A Benders decomposition for maximum simulated likelihood estimation of advanced discrete choice models

Tom Häring, Claudia Bongiovanni, Michel Bierlaire

TRANSP-OR Laboratory

EPFL

Contents

Why maximum likelihood estimation (MLE)?

- **MLE** is for example used to estimate the parameters of **discrete choice models**

hEART 2022

TRISTAN XI

ICMC 2022

Why maximum likelihood estimation (MLE)?

- For each **individual *n***, every **alternative *i*** has an associated **utility**:

parameters to be estimated

random error term

$$U_{in} = U_{in}(\beta, x, \varepsilon_{in})$$

exogenous attributes

- Assumptions: I.) **linear** in parameters
  II.) we can **draw** from error terms

Why maximum likelihood estimation (MLE)?

- For each **individual *n***, every **alternative *i*** has an associated **utility**:

deterministic part

$$U_{in} = \sum_k \beta_k x_{ink} + \epsilon_{in} = V_{in} + \epsilon_{in}$$

stochastic part

- Behavioral assumption: the individual chooses the alternative with **the highest utility**

Why maximum likelihood estimation (MLE)?

- Data: **observed choices** y_{in} (= 1 if ind. *n* chose alternative *i*, else = 0)
- Find parameters β_k such that the **likelihood** of this outcome is **maximized**
- **Log-Likelihood function:**

$$\ln\left(\prod_n \prod_i P_n(i)^{y_{in}}\right) = \sum_n \sum_i y_{in} \ln P_n(i)$$

where

$$P_n(i) = \mathbb{P}(V_{in} + \epsilon_{in} \geq V_{jn} + \epsilon_{jn} \forall j \in J)$$

Why simulated MLE?

- DCMs model choices **realistically** [1], but in general lead to **non-convex** probabilities [2]

  ➡ No global optimality certificates, **danger of local optima**

  ➡ Non-convex solver ≈ **Blackbox**

- **Simulation** mitigates this by giving a **linear** approximation [3] and allows DCMs to be easily **integrated** in optimization programs [2]

---

*[1] **Bierlaire**: Discrete choice models (1998)*
*[2] **Pacheco:** Integrating advanced discrete choice models in mixed integer linear optimization (2021)*
*[3] **Train:** Discrete choice methods with simulation (2009)*

Why simulated MLE?

- How:
  - **Simulate** R scenarios, utilities become **deterministic**:

$$U_{inr} = V_{in} + \epsilon_{inr}$$

Draw from distribution

  - Let ω_{inr} be the **choice variables**
  - **Approximated** probabilities:

$$\widehat{P}_n(i) = \frac{1}{R}\sum_{r=0}^{R-1} \omega_{inr}$$

Meritxell Pacheco: *A general framework for the integration of complex choice models into mixed integer optimization (2020)*

Why a mixed integer linear program (MILP)?

- Allow inclusion of **integer variables** in estimation procedure
  - ➢Model **advanced** DCMs, e. g. **latent variables** / **classes**
  - ➢Additional features, e. g. **automatic / assisted specification**
- Vast literature on efficient **modeling** & **performance**
- Gives **control** over **optimization process**: information on **bounds, optimality gaps, user-generated cuts**, etc.

Simulated MLE as an MILP

- **Objective**: max Log-Likellihood $$\sum_n \sum_i y_{in} \ln P_n(i)$$

max sim. Log-Likelihood $$\sum_{in} y_{in} \ln \sum_{r=0}^{R-1} \omega_{inr} - y_{in} \ln R$$

$$s_{in} = \sum_r \omega_{inr}$$

$$z_{in} \leq L_r - K_r s_{in}$$

max $$\sum_n \sum_i y_{in} z_{in}$$

Lurkin, Fernandez and Bierlaire: *A MILP formulation for the maximum likelihood estimation of continuous and discrete parameters in choice models (2018)*

