Mixed Integer Programming Approaches for Real-Time Consumer Preference Elicitation (and Causal Inference)

Juan Pablo Vielma

Massachusetts Institute of Technology

Operations Management/Management Science Workshop
Booth School of Business, University of Chicago.

Chicago, IL, October, 2017.

(Nonlinear) Mixed Integer Programming (MIP)

$$\min \quad f(x)$$
$$s.t.$$
$$x \in C$$
$$x_i \in \mathbb{Z} \quad i \in I$$

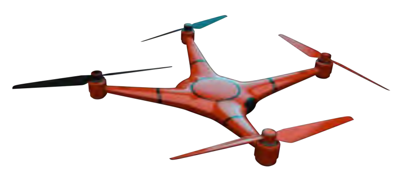
Mostly convex f and C.

http://www.gurobi.com/company/example-customers

Feature	Chewbacca	BB-8
Wookiee	Yes	No
Droid	No	Yes
Blaster	Yes	No
I would buy toy	☑	☐

Marketing
And
Experimental Design

Causal Inference for Educational Impact of 2010 Chilean Earthquake



“Infinite”-Dimensional MIP and Control or Aerial Drones

50+ Years of MIP = Significant Solver Speedups

- Algorithmic Improvements (**Machine Independent**):
  - Commercial Solver Speedup ≈ **1.9 x / year**

GUROBI OPTIMIZATION IBM CPLEX FICO

- Mostly linear, but also quadratic:
  - Gurobi v6.0 (2014) – v6.5 (2015) quadratic: **4.43 x** (**V**., Dunning, **Huchette**, Lubin, 2015)
- Also great "open-source" solvers

SCIP COIN|OR CBC GNU GLPK

- Emerging: General Convex Nonlinear (e.g. SDP)

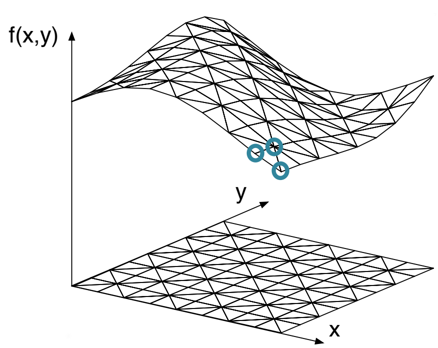
KNITRO < COIN|OR Bonmin < 

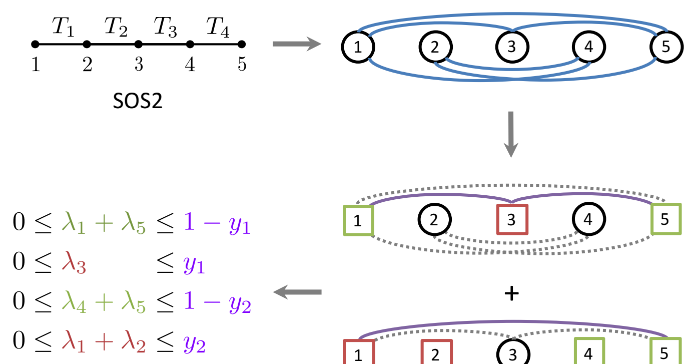
MIP Modelling and Advanced Formulations

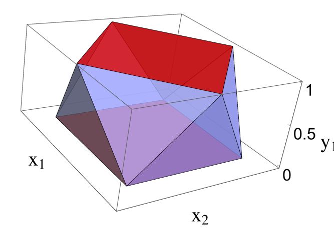
- MIP Representability: What can be modeled with MIP?
  - Linear: Jeroslow & Lowe ‘80s … Basu, Martin, Ryan and Wang ’17
  - Convex Nonlinear:
    - MIP formulation for the set of **Prime Numbers**

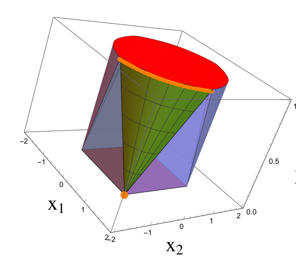
      ✓ Non-Convex Polynomial MIP formulation (Jones et al. ’76)

      ✗ Convex of any kind (Lubin, Zadik and **V.** ’17)
- Linear/nonlinear formulation techniques:









Accessing Solvers = Modelling Languages

- User-friendly algebraic modelling languages (AML):

GAMS
AMPL

Standalone and Fast

PYOMO ↔ python™
CVX RESEARCH ↔ MATLAB®

Based on General Language and Versatile

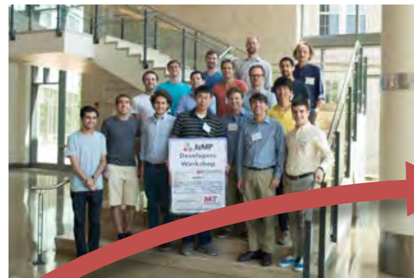
- 21st Century AML: JuMP ↔ julia
  - Free and Open-Source
  - Easy to use, but as advanced as proprietary C/C++ interphases
  - As fast as standalone AMLs and C/C++ interphases

JuMP

Created by students

Iain Dunning, Miles Lubin
and **Joey Huchette**
2016 ICS Prize

Community Developers



Google Research

OPERATIONS RESEARCH CENTER

Software Engineer

Jarrett Revels

MIT MANAGEMENT SLOAN SCHOOL

JuMP-Suit?

Juan Pablo Vielma

MIT MANAGEMENT SLOAN SCHOOL

Outline

- MIP and Consumer Preference Elicitation
  - Direct improvement from MIP formulation
  - Fast, versatile and efficient learning

- MIP and Causal Inference
  - Indirect improvement from MIP formulation
  - Right formulation brings you back to solving the problem you really wanted to solve

Mixed Integer Programming
(joint work with Joey Huchette)

and

Consumer Preference Elicitation
(joint work with Denis Saure)

Adaptive Choice-Based Conjoint Analysis

Feature	SX530	RX100
Zoom	50x	3.6x
Prize	$249.99	$399.99
Weight	15.68 ounces	7.5 ounces
Prefer	☑	☐

Feature	TG-4	G9
Waterproof	Yes	No
Prize	$249.99	$399.99
Weight	7.36 lb	7.5 lb
Prefer	☐	☑

Feature	TG-4	Galaxy 2
Waterproof	Yes	No
Prize	$249.99	$399.99
Viewfinder	Electronic	Optical
Prefer	☑	☐

Estimate of preference **parameter**

- Today: Minimize **variance** of **parameter** estimates

Parametric Model = Logistic Regression

Feature	Chewbacca	BB-8
Wookiee	Yes	No
Droid	No	Yes
Blaster	Yes	No
Prefer?	☑ ✓	☐
	x^1	x^2

Product profile

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = x^2$$

MNL Random **Linear** Utility

$$U_j = \underbrace{\beta \cdot x^j}_{\sum_{j=1}^{d} \beta_i x_i^j} + \epsilon_j$$

$$\longleftrightarrow \quad z = x^1 - x^2$$

Question:

