\max \boldsymbol{U}$$

Let $\boldsymbol{\theta}$ denote the set of model parameters and assume a distribution for ε is fixed.

1. calculate choice probabilities $P_{nj}(\boldsymbol{X}_n, \boldsymbol{\theta})$ for each decider n and alternative j
2. build log-likelihood function $\log \mathcal{L}(\theta \mid \boldsymbol{X}, y) = \sum_{n,j} 1(y_n = j) \log P_{nj}(\boldsymbol{X}_n, \theta)$
3. maximize $\log \mathcal{L}(\theta \mid \boldsymbol{X}, y)$ with respect to θ

Faster and more robust MLE for RUMs

$$\rightarrow \quad \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\beta + \gamma) + \varepsilon \quad \rightarrow \quad \text{choice } y = \arg\max \boldsymbol{U}$$

Let $\boldsymbol{\theta}$ denote the set of model parameters and assume a distribution for ε is fixed.

1. calculate choice probabilities $P_{nj}(\boldsymbol{X}_n, \boldsymbol{\theta})$ for each decider n and alternative j
2. build log-likelihood function $\log \mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{X}, y) = \sum_{n,j} 1(y_n = j) \log P_{nj}(\boldsymbol{X}_n, \boldsymbol{\theta})$
3. maximize $\log \mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{X}, y)$ with respect to $\boldsymbol{\theta}$

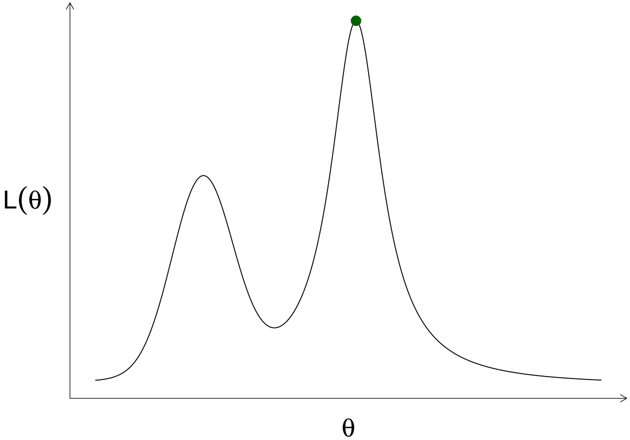
Faster and more robust MLE for RUMs

$$\rightarrow \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} = \boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\beta + \gamma) + \varepsilon \quad \rightarrow \quad \text{choice } y = \arg\max \boldsymbol{U}$$

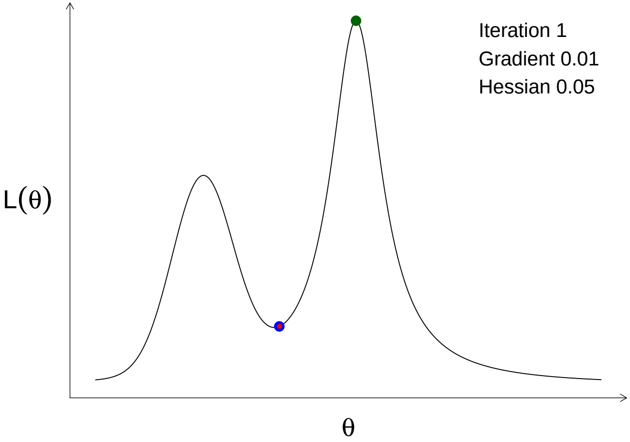
Let $\boldsymbol{\theta}$ denote the set of model parameters and assume a distribution for ε is fixed.

1. calculate choice probabilities $P_{nj}(\boldsymbol{X}_n, \boldsymbol{\theta})$ for each decider n and alternative j
2. build log-likelihood function $\log \mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{X}, y) = \sum_{n,j} 1(y_n = j) \log P_{nj}(\boldsymbol{X}_n, \boldsymbol{\theta})$
3. maximize $\log \mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{X}, y)$ with respect to $\boldsymbol{\theta}$

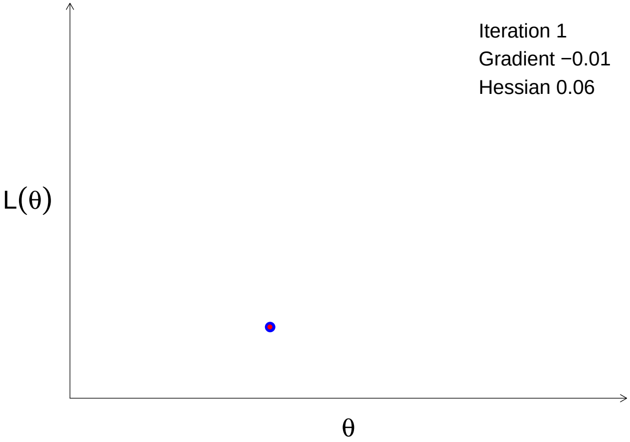
Faster and more robust MLE for RUMs



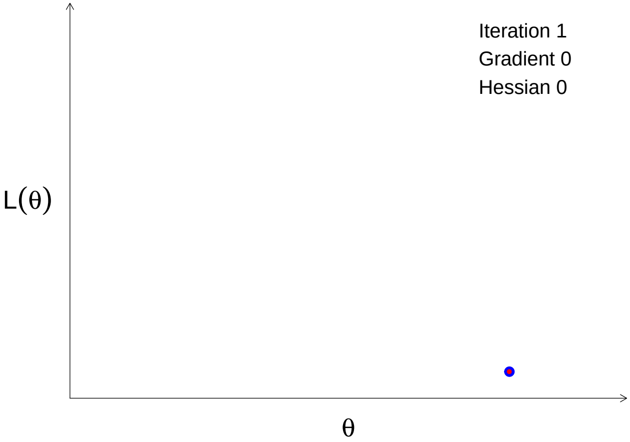
Faster and more robust MLE for RUMs



Faster and more robust MLE for RUMs



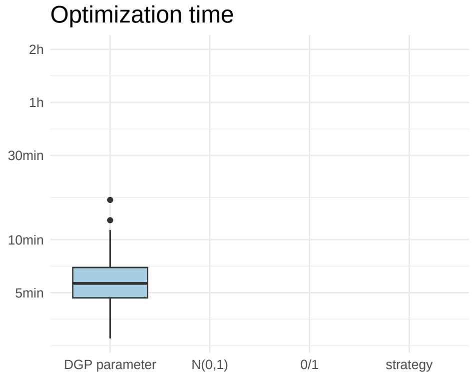
Faster and more robust MLE for RUMs



Outline of this talk

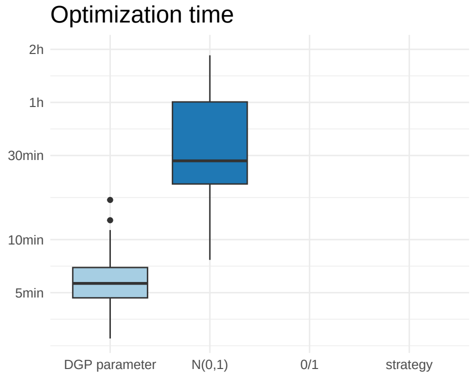
100 data sets simulated from a probit model with $\theta_i \sim N(0,1)$:

$N = 200$ deciders, $T = 10$ choices each, $J = 5$ alternatives, 5 covariates with normal random effects



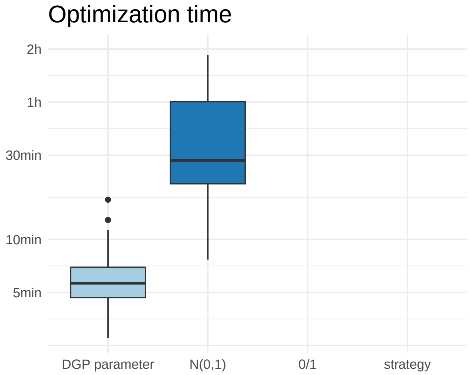
100 data sets simulated from a probit model with $\theta_i \sim N(0,1)$:

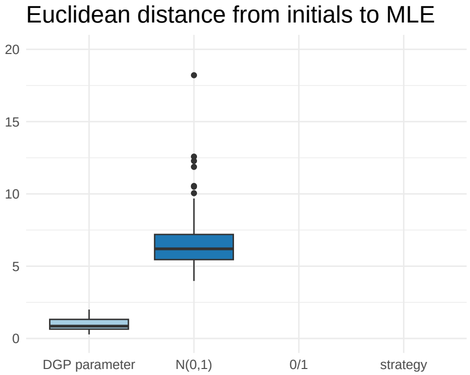
$N = 200$ deciders, $T = 10$ choices each, $J = 5$ alternatives, 5 covariates with normal random effects



100 data sets simulated from a probit model with $\theta_i \sim N(0,1)$:

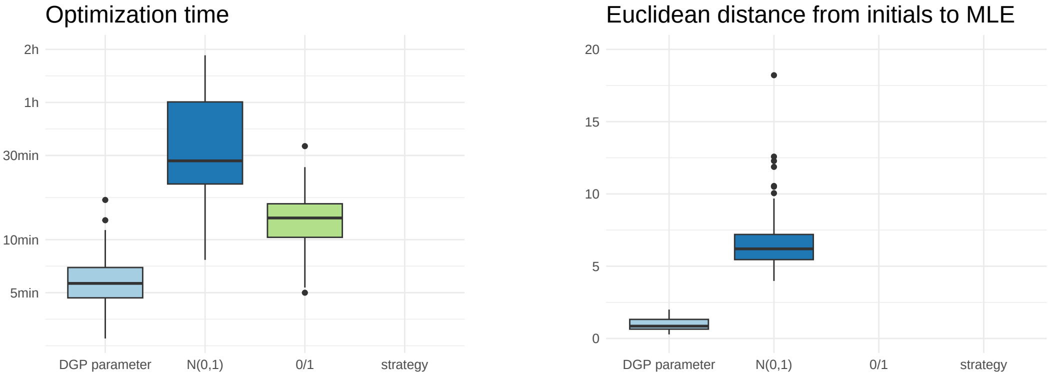
$N = 200$ deciders, $T = 10$ choices each, $J = 5$ alternatives, 5 covariates with normal random effects