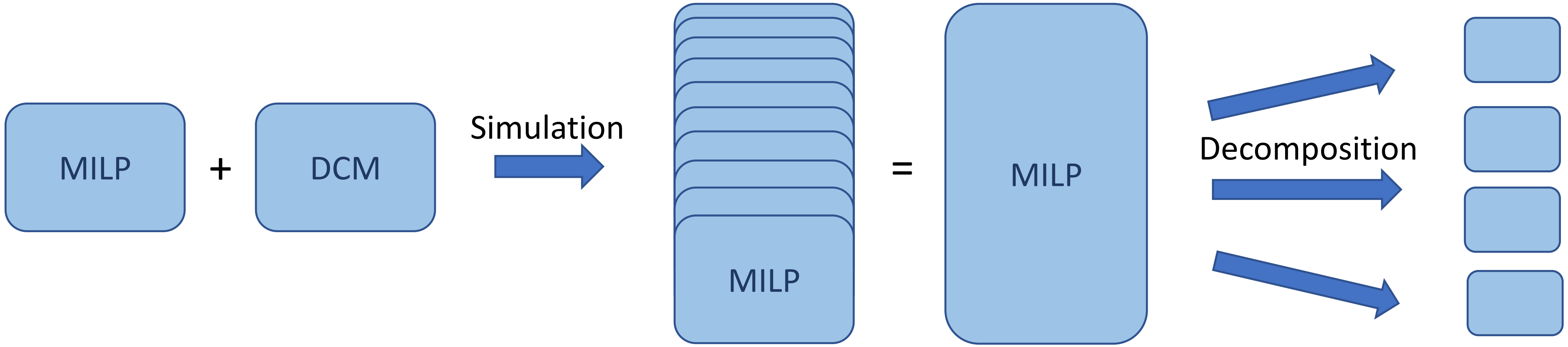
Simulated MLE as an MILP

- **Constraints**:

$$
\begin{aligned}
\sum_i \omega_{inr} &= 1 && \forall n, r \\
U_{inr} &= \sum_k \beta_k x_{ink} + \epsilon_{inr} && \forall i, n, r \\
U_{nr} &\geq U_{inr} && \forall i, n, r \\
U_{nr} &= \sum_i U_{inr}\omega_{inr} && \forall n, r \\
s_{in} &= \sum_r \omega_{inr} && \forall i, n \\
z_{in} &\leq L_r - K_r s_{in} && \forall i, n \\
\omega_{inr} &\in \{0, 1\} \\
\beta, s, z, U, U &\in \mathbb{R}
\end{aligned}
$$