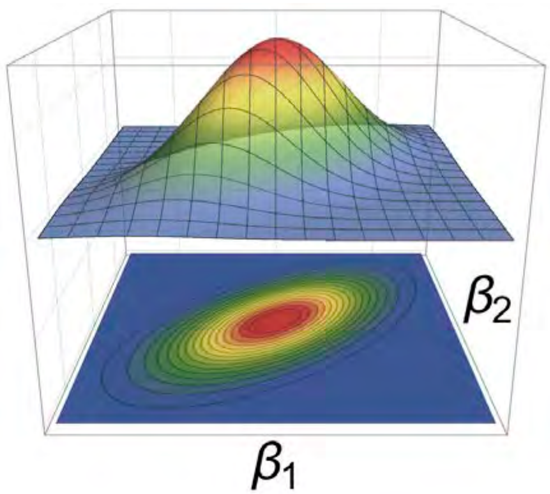
$$x^1 \succ x^2 \Leftrightarrow \quad U_1 \text{ “>” } U_2$$

$$\Leftrightarrow \beta \cdot z \text{ “>” } 0$$

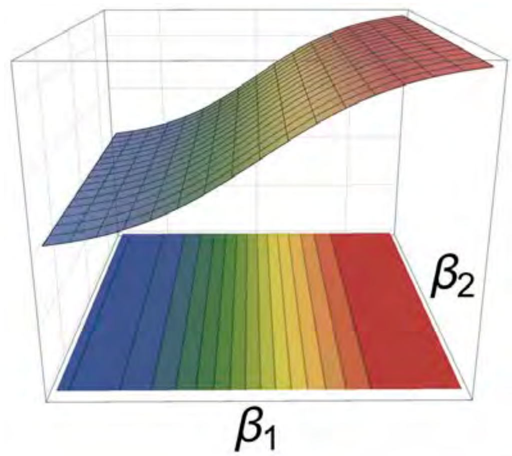
$$\mathbb{P}\left(x^1 \succ x^2 \mid \beta\right) = \frac{1}{1 + e^{-\beta \cdot z}}$$

Bayesian Model with Normal Prior

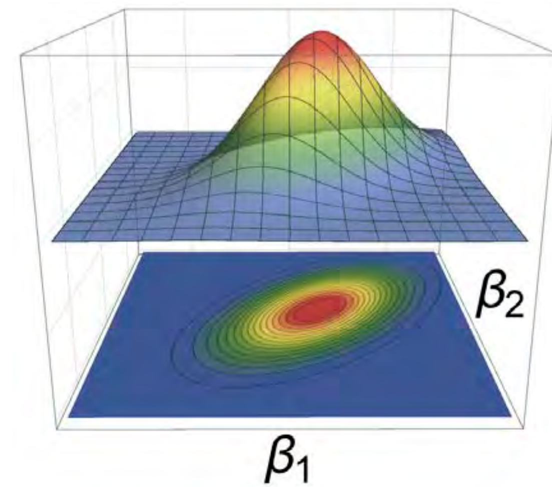
Prior distribution



Answer likelihood



Posterior distribution



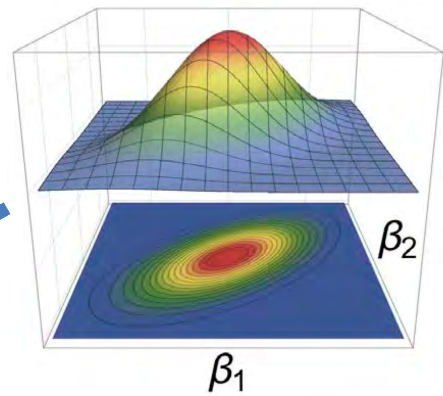
$$\beta \sim N(\mu, \Sigma) \qquad L(y \mid \beta, z) \qquad g(\beta \mid y, z)$$

$$y = \text{sign}(\beta \cdot z) \qquad L(y \mid \beta, z) = \left(1 + e^{-y\beta \cdot z}\right)^{-1}$$

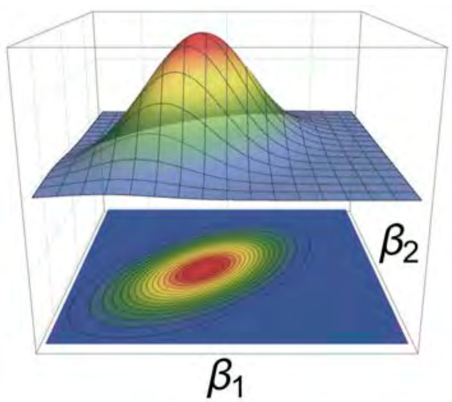
$$g(\beta \mid y, z) \propto \phi(\beta \,;\, \mu, \Sigma)\, L(y \mid \beta, z)$$

D-Error and Expected Posterior Variance

$$f(z,\mu,\Sigma) := \mathbb{E}_{y,\beta}\left\{(\det \operatorname{cov}(\beta \mid y, z))^{1/m}\right\}$$

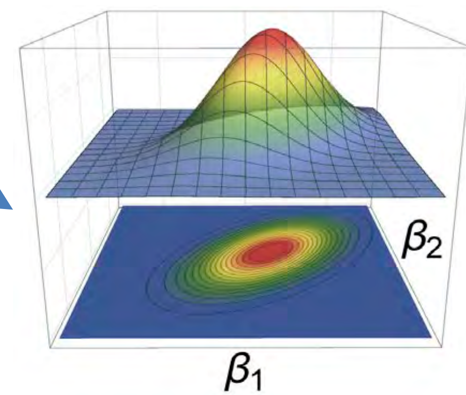


$y < 0$ $\quad$ $y > 0$



$\operatorname{cov}(\beta) = \Sigma_1$

$$\beta \sim N(\mu, \Sigma)$$



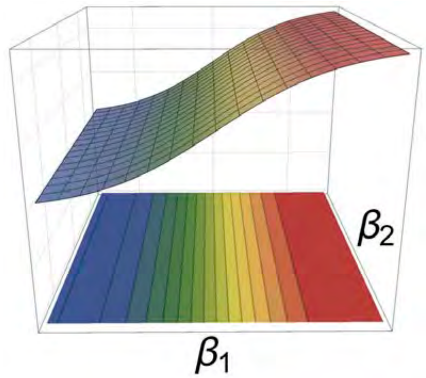
$\operatorname{cov}(\beta) = \Sigma_2$

$$\min_{z \in \{-1,0,1\}^n \setminus \{\mathbf{0}\}} f(z,\mu,\Sigma)$$

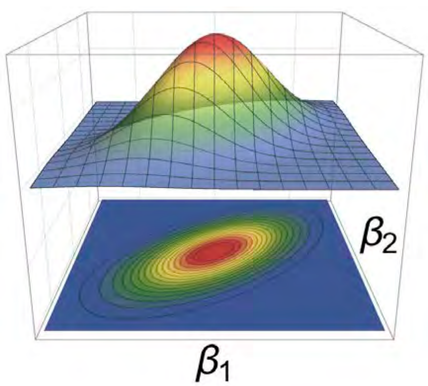
- $f(z,\mu,\Sigma)$ is hard to evaluate, non-convex and n large

1st Step: Moment-Matching Approximate Bayes

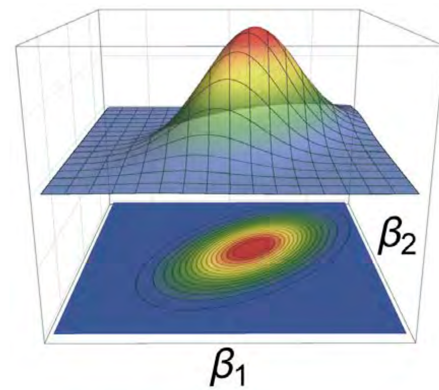
Answer likelihood



Prior distribution



Posterior distribution



Feature	TG-4	Galaxy 2
Waterproof	Yes	No
Prize	$249.99	$399.99
Viewfinder	Electron	Optical
Prefer	✓	☐

$$\beta \sim N\left(\mu^i, \Sigma^i\right)$$

$$\beta \overset{approx.}{\sim} N\left(\mu^{i+1}, \Sigma^{i+1}\right)$$

- $\mu^{i+1} = \mathbb{E}\left(\beta \,|\, y, x^1, x^2\right)$
- $\Sigma^{i+1} = \text{cov}\left(\beta \,|\, y, x^1, x^2\right)$

