A Partial-Order Based Model to Estimate Individual Preferences using Panel Data

Srikanth Jagabathula
NYU Stern School of Business, New York

Joint with:

Gustavo Vulcano
NYU Stern School of Business
School of Business, Torquato di Tella University, Argentina

Dmitry Mitrofanov
NYU Stern School of Business

Effective operational decisions require accurate demand predictions

Goal: predict customer's choice from a menu of items

Existing methods in operations

Data
- sales transactions
- product availability

Models
choice models:
accounting for substitution

substitution rates:
15 - 45% in airlines [Ja et al '01]
45% in retail [Gruen at al. '02]

Limitations

ignores
- repeated customer visits
- *considered* items

Our focus:
personalized demand predictions using panel data

Panel data sales transactions tagged by customer id

Challenges

- very few observations per customer
- product unavailability [stock-outs]
- switching [price/quantity promotion]
- latent consideration sets

Our focus:
personalized demand predictions using panel data

Panel data sales transactions tagged by customer id

Challenges

- very few observations per customer
- product unavailability [stock-outs]
- switching [price/quantity promotion]
- latent consideration sets

Our approach

use partial orders to efficiently extract preference info

Our focus:
personalized demand predictions using panel data

Panel data sales transactions tagged by customer id

Challenges

- very few observations per customer
- product unavailability [stock-outs]
- switching [price/quantity promotion]
- latent consideration sets

Our approach

choice model
both utility- and rank-based

Our focus:
personalized demand predictions using panel data

Panel data — sales transactions tagged by customer id

Challenges

- very few observations per customer
- product unavailability [stock-outs]
- switching [price/quantity promotion]
- latent consideration sets

Our approach

behavioral model

Our focus:
personalized demand predictions using panel data

Panel data sales transactions tagged by customer id

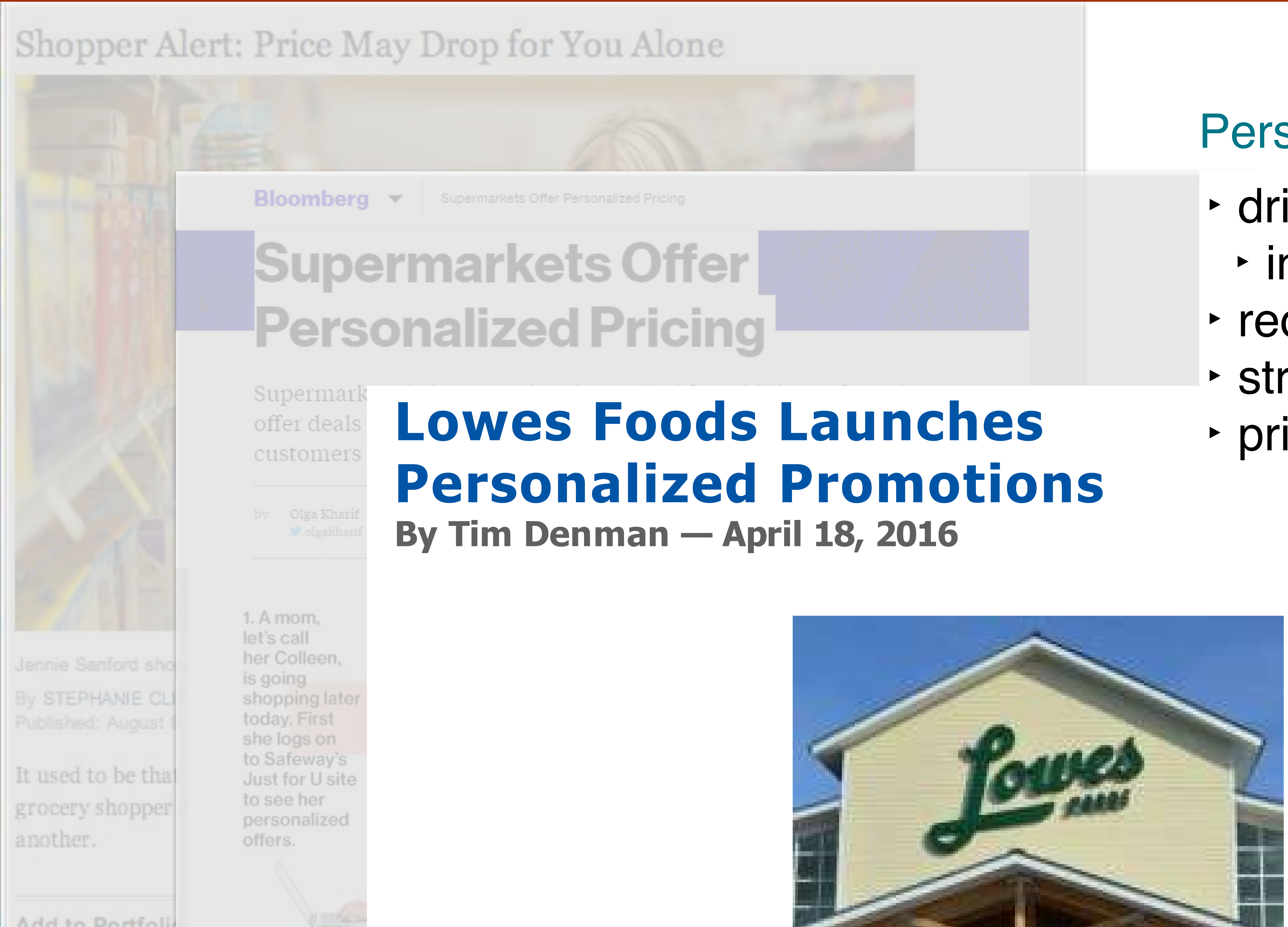
Challenges

- very few observations per customer
- product unavailability [stock-outs]
- switching [price/quantity promotion]
- latent consideration sets

Contributions

- methodology for accurate predictions
- performance gains on IRI dataset

Application to personalized promotions



Personalized promotions:

- drive up sales
  - increase both visits & basket size
- reduce competition
- stronger cust. relationship
- price discrimination

Related work: Rich work spanning various areas

Operations:

Estimation: [Farias, Jagabathula, Shah ’12], [Haensel, Koole ’11], [van Ryzin, Vulcano ’13], [Jagabathula, Rusmevichientong ’13], …
Optimization: [Mahajan, van Ryzin ’01], [Farias, Jagabathula, Shah ’13], [Papers by Rusmevichientong, Topaloglu, and Gallego], …
Airline RM: [Zhang, Cooper ’06], [Chaneton, Vulcano ’11], [Kunnumkal ’12], …

Focus is on
- decision-making
- estimation from aggregated transactions

our focus: panel data & individual predictions

Marketing:

[Allenby, Lenk ’95], [Chib, Seetharaman, Strijnev ’04], [Chintagunta ’92], [Erdem ’93], [Erdem, Imai, Keane ’03], [Guadagni, Little ’83], [Hendel, Nevo ’06], [Jedidi, Mela, Gupta ’99], [Lattin ’87], …

how covariates affect choice

we complement by enriching inferences through partial orders

Machine Learning/Statistics:

[Lu, Boutilier ’11], [Lebanon, Mao ’08], [Guiver, Snelson ’09], [Jagabathula, Shah ’08], [Jagabathula, Shah ’11], [Meila, Chen ’10], [Meila, et.al. ’07], [Kondor, Barbosa ’10], [Huang, Guestrin ’10], …

methodological connections

we borrow and contribute

A Partial-Order Based Model to Estimate Individual Preferences Using Panel Data

1. Model
   assumptions, data for estimation
2. Modular inference framework
   consideration set, clustering of DAGs, ML estimation
3. Numerical results
   results on real-world panel data from IRI dataset
4. Summary/Conclusions
   takeaway messages

A Partial-Order Based Model to Estimate Individual Preferences Using Panel Data

1. Model
   assumptions, data for estimation

2. Modular inference framework
   consideration set, clustering of DAGs, ML estimation

3. Numerical results
   results on real-world panel data from IRI dataset

4. Summary/Conclusions
   takeaway messages

Panel data: sales transactions tagged with customer ids

n products — mutually substitutable; 0 is no-purchase

m customers — make repeated purchases

Observations: for each customer and time period

offer set — purchased product

	offer set S_t = filled dots	purchase i_t	
t = 1	1 2 3 4 5 6 7	3	$S_1 = \{1,2,3,4,6,7\}$
t = 2	1 2 3 4 5 6 7	3	$S_2 = \{1,2,3,6,7\}$
t = 3	1 2 3 4 5 6 7	2	$S_3 = \{1,2,6,7\}$

IRI Academic Data Set

Weekly store sales and consumer panel data

31
product categories

11 years
2001 - 2011

We analyzed year 2007 panel data

29
product categories

84K
user-category combinations

1.2M
transactions across 52 weeks

Training data summary

- aggregate items by vendors
- training: first 26 weeks
- retain customers ≥ 2 sales

Category name	# Vendors	# Customers retained
Beer	67	1154
Blades	9	243
Carbonated Beverages	46	4387
Cigarettes	13	307
Coffee	59	2255
Cold Cereal	39	3998
Deodorant	32	653
Diapers	4	173
Facial Tissue	10	2063
Frozen Dinners/Entrees	77	3288
Frozen Pizza	38	2946
Household Cleaners	68	1699
Hot dogs	41	2187
Laundry Detergent	18	2181
Margarine/Butter	16	2750
Mayonnaise	14	2386
Milk	33	4652
Mustard	52	2515
Paper Towels	11	2051
Peanut Butter	19	1923
Salt Snacks	95	4446
Shampoo	41	738
Soup	90	4322
Spaghetti/Italian Sauce	52	2698
Sugar Substitutes	10	308
Toilet Tissue	11	2817
Toothbrushes	36	499
Toothpaste	25	1186
Yogurt	26	3491

Beer	67	1154
Blades	9	243
rbonated Beverages	46	4387
Cigarettes	13	307
Coffee	59	2255
Cold Cereal	39	3998
Deodorant	32	653
Diapers	4	173
Facial Tissue	10	2063
ozen Dinners/Entrees	77	3288
Frozen Pizza	38	2946
lousehold Cleaners	68	1699
Hot dogs	41	2187
Laundry Detergent	18	2181
Margarine/Butter	16	2750
Mayonnaise	14	2386

Training data summary

- aggregate items by vendors
- training: first 26 weeks
- retain customers $\geq$ 2 sales