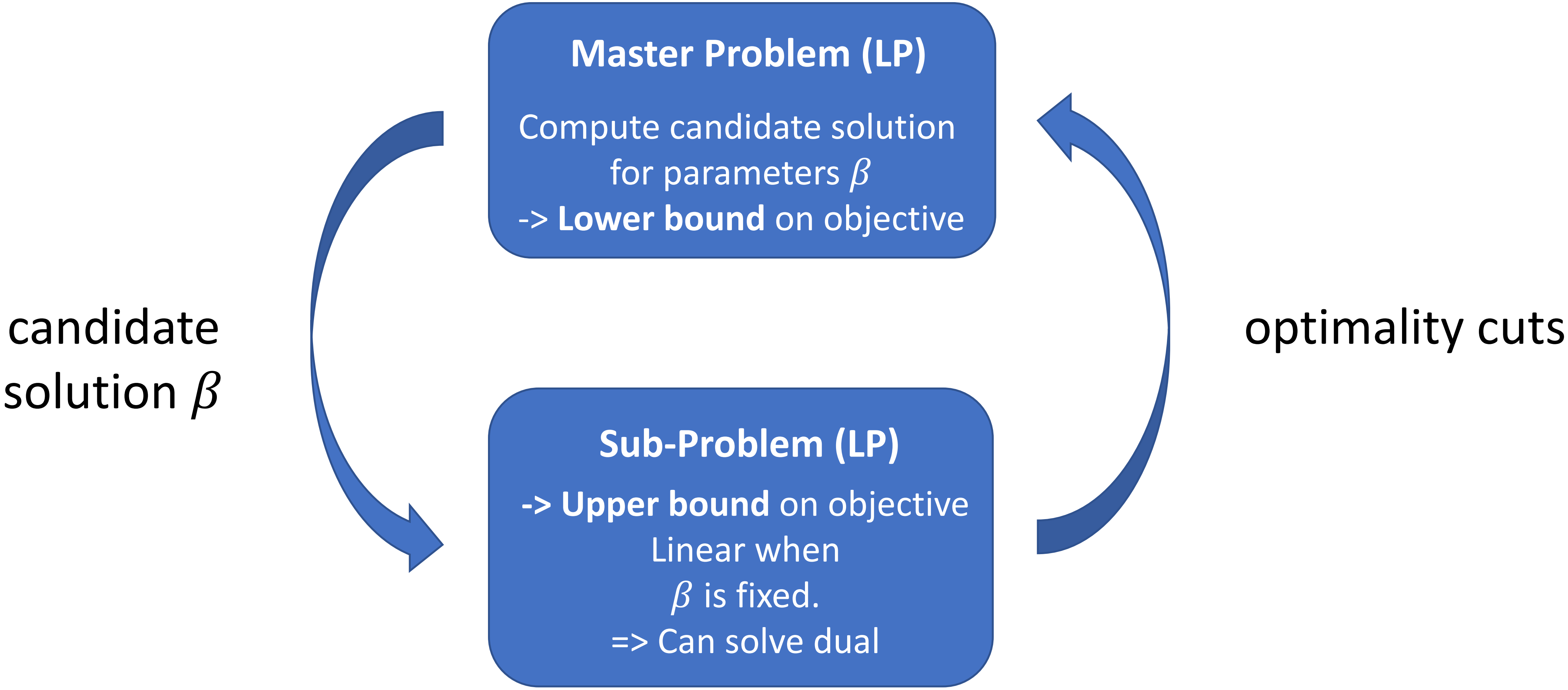
Why decomposition?

- Problem: Simulation **increases problem size** by solving **many scenarios**
  ➡ **only small instances** can be solved in reasonable time [1]
- To solve large MILPs efficiently we consider **decomposition methods**



[1] ***Pacheco:*** *Integrating advanced discrete choice models in mixed integer linear optimization (2021)*

The Benders decomposition



The Benders decomposition

- For a **fixed** β_k the rest of the MILP becomes a **Knapsack-problem => totally unimodular:**
  - Utilities become fixed $U_{inr} = \sum_k \beta_k^{\text{fixed}} x_{ink} + \epsilon_{inr}$
  - Now:

$$\left.\begin{aligned} U_{nr} &= \sum_i U_{inr}\omega_{inr} \\ U_{nr} &\geq U_{inr} \\ \sum_i \omega_{inr} &= 1 \\ \omega_{inr} &\in [0, 1] \end{aligned}\right\}$$

$\omega_{i^*nr} = 1$

for the alternative i^*

with highest utility

The Benders decomposition

- Start with **initial guess** for the variable to be fixed
- **Subproblems:**

  - relax **integral domains**: $\omega_{inr} \in [0,1]$

  - add **constraints**: $\beta_k = \beta_k^{\text{fixed}} \qquad (\varphi_k^{\beta})$

    ➡ RHS of primal = objective of dual = $\ldots + \sum_k \varphi_k^{\beta} \beta_k^{\text{fixed}}$

  - **solve dual**, get optimal values for φ_k^{β}

The Benders decomposition

- **Solve master problem:**

$$\min_{\beta,\mathcal{L}} \quad \mathcal{L}$$

$$\text{s.t.} \quad \mathcal{L} \geqslant \mathcal{L}^* + \sum_n \sum_k \varphi_{nk}^{\beta} (\beta_k - \beta_k^{\text{fixed}})$$

Benders cut

objective value of primal / dual

Replace fixed variable value by actual MP variable

The Benders decomposition

- **Typically:**
  - The variable to be fixed is **integer**, so that the subproblems are linear
  - Thus **MP is an integer program (bottleneck!)**

- **But in our case:**
  - The variable to be fixed is **continous**, but thanks to TU-ness the subproblems are (*technically*) still linear!
  - Thus **SP is a linear program!**

From solving an MILP to iteratively solving LP's!