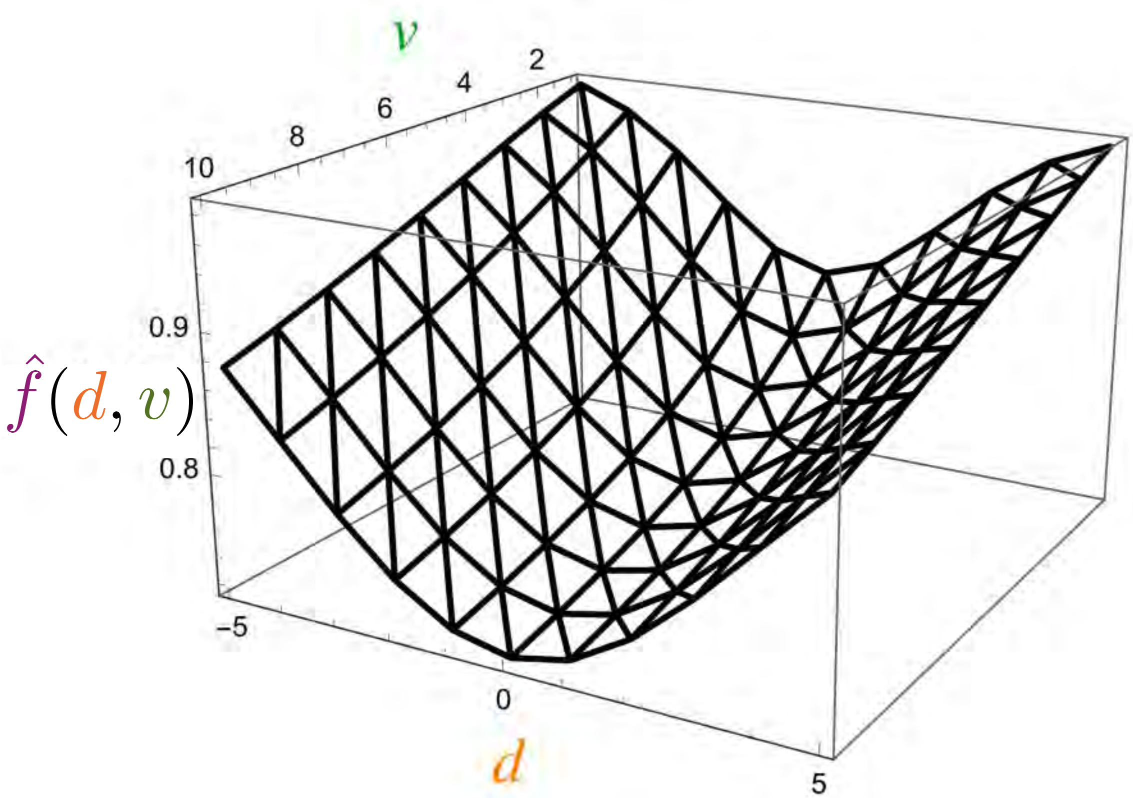
- Linear Algebra + 1-d numerical integration (e.g. BDA3)

2^{nd} Step: More Linear Algebra from V. and S. '16

- D-efficiency $f(z)$ = Non-convex function $f(d, v)$ of

$$\text{mean:} \quad d := \mu \cdot z$$

$$\text{variance:} \quad v := z' \cdot \sum \cdot z$$



Can evaluate $f(d, v)$ with 1-dim integral 🙂

Piecewise Linear (PWL) Interpolation $\hat{f}(d, v)$

Balances known criteria:

- minimize mean of question (no clear expected answer)
- maximize variance of question (uncertainty in expected answer)

3rd Step: "Almost" Direct Linear MIP Formulation

$z = x^1 - x^2$

MIP formulation for PWL function

$$\min \quad \hat{f}(d, v) \quad ✓$$

$$s.t.$$

$$\mu \cdot (x^1 - x^2) = d$$

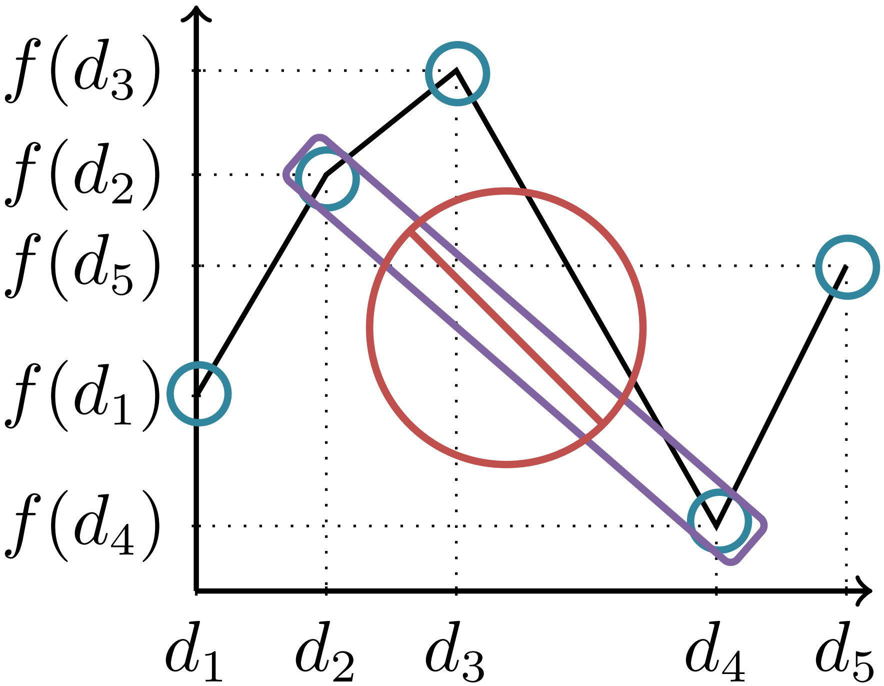
$$(x^1 - x^2)' \cdot \Sigma \cdot (x^1 - x^2) = v \quad ✓$$

linearize $x_i^k \cdot x_j^l$ $\quad \|x^1 - x^2\|_2^2 \geq 1 \quad (x^1 \neq x^2)$

$$x^1, x^2 \in \{0, 1\}^n$$

Simple Formulation for Univariate Functions

$$z = f(x)$$



$$\text{Size} = O(\#\text{ of segments})$$

Non-Ideal: Fractional Extreme Points

$$\begin{pmatrix} x \\ z \end{pmatrix} = \sum_{j=1}^{5} \begin{pmatrix} d_j \\ f(d_j) \end{pmatrix} \lambda_j$$