After pre-processing

64K
user-category combinations

1.1M
transactions across 52 weeks

Average across 29 categories

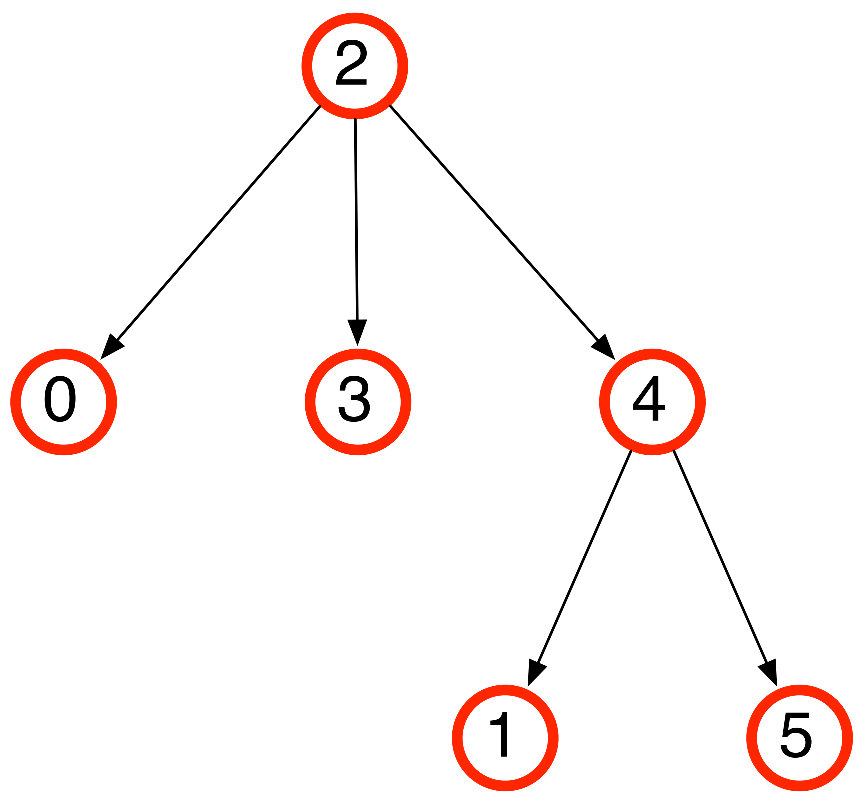
36
vendors

2.2K
customers

Choice model: preferences consistent across purchase instances

Customers characterized by
- partial-order (DAG)
- distribution over rankings

- samples rank list consistent with DAG
- purchases most preferred offered item



Directed Acyclic Graph (DAG)

2 → 3
2 → 0
2 → 4
4 → 1
4 → 5
2 → 1
2 → 5

bag of comparisons

Offer set: {1, 2, 4}

Purchase: 2

2 chosen *every* instance

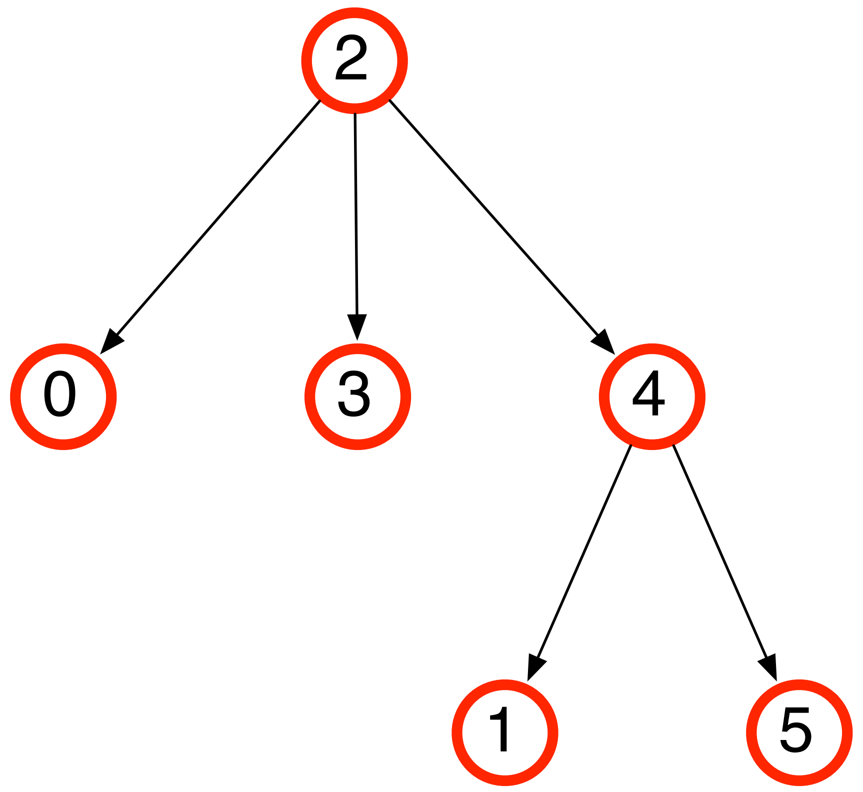
2	4
↓	↓
3	3
↓	↓
4	2
↓	↓
1	1
↓	↓
0	0
↓	↓
5	5

Choice model: preferences consistent across purchase instances

Customers characterized by
- partial-order (DAG)
- distribution over rankings

- samples rank list consistent with DAG
- purchases most preferred offered item

Subsumes existing methods with empty partial order



$2 \rightarrow 3$
$2 \rightarrow 0$
$2 \rightarrow 4$
$4 \rightarrow 1$
$4 \rightarrow 5$
$2 \rightarrow 1$
$2 \rightarrow 5$

Offer set: {1, 3, 4}

Purchase: 3 or 4

choice depends on sampled rank list

$2 \rightarrow 3 \rightarrow 4 \rightarrow 1 \rightarrow 0 \rightarrow 5$

$2 \rightarrow 4 \rightarrow 3 \rightarrow 1 \rightarrow 0 \rightarrow 5$

Model benefits: efficiently extracts preference information from data

Infer preference relations with fewer samples

RUM model		Strong preferences

Rank lists differ across periods	Rank lists consistent with partial order	Rank list same across periods
prefers 1 or 2 each period		prefers 1 over 2 every period
many samples needed to estimate	fewer samples needed to estimate	only few samples needed to estimate

Model benefits: efficiently extracts preference information from data

Infer preference relations with fewer samples

Heterogeneity through DAGs yields parsimonious models

Two customer types:

type 1	type 2
empty DAG	$2 \rightarrow 3 \rightarrow 4$
random choice	random choice from 1, 2, and 5

classical	need two MNL
DAGs	one MNL and two DAGs

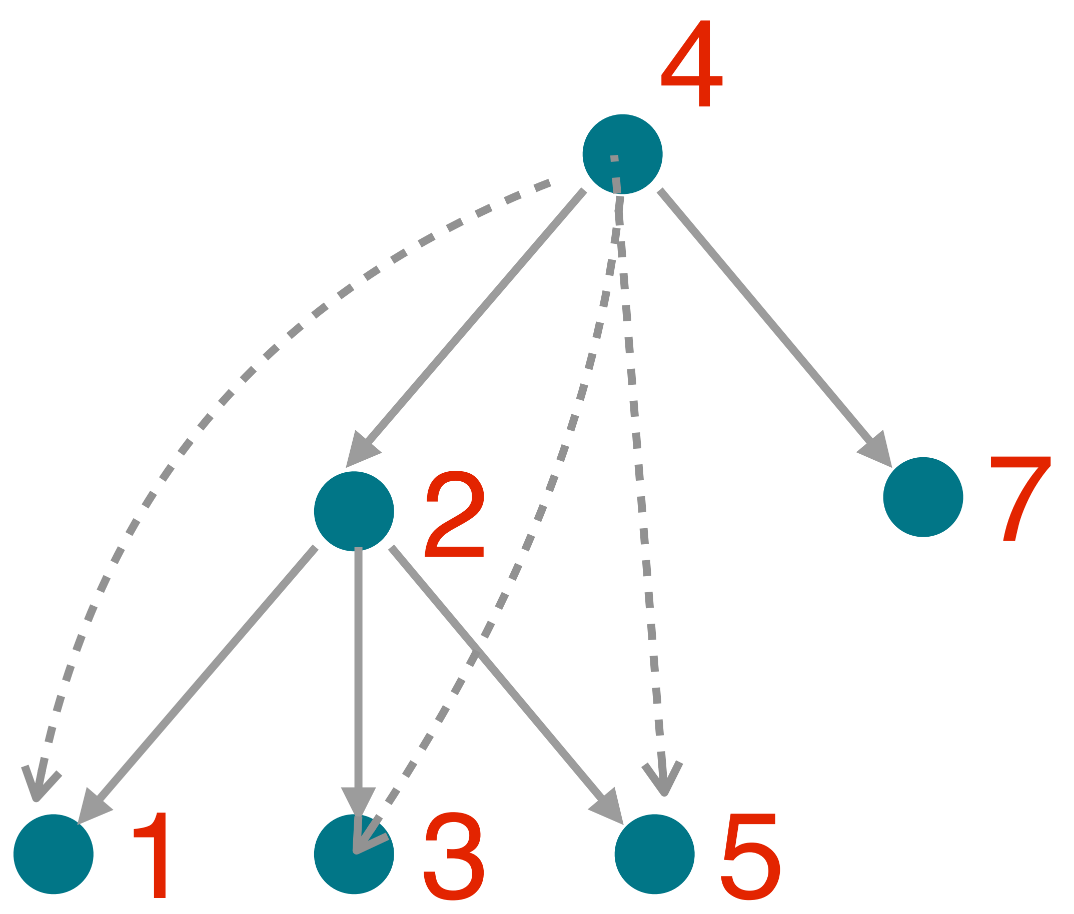
Model benefits: efficiently extracts preference information from data

Infer preference relations with fewer samples

Heterogeneity through DAGs yields parsimonious models

Enrich revealed comparisons with inferred comparisons

	purchase	offer set
$t = 1$	4	{2, 4, 7}
$t = 2$	2	{1, 2, 3, 5}

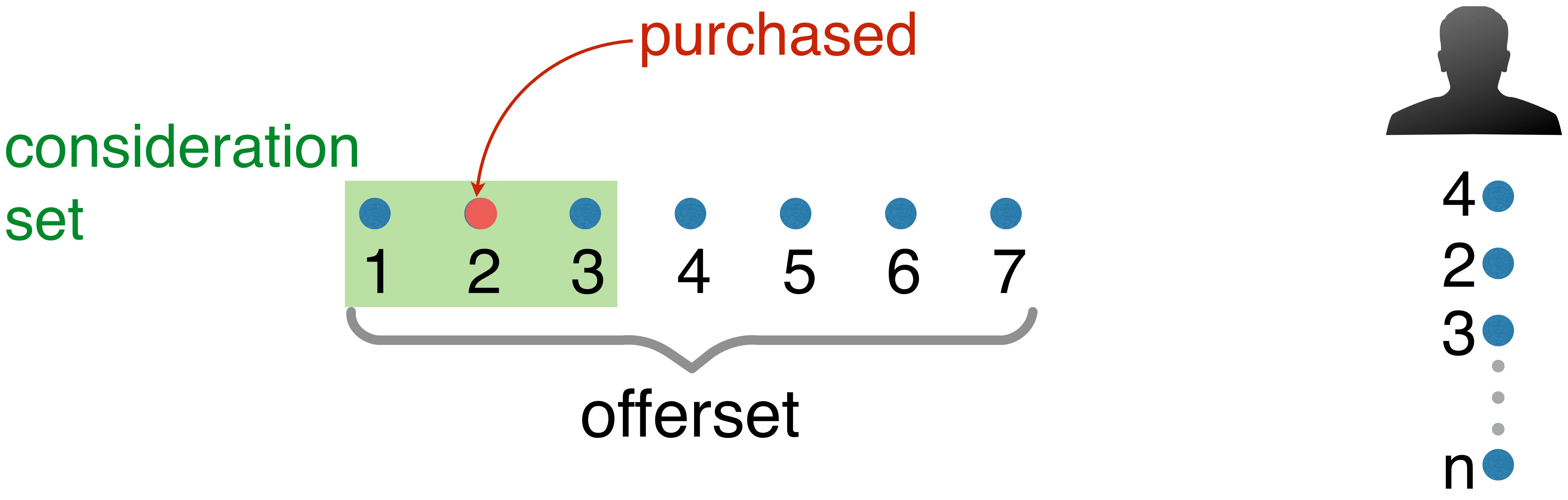


transitivity allows inferring edges not "directly revealed"

enriched preference data ➡ more accurate predictions

Consideration set:
customers may not consider all the products on offer

Customers may choose from smaller consideration set



Avoids inferring spurious preference relations

Model: Customers sample rank lists from a distribution

Market with K segments	type k samples DAG from distribution f_k uses f_k to sample rank lists consistent with DAG
Sales over T periods	product universe (almost) constant customer population (almost) constant
Objective	predict purchase of each customer