The Benders decomposition

- **Difficulty:**

Simply adding the constraint $\beta_k = \beta_k^{\text{fixed}}$ **does not work in our case** because of the **non-linearity** of the problem

The Benders decomposition

- **Constraints**:

Goal: linear in $\boldsymbol{\beta_k}$

$$
\begin{aligned}
\sum_i \omega_{inr} &= 1 && \forall n, r \\
U_{inr} &= \sum_k \beta_k x_{ink} + \epsilon_{inr} && \forall i, n, r \\
U_{nr} &\geq U_{inr} && \forall i, n, r \\
U_{nr} &= \sum_i U_{inr}\omega_{inr} && \forall n, r \\
\beta_k &= \beta_k^{\text{fixed}} && \forall k \\
s_{in} &= \sum_r \omega_{inr} && \forall i, n \\
z_{in} &\leq L_r - K_r s_{in} && \forall i, n \\
\omega_{inr} &\in [0, 1] \\
\beta, s, z, U, U &\in \mathbb{R}
\end{aligned}
$$

Non-linear!

The Benders decomposition

- **Constraints**:

$$
\begin{aligned}
\sum_i \omega_{inr} &= 1 \\
U_{inr} &= \sum_k \beta_k x_{ink} + \epsilon_{inr} \\
U_{nr} &\geq U_{inr} \\
U_{nr} &= \sum_i U_{inr} \omega_{inr} \\
\beta_{\boldsymbol{k}} &= \beta_{\boldsymbol{k}}^{\text{fixed}}
\end{aligned}
$$

The Benders decomposition

- **Constraints**:

$$\begin{aligned}
\sum_i \omega_{inr} &= 1 \\
U_{inr} &= \sum_k \beta_k x_{ink} + \epsilon_{inr} \\
U_{nr} &\geq U_{inr} \\
U_{nr} &= \sum_i U_{inr}^{\text{fixed}} \omega_{inr} \\
\beta_k &= \beta_k^{\text{fixed}}
\end{aligned}$$

Disconnected!

The Benders decomposition

- **Constraints**:

$$
\begin{aligned}
\sum_i \omega_{inr} &= 1 \\
U_{inr} &= \sum_k \beta_k x_{ink} + \epsilon_{inr} \\
U_{nr} &\geq U_{inr} \\
U_{nr} &= \sum_i \omega_{inr} \Big[\sum_k \beta_k x_{ink} + \epsilon_{inr} \Big] \\
\beta_k &= \beta_k^{\text{fixed}}
\end{aligned}
$$

The Benders decomposition

- **Constraints**:

$$
\begin{aligned}
\sum_i \omega_{inr} &= 1 \\
U_{inr} &= \sum_k \beta_k x_{ink} + \epsilon_{inr} \\
U_{nr} &\geq U_{inr} \\
U_{nr} &= \sum_i \Big[\sum_k (\omega_{inr}\beta_k) x_{ink} + \omega_{inr}\epsilon_{inr} \Big] \\
\beta_k &= \beta_k^{\text{fixed}}
\end{aligned}
$$

The Benders decomposition

- **Constraints**:

$$
\begin{aligned}
\sum_i \omega_{inr} &= 1 \\
U_{inr} &= \sum_k \beta_k x_{ink} + \epsilon_{inr} \\
U_{nr} &\geq U_{inr} \\
U_{nr} &= \sum_i \Big[\sum_k \eta_{inrk} x_{ink} + \omega_{inr} \epsilon_{inr} \Big] \\
\eta_{inrk} &= \beta_k \omega_{inr} \\
\beta_k &= \beta_k^{\text{fixed}}
\end{aligned}
$$