$$1 = \sum_{j=1}^{5} \lambda_j, \quad \lambda_j \geq 0$$

$$y \in \{0,1\}^4, \quad \sum_{i=1}^{4} y_i = 1$$

$$0 \leq \lambda_1 \leq y_1$$

$$0 \leq \lambda_2 \leq y_1 + y_2$$

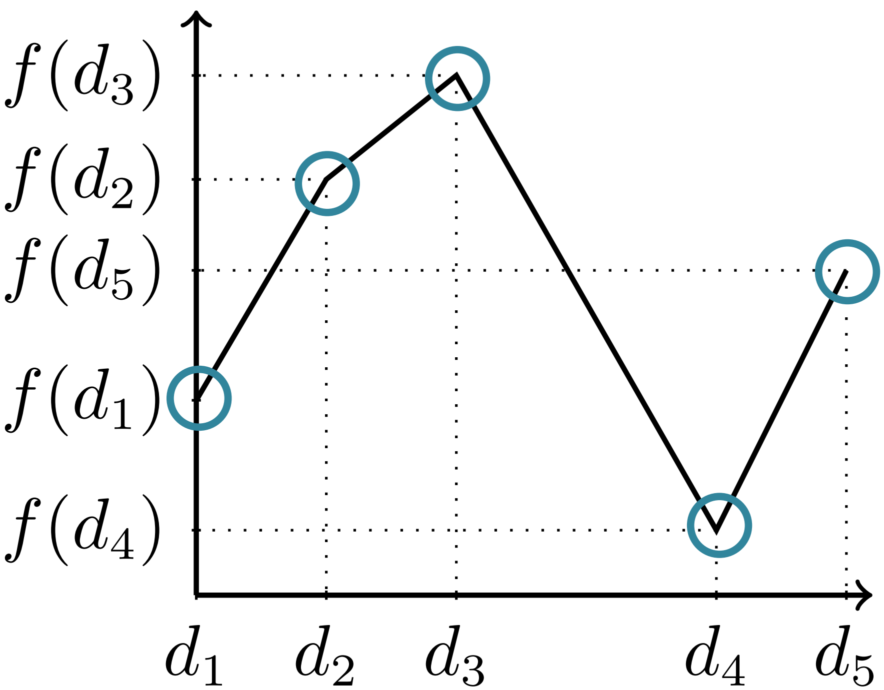
$$0 \leq \lambda_3 \leq y_2 + y_3$$

$$0 \leq \lambda_4 \leq y_3 + y_4$$

$$0 \leq \lambda_5 \leq y_4$$

Advanced Formulation for Univariate Functions

$$z = f(x)$$



$$\binom{x}{z} = \sum_{j=1}^{5} \binom{d_j}{f(d_j)} \lambda_j$$

$$1 = \sum_{j=1}^{5} \lambda_j, \quad \lambda_j \geq 0$$

$$y \in \{0, 1\}^2$$

$$0 \leq \lambda_1 + \lambda_5 \leq 1 - y_1$$

$$0 \leq \lambda_3 \leq y_1$$

$$0 \leq \lambda_4 + \lambda_5 \leq 1 - y_2$$

$$0 \leq \lambda_1 + \lambda_2 \leq y_2$$

$$\text{Size} = O(\log_2 \# \text{ of segments})$$

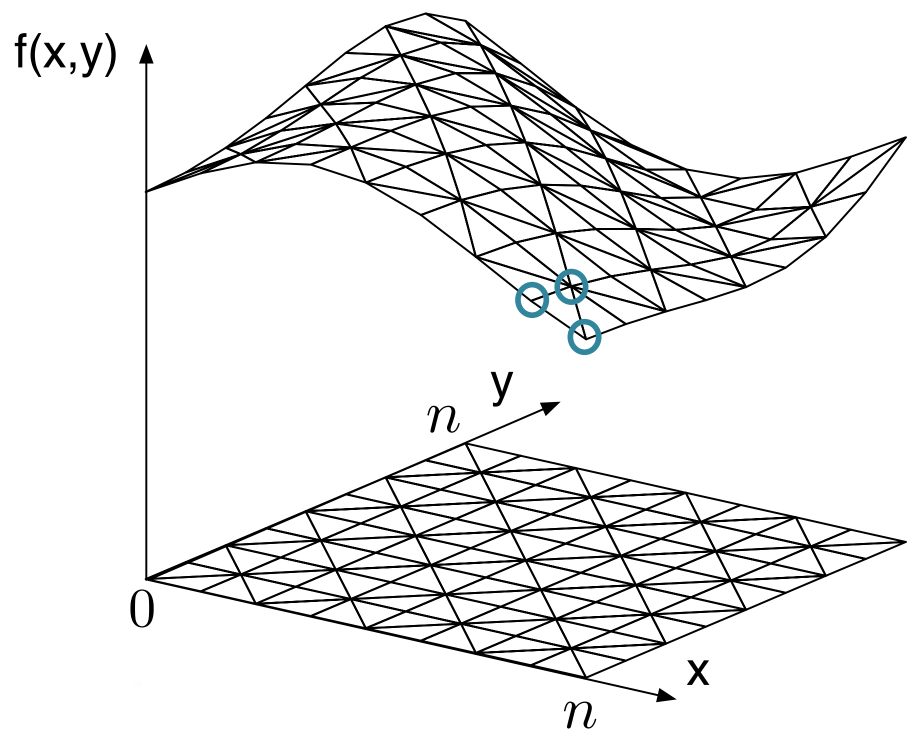
Ideal: Integral Extreme Points

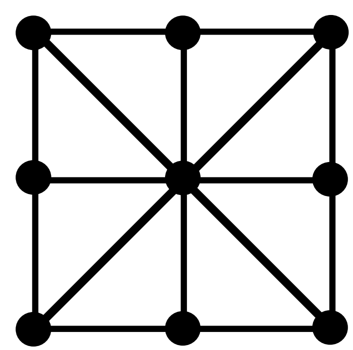
Significant computational advantage

V. and Nemhauser, 2011.

Technique Also Works for Multivariate Functions

- Union Jack triangulation (V. and Nemhauser, 2011)
  - $\text{Size} = 4\lceil \log_2 n \rceil + 2$
- For general triangulations (Huchette and V., 2016, 2017)
  - $\text{Size} \leq 4\lceil \log_2 n \rceil + 6$
  - Based on finding a bi-clique cover of an auxiliary graph
    - Can use a MIP to find the smallest formulation!





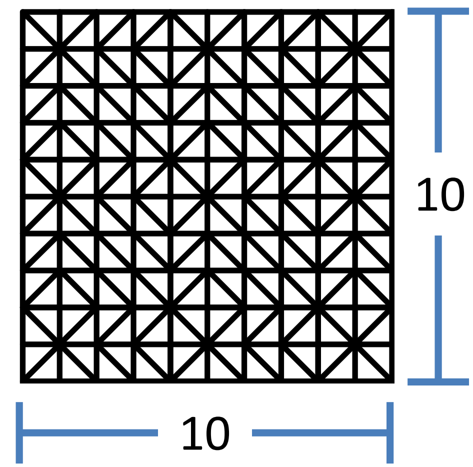
Easy to Build through julia & JuMP

- PiecewiseLinearOpt.jl (**Huchette** and V. 2017)

$$\begin{aligned} \min \quad & \exp(x+y) \\ s.t. \quad & \\ & x, y \in [0,1] \end{aligned}$$

Automatically select Δ

Automatically construct formulation (easily chosen)



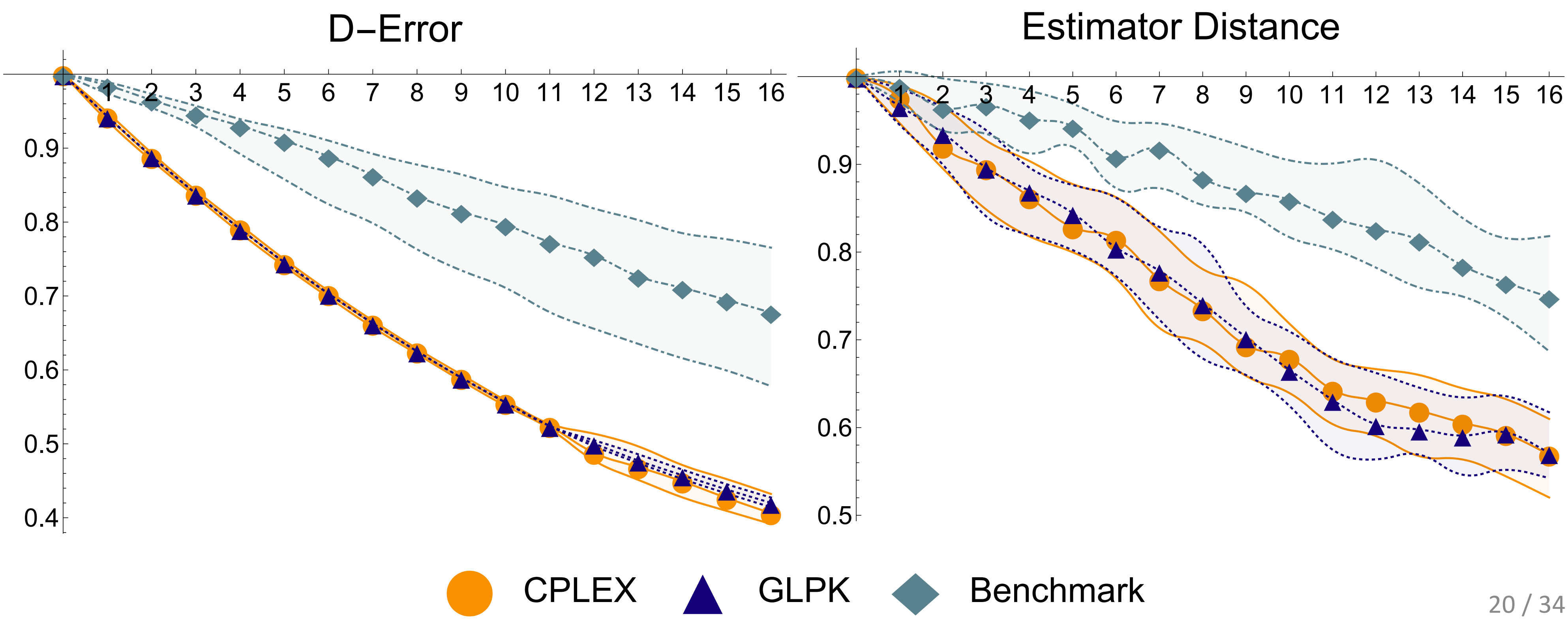

```
using JuMP, PiecewiseLinearOpt
m = Model()
@variable(m, x)
@variable(m, y)

z = piecewiselinear(m, x, y, 0:0.1:1, 0:0.1:1, (u,v) -> exp(u+v))
@objective(m, Min, z)
```