A Partial-Order Based Model to Estimate Individual Preferences Using Panel Data

1. Model
   assumptions, data for estimation

2. Modular inference framework
   consideration set, clustering of DAGs, ML estimation

3. Numerical results
   results on real-world panel data from IRI dataset

4. Summary/Conclusions
   takeaway messages

Three-step modular framework for inference

			technique
Step 1	Construct customer DAGs		model for consideration set
		DAGs	
Step 2	Cluster the DAGs		clustering technique
		DAG clusters	
Step 3	Fit choice models to clusters		Approx ML estimation
		choice models	

Steps do not depend on assumptions of previous steps

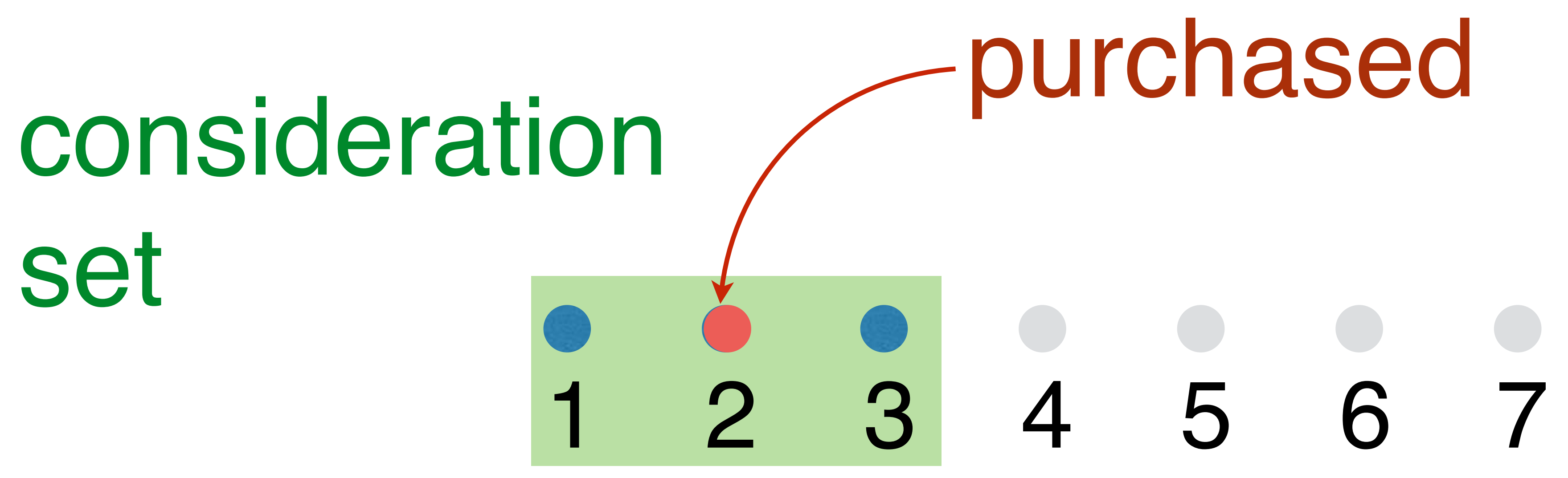
Three-step modular framework for inference

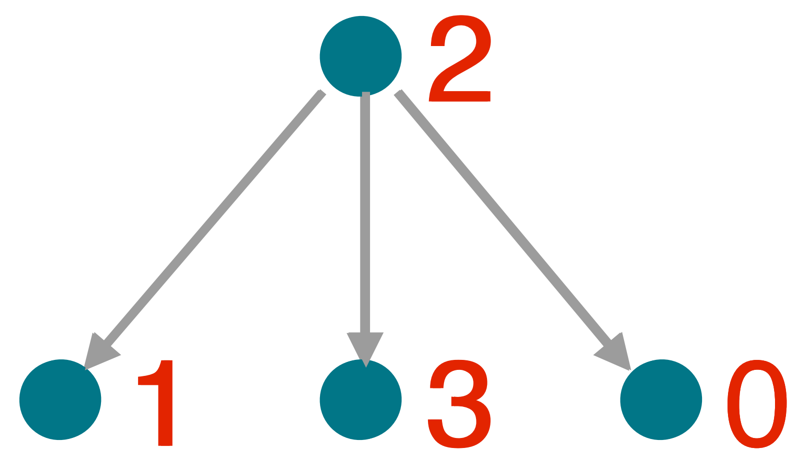
		technique
Step 1	Construct customer DAGs	model for consideration set
	DAGs	
Step 2	Cluster the DAGs	clustering technique
	DAG clusters	
Step 3	Fit choice models to clusters	Approx ML estimation
	choice models	

Steps do not depend on assumptions of previous steps

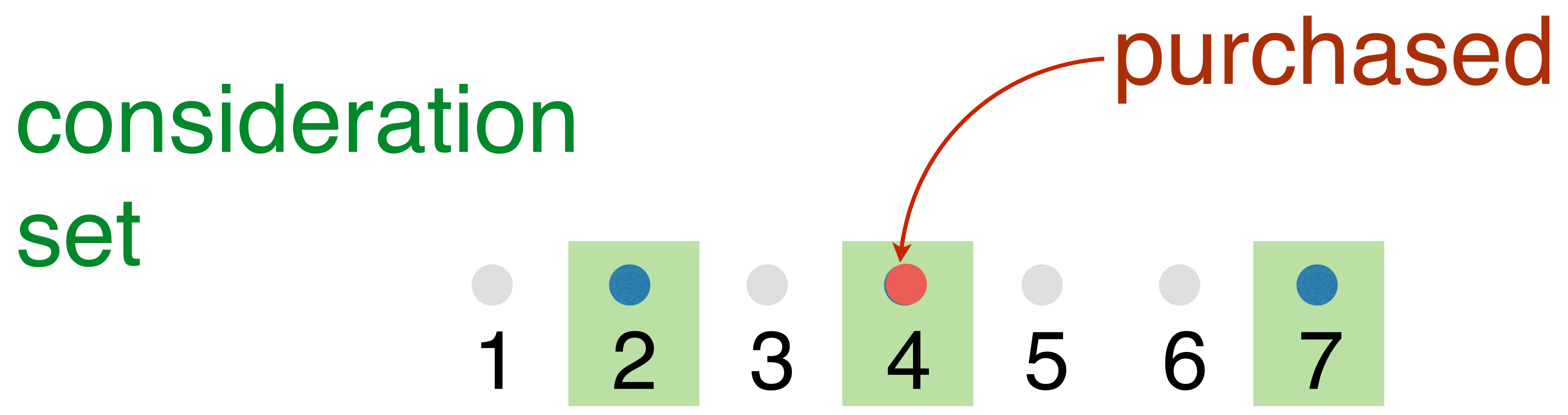
Standard assumption: consider everything on offer

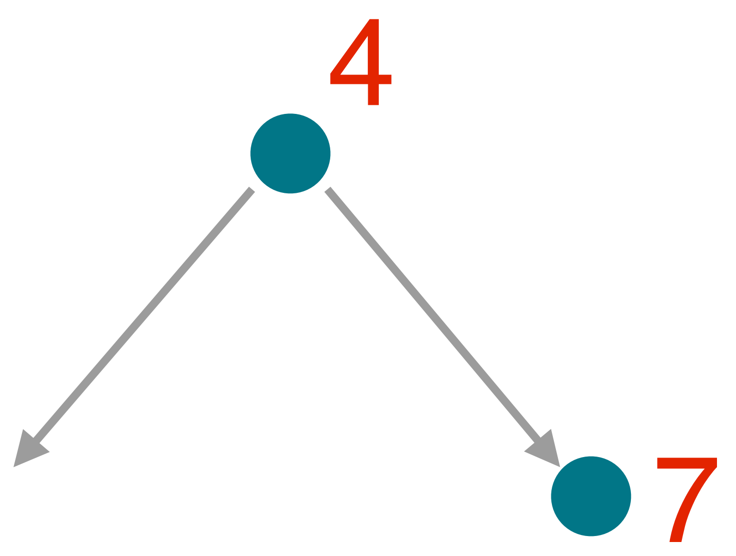
t = 1; offer set = {1, 2, 3}





t = 2; offer set = {2, 4, 7}





Inertial assumption: effect of stock-outs, promotions, and brand inertia captured

Standard assumption	potentially unrealistic for frequently purchased products
Behavioral principle	customers do not consider unless they have to captures brand choice inertia (short-term loyalty) [Jeuland ’79]