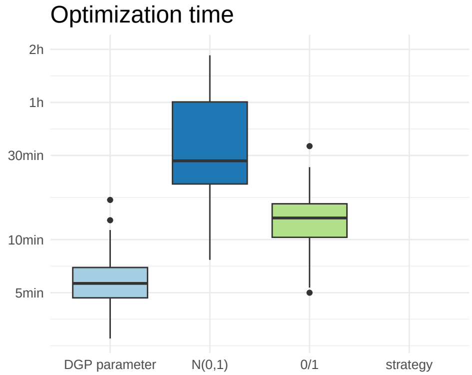
100 data sets simulated from a probit model with $\theta_i \sim N(0,1)$:

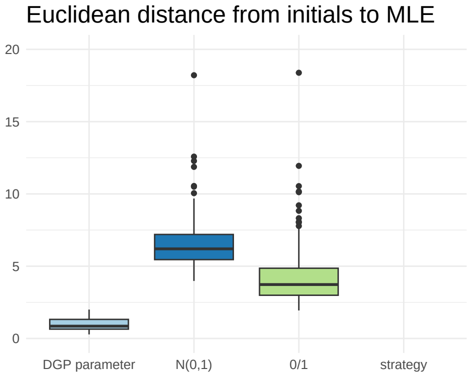
$N = 200$ deciders, $T = 10$ choices each, $J = 5$ alternatives, 5 covariates with normal random effects



100 data sets simulated from a probit model with $\theta_i \sim N(0,1)$:

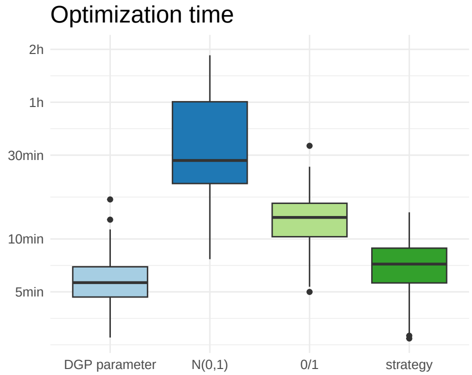
$N = 200$ deciders, $T = 10$ choices each, $J = 5$ alternatives, 5 covariates with normal random effects

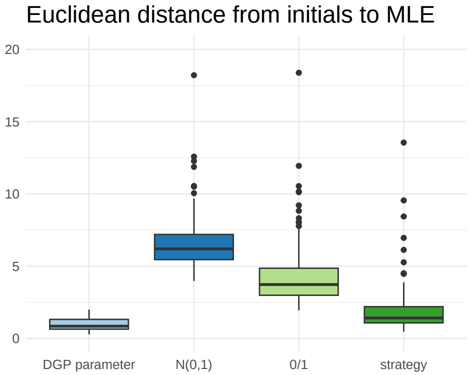




100 data sets simulated from a probit model with $\theta_i \sim N(0,1)$:

$N = 200$ deciders, $T = 10$ choices each, $J = 5$ alternatives, 5 covariates with normal random effects





Outline

Different parameters, different strategies

$$\boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\boldsymbol{\beta} + \boldsymbol{\gamma}) + \boldsymbol{\varepsilon}$$

Parameter	Initialization strategy
$\boldsymbol{\beta}$ is alternative-varying and	
$\boldsymbol{X}$ is alternative-varying	constant utility direction
$\boldsymbol{X}$ is alternative-constant (ASCs $\boldsymbol{\mu}$ is special case)	minimize choice frequency prediction error
$\boldsymbol{\beta}$ is alternative-constant	linear probability model
covariance $\boldsymbol{\Sigma}$ for $\boldsymbol{\varepsilon}$	MCMC
covariance $\boldsymbol{\Omega}$ for $\boldsymbol{\gamma}$	MCMC

Different parameters, different strategies

$$\boldsymbol{U} = \mu + \boldsymbol{X}(\beta + \gamma) + \varepsilon$$

Parameter	Initialization strategy
β is alternative-varying and	
▶ $\boldsymbol{X}$ is alternative-varying	constant utility direction
$\boldsymbol{X}$ is alternative-constant (ASCs μ is special case)	minimize choice frequency prediction error
β is alternative-constant	linear probability model
covariance Σ for ε	MCMC
covariance Ω for γ	MCMC

$$\boldsymbol{U} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

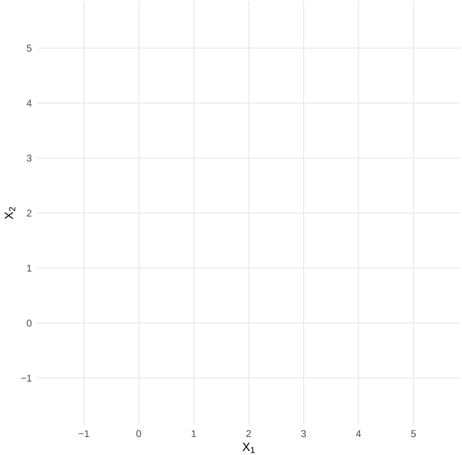
$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

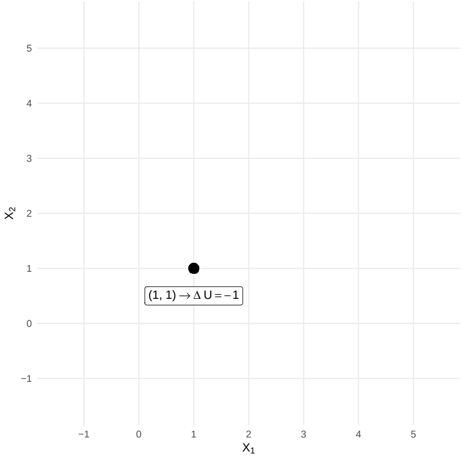
$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$



$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

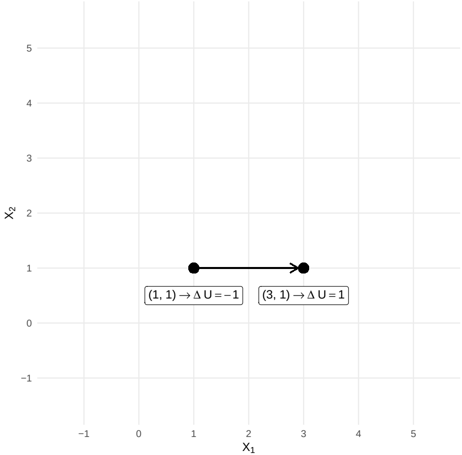
$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$



$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

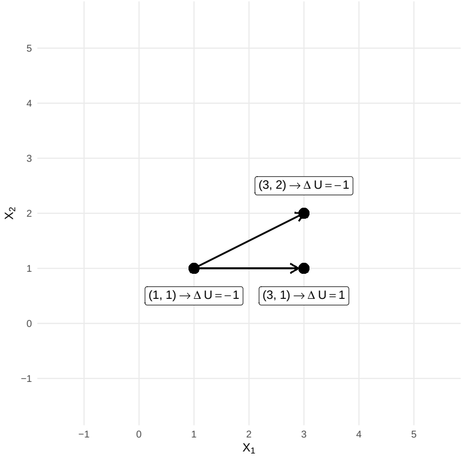
$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$



$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

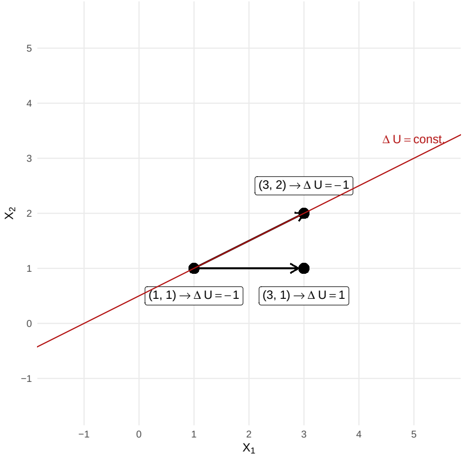
$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$



$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

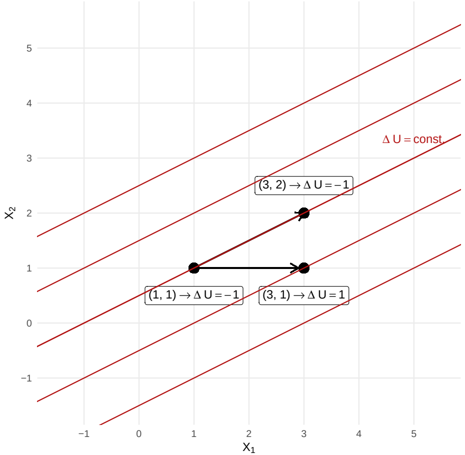


$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$$\Delta U = \text{const. in direction } \overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$$

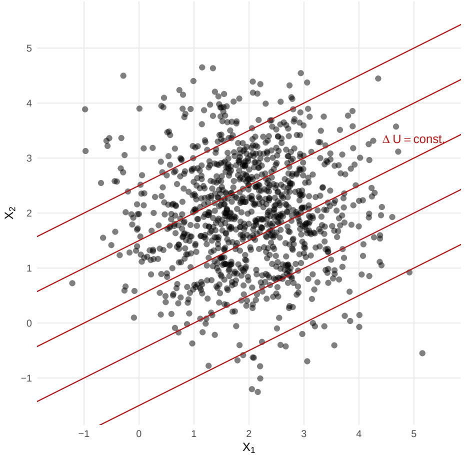


$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$\Delta U = \text{const.}$ in direction $\overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$

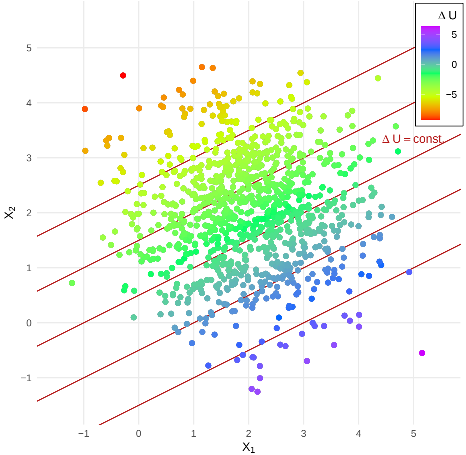


$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$$\Delta U = \text{const. in direction } \overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$$