Easy to Build through julia & JuMP

```julia
function getquestion(μ,Σ,variancefuction)
  n = size(Σ,1)
  m = Model()
  # define variables for linearization
  @variable(m, 0 <= x[1:n] <= 1, Int)
  @variable(m, 0 <= y[1:n] <= 1, Int)
  # x ≠ y
  @constraint(m, linquad(m,(x-y)⋅(x-y)) >= 1)
  # v = x-y, β ~ 𝒩(μ,Σ), v⋅β ~ 𝒩(μᵥ,σ²), μᵥ = μ⋅v, σ² = v'*Σ*v
  @variable(m, μᵥ)
  @constraint(m, μᵥ == μ⋅(x-y) )
  @variable(m, σ² >=0)
  @constraint(m, σ² == linquad(m,(x-y)⋅(Σ*(x-y))))
  # (x-y)'*Σ*(x-y) <= eigmax(Σ) ||x-y||₂ <= eigmax(Σ)*n
  σ̄² = eigmax(Σ)*n
  # (x-y)'*Σ*(x-y) >= eigmin(Σ) ||x-y||₂ >= eigmin(Σ)   ( x ≠ y )
  σ̲² = eigmin(Σ)
  μ̄ᵥ = norm(μ,1)
  μᵥnpoints = 2^k - 1
  μᵥpoints = 0:μ̄ᵥ/μᵥnpoints:μ̄ᵥ+(μ̄ᵥ/μᵥnpoints)/2
  σ²range = σ̄² - σ̲²
  σ²npoints = 2^k-1
  σ²points = σ̲²:σ²range/σ²npoints:σ̄²+(σ²range/σ²npoints)/2
  pwl = BivariatePWLFunction(μᵥpoints, σ²points, (μᵥ,σ²) -> variancefuction(μᵥ,sqrt(σ²)))
  obj = piecewiselinear(m, μᵥ, σ², pwl)
  @objective(m, Min,  obj )
  status = solve(m)
  return [  round(Int64,getvalue(x)),  round(Int64,getvalue(y))]
end
```

MIP v/s Best Benchmark (Toubia et al. '03,'04)

- 16 questions, 2 options, 12 features, 100 individual β^* sampled from known prior N(μ,Σ)
- Best **benchmark** v/s MIP + Moment Matching
- CPLEX: ≤ 1 s (0.2 s Avg.), GLPK: ≤ 5 s (1.7 s Avg.)



Easy To Add Questionnaire Rules

Feature	Chewbacca	BB-8
Wookiee	Yes	No
Droid	No	Yes
Blaster	Yes	No
Prefer?	☑	☐

Product profile

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = x^2$$

- Realism is important: Wookiees are not Droids!

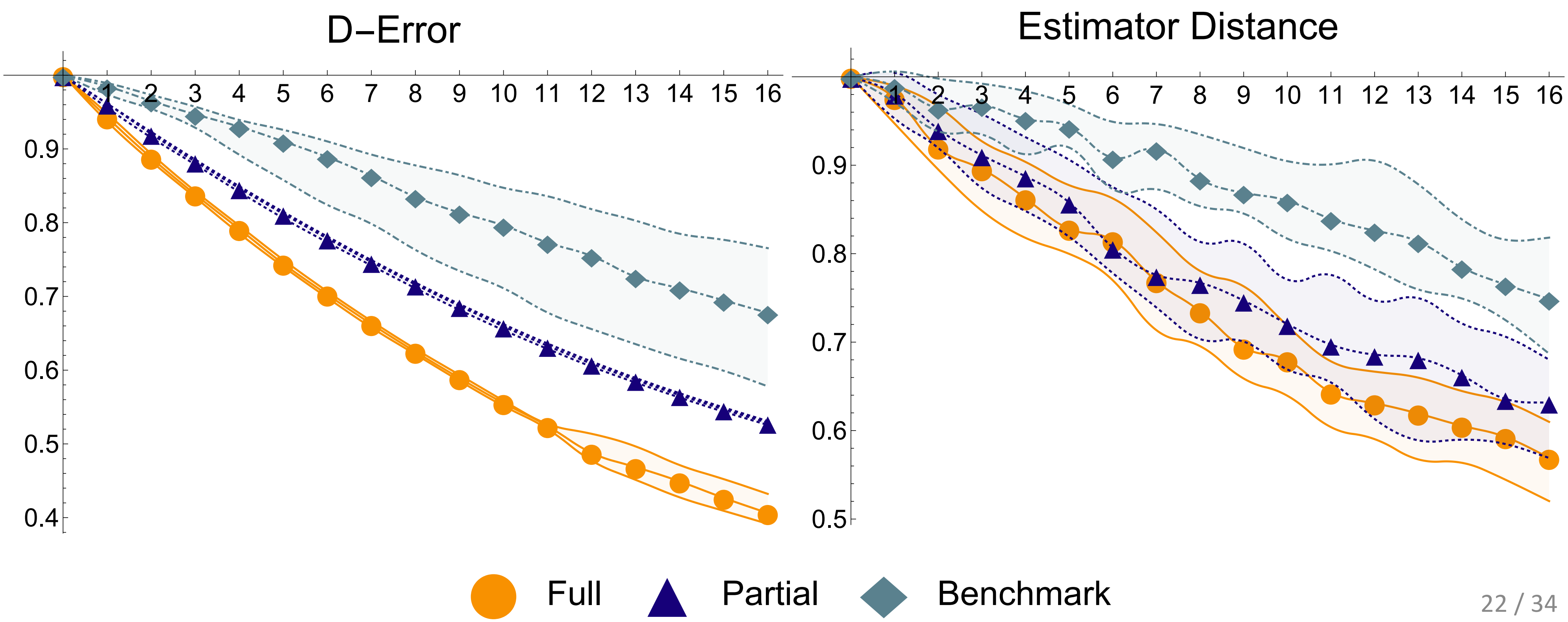
$$x^1_{\text{Wookie}} + x^1_{\text{Droid}} \leq 1$$

- Partial Profiles:
  - Limit # of feature differences and assume those not shown are the same (e.g. both are members of the resistance)

$$\left\| x^1 - x^2 \right\|_1 \leq 3$$

Full v/s Partial Profiles (# Feature Differences)

- 16 questions, 2 options, 12 features, 100 individual β^* sampled from known prior N(μ,Σ)
- Best **benchmark** v/s MIP + Moment Matching (CPLEX)
- Full: ≤ 1 s (0.2 s Avg.), Partial (5 diff.): ≤ 66 s (8 s Avg.)