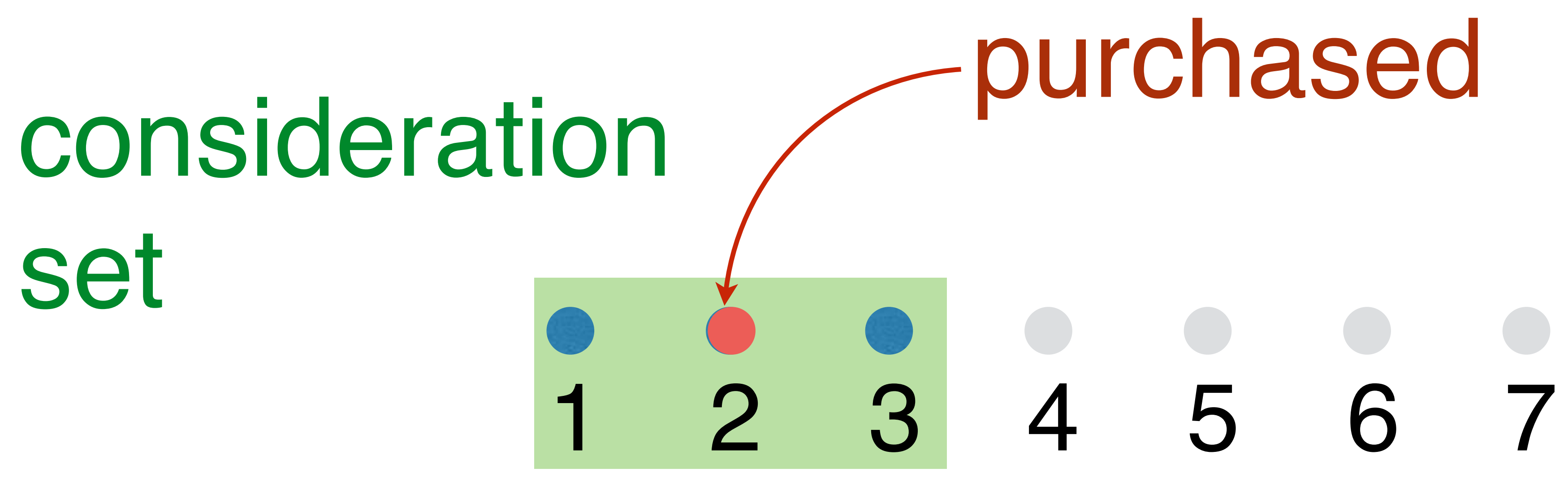
Behavioral rules

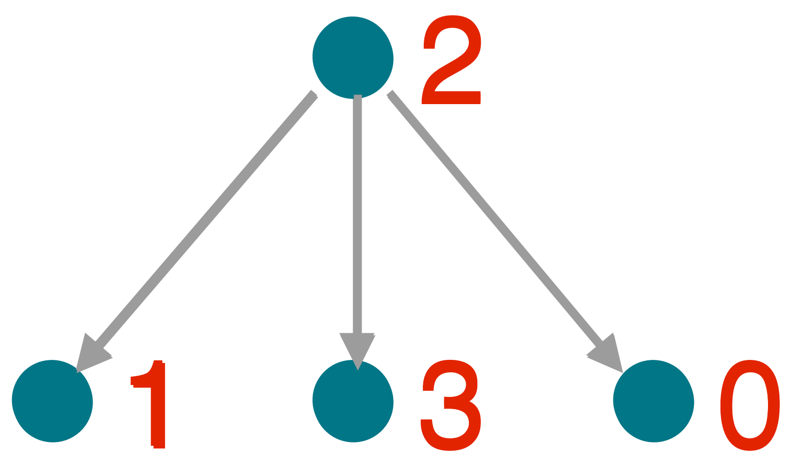
to infer preferences among full-priced items

prev purchase	consideration set
stocked-out	offer set
in-stock	prev purchase + promoted items

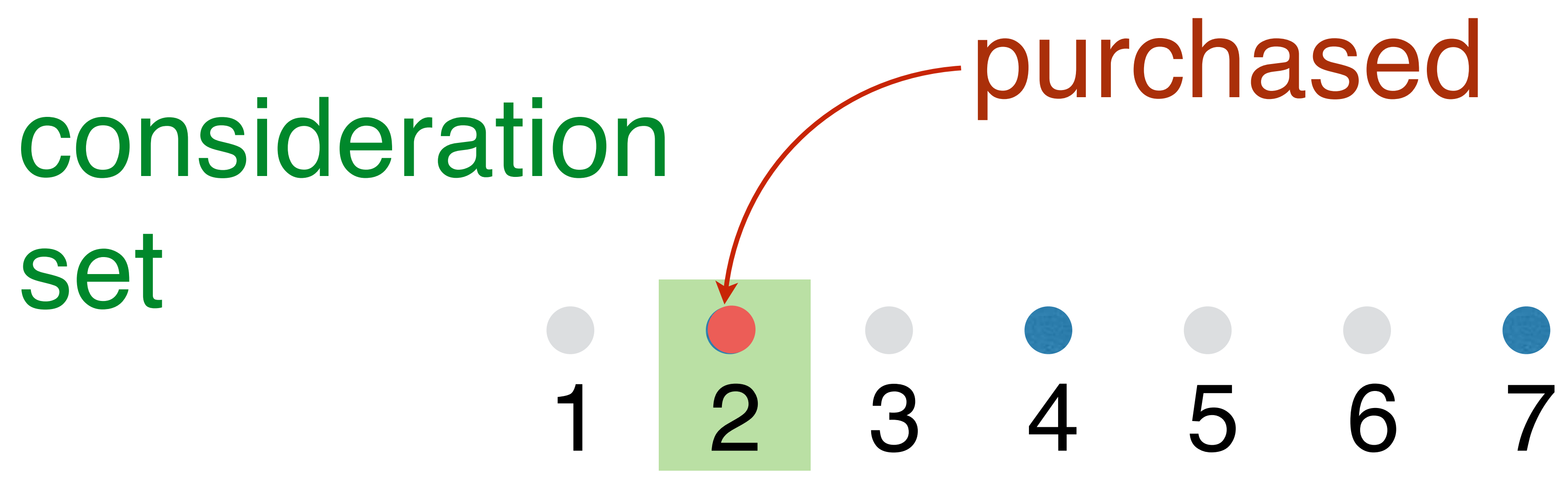
Inertial assumption: example

t = 1; offer set = {1, 2, 3}



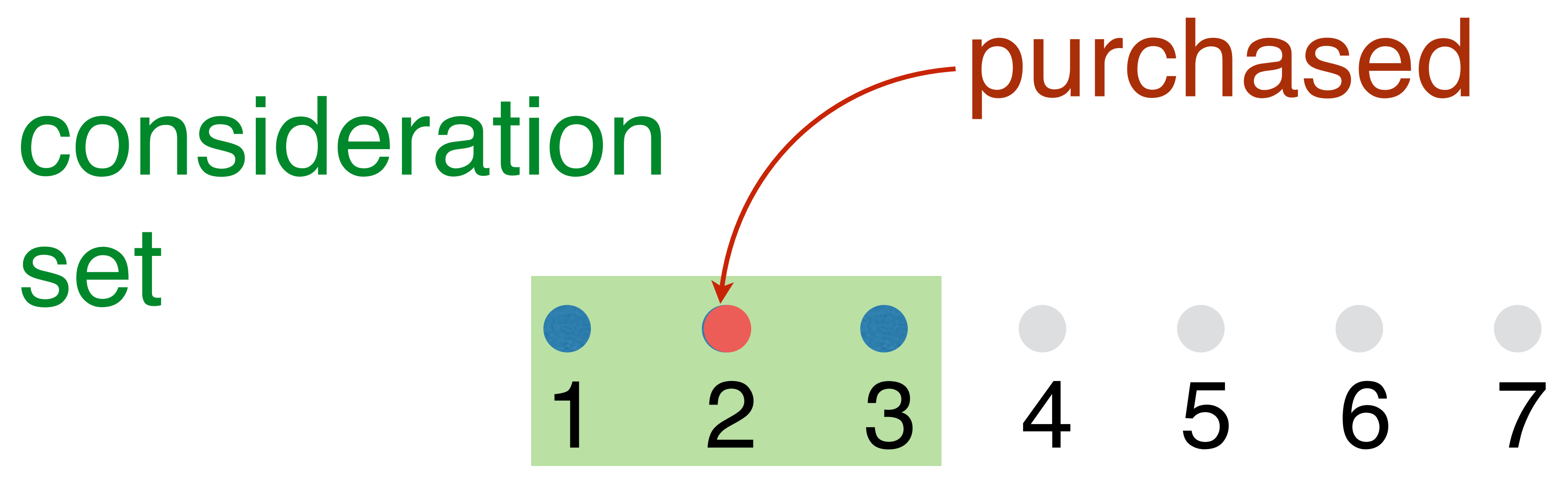


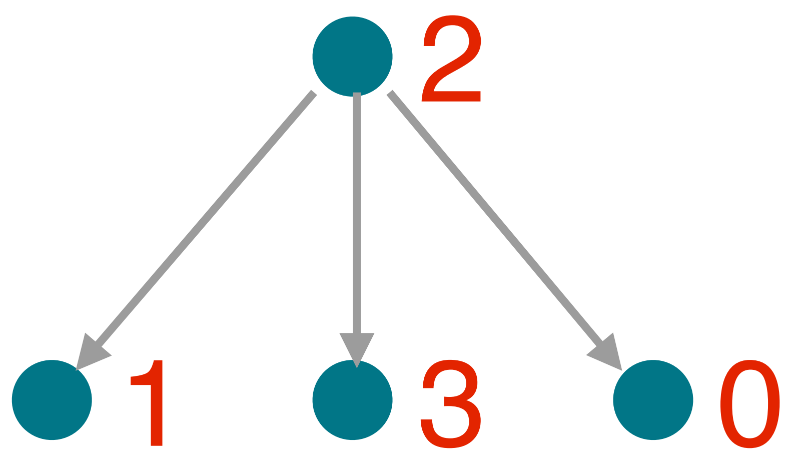
t = 2; offer set = {2, 4, 7}



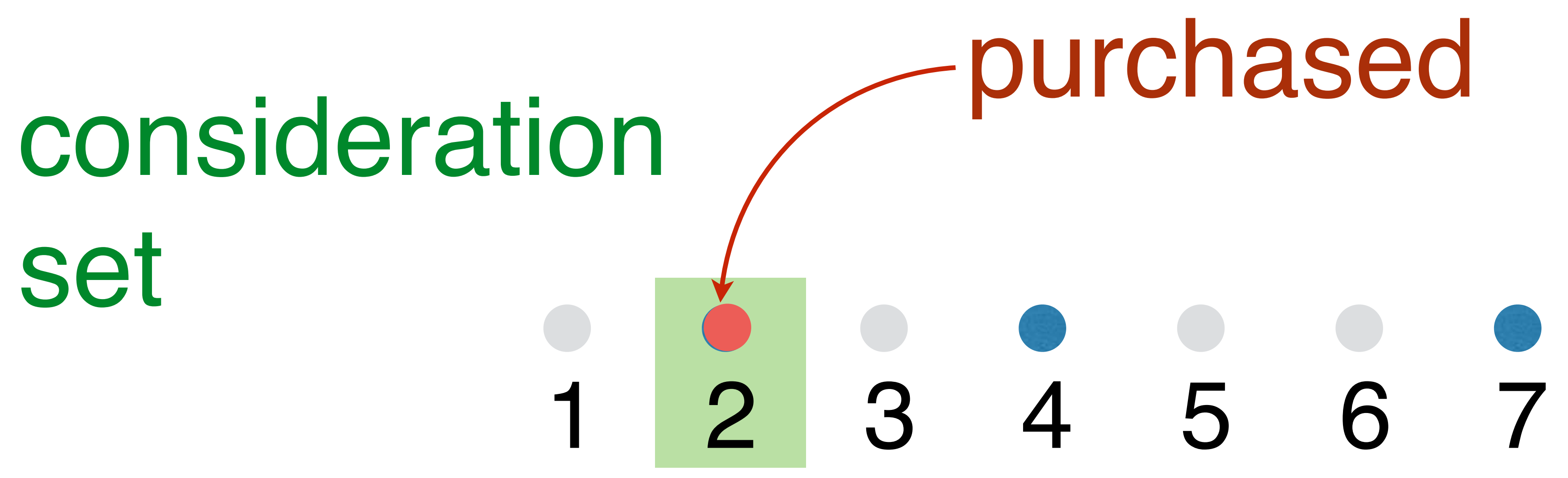
Inertial assumption: example (contd.)