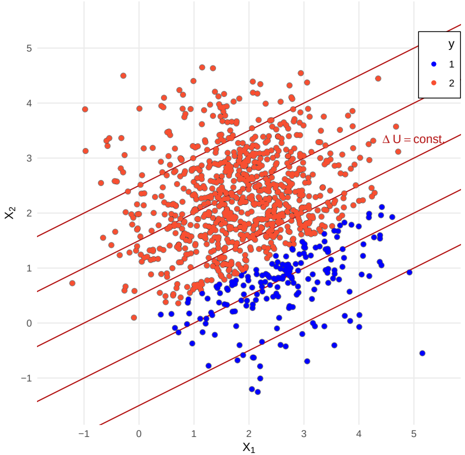
UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$\Delta U = \text{const.}$ in direction $\overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$

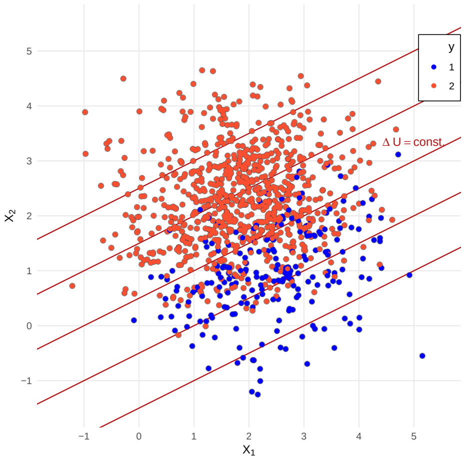


$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$$\Delta U = \text{const. in direction } \overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$$



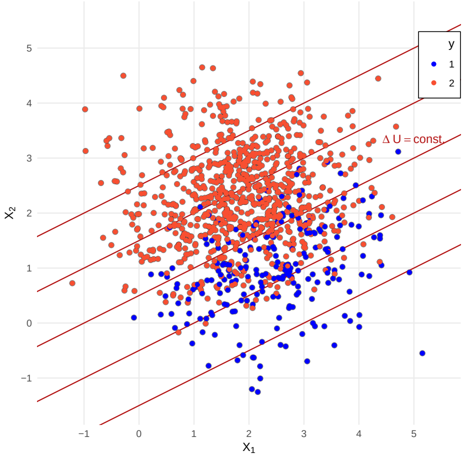
$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$\Delta U = \text{const.}$ in direction $\overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$

identify direction as the kernel of $\text{Cov}(y, \boldsymbol{X})$



$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

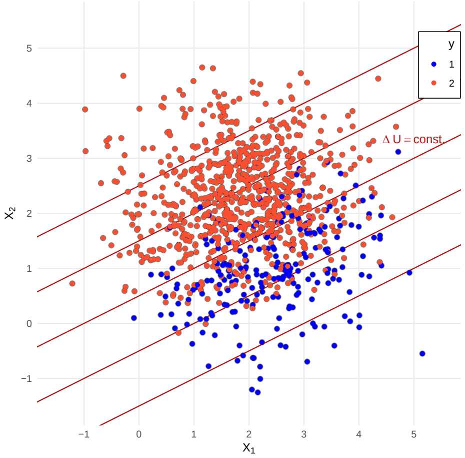
$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$\Delta U = \text{const.}$ in direction $\overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$

identify direction as the kernel of $\text{Cov}(y, \boldsymbol{X})$

this yields a consistent initial estimator $\widehat{\beta}$



💡 Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)

UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, 1.000 replications

N	1.000	■	10.000	□
J	4	■	8	□
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	□

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

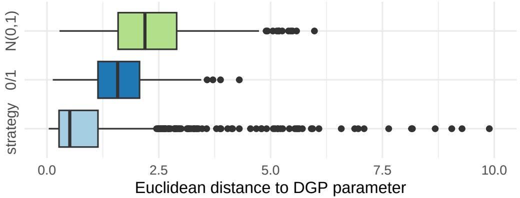
$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$$\Delta U = \text{const. in direction } \overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$$

identify direction as the kernel of $\text{Cov}(y, \boldsymbol{X})$

this yields a consistent initial estimator $\widehat{\beta}$

Average computation time: < 1 second



Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)

UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$$\Delta U = \text{const. in direction } \overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$$

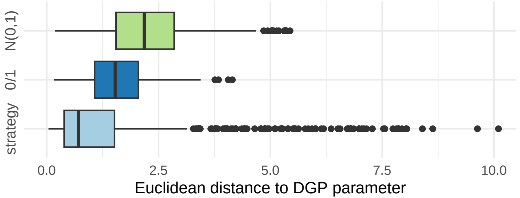
identify direction as the kernel of $\text{Cov}(y, \boldsymbol{X})$

this yields a consistent initial estimator $\widehat{\beta}$

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, 1.000 replications

N	1.000	■	10.000	□
J	4	■	8	□
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	□	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	■

Average computation time: < 1 second



💡 Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)

UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$\Delta U =$ const. in direction $\overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$

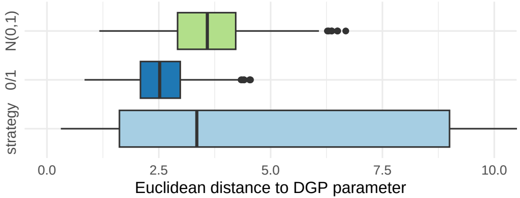
identify direction as the kernel of $\text{Cov}(y, \boldsymbol{X})$

this yields a consistent initial estimator $\widehat{\beta}$

Simulation: $\beta_i \sim \mathcal{N}(0, 1)$, 1.000 replications

N	1.000	■	10.000	□
J	4	□	8	■
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	□

Average computation time: <1 second



💡 Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)

UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, 1.000 replications

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

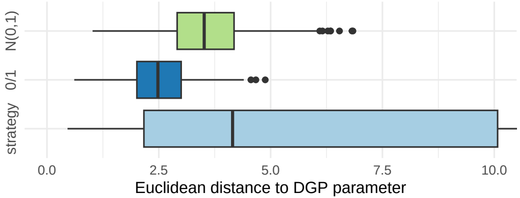
$\Delta U =$ const. in direction $\overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$

identify direction as the kernel of $\text{Cov}(y, \boldsymbol{X})$

this yields a consistent initial estimator $\widehat{\beta}$

N	1.000	■	10.000	□
J	4	□	8	■
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	□	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	■

Average computation time: < 1 second



Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$$\Delta U = \text{const. in direction } \overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$$

identify direction as the kernel of $\text{Cov}(y, \boldsymbol{X})$

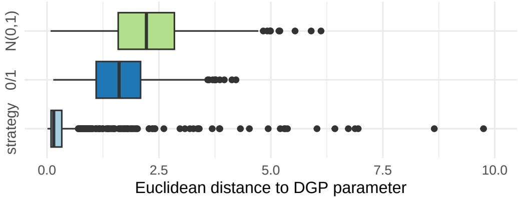
this yields a consistent initial estimator $\widehat{\beta}$

💡 Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, 1.000 replications

N	1.000	☐	10.000	■
J	4	■	8	☐
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	☐

Average computation time: <1 second



UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$\Delta U =$ const. in direction $\overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$

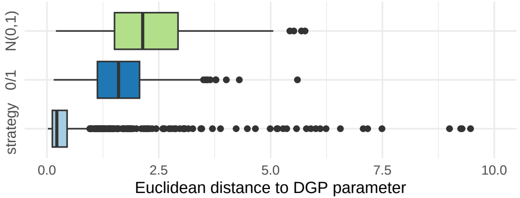
identify direction as the kernel of $\text{Cov}(y, \boldsymbol{X})$

this yields a consistent initial estimator $\widehat{\beta}$

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, 1.000 replications

N	1.000	☐	10.000	■
J	4	■	8	☐
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	☐	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	■

Average computation time: < 1 second



Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$\Delta U =$ const. in direction $\overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$

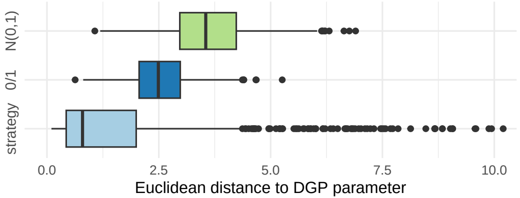
identify direction as the kernel of $\text{Cov}(y, \boldsymbol{X})$

this yields a consistent initial estimator $\widehat{\beta}$

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, 1.000 replications

N	1.000	☐	10.000	■
J	4	☐	8	■
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	☐

Average computation time: $<$ 1 second



💡 Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } \Delta U = U_1 - U_2 \geq 0 \\ 2, & \text{otherwise} \end{cases}$$

$\Delta U =$ const. in direction $\overrightarrow{\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}} = \overrightarrow{\begin{pmatrix} 1/\beta_1 \\ 1/\beta_2 \end{pmatrix}}$

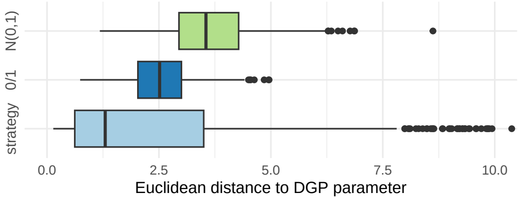
identify direction as the kernel of $\text{Cov}(y, \boldsymbol{X})$

this yields a consistent initial estimator $\widehat{\beta}$

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, 1.000 replications

N	1.000	□	10.000	■
J	4	□	8	■
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	□	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	■

Average computation time: <1 second



Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)

$$\boldsymbol{U} = \mu + \boldsymbol{X}(\beta + \gamma) + \varepsilon$$

Parameter	Initialization strategy
β is alternative-varying and	
$\boldsymbol{X}$ is alternative-varying	constant utility direction
▶ $\boldsymbol{X}$ is alternative-constant (ASCs μ is special case)	minimize choice frequency prediction error
β is alternative-constant	linear probability model
covariance Σ for ε	MCMC
covariance Ω for γ	MCMC

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } U_1 = \max \boldsymbol{U} \\ 2, & \text{if } U_2 = \max \boldsymbol{U} \\ 3, & \text{if } U_3 = \max \boldsymbol{U} \end{cases}$$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

$$\boldsymbol{y}^d = \begin{cases} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, & \text{if } U_1 = \max \boldsymbol{U} \\ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, & \text{if } U_2 = \max \boldsymbol{U} \\ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, & \text{if } U_3 = \max \boldsymbol{U} \end{cases}$$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

Initialization strategy:

1. assume $\boldsymbol{\Sigma}$ is known, or choose $\boldsymbol{\Sigma} = I_J$ else
2. let $\overline{\boldsymbol{y}^d}$ be the average of $\boldsymbol{y}^d$
3. find $\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \| P(\boldsymbol{\beta}, \boldsymbol{\Sigma}) - \overline{\boldsymbol{y}^d} \|_2$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