MIP and Causal Inference

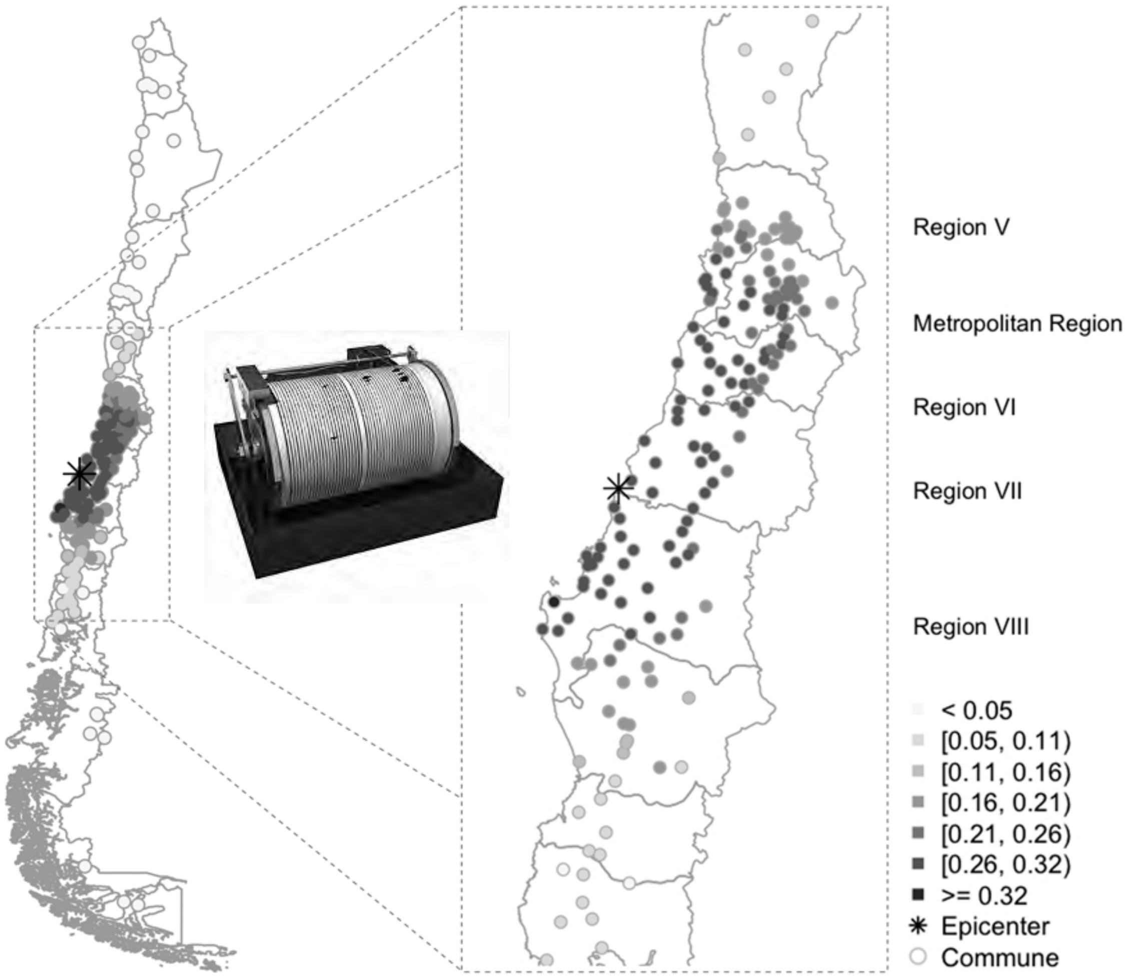
Joint work with Magdalena Bennett and Jose Zubizarreta

Educational Impact of 2010 Chilean Earthquake

- 6th Strongest in Recorded History (8.8)

- Impact on Educational Achievement (PSU Scores)?

Very High Quality Data is Available



Earthquake Intensity

SIMCE
Sistema de Medición de Calidad de la Educación

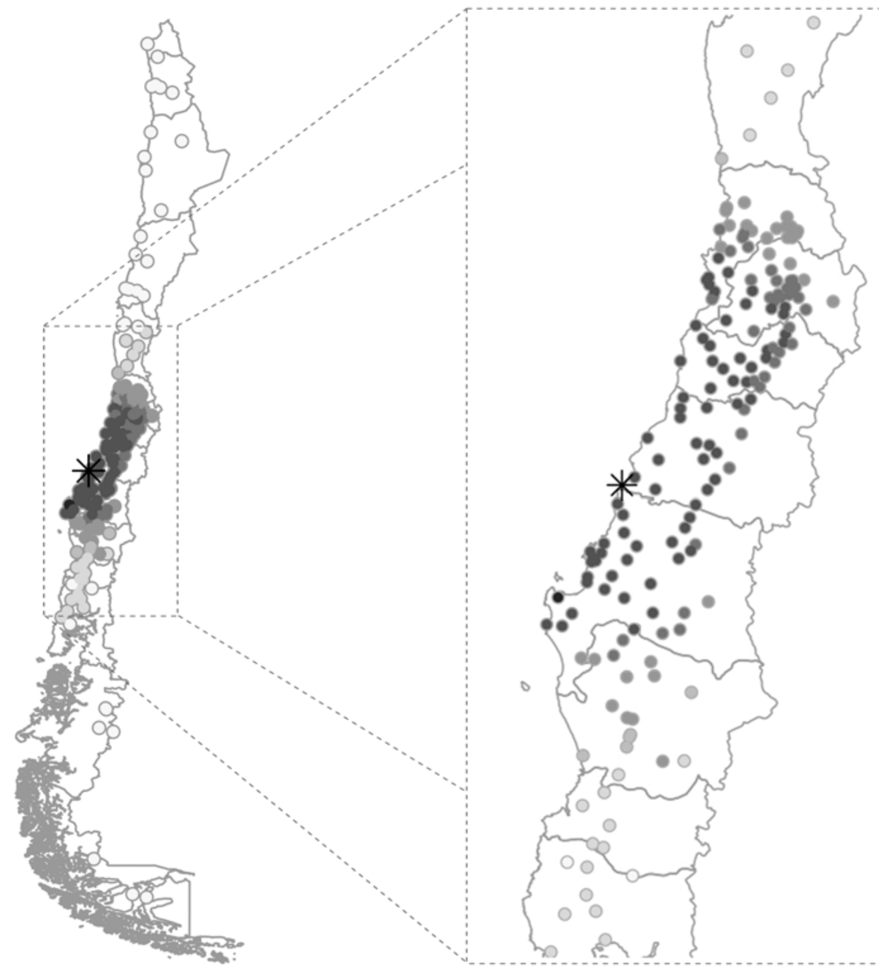
Agencia de Calidad de la Educación
Gobierno de Chile

Test Scores and Demographic Info

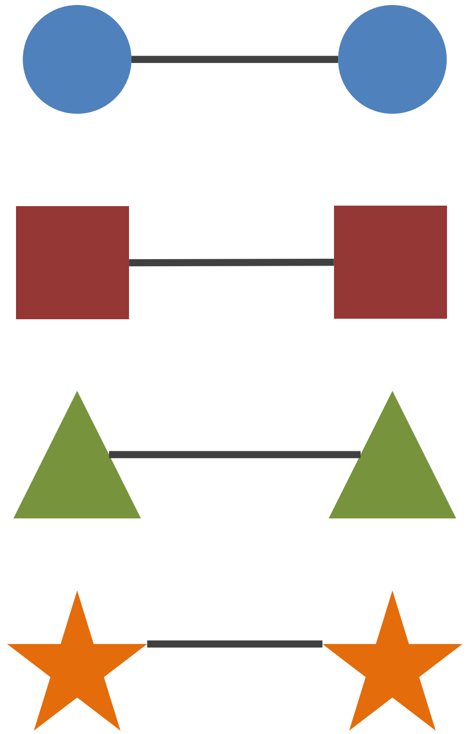
Covariate Balance Important for Inference