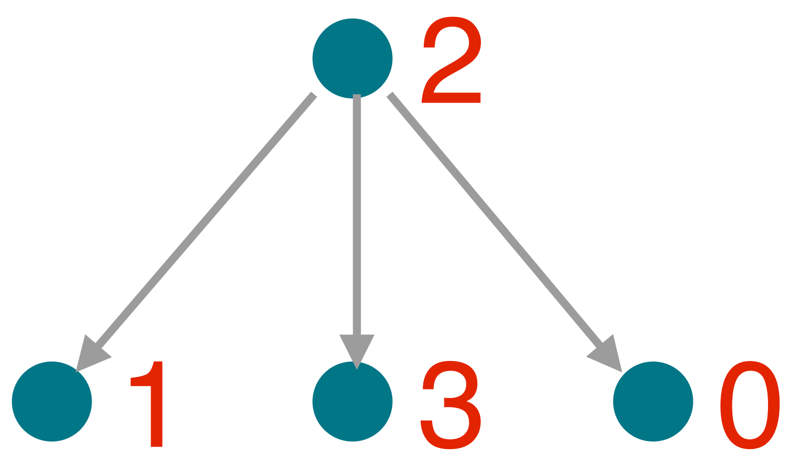
t = 1; offer set = {1, 2, 3}





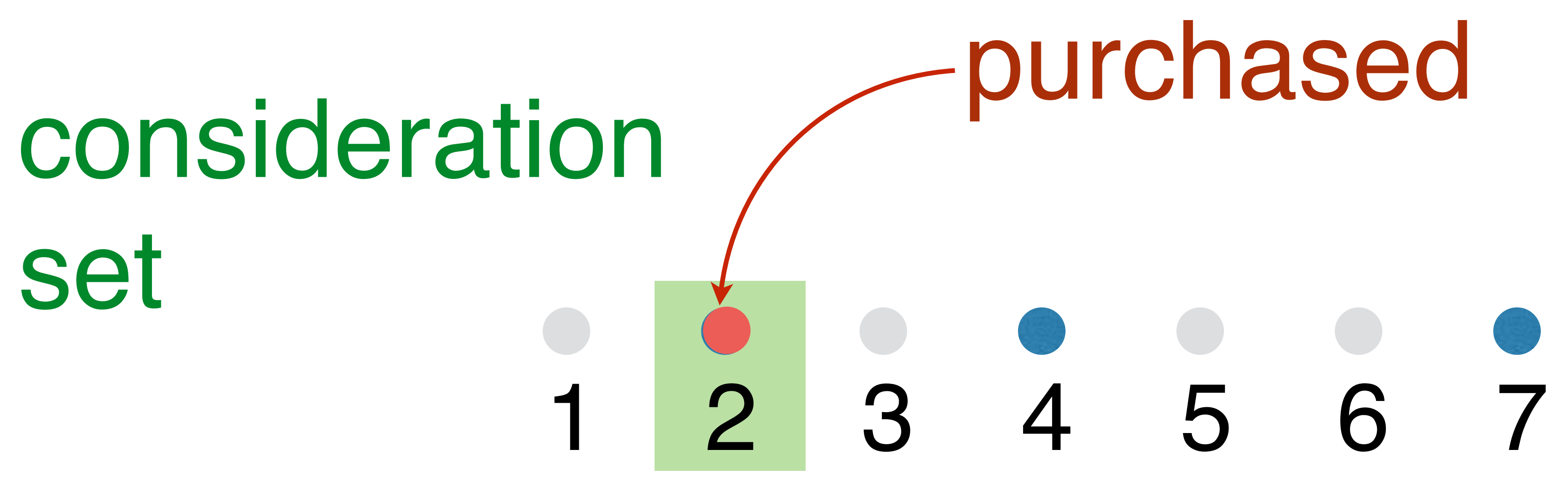
t = 2; offer set = {2, 4, 7}

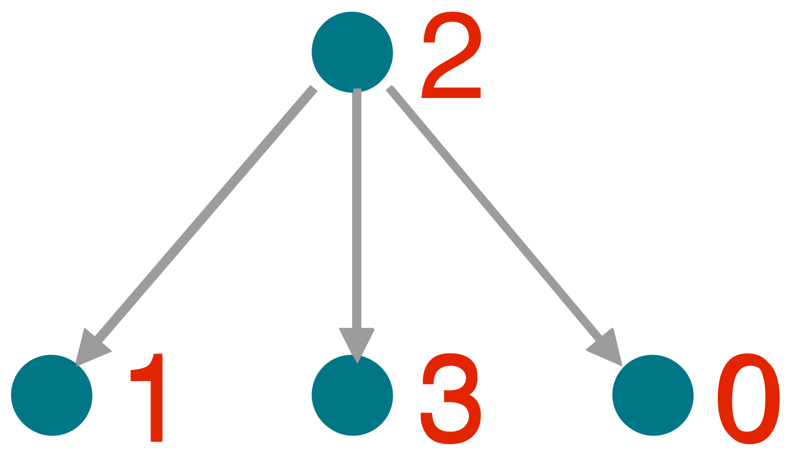




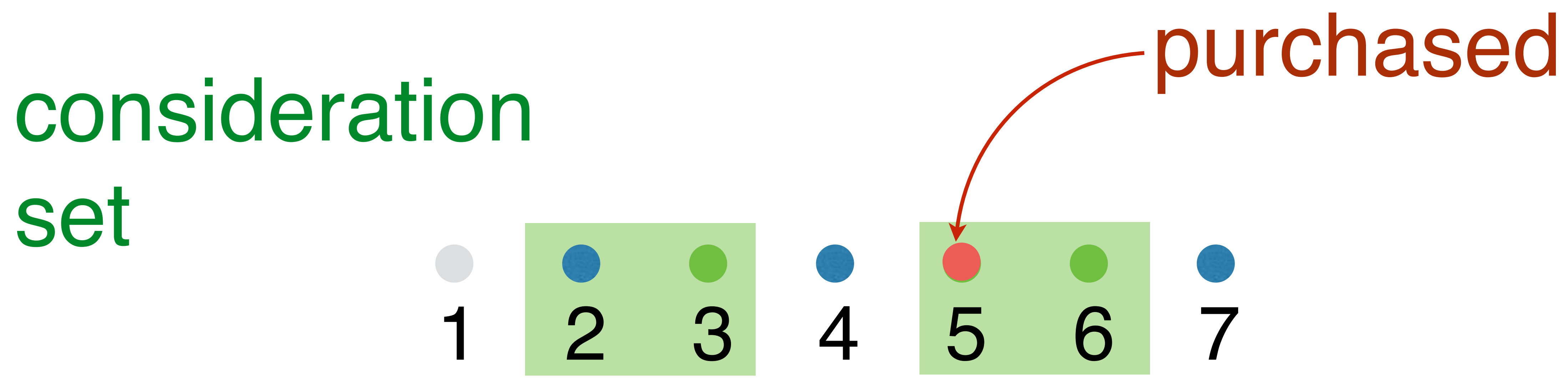
Inertial assumption: example (contd.)

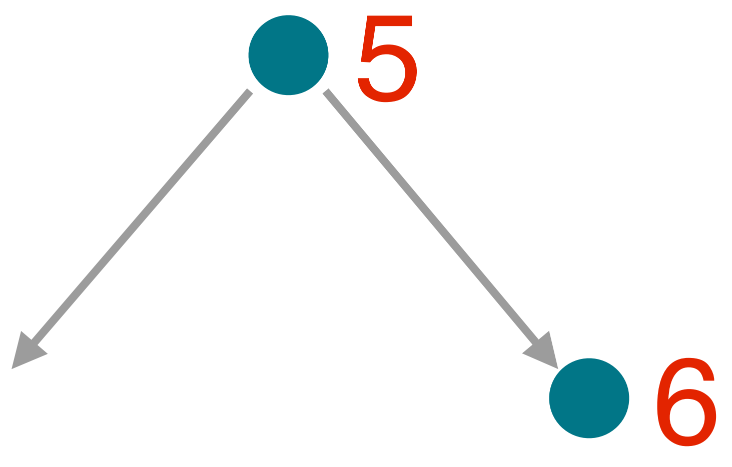
t = 2; offer set = {2, 4, 7}





t = 3; offer set = {2, 3, 4, 5, 6}; promoted items = {3, 5, 6}





Censored assumption: allow deviations from the inertial assumption

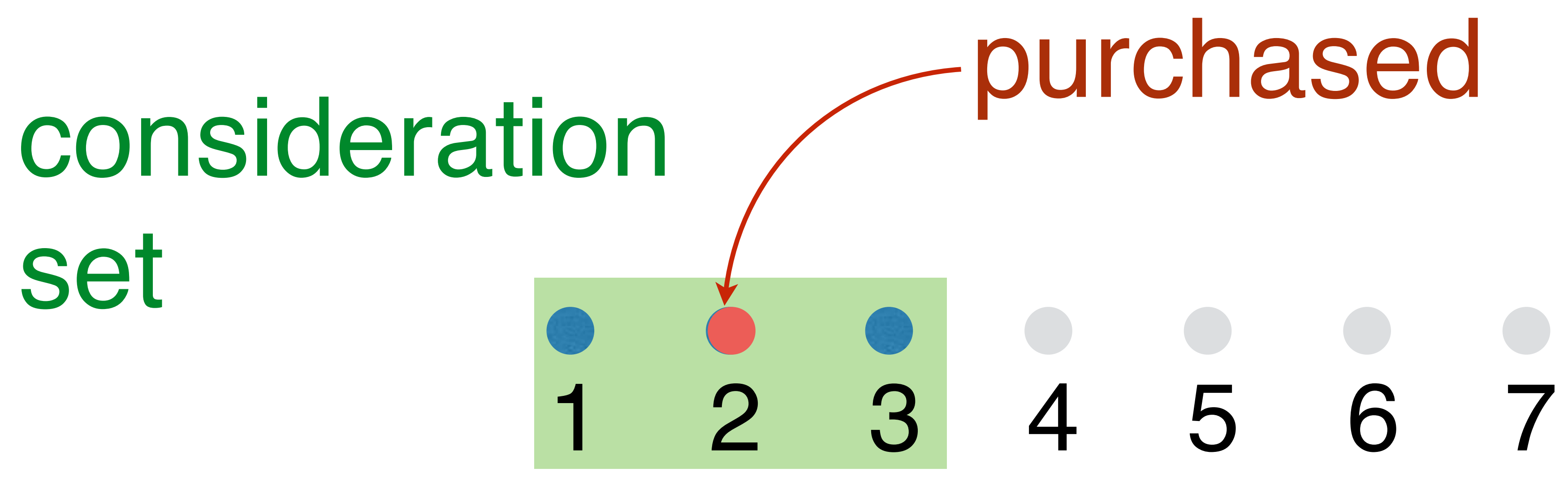
purchase unexplained by inertial assumption → consideration set unobserved/censored