Initialization strategy:

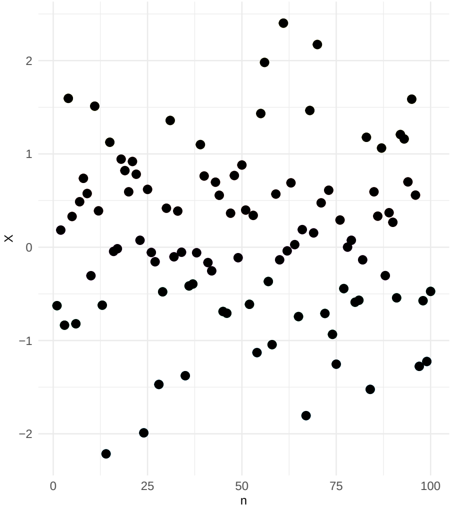
1. assume $\boldsymbol{\Sigma}$ is known, or choose $\boldsymbol{\Sigma} = I_J$ else
2. let $\overline{\boldsymbol{y}^d}$ be the average of $\boldsymbol{y}^d$
3. find $\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|P(\boldsymbol{\beta}, \boldsymbol{\Sigma}) - \overline{\boldsymbol{y}^d}\|_2$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

Initialization strategy:

1. assume Σ is known, or choose $\Sigma = I_J$ else
2. let $\overline{\boldsymbol{y}^d}$ be the average of $\boldsymbol{y}^d$
3. find $\widehat{\beta} = \arg\min_{\beta} \|P(\beta, \Sigma) - \overline{\boldsymbol{y}^d}\|_2$

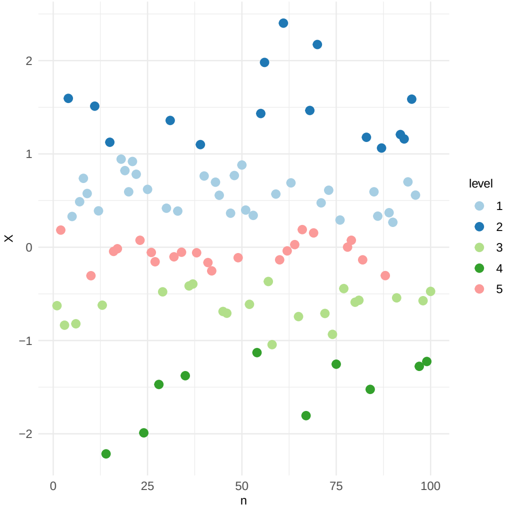


$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

Initialization strategy:

1. assume $\boldsymbol{\Sigma}$ is known, or choose $\boldsymbol{\Sigma} = I_J$ else
2. let $\overline{\boldsymbol{y}^d}$ be the average of $\boldsymbol{y}^d$
3. find $\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|P(\boldsymbol{\beta}, \boldsymbol{\Sigma}) - \overline{\boldsymbol{y}^d}\|_2$

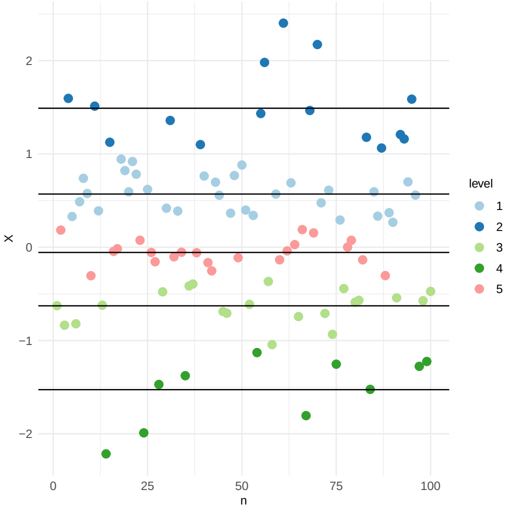


$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

Initialization strategy:

1. assume $\boldsymbol{\Sigma}$ is known, or choose $\boldsymbol{\Sigma} = I_J$ else
2. let $\overline{\boldsymbol{y}^d}$ be the average of $\boldsymbol{y}^d$
3. find $\widehat{\beta} = \arg\min_{\beta} \| P(\beta, \boldsymbol{\Sigma}) - \overline{\boldsymbol{y}^d} \|_2$



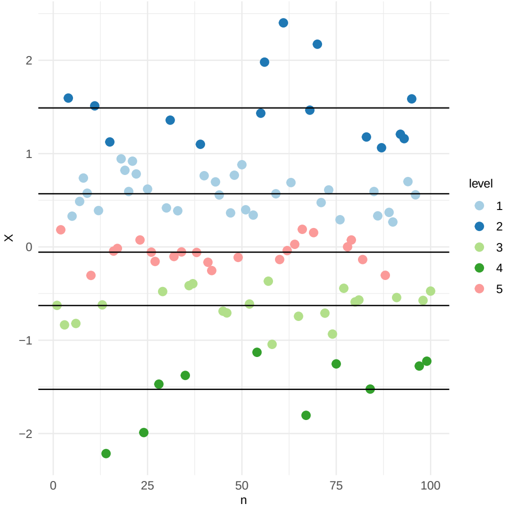
UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

Initialization strategy:

1. assume $\boldsymbol{\Sigma}$ is known, or choose $\boldsymbol{\Sigma} = I_J$ else
2. localize X, for each level $L_i \neq 0$:
   - 2.1 let $\overline{\boldsymbol{y}_i^d}$ be the average of $\boldsymbol{y}_i^d$
   - 2.2 find $\widehat{\boldsymbol{\beta}}_i = \arg\min_{\boldsymbol{\beta}} \|P(\boldsymbol{\beta}, \boldsymbol{\Sigma}) - \overline{\boldsymbol{y}_i^d}\|_2$
   - 2.3 $\widehat{\boldsymbol{\beta}}_i \leftarrow \widehat{\boldsymbol{\beta}}_i / L_i$
3. $\widehat{\boldsymbol{\beta}} = \overline{\widehat{\boldsymbol{\beta}}_i}$



$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

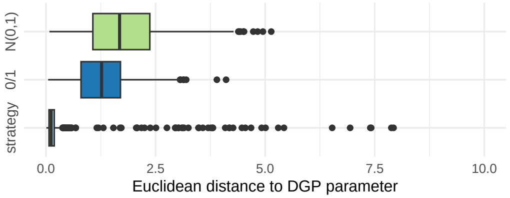
Initialization strategy:

1. assume Σ is known, or choose $\Sigma = I_J$ else
2. localize X, for each level $L_i \neq 0$:
   2.1 let $\overline{\boldsymbol{y}_i^d}$ be the average of $\boldsymbol{y}_i^d$
   2.2 find $\widehat{\beta}_i = \arg\min_{\beta} \|P(\beta, \Sigma) - \overline{\boldsymbol{y}_i^d}\|_2$
   2.3 $\widehat{\beta}_i \leftarrow \widehat{\beta}_i / L_i$
3. $\widehat{\beta} = \overline{\widehat{\beta}_i}$

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, $N = 1.000$, 1.000 rep.

J	4	■	8	□
levels	1	■	N	□
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	□

Average computation time: < 1 second



UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

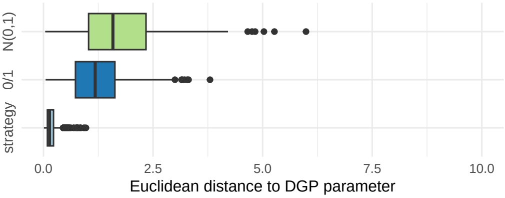
Initialization strategy:

1. assume Σ is known, or choose $\Sigma = I_J$ else
2. localize X, for each level $L_i \neq 0$:
   - 2.1 let $\overline{\boldsymbol{y}_i^d}$ be the average of $\boldsymbol{y}_i^d$
   - 2.2 find $\widehat{\beta}_i = \arg\min_{\beta} \|P(\beta, \Sigma) - \overline{\boldsymbol{y}_i^d}\|_2$
   - 2.3 $\widehat{\beta}_i \leftarrow \widehat{\beta}_i / L_i$
3. $\widehat{\beta} = \overline{\widehat{\beta}_i}$

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, $N = 1.000$, 1.000 rep.

J	4	■	8	□
levels	1	■	N	□
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	□	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	■

Average computation time: < 1 second



UNIVERSITÄT BIELEFELD

Faculty of Business Administration and Economics

$$U = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

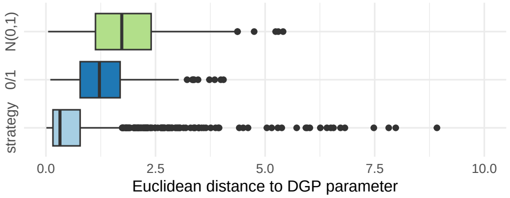
Initialization strategy:

1. assume Σ is known, or choose $\Sigma = I_J$ else
2. localize X, for each level $L_i \neq 0$:
   - 2.1 let $\overline{y_i^d}$ be the average of y_i^d
   - 2.2 find $\widehat{\beta}_i = \arg\min_{\beta} \|P(\beta, \Sigma) - \overline{y_i^d}\|_2$
   - 2.3 $\widehat{\beta}_i \leftarrow \widehat{\beta}_i / L_i$
3. $\widehat{\beta} = \overline{\widehat{\beta}_i}$

Simulation: $\beta_i \sim \mathcal{N}(0, 1)$, $N = 1.000$, 1.000 rep.

J	4	■	8	□
levels	1	□	N	■
Σ	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	□

Average computation time: < 1 second



UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

$$U = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

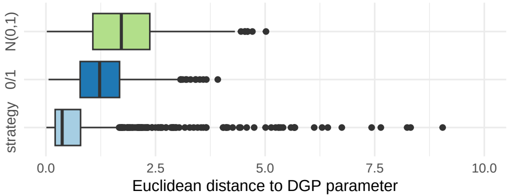
Initialization strategy:

1. assume Σ is known, or choose $\Sigma = I_J$ else
2. localize X, for each level $L_i \neq 0$:
   2.1 let $\overline{y_i^d}$ be the average of y_i^d
   2.2 find $\widehat{\beta}_i = \arg\min_{\beta} \|P(\beta, \Sigma) - \overline{y_i^d}\|_2$
   2.3 $\widehat{\beta}_i \leftarrow \widehat{\beta}_i / L_i$
3. $\widehat{\beta} = \overline{\widehat{\beta}_i}$

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, $N = 1.000$, 1.000 rep.