- Dose 1 = Control = Affected by Earthquake
- Dose 2 = Treatment = Not affected by Earthquake

Covariate	Dose 1	Dose 2
Gender		
Male	462	462
Female	538	538
School SES		
Low	75	75
Mid-low	327	327
Medium	294	294
Mid-high	189	189
High	115	115
Mother's education		
Primary	335	335
Secondary	426	426
Technical	114	114
College	114	114
Missing	11	11
⋮		

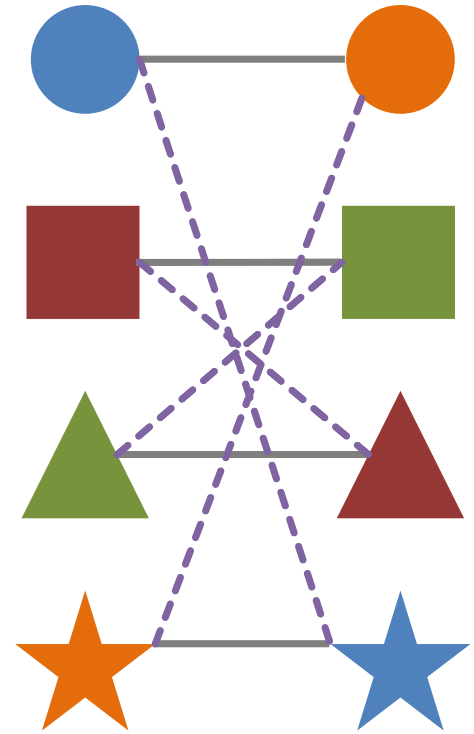


Traditional Matching: Exact v/s Fine Balance



Exact Matching

Match units with same category in all covariates



Fine Balance

Different matches for different covatiates

Matching v/s MIP

- Matching: # Variables = (# treated) × (# controls)
- Simple MIP:
  - Just count & balance units in each category/covariate
  - # Variables = (# treated) + (# controls)
  - Can use known results (Balas and Pulleyblank , 1983) to show MIP formulation is as strong as matching formulation
  - Can show MIP formulation is integral for 2 covariates
  - Problem is NP-hard for > 2 covariates
  - Usually very fast solve times: cuppa coffee time ≈ 5 min
  - More flexible…. doses and representability!!

Multiple Doses + Representation

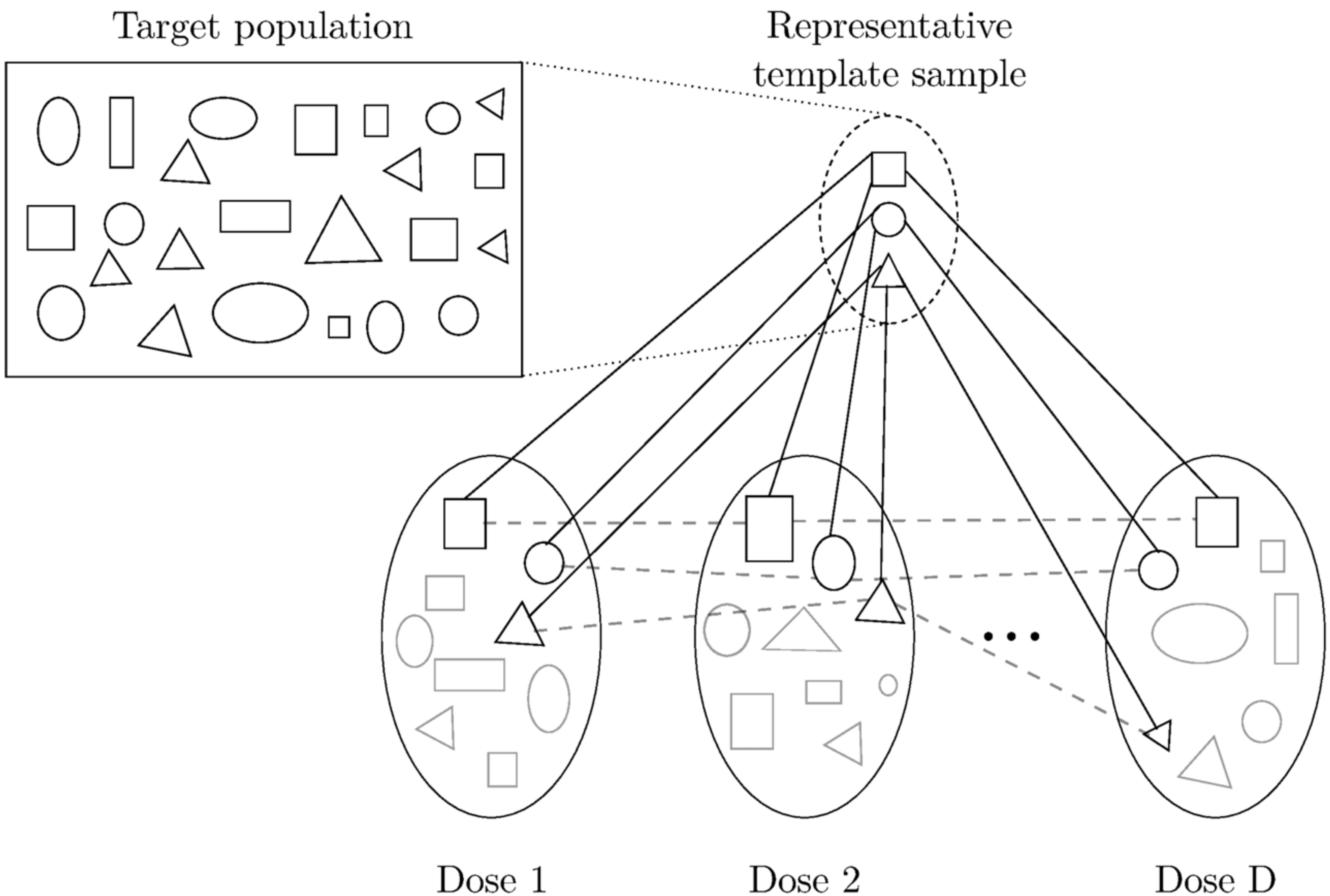
- Base, Medium and High

Covariate	Dose 1	Dose 2	Dose 3
Gender			
Male	462	462	462
Female	538	538	538
School SES			
Low	75	75	75
Mid-low	327	327	327
Medium	294	294	294
Mid-high	189	189	189
High	115	115	115
Mother's education			
Primary	335	335	335
Secondary	426	426	426
Technical	114	114	114
College	114	114	114
Missing	11	11	11
⋮			