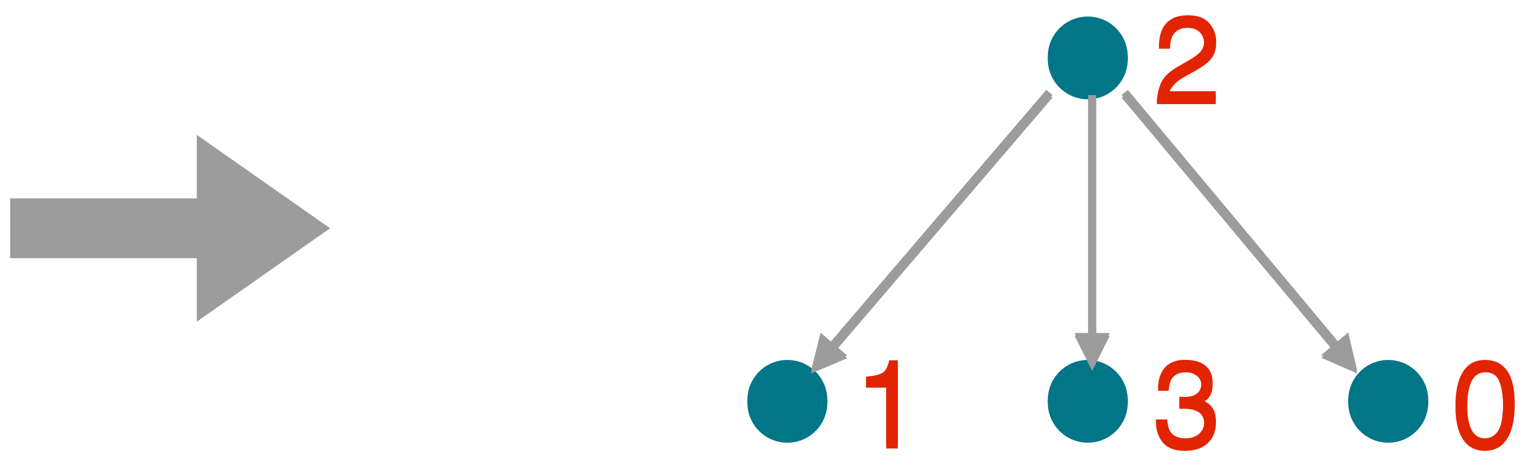
RULE

update "prev purchase" (sticky product); don't add edges to DAG

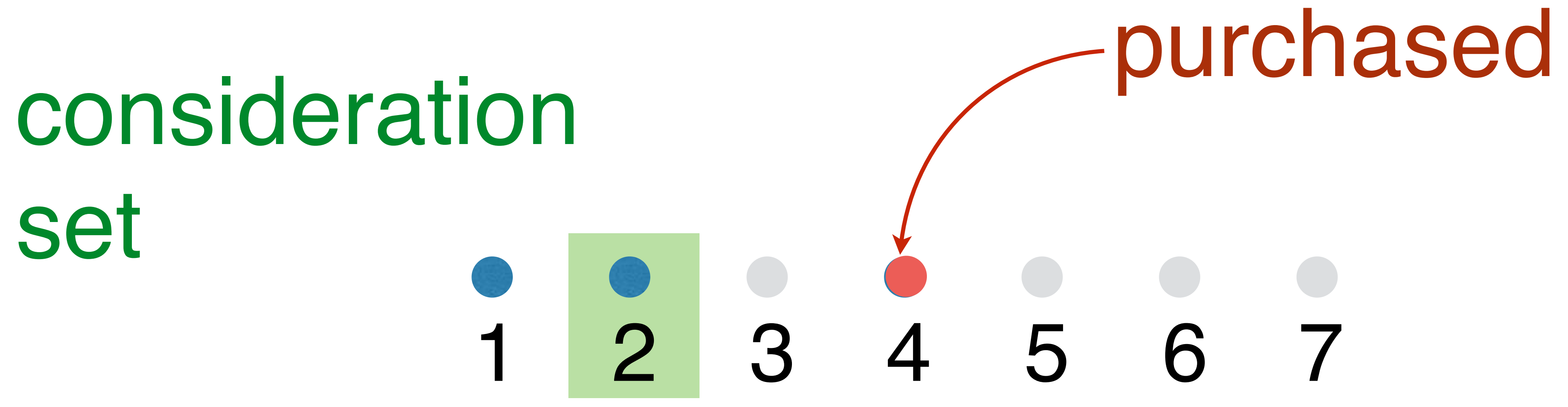
Censored assumption: example

t = 1; offer set = {1, 2, 3}





t = 2; offer set = {1, 2, 4}



DAG does not change

Category name	# Vendors	# Customers retained	Std.	Inert.	Cens.
Beer	67	1154	36%	41%	85%
Blades	9	243	50%	55%	91%
Carbonated Beverages	46	4387	14%	15%	39%
Cigarettes	13	307	75%	76%	99%
Coffee	59	2255	36%	47%	84%
Cold Cereal	39	3998	13%	17%	49%
Deodorant	32	653	42%	49%	89%
Diapers	4	173	51%	54%	88%
Facial Tissue	10	2063	44%	49%	89%
Frozen Dinners/Entrees	77	3288	21%	28%	55%
Frozen Pizza	38	2946	29%	40%	63%
Household Cleaners	68	1699	15%	20%	90%
Hot dogs	41	2187	37%	51%	92%
Laundry Detergent	18	2181	46%	61%	89%
Margarine/Butter	16	2750	43%	50%	90%
Mayonnaise	14	2386	70%	72%	96%
Milk	33	4652	33%	36%	72%
Mustard	52	2515	19%	35%	89%
Paper Towels	11	2051	33%	49%	87%
Peanut Butter	19	1923	54%	61%	95%
Salt Snacks	95	4446	12%	14%	42%
Shampoo	41	738	37%	46%	83%
Soup	90	4322	17%	22%	55%
Spaghetti/Italian Sauce	52	2698	38%	50%	81%
Sugar Substitutes	10	308	80%	81%	98%
Toilet Tissue	11	2817	41%	55%	87%
Toothbrushes	36	499	38%	48%	92%
Toothpaste	25	1186	51%	58%	89%
Yogurt	26	3491	33%	39%	55%

Avg. number of customers with non-empty DAGs

Beer	67	1154	36%	41%	85%
Blades	9	243	50%	55%	91%
bonated Beverages	46	4387	14%	15%	39%
Cigarettes	**13**	**307**	**75%**	**76%**	**99%**
Coffee	**59**	**2255**	**36%**	**47%**	**84%**
Cold Cereal	39	3998	13%	17%	49%
Deodorant	32	653	42%	49%	89%
Diapers	4	173	51%	54%	88%
Facial Tissue	10	2063	44%	49%	89%
en Dinners/Entrees	77	3288	21%	28%	55%
Frozen Pizza	38	2946	29%	40%	63%
usehold Cleaners	68	1699	15%	20%	90%
Hot dogs	41	2187	37%	51%	92%
undry Detergent	18	2181	46%	61%	89%
argarine/Butter	16	2750	43%	50%	90%
Mayonnaise	14	2386	70%	72%	96%
Milk	33	4652	33%	36%	72%
Mustard	52	2515	19%	35%	89%
Paper Towels	11	2051	33%	49%	87%

Avg. number of customers with non-empty DAGs

31%

standard

39%

inertial

71%

censored

Category name	# Vendors	# Customers retained	Std.	Inert.	Cens.
Beer	67	1154	52	25	43
Blades	9	243	5	3	4
Carbonated Beverages	46	4387	20	8	16
Cigarettes	13	307	8	7	8
Coffee	59	2255	26	16	22
Cold Cereal	39	3998	22	12	19
Deodorant	32	653	18	11	15
Diapers	4	173	4	3	3
Facial Tissue	10	2063	5	4	5
Frozen Dinners/Entrees	77	3288	41	23	40
Frozen Pizza	38	2946	18	12	18
Household Cleaners	68	1699	43	28	35
Hot dogs	41	2187	21	13	18
Laundry Detergent	18	2181	13	7	12
Margarine/Butter	16	2750	12	9	11
Mayonnaise	14	2386	8	4	7
Milk	33	4652	18	10	15
Mustard	52	2515	22	13	19
Paper Towels	11	2051	8	6	8
Peanut Butter	19	1923	10	6	9
Salt Snacks	95	4446	38	21	33
Shampoo	41	738	25	15	21
Soup	90	4322	42	18	37
Spaghetti/Italian Sauce	52	2698	23	13	20
Sugar Substitutes	10	308	6	5	5
Toilet Tissue	11	2817	9	6	8
Toothbrushes	36	499	19	13	17
Toothpaste	25	1186	15	7	13
Yogurt	26	3491	11	7	11

Avg. number of edges in the DAGs

Blades	9	243	5	3	4
onated Beverages	46	4387	20	8	16
Cigarettes	13	307	8	7	8
Coffee	59	2255	26	16	22
Cold Cereal	39	3998	22	12	19
Deodorant	32	653	18	11	15
Diapers	4	173	4	3	3
Facial Tissue	10	2063	5	4	5
en Dinners/Entrees	77	3288	41	23	40
Frozen Pizza	38	2946	18	12	18
usehold Cleaners	68	1699	43	28	35
Hot dogs	41	2187	21	13	18
undry Detergent	18	2181	13	7	12
argarine/Butter	16	2750	12	9	11
Mayonnaise	14	2386	8	4	7
Milk	33	4652	18	10	15
Mustard	52	2515	22	13	19