J	4	■	8	□
levels	1	□	N	■
Σ	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	□	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	■

Average computation time: < 1 second



$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

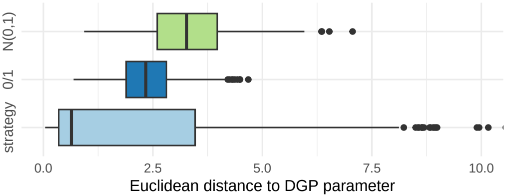
Initialization strategy:

1. assume Σ is known, or choose $\Sigma = I_J$ else
2. localize X, for each level $L_i \neq 0$:
   2.1 let $\overline{\boldsymbol{y}_i^d}$ be the average of $\boldsymbol{y}_i^d$
   2.2 find $\widehat{\beta}_i = \arg\min_{\beta} \|P(\beta, \Sigma) - \overline{\boldsymbol{y}_i^d}\|_2$
   2.3 $\widehat{\beta}_i \leftarrow \widehat{\beta}_i / L_i$
3. $\widehat{\beta} = \overline{\widehat{\beta}_i}$

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, $N = 1.000$, 1.000 rep.

J	4	☐	8	■
levels	1	■	N	☐
Σ	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	☐

Average computation time: 2 seconds



$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

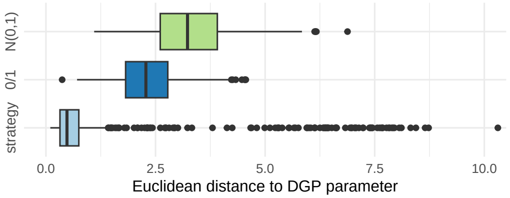
Initialization strategy:

1. assume Σ is known, or choose $\Sigma = I_J$ else
2. localize X, for each level $L_i \neq 0$:
   - 2.1 let $\overline{\boldsymbol{y}_i^d}$ be the average of $\boldsymbol{y}_i^d$
   - 2.2 find $\widehat{\beta}_i = \arg\min_{\beta} \|P(\beta, \Sigma) - \overline{\boldsymbol{y}_i^d}\|_2$
   - 2.3 $\widehat{\beta}_i \leftarrow \widehat{\beta}_i / L_i$
3. $\widehat{\beta} = \overline{\widehat{\beta}_i}$

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, $N = 1.000$, 1.000 rep.

J	4	☐	8	■
levels	1	■	N	☐
Σ	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	☐	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	■

Average computation time: 2 seconds



$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

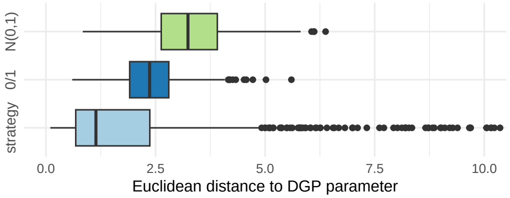
Initialization strategy:

1. assume Σ is known, or choose $\Sigma = I_J$ else
2. localize X, for each level $L_i \neq 0$:
   - 2.1 let $\overline{\boldsymbol{y}_i^d}$ be the average of $\boldsymbol{y}_i^d$
   - 2.2 find $\widehat{\beta}_i = \arg\min_{\beta} \|P(\beta, \Sigma) - \overline{\boldsymbol{y}_i^d}\|_2$
   - 2.3 $\widehat{\beta}_i \leftarrow \widehat{\beta}_i / L_i$
3. $\widehat{\beta} = \overline{\widehat{\beta}_i}$

Simulation: $\beta_i \sim \mathcal{N}(0, 1)$, $N = 1.000$, 1.000 rep.

J	4	☐	8	■
levels	1	☐	N	■
Σ	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	☐

Average computation time: 2 seconds



$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{pmatrix} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

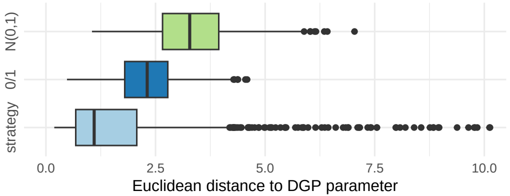
Initialization strategy:

1. assume Σ is known, or choose $\Sigma = I_J$ else
2. localize X, for each level $L_i \neq 0$:
   - 2.1 let $\overline{\boldsymbol{y}_i^d}$ be the average of $\boldsymbol{y}_i^d$
   - 2.2 find $\widehat{\beta}_i = \arg\min_{\beta} \| P(\beta, \Sigma) - \overline{\boldsymbol{y}_i^d} \|_2$
   - 2.3 $\widehat{\beta}_i \leftarrow \widehat{\beta}_i / L_i$
3. $\widehat{\beta} = \overline{\widehat{\beta}_i}$

Simulation: $\beta_i \sim \mathcal{N}(0, 1)$, $N = 1.000$, 1.000 rep.

J	4	☐	8	■
levels	1	☐	N	■
Σ	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	☐	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	■

Average computation time: 2 seconds



$$\boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\boldsymbol{\beta} + \boldsymbol{\gamma}) + \boldsymbol{\varepsilon}$$

Parameter	Initialization strategy
$\boldsymbol{\beta}$ is alternative-varying and	
$\boldsymbol{X}$ is alternative-varying	constant utility direction
$\boldsymbol{X}$ is alternative-constant (ASCs $\boldsymbol{\mu}$ is special case)	minimize choice frequency prediction error
► $\boldsymbol{\beta}$ is alternative-constant	linear probability model
covariance Σ for $\boldsymbol{\varepsilon}$	MCMC
covariance Ω for $\boldsymbol{\gamma}$	MCMC

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\boldsymbol{U} \;=\; \boldsymbol{X}\boldsymbol{\beta} \;+\; \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\underbrace{\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}}_{J=2} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \underbrace{\beta_1}_{P=1} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\underbrace{\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}}_{J=2} = \begin{pmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{pmatrix} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}}_{P=2} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\underbrace{\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}}_{J=2} = \begin{pmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{pmatrix} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}}_{P=2} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$y = \begin{cases} 1, & \text{if } U_1 = \max \boldsymbol{U} \\ 2, & \text{if } U_2 = \max \boldsymbol{U} \end{cases}$$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

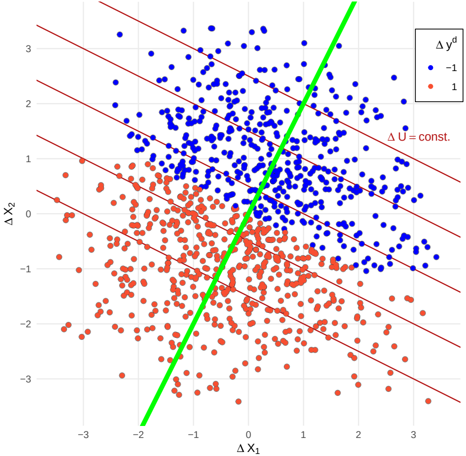
$$\underbrace{\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}}_{J=2} = \begin{pmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{pmatrix} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}}_{P=2} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$\boldsymbol{y}^d = \begin{cases} \binom{1}{0}, & \text{if } U_1 = \max \boldsymbol{U} \\ \binom{0}{1}, & \text{if } U_2 = \max \boldsymbol{U} \end{cases}$$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\underbrace{\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}}_{J=2} = \begin{pmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{pmatrix} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}}_{P=2} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$\boldsymbol{y}^d = \begin{cases} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, & \text{if } U_1 = \max \boldsymbol{U} \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, & \text{if } U_2 = \max \boldsymbol{U} \end{cases}$$



$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\boldsymbol{y}^d = \boldsymbol{X}\alpha + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{I})$$

$$\underbrace{\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}}_{J=2} = \begin{pmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{pmatrix} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}}_{P=2} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$\boldsymbol{y}^d = \begin{cases} \binom{1}{0}, & \text{if } U_1 = \max \boldsymbol{U} \\ \binom{0}{1}, & \text{if } U_2 = \max \boldsymbol{U} \end{cases}$$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\underbrace{\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}}_{J=2} = \begin{pmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{pmatrix} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}}_{P=2} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$\boldsymbol{y}^d = \begin{cases} \binom{1}{0}, & \text{if } U_1 = \max \boldsymbol{U} \\ \binom{0}{1}, & \text{if } U_2 = \max \boldsymbol{U} \end{cases}$$

$$\boldsymbol{y}^d = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{I})$$

- find $\widehat{\boldsymbol{\alpha}}$ via OLS
- $\widehat{\boldsymbol{\alpha}}/\|\widehat{\boldsymbol{\alpha}}\|$ is consistent for $\boldsymbol{\beta}/\|\boldsymbol{\beta}\|$

  Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)
- robust to $\boldsymbol{\Sigma}$ and the existence of additional independent regressors

$$\boldsymbol{U} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

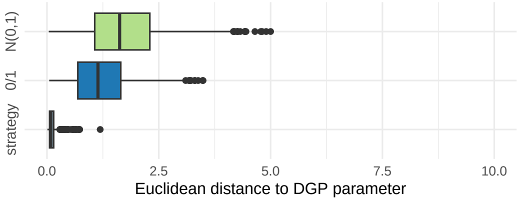
$$\boldsymbol{y}^d = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{I})$$

- ☐ find $\widehat{\boldsymbol{\alpha}}$ via OLS
- ☐ $\widehat{\boldsymbol{\alpha}}/\|\widehat{\boldsymbol{\alpha}}\|$ is consistent for $\boldsymbol{\beta}/\|\boldsymbol{\beta}\|$
  - Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)
- ☐ robust to $\boldsymbol{\Sigma}$ and the existence of additional independent regressors

Simulation: $\beta_i \sim \mathcal{N}(0, 1)$, $J = 3$, 1.000 replications

N	1.000	■	10.000	□
P	3	■	10	□
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$	□

Average computation time: <1 second



UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

$\epsilon \sim \Sigma$

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

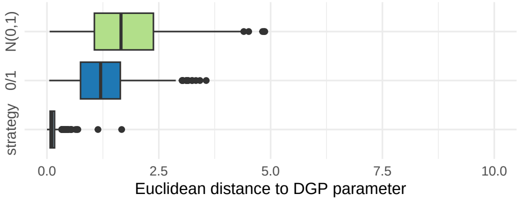
$$\boldsymbol{y}^d = \boldsymbol{X}\alpha + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{I})$$