The Benders decomposition

- **Constraints**:

$$
\begin{aligned}
\sum_i \omega_{inr} &= 1 \\
U_{inr} &= \sum_k \beta_k x_{ink} + \epsilon_{inr} \\
U_{nr} &\geq U_{inr} \\
U_{nr} &= \sum_i \left[\sum_k \eta_{inrk} x_{ink} + \omega_{inr} \epsilon_{inr} \right] \\
\psi_{inrk} &= \frac{1}{2}(\beta_k + \omega_{inr}) \\
\phi_{inrk} &= \frac{1}{2}(\beta_k - \omega_{inr}) \\
\eta_{inrk} &= \psi_{inrk}^2 - \phi_{inrk}^2 \\
\beta_k &= \beta_k^{\text{fixed}}
\end{aligned}
$$

$\eta_{inrk} = \psi_{inrk}^2 - \phi_{inrk}^2$ → piece-wise linear approximations

$\beta_k = \beta_k^{\text{fixed}}$ → **Does not preserve total unimodularity**

The Benders decomposition

- **Constraints**:

$$
\begin{aligned}
\sum_i \omega_{inr} &= 1 \\
U_{inr} &= \sum_k \beta_k x_{ink} + \epsilon_{inr} \\
U_{nr} &\geq U_{inr} \\
U_{nr} &= \sum_i U_{inr} \omega_{inr} \\
\beta_k &= \beta_k^{\text{fixed}}
\end{aligned}
$$

Convex?

The Benders decomposition

- We design a **quasi**-linearization:

$$\begin{aligned} \eta_{inrk} &= \beta_k \, \omega_{inr} \\ \beta_k &= \beta_k^{\text{fixed}} \end{aligned}$$

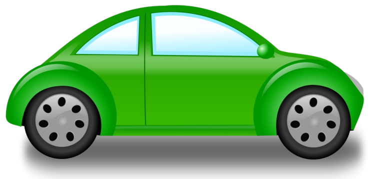
$\Rightarrow$

$$\begin{aligned} \chi_{inr} + \omega_{inr} &= 1 \\ \eta_{inrk} + \beta_k^{\text{fixed}} \chi_{inr} &= \beta_k^{\text{fixed}} \\ \sum_i \eta_{inrk} &= \beta_k \end{aligned}$$

Application to a mode choice problem

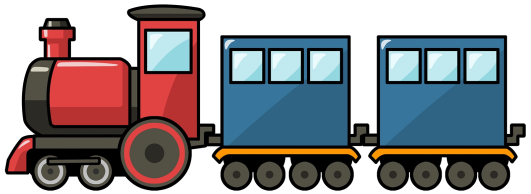
- Dataset: **RP** data on **mode choice**, Netherlands, 1987
- Simple **binary logit model**:

  choice between two modes – **car** and **rail**



$$U_{\mathrm{car},n} = \beta_{time} * \mathrm{traveltime}_{\mathrm{car}}$$

$$U_{\mathrm{rail},n} = \beta_{time} * \mathrm{traveltime}_{\mathrm{rail}}$$



- Compare **decomposition** vs. **undecomposed MILP**

N	R	sLL-M	sLL-D	Gap [%]	T-M	T-D
20	50	-12.607	-12.658	-0.40	64.942	10.061
20	100	-12.212	-12.258	-0.38	403.694	9.902
20	200	-12.283	-12.648	-2.97	1117.064	16.939
50	50	-30.848	-31.030	-0.59	286.679	29.780
50	100	-30.461	-31.040	-1.90	1558.604	65.006
50	200	-30.566	-30.692	-0.41	5375.655	98.206
100	50	-65.204	-65.801	-0.92	2820.229	28.781
100	100	-65.784	-67.419	-2.49	4346.067	274.163
100	200	-65.699	-66.018	-0.49	10800+	295.741
200	50	-123.551	-124.027	-0.39	1476.185	120.579
200	100	-124.000	-124.243	-0.20	10800+	327.253
200	200	-124.707	-124.106	0.48	10800+	1262.755