Avg. number of edges in the DAGs

19

standard

11

inertial

17

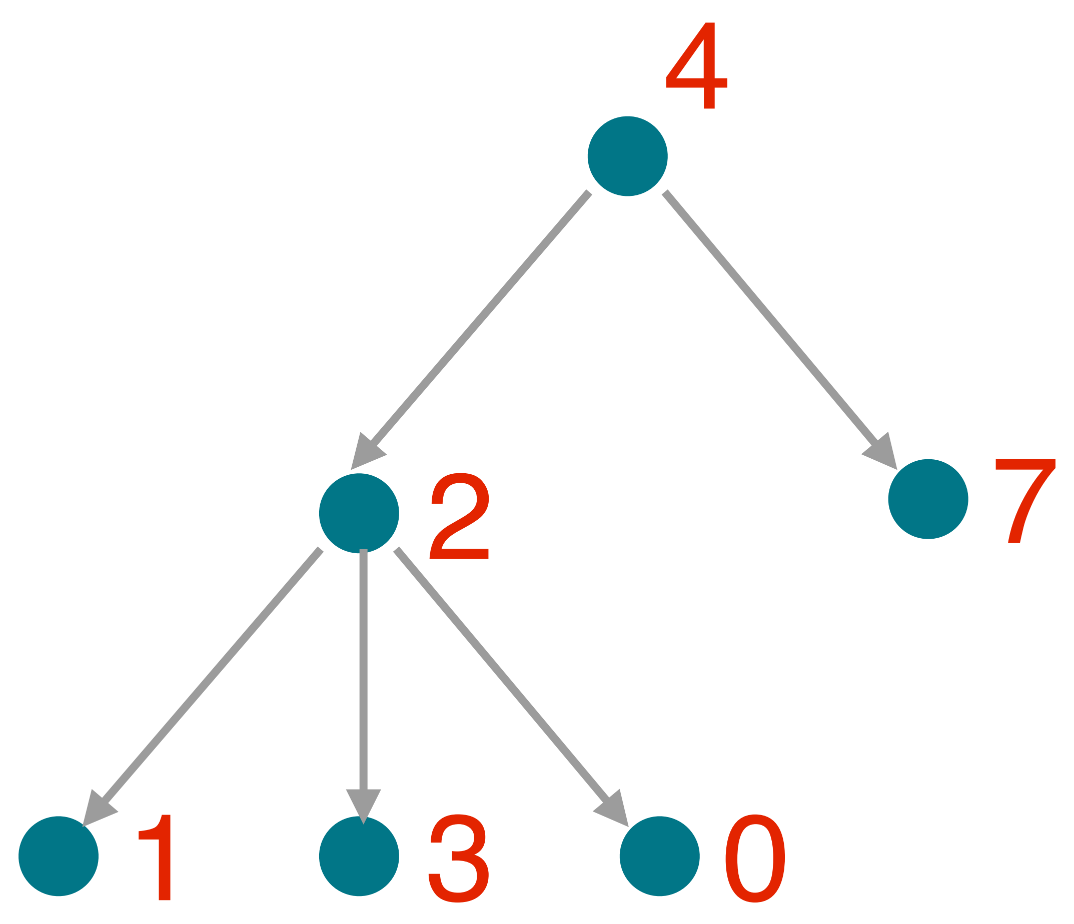
censored

Three-step modular framework for inference

		technique
Step 1	Construct customer DAGs	model for consideration set
	↓ DAGs	
Step 2	Cluster the DAGs	clustering technique
	↓ DAG clusters	
Step 3	Fit choice models to clusters	Approx ML estimation
	↓ choice models	

Steps do not depend on assumptions of previous steps

Clustering of user DAGs: partial orders in same cluster are "close" to central ranking



Each customer represented by a DAG

Clustering assumptions

- Each customer mapped to a single cluster
- Every cluster represented by a central ranking
- Customer DAGs in same cluster "close" to central ranking

Distance between DAG of customer central ranking

(defined over resp. transitive closures)

$$d(c, \sigma) = \text{(\# disagreements)} - \text{(\# agreements)}$$
$$= 2\,\text{(\# disagreements)} - \text{(\# edges in DAG of } c\text{)}$$

MIP to cluster DAGs and find centroid ranking of each cluster

indicator that edge *(i,j)* in *c* is a disagreement with centroid in an allocation

$$\min \sum_{c} \sum_{(i,j)\in c} (2\boxed{w_{ijc}} - m_c)$$

$$\text{s.t.} \sum_{k=1}^{K} \boxed{T_{ck}} = 1,$$

indicator that *c* is assigned to cluster *k*

counting disagreements $\boxed{w_{ijc} \geq \delta_{jik} + T_{ck} - 1,} \forall (i,j) \in c, \forall k,$

total order between *i* and *j* $\boxed{\delta_{ijk} + \delta_{jik} = 1, \forall i, j, k,}$

transitivity constraint $\boxed{\delta_{ijk} \geq \delta_{irk} + \delta_{rjk} - 1,} \forall i, j, k,$

$$\left.\begin{aligned} 0 \leq w_{ijc} &\leq 1, \forall i, j, c, \\ \delta_{ijk} &\in \{0,1\} \\ T_{ck} &\in \{0,1\}, \forall c, k. \end{aligned}\right\}$$

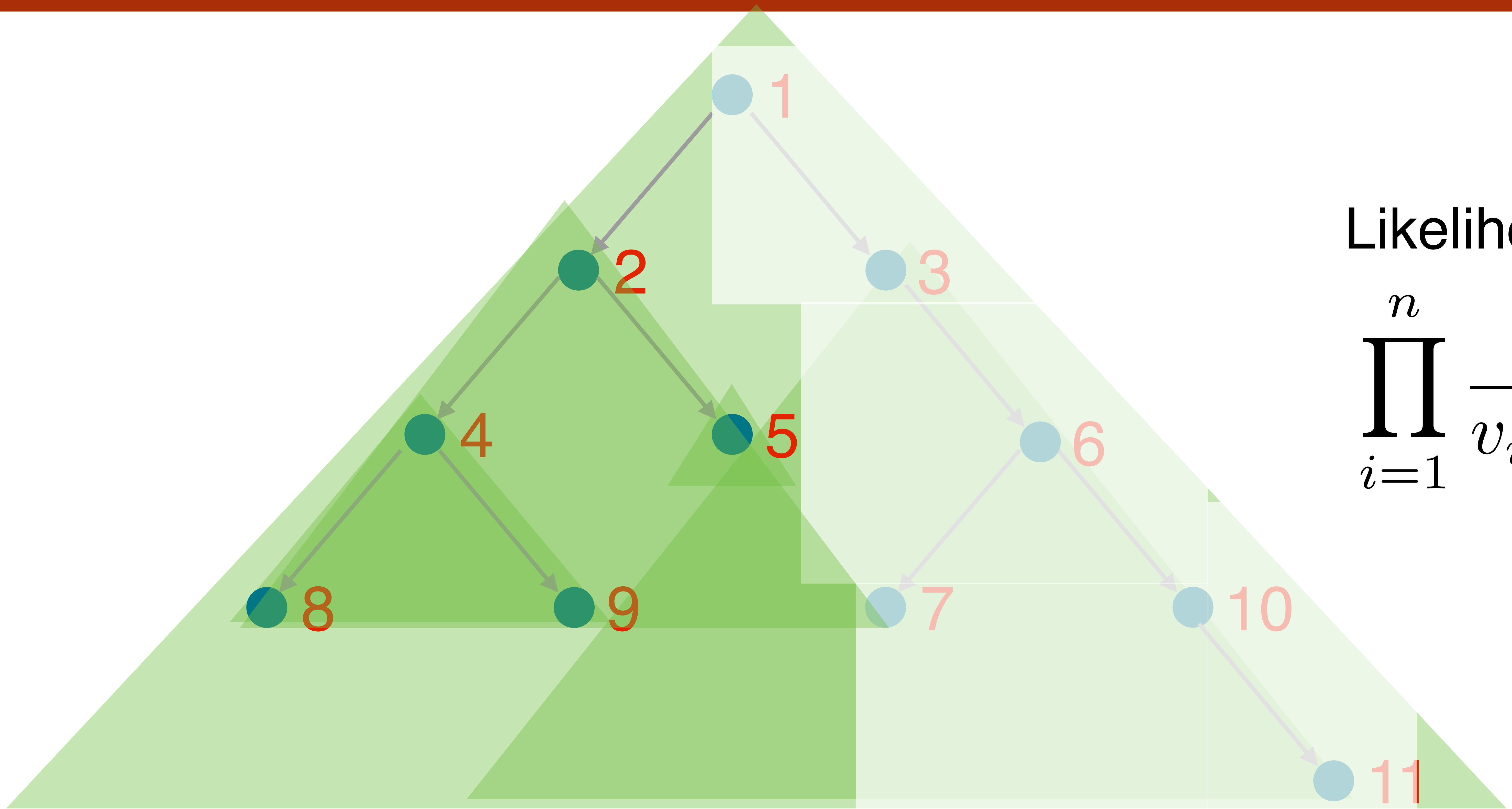
bounds and integrality constraints

Three-step modular framework for inference

		technique
Step 1	Construct customer DAGs	model for consideration set
	↓ DAGs	
Step 2	Cluster the DAGs	clustering technique
	↓ DAG clusters	
Step 3	Fit choice models to clusters	Approx ML estimation
	↓ choice models	

Steps do not depend on assumptions of previous steps

MNL model has closed-form expression for forest of trees with one root



Likelihood of DAG

$$\prod_{i=1}^{n} \frac{v_i}{v_i + \sum_{j \in \boxed{T_i}} v_j}$$

subtree of node *i*

prob of tree = (prob of subtree 2) x (prob of subtree 3) x $\frac{v_1}{v_1 + \cdots + v_{11}}$

prob of subtree 2 = (prob of subtree 4) x (prob of subtree 5) x $\frac{v_2}{v_2 + v_4 + v_5 + v_8 + v_9}$

⋮

A Partial-Order Based Model to Estimate Individual Preferences Using Panel Data

1. Model
   assumptions, data for estimation

2. Modular inference framework
   consideration set, clustering of DAGs, ML estimation

3. Numerical results
   results on real-world panel data from IRI dataset

4. Summary/Conclusions
   takeaway messages

Benchmark models compared: LC-MNL and RPL

k-latent class MNL (LC-MNL)

- sample class membership; follow MNL for that class
- # of parameters = $k\ n$
- EM-based regularized max likelihood estimation
- best performance up to $k = 10$ reported

Random Parameters Logit (RPL)

- sample MNL parameters ~ multivariate normal
- # of parameters = $2\ n$
- Max simulated likelihood estimation (MSLE)
- computationally intensive

Experiments conducted: one step-ahead prediction

U = user set N = product set T = # of discrete time periods

For any $t = 1, 2, 3, \ldots, T$

Given

everything until time period t

(offer sets, promoted items, purchases of users in U)

S_{t+1} = offer set in period $t+1$

P_{t+1} = promoted items in period $t+1$

U_{t+1} = users purchasing in period $t+1$

Prediction

$f_u(i, t+1) = 1$ if i has highest choice probability for u in period $t+1$

for all $u \in U_{t+1}$, $i \in S_{t+1}$

# of purchases of i by u

X2 score = $\frac{1}{|U||N|} \sum_{u \in U, i \in N} \frac{(n_{ui} - \hat{n}_{ui})^2}{0.5 + \hat{n}_{ui}}$, $\hat{n}_{ui} = \sum_{t \in T} f_u(i, t)$ similar to χ^2 score $\frac{(O-E)^2}{E}$

lower is better

miss rate = $\frac{1}{|U||T|} \sum_{u \in U, t \in T} I[f_u(a_{j_{u,t}}) \neq 1]$