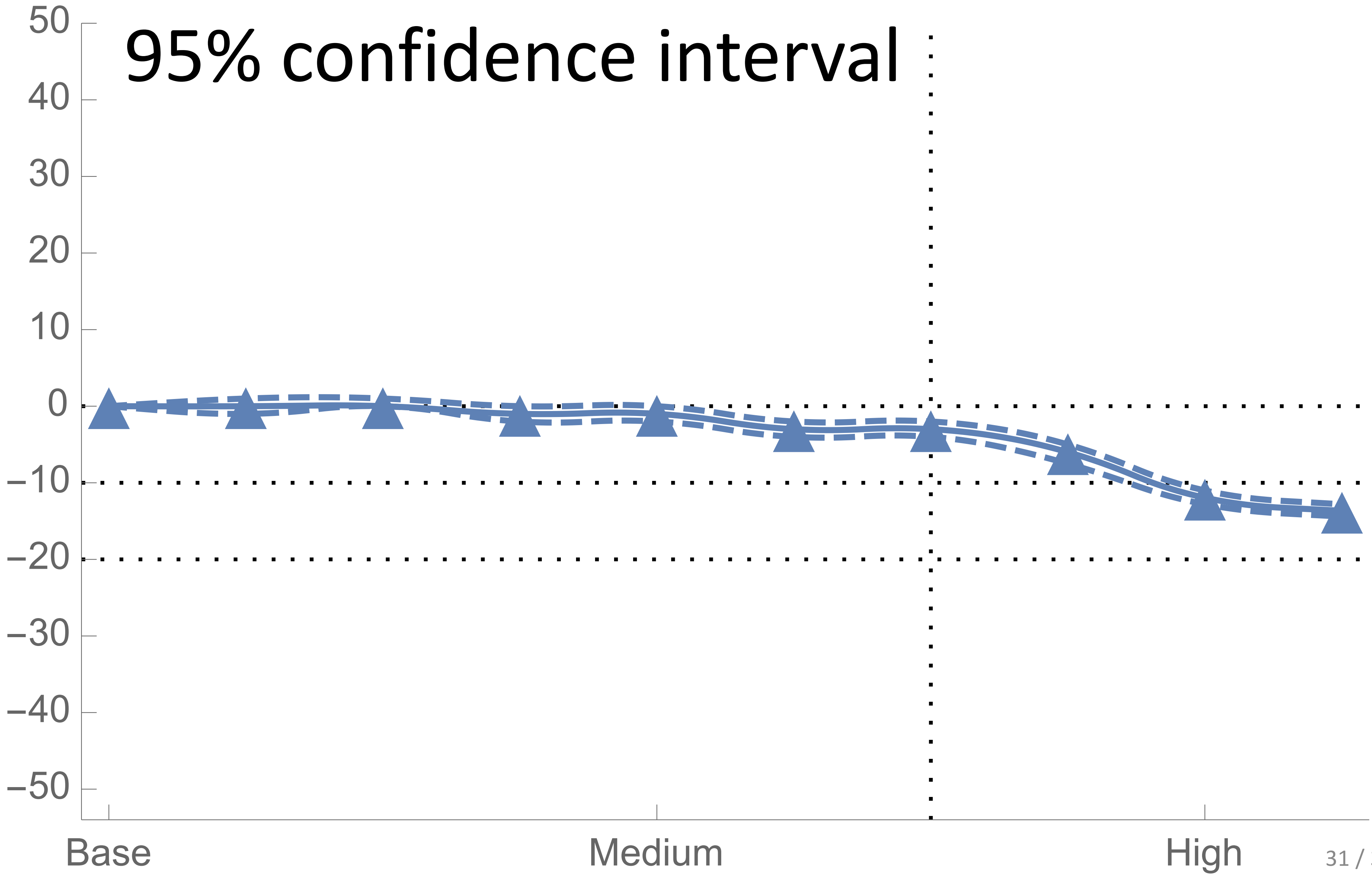
- Whole population

Covariate	Population	Template
Father's education		
Secondary	0.39	0.40
Technical	0.09	0.09
College	0.15	0.14
Missing	0.05	0.04
Mother's education		
Secondary	0.41	0.43
Technical	0.13	0.12
College	0.12	0.12
Missing	0.01	0.01
Household income		
100-200	0.26	0.26
200-400	0.30	0.31
400-600	0.13	0.12
600-1400	0.13	0.14
1400 or more	0.09	0.09
Missing	0.01	0.02

Template + Multiple Doses

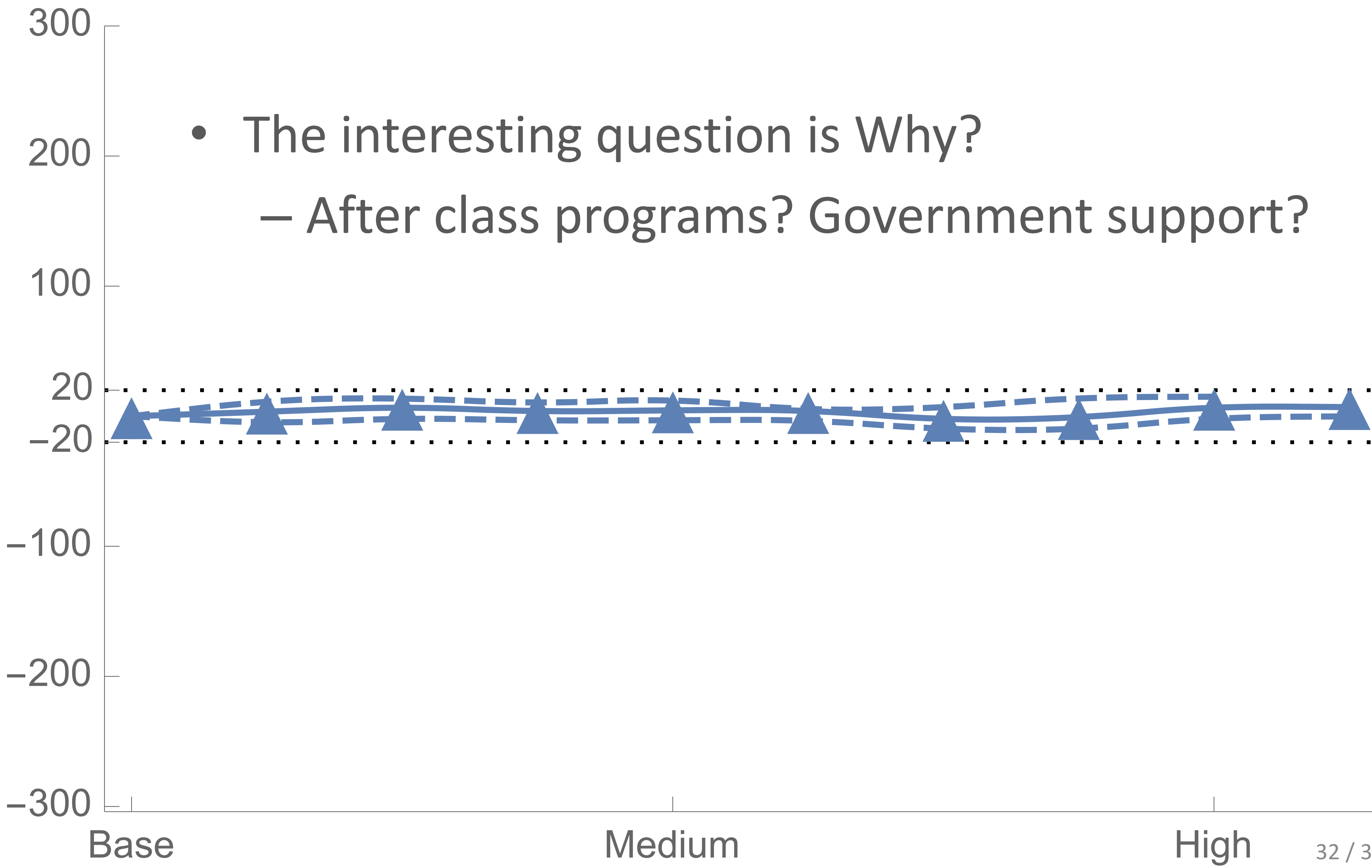


Relative (To no Quake) Attendance Impact [%]



Relative (To no Quake) PSU Score Impact (150—850)

- The interesting question is Why?
  - After class programs? Government support?



Summary

- Advances in MIP
  - Advanced Formulations
  - Advanced Solvers
  - Easy Access Through JuMP
- Direct advantage (Choice Based Conjoint Analysis)
  - Real-time and versatile adaptive questionnaires
  - Cut number of questions in half
  - 20% improvement in estimated parameter quality
  - Market-share predictions cut in half
- Indirect advantage (Causal Inference)
  - Good and flexible formulations can bring you back to solving the problem you really wanted to solve