N	R	β - M	β - D	Gap [%]	T - M	T - D
20	50	-1.048	-0.97	7.44	65	10
20	100	-1.143	-1.11	2.89	404	10
20	200	-1.182	-2.16	-82.74	1117	17
50	50	-1.223	-0.935	23.55	287	30
50	100	-1.223	-1.783	-45.79	1559	65
50	200	-1.223	-1.307	-6.87	5376	98
100	50	-0.889	-0.612	31.16	2820	29
100	100	-0.943	-0.451	52.17	4346	274
100	200	-0.899	-0.85	5.45	10800	296
200	50	-1.39	-1.322	4.89	1476	121
200	100	-1.49	-1.393	6.51	10800	327
200	200	-1.021	-1.377	-34.87	10800	1263

Application to a mode choice problem

- First **conjecture**: gaps are caused by **log-linearization** in MSLE
- **Remedy:** apply decomposition to *continuous pricing problem (CPP)*

→ Almost **equivalent** problem structure, **no log-linearization**

Application to a continouos pricing problem

- Continuous pricing problem:

$$\max_{p,\omega,U,H} \sum_n \sum_r \sum_i \frac{1}{R} \theta_{in} p_i \omega_{inr}$$

s.t.

$$\begin{aligned}
\sum_i \omega_{inr} &= 1 && \forall n, r \\
H_{nr} &= \sum_i U_{inr} \omega_{inr} && \forall n, r \\
H_{nr} &\geq U_{inr} && \forall i, n, r \\
U_{inr} &= \sum_{k \neq l} \beta_k x_{ink} + \beta_l p_i + \varepsilon_{inr} && \forall i, n, r \\
\omega &\in \{0, 1\} \\
p, U, H &\in \mathbb{R}
\end{aligned}$$

Application to a continouos pricing problem

N	R	obj-MILP	obj-D	Gap [%]	P-MILP	P-D	Gap [%]	T-MILP	T-D
20	50	216.407	209.196	3.33	28.475	30.764	-8.04	7	11
20	100	202.642	201.712	0.46	28.302	26.576	6.1	37	21
20	200	200.901	200.185	0.36	30.03	28.721	4.36	205	49
50	50	440.686	437.243	0.78	28.579	29.989	-4.94	55	27
50	100	431.088	426.669	1.03	28.99	27.778	4.18	241	62
50	200	429.605	429.108	0.12	28.574	28.655	-0.28	1022	163
100	50	990.026	988.732	0.13	29.118	28.944	0.6	252	31
100	100	977.606	976.149	0.15	30.099	29.925	0.58	1224	69
100	200	978.589	976.932	0.17	30.106	30.185	-0.26	3039	304
200	50	1906.696	1904.189	0.13	28.977	28.678	1.03	1144	65
200	100	1882.793	1877.641	0.27	29.277	30.052	-2.65	4104	359
200	200	1873.964	1871.614	0.13	29.276	29.343	-0.23	10811	690

Large number of draws (MSLE)

N	R	sLL-M	sLL-D	Gap [%]	T-M	T-D
50	20	-29.417	-29.908	1.67	22	6
50	50	-29.294	-31.173	6.41	279	26
50	100	-28.885	-29.42	1.85	1375	42
50	150	-29.973	-30.092	0.4	2852	70
50	200	-30.091	-30.101	0.03	10800	131
50	250	-30.741	-30.775	0.11	10800	156
50	300	-30.837	-30.843	0.02	10800	133
50	400	-30.632	-30.638	0.02	10800	130
50	600	-30.479	-30.51	0.1	10800	289
50	800		-32.035		10800	319
50	1000		-30.523		10800	349

Ideas for future work

- Improving Benders:
  - Find a better way to linearize the product (?)
  - Find a convex-quadratic formulation (?)
- Investigate column generation
- Combined column generation + Benders approach


Thanks!