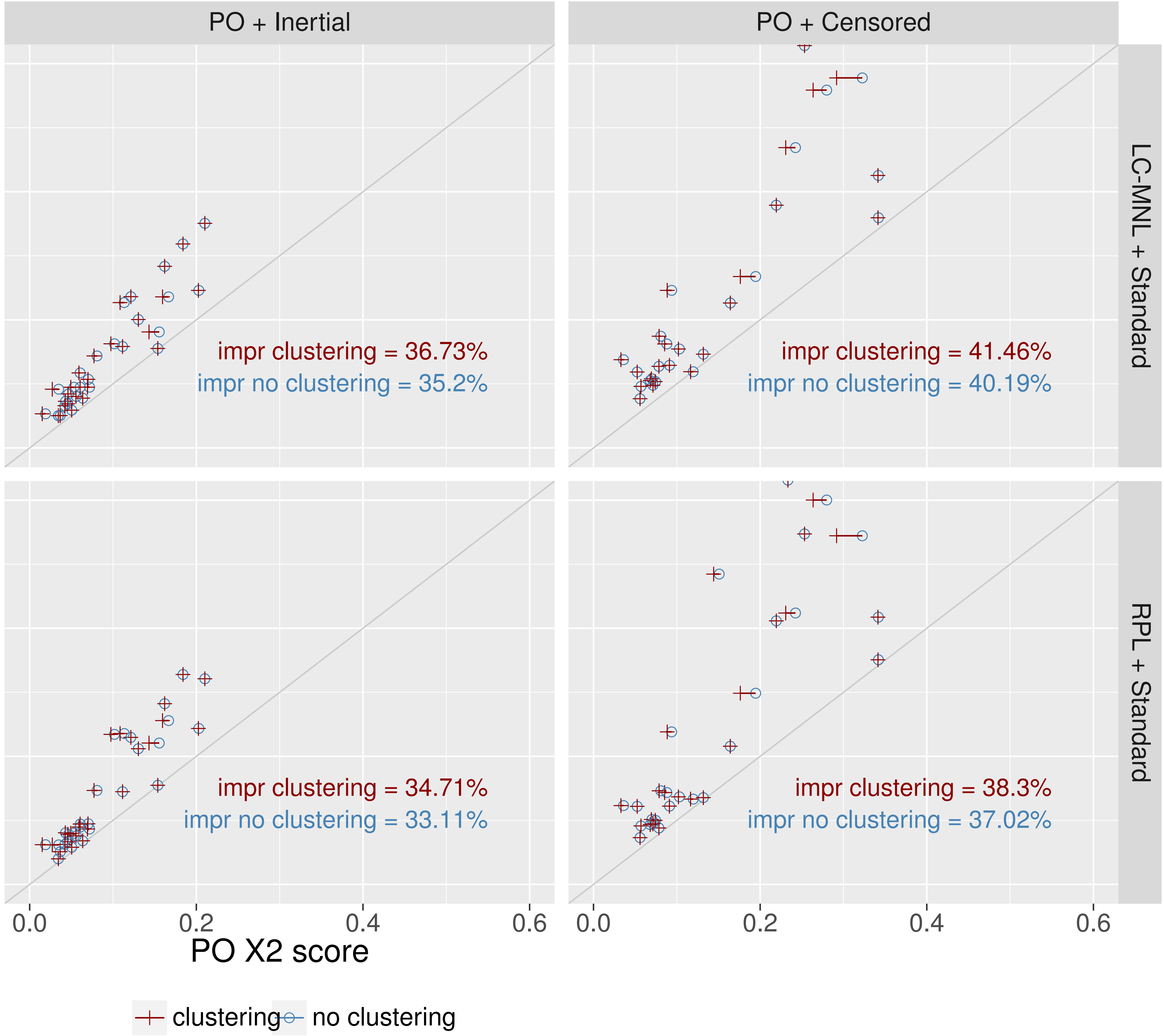
obs. purchase in t of u

Lower is better

improvement attributed to DAGs

- benchmarks have more parameters
- time: our method ~ 10 secs, RPL ~ 67 minutes

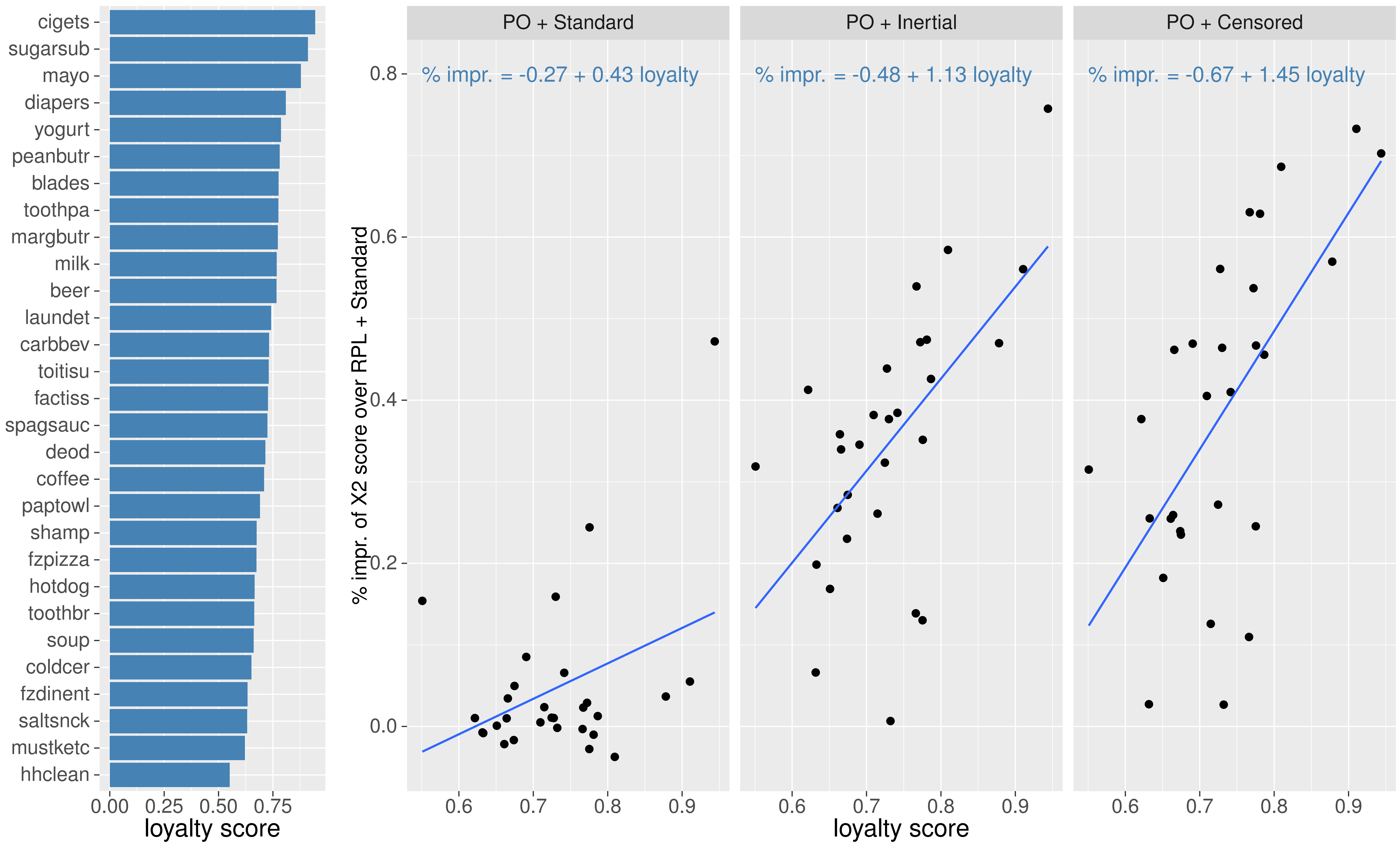
Lower is better



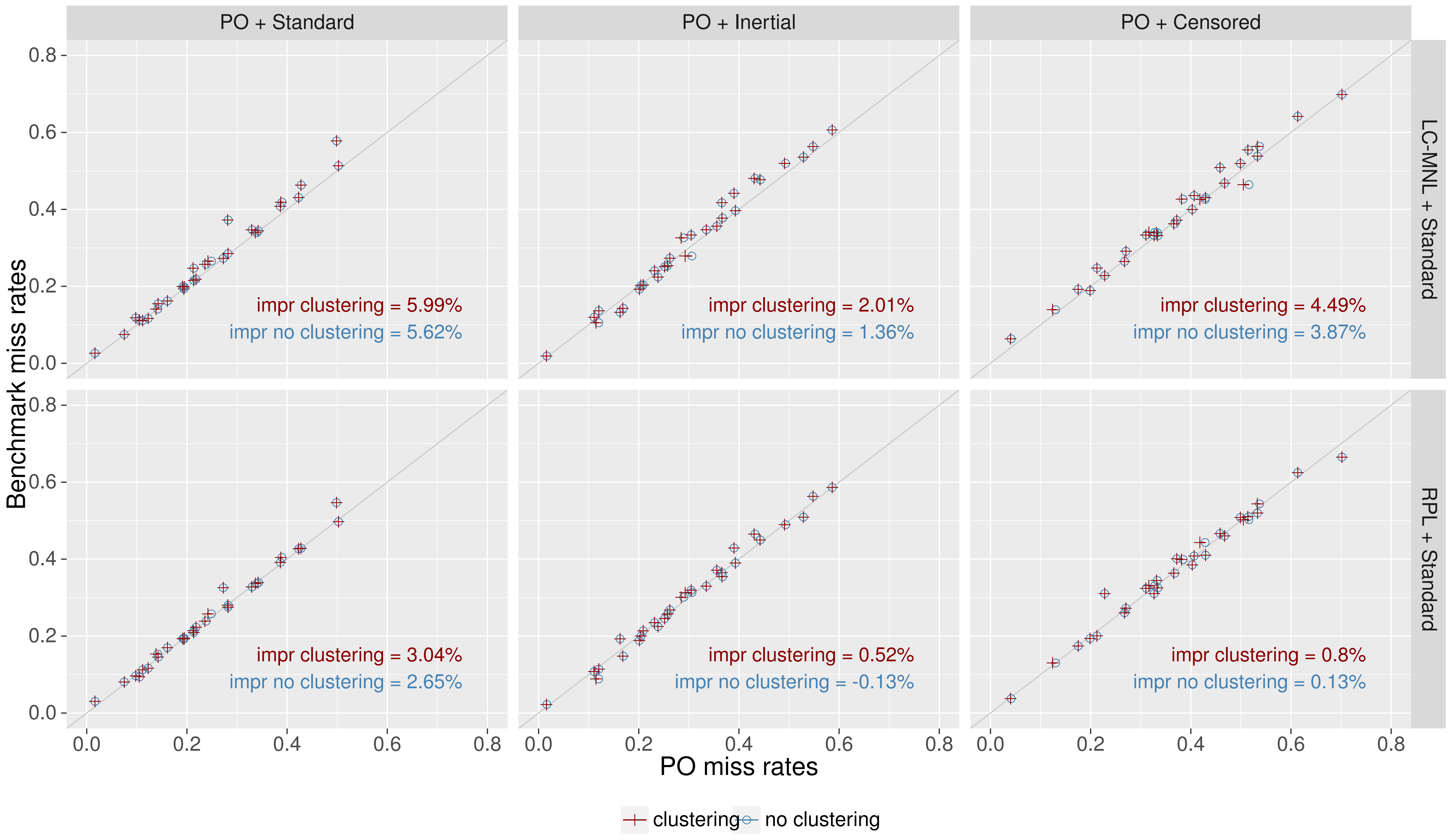
Boost due to

- accounting for behavioral effects
- clustering of DAGs