- ☐ find $\widehat{\alpha}$ via OLS
- ☐ $\widehat{\alpha}/\|\widehat{\alpha}\|$ is consistent for $\beta/\|\beta\|$
  - Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)
- ☐ robust to Σ and the existence of additional independent regressors

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, $J = 3$, 1.000 replications

N	1.000	■	10.000	☐
P	3	■	10	☐
Σ	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	☐	$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$	■

Average computation time: $<$ 1 second



UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, $J = 3$, 1.000 replications

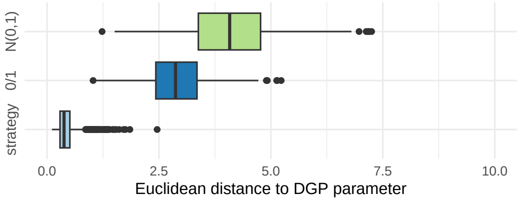
N	1.000	■	10.000	□
P	3	□	10	■
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$	□

$$\boldsymbol{U} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\boldsymbol{y}^d = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{I})$$

- ☐ find $\widehat{\boldsymbol{\alpha}}$ via OLS
- ☐ $\widehat{\boldsymbol{\alpha}}/\|\widehat{\boldsymbol{\alpha}}\|$ is consistent for $\boldsymbol{\beta}/\|\boldsymbol{\beta}\|$
  - Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)
- ☐ robust to $\boldsymbol{\Sigma}$ and the existence of additional independent regressors

Average computation time: < 1 second



$$U = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

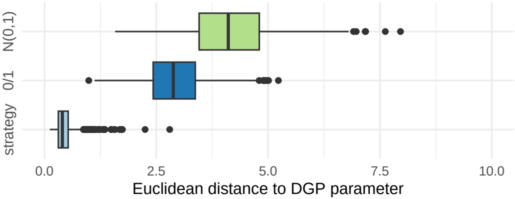
$$y^d = X\alpha + u, \quad u \sim \mathcal{N}(0, I)$$

- ☐ find $\widehat{\alpha}$ via OLS
- ☐ $\widehat{\alpha}/\|\widehat{\alpha}\|$ is consistent for $\beta/\|\beta\|$
  - Consistency as $N \to \infty$ and under a technical assumption on X (normality is sufficient)
- ☐ robust to Σ and the existence of additional independent regressors

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, $J = 3$, 1.000 replications

N	1.000	■	10.000	☐
P	3	☐	10	■
Σ	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	☐	$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$	■

Average computation time: < 1 second



$$U = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

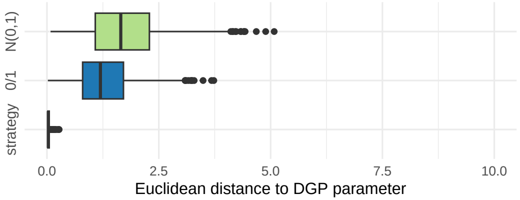
$$y^d = X\alpha + u, \quad u \sim \mathcal{N}(0, I)$$

- ☐ find $\hat{\alpha}$ via OLS
- ☐ $\hat{\alpha}/\|\hat{\alpha}\|$ is consistent for $\beta/\|\beta\|$
  - Consistency as $N \to \infty$ and under a technical assumption on X (normality is sufficient)
- ☐ robust to Σ and the existence of additional independent regressors

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, $J = 3$, 1.000 replications

N	1.000	☐	10.000	■
P	3	■	10	☐
Σ	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$	☐

Average computation time: < 1 second



UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

$\epsilon \sim \Sigma$

$$\boldsymbol{U} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

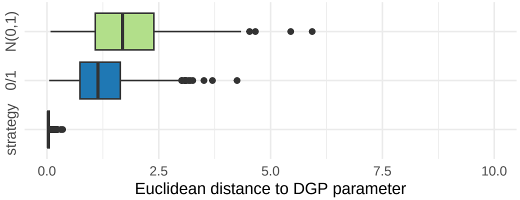
$$\boldsymbol{y}^d = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{I})$$

- ☐ find $\widehat{\boldsymbol{\alpha}}$ via OLS
- ☐ $\widehat{\boldsymbol{\alpha}}/\|\widehat{\boldsymbol{\alpha}}\|$ is consistent for $\boldsymbol{\beta}/\|\boldsymbol{\beta}\|$
  - Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)
- ☐ robust to $\boldsymbol{\Sigma}$ and the existence of additional independent regressors

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, $J = 3$, 1.000 replications

N	1.000	☐	10.000	■
P	3	■	10	☐
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	☐	$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$	■

Average computation time: <1 second



$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

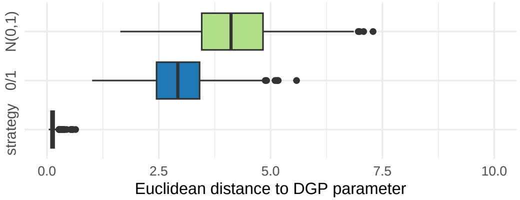
$$\boldsymbol{y}^d = \boldsymbol{X}\alpha + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{I})$$

- ☐ find $\widehat{\alpha}$ via OLS
- ☐ $\widehat{\alpha}/\|\widehat{\alpha}\|$ is consistent for $\beta/\|\beta\|$
  - Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)
- ☐ robust to Σ and the existence of additional independent regressors

Simulation: $\beta_i \sim \mathcal{N}(0, 1)$, $J = 3$, 1.000 replications

N	1.000	☐	10.000	■
P	3	☐	10	■
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$	☐

Average computation time: $<$ 1 second



UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

Simulation: $\beta_i \sim \mathcal{N}(0,1)$, $J = 3$, 1.000 replications

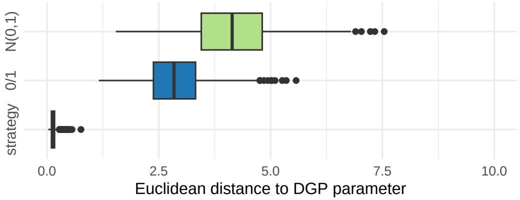
N	1.000	☐	10.000	■
P	3	☐	10	■
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	☐	$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$	■

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

$$\boldsymbol{y}^d = \boldsymbol{X}\alpha + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{I})$$

- ☐ find $\widehat{\alpha}$ via OLS
- ☐ $\widehat{\alpha}/\|\widehat{\alpha}\|$ is consistent for $\beta/\|\beta\|$
  - Consistency as $N \to \infty$ and under a technical assumption on $\boldsymbol{X}$ (normality is sufficient)
- ☐ robust to $\boldsymbol{\Sigma}$ and the existence of additional independent regressors

Average computation time: <1 second

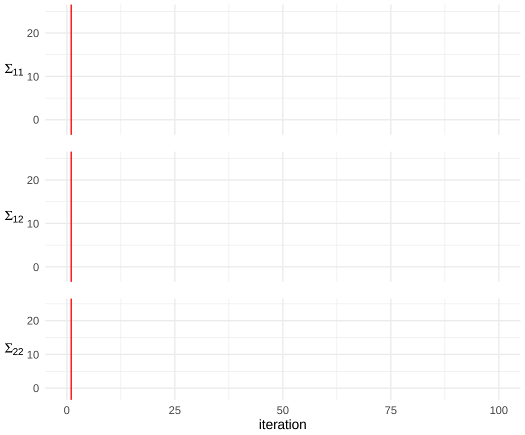


$$\boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\beta + \gamma) + \varepsilon$$

Parameter	Initialization strategy
β is alternative-varying and	
$\boldsymbol{X}$ is alternative-varying	constant utility direction
$\boldsymbol{X}$ is alternative-constant (ASCs $\boldsymbol{\mu}$ is special case)	minimize choice frequency prediction error
β is alternative-constant	linear probability model
▶ covariance Σ for ε	MCMC
covariance Ω for γ	MCMC

$$\boldsymbol{U} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma})$$

1. assume $\boldsymbol{\beta}$ is known
2. for $\approx$ 100 iterations:
   2.1 $\boldsymbol{U} \mid \boldsymbol{\Sigma}, \boldsymbol{\beta}, \boldsymbol{X}, y \sim$ truncated $\mathcal{N}$
   2.2 $\boldsymbol{\Sigma} \mid \boldsymbol{\beta}, \boldsymbol{U} \sim \mathcal{W}^{-1}$
3. derive $\widehat{\boldsymbol{\Sigma}}$ as sample average of $\boldsymbol{\Sigma}$



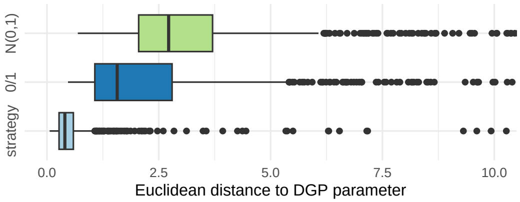
$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

1. assume β is known
2. for $\approx$ 100 iterations:
   2.1 $\boldsymbol{U} \mid \Sigma, \beta, \boldsymbol{X}, y \sim$ truncated $\mathcal{N}$
   2.2 $\Sigma \mid \beta, \boldsymbol{U} \sim \mathcal{W}^{-1}$
3. derive $\widehat{\Sigma}$ as sample average of Σ

Simulation: $\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(J \cdot \boldsymbol{I}, J+1)$, $P = 5$, 1.000 rep.

N	1.000	■	10.000	□
J	4	■	8	□

Average computation time: 3 seconds



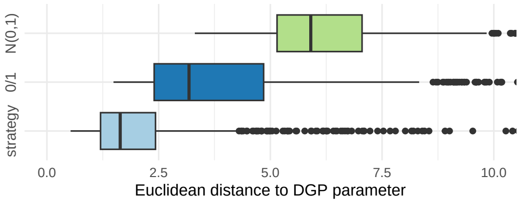
$$U = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

1. assume β is known
2. for ≈ 100 iterations:
   2.1 $U \mid \Sigma, \beta, X, y \sim$ truncated $\mathcal{N}$
   2.2 $\Sigma \mid \beta, U \sim \mathcal{W}^{-1}$
3. derive $\hat{\Sigma}$ as sample average of Σ

Simulation: $\Sigma \sim \mathcal{W}^{-1}(J \cdot I, J + 1)$, $P = 5$, 1.000 rep.

N	1.000	■	10.000	□
J	4	□	8	■

Average computation time: 3 seconds



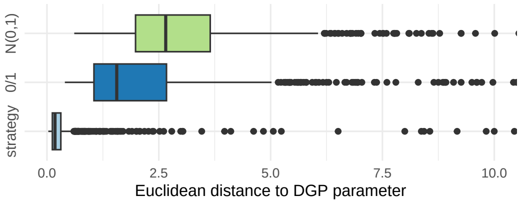
Simulation: $\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(J \cdot \boldsymbol{I}, J+1)$, $P = 5$, 1.000 rep.

N	1.000	☐	10.000	■
J	4	■	8	☐

$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

Average computation time: 30 seconds

1. assume β is known
2. for ≈ 100 iterations:
   2.1 $\boldsymbol{U} \mid \Sigma, \beta, \boldsymbol{X}, y \sim$ truncated $\mathcal{N}$
   2.2 $\Sigma \mid \beta, \boldsymbol{U} \sim \mathcal{W}^{-1}$
3. derive $\widehat{\Sigma}$ as sample average of Σ



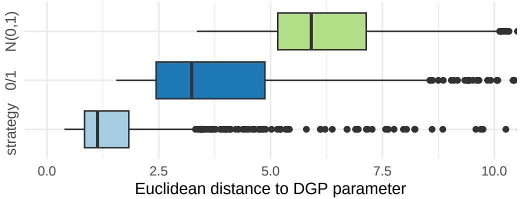
$$\boldsymbol{U} = \boldsymbol{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

1. assume β is known
2. for $\approx$ 100 iterations:
   2.1 $\boldsymbol{U} \mid \Sigma, \beta, \boldsymbol{X}, y \sim$ truncated $\mathcal{N}$
   2.2 $\Sigma \mid \beta, \boldsymbol{U} \sim \mathcal{W}^{-1}$
3. derive $\widehat{\Sigma}$ as sample average of Σ

Simulation: $\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(J \cdot \boldsymbol{I}, J + 1)$, $P = 5$, 1.000 rep.

N	1.000	☐	10.000	■
J	4	☐	8	■

Average computation time: 30 seconds



$$\boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\boldsymbol{\beta} + \boldsymbol{\gamma}) + \boldsymbol{\varepsilon}$$

Parameter	Initialization strategy
$\boldsymbol{\beta}$ is alternative-varying and	
$\boldsymbol{X}$ is alternative-varying	constant utility direction
$\boldsymbol{X}$ is alternative-constant (ASCs $\boldsymbol{\mu}$ is special case)	minimize choice frequency prediction error
$\boldsymbol{\beta}$ is alternative-constant	linear probability model
covariance $\boldsymbol{\Sigma}$ for $\boldsymbol{\varepsilon}$	MCMC
covariance $\boldsymbol{\Omega}$ for $\boldsymbol{\gamma}$	MCMC

$$\boldsymbol{U} = \boldsymbol{X}(\beta + \gamma) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma}),$$
$$\gamma \sim \mathcal{N}(0, \boldsymbol{\Omega})$$

$$\boldsymbol{U} = \boldsymbol{X}(\underbrace{\beta + \gamma}_{\beta_n}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma}),$$

$$\beta_n \sim \mathcal{N}(\beta, \boldsymbol{\Omega})$$

$$\boldsymbol{U} = \boldsymbol{X}(\underbrace{\boldsymbol{\beta} + \boldsymbol{\gamma}}_{\boldsymbol{\beta}_n}) + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma}), \quad \boldsymbol{\beta}_n \sim \mathcal{N}(\boldsymbol{\beta}, \boldsymbol{\Omega})$$

1. assume $\boldsymbol{\beta}$ is known
2. for $\approx$ 200 iterations:
   2.1 $\boldsymbol{U} \mid \boldsymbol{\Sigma}, (\boldsymbol{\beta}_n), \boldsymbol{X}, y \sim$ truncated $\mathcal{N}$
   2.2 $\boldsymbol{\beta}_n \mid \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{U} \sim \mathcal{N}$ for each n
   2.3 $\boldsymbol{\Sigma} \mid (\boldsymbol{\beta}_n), \boldsymbol{U} \sim \mathcal{W}^{-1}$
   2.4 $\boldsymbol{\Omega} \mid (\boldsymbol{\beta}_n), \boldsymbol{\beta} \sim \mathcal{W}^{-1}$
3. derive $\widehat{\boldsymbol{\Sigma}}$, $\widehat{\boldsymbol{\Omega}}$ as sample average of $\boldsymbol{\Sigma}$, $\boldsymbol{\Omega}$

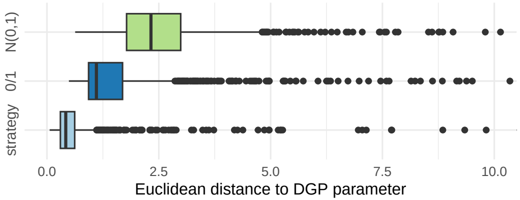
$$\boldsymbol{U} = \boldsymbol{X}(\underbrace{\beta + \gamma}_{\beta_n}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma),$$
$$\beta_n \sim \mathcal{N}(\beta, \Omega)$$

1. assume β is known
2. for ≈ 200 iterations:
   2.1 $\boldsymbol{U} \mid \Sigma, (\beta_n), \boldsymbol{X}, y \sim$ truncated $\mathcal{N}$
   2.2 $\beta_n \mid \beta, \Omega, \boldsymbol{U} \sim \mathcal{N}$ for each n
   2.3 $\Sigma \mid (\beta_n), \boldsymbol{U} \sim \mathcal{W}^{-1}$
   2.4 $\Omega \mid (\beta_n), \beta \sim \mathcal{W}^{-1}$
3. derive $\widehat{\Sigma}$, $\widehat{\Omega}$ as sample average of Σ, Ω

Simulation: $\boldsymbol{\Omega} \sim \mathcal{W}^{-1}(P \cdot \boldsymbol{I}, P+1)$, $J = 3$, $T = 10$

N	100	■	500	□
P	3	■	5	□
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	□

Average computation time: 3 seconds



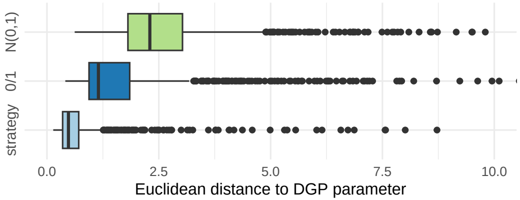
$$\boldsymbol{U} = \boldsymbol{X}(\underbrace{\beta + \gamma}_{\beta_n}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma), \quad \beta_n \sim \mathcal{N}(\beta, \Omega)$$

1. assume β is known
2. for ≈ 200 iterations:
   2.1 $\boldsymbol{U} \mid \Sigma, (\beta_n), \boldsymbol{X}, y \sim$ truncated $\mathcal{N}$
   2.2 $\beta_n \mid \beta, \boldsymbol{\Omega}, \boldsymbol{U} \sim \mathcal{N}$ for each n
   2.3 $\boldsymbol{\Sigma} \mid (\beta_n), \boldsymbol{U} \sim \mathcal{W}^{-1}$
   2.4 $\boldsymbol{\Omega} \mid (\beta_n), \beta \sim \mathcal{W}^{-1}$
3. derive $\widehat{\Sigma}$, $\widehat{\Omega}$ as sample average of Σ, Ω

Simulation: $\boldsymbol{\Omega} \sim \mathcal{W}^{-1}(P \cdot \boldsymbol{I}, P+1)$, $J = 3$, $T = 10$

N	100	■	500	□
P	3	■	5	□
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	□	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	■

Average computation time: 3 seconds



UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

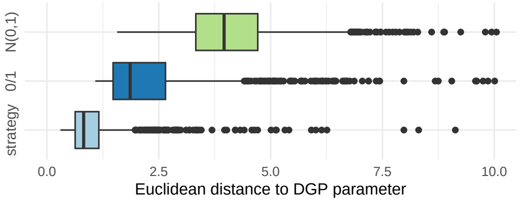
$$\boldsymbol{U} = \boldsymbol{X}(\underbrace{\beta + \gamma}_{\beta_n}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma), \quad \beta_n \sim \mathcal{N}(\beta, \Omega)$$

1. assume β is known
2. for ≈ 200 iterations:
   2.1 $\boldsymbol{U} \mid \Sigma, (\beta_n), \boldsymbol{X}, y \sim$ truncated $\mathcal{N}$
   2.2 $\beta_n \mid \beta, \boldsymbol{\Omega}, \boldsymbol{U} \sim \mathcal{N}$ for each n
   2.3 $\boldsymbol{\Sigma} \mid (\beta_n), \boldsymbol{U} \sim \mathcal{W}^{-1}$
   2.4 $\boldsymbol{\Omega} \mid (\beta_n), \beta \sim \mathcal{W}^{-1}$
3. derive $\widehat{\Sigma}$, $\widehat{\Omega}$ as sample average of Σ, Ω

Simulation: $\boldsymbol{\Omega} \sim \mathcal{W}^{-1}(P \cdot \boldsymbol{I}, P + 1)$, $J = 3$, $T = 10$

N	100	■	500	□
P	3	□	5	■
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	□

Average computation time: 3 seconds



UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

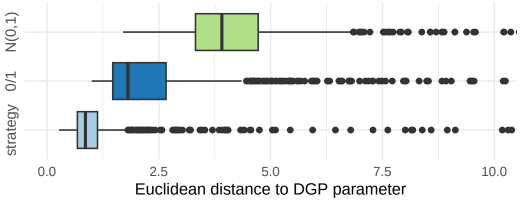
$$\boldsymbol{U} = \boldsymbol{X}(\underbrace{\beta + \gamma}_{\beta_n}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma), \quad \beta_n \sim \mathcal{N}(\beta, \Omega)$$

1. assume β is known
2. for $\approx$ 200 iterations:
   2.1 $\boldsymbol{U} \mid \Sigma, (\beta_n), \boldsymbol{X}, y \sim$ truncated $\mathcal{N}$
   2.2 $\beta_n \mid \beta, \Omega, \boldsymbol{U} \sim \mathcal{N}$ for each n
   2.3 $\Sigma \mid (\beta_n), \boldsymbol{U} \sim \mathcal{W}^{-1}$
   2.4 $\Omega \mid (\beta_n), \beta \sim \mathcal{W}^{-1}$
3. derive $\widehat{\Sigma}$, $\widehat{\Omega}$ as sample average of Σ, Ω

Simulation: $\boldsymbol{\Omega} \sim \mathcal{W}^{-1}(P \cdot \boldsymbol{I}, P + 1)$, $J = 3$, $T = 10$

N	100	■	500	□
P	3	□	5	■
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	□	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	■

Average computation time: 3 seconds



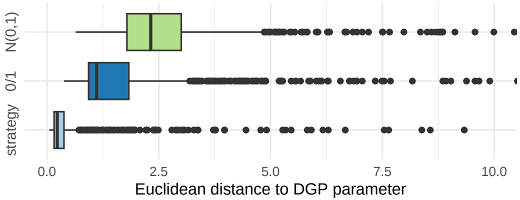
$$U = X(\underbrace{\beta + \gamma}_{\beta_n}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma), \quad \beta_n \sim \mathcal{N}(\beta, \Omega)$$

1. assume β is known
2. for ≈ 200 iterations:
   2.1 $U \mid \Sigma, (\beta_n), X, y \sim$ truncated $\mathcal{N}$
   2.2 $\beta_n \mid \beta, \Omega, U \sim \mathcal{N}$ for each n
   2.3 $\Sigma \mid (\beta_n), U \sim \mathcal{W}^{-1}$
   2.4 $\Omega \mid (\beta_n), \beta \sim \mathcal{W}^{-1}$
3. derive $\hat{\Sigma}$, $\hat{\Omega}$ as sample average of Σ, Ω

Simulation: $\Omega \sim \mathcal{W}^{-1}(P \cdot I, P+1)$, $J = 3$, $T = 10$

N	100	☐	500	■
P	3	■	5	☐
Σ	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	☐

Average computation time: 12 seconds



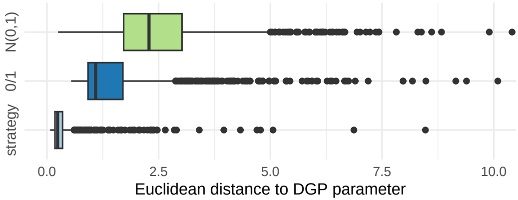
$$\boldsymbol{U} = \boldsymbol{X}(\underbrace{\beta + \gamma}_{\beta_n}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma), \quad \beta_n \sim \mathcal{N}(\beta, \boldsymbol{\Omega})$$

1. assume β is known
2. for ≈ 200 iterations:
   2.1 $\boldsymbol{U} \mid \Sigma, (\beta_n), \boldsymbol{X}, y \sim$ truncated $\mathcal{N}$
   2.2 $\beta_n \mid \beta, \boldsymbol{\Omega}, \boldsymbol{U} \sim \mathcal{N}$ for each n
   2.3 $\Sigma \mid (\beta_n), \boldsymbol{U} \sim \mathcal{W}^{-1}$
   2.4 $\boldsymbol{\Omega} \mid (\beta_n), \beta \sim \mathcal{W}^{-1}$
3. derive $\widehat{\Sigma}$, $\widehat{\boldsymbol{\Omega}}$ as sample average of Σ, $\boldsymbol{\Omega}$

Simulation: $\boldsymbol{\Omega} \sim \mathcal{W}^{-1}(P \cdot \boldsymbol{I}, P+1)$, $J = 3$, $T = 10$

N	100	☐	500	■
P	3	■	5	☐
Σ	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	☐	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	■

Average computation time: 12 seconds



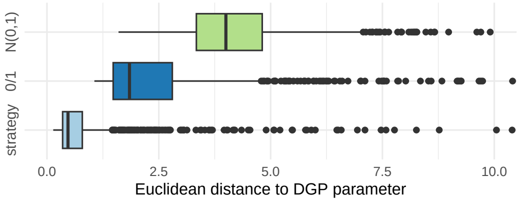
$$\boldsymbol{U} = \boldsymbol{X}(\underbrace{\beta + \gamma}_{\beta_n}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma), \quad \beta_n \sim \mathcal{N}(\beta, \boldsymbol{\Omega})$$

1. assume β is known
2. for ≈ 200 iterations:
   2.1 $\boldsymbol{U} \mid \Sigma, (\beta_n), \boldsymbol{X}, y \sim$ truncated $\mathcal{N}$
   2.2 $\beta_n \mid \beta, \boldsymbol{\Omega}, \boldsymbol{U} \sim \mathcal{N}$ for each n
   2.3 $\Sigma \mid (\beta_n), \boldsymbol{U} \sim \mathcal{W}^{-1}$
   2.4 $\boldsymbol{\Omega} \mid (\beta_n), \beta \sim \mathcal{W}^{-1}$
3. derive $\widehat{\Sigma}$, $\widehat{\boldsymbol{\Omega}}$ as sample average of Σ, $\boldsymbol{\Omega}$

Simulation: $\boldsymbol{\Omega} \sim \mathcal{W}^{-1}(P \cdot \boldsymbol{I}, P+1)$, $J = 3$, $T = 10$

N	100	☐	500	■
P	3	☐	5	■
Σ	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	■	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	☐

Average computation time: 12 seconds



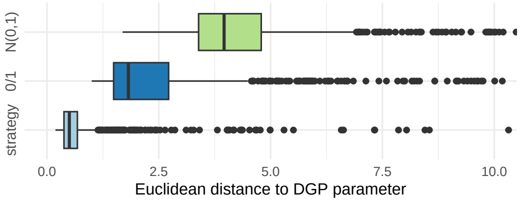
$$\boldsymbol{U} = \boldsymbol{X}(\underbrace{\beta + \gamma}_{\beta_n}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma), \quad \beta_n \sim \mathcal{N}(\beta, \Omega)$$

1. assume β is known
2. for ≈ 200 iterations:
   2.1 $\boldsymbol{U} \mid \Sigma, (\beta_n), \boldsymbol{X}, y \sim$ truncated $\mathcal{N}$
   2.2 $\beta_n \mid \beta, \Omega, \boldsymbol{U} \sim \mathcal{N}$ for each n
   2.3 $\Sigma \mid (\beta_n), \boldsymbol{U} \sim \mathcal{W}^{-1}$
   2.4 $\Omega \mid (\beta_n), \beta \sim \mathcal{W}^{-1}$
3. derive $\widehat{\Sigma}$, $\widehat{\Omega}$ as sample average of Σ, Ω

Simulation: $\boldsymbol{\Omega} \sim \mathcal{W}^{-1}(P \cdot \boldsymbol{I}, P+1)$, $J = 3$, $T = 10$

N	100	□	500	■
P	3	□	5	■
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \\ \vdots & & \ddots \end{pmatrix}$	□	$\begin{pmatrix} 3 & 1 & \cdots \\ 1 & 3 & \\ \vdots & & \ddots \end{pmatrix}$	■

Average computation time: 12 seconds



Outline

Putting them together

We can combine the strategies in the general "$\boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{X}(\boldsymbol{\beta} + \boldsymbol{\gamma}) + \boldsymbol{\varepsilon}$" case:

1. orthogonalize the regressors using the Frisch–Waugh–Lovell (FWL) theorem
2. get an initial estimate for $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$:
   2.1 apply the presented strategies separately
   2.2 estimate a common utility scale via a linear probability model
3. perform MCMC jointly for $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$

Outline

Empirical application

Stated-choice experiment by Helveston et al. (2015) on car purchase decision in USA:

Attribute*	Option 1	Option 2	Option 3
Vehicle Type	Conventional 300 mile range on 1 tank	Plug-In Hybrid & 300 mile range on 1 tank (first 40 miles electric)	Electric 75 mile range on full charge
Brand	German	American	Japanese
Purchase Price	$18,000	$32,000	$24,000
Fast Charging Capability	--	Not Available	Available
Operating Cost (Equivalent Gasoline Fuel Efficiency)	19 cents per mile (20 MPG equivalent)	12 cents per mile (30 MPG equivalent)	6 cents per mile (60 MPG equivalent)
0 to 60 mph Acceleration Time**	8.5 seconds (Medium-Slow)	8.5 seconds (Medium-Slow)	7 seconds (Medium-Fast)
	○	○	X

Empirical application

Stated-choice experiment by Helveston et al. (2015) on car purchase choice (USA):

- $N = 384$ deciders
- a total of 5760 decisions
- $J = 3$ alternatives
- $P = 16$ alternative-specific regressors
- we assumed $\varepsilon \sim$ iid $\mathcal{N}(0, \sqrt{1.6})$

Empirical application

Stated-choice experiment by Helveston et al. (2015) on car purchase choice (USA):

- $N = 384$ deciders
- a total of 5760 decisions
- $J = 3$ alternatives
- $P = 16$ alternative-specific regressors
- we assumed $\varepsilon \sim$ iid $\mathcal{N}(0, \sqrt{1.6})$

- With random initial values, only 15% of runs converged. This took a total of 3 hours.
- Informed initial values using the presented strategy led to a single run that converged in 2 minutes to the estimates reported by Helveston et al. (2015).

Empirical application

Stated-choice experiment by Helveston et al. (2015) on car purchase choice (USA):

- $N = 384$ deciders
- a total of 5760 decisions
- $J = 3$ alternatives
- $P = 16$ alternative-specific regressors
- we assumed $\varepsilon \sim$ iid $\mathcal{N}(0, \sqrt{1.6})$

- With random initial values, only 15% of runs converged. This took a total of 3 hours.
- Informed initial values using the presented strategy led to a single run that converged in 2 minutes to the estimates reported by Helveston et al. (2015).

UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

Outline

Takeaways

- RUMs are widely used in discrete choice applications, but MLE quickly becomes computational challenging (with uninformed initialization)
- We can improve optimization time and convergence rate by using consistent and numerically fast initial estimators:
  1. constant utility direction
  2. minimize choice frequency prediction error
  3. linear probability model
  4. MCMC

Takeaways

- RUMs are widely used in discrete choice applications, but MLE quickly becomes computational challenging (with uninformed initialization)

- We can improve optimization time and convergence rate by using consistent and numerically fast initial estimators:
  1. constant utility direction
  2. minimize choice frequency prediction error
  3. linear probability model
  4. MCMC

Thanks for your attention! Do you have any questions or comments?

lennart.oelschlaeger@uni-bielefeld.de
loelschlaeger.de/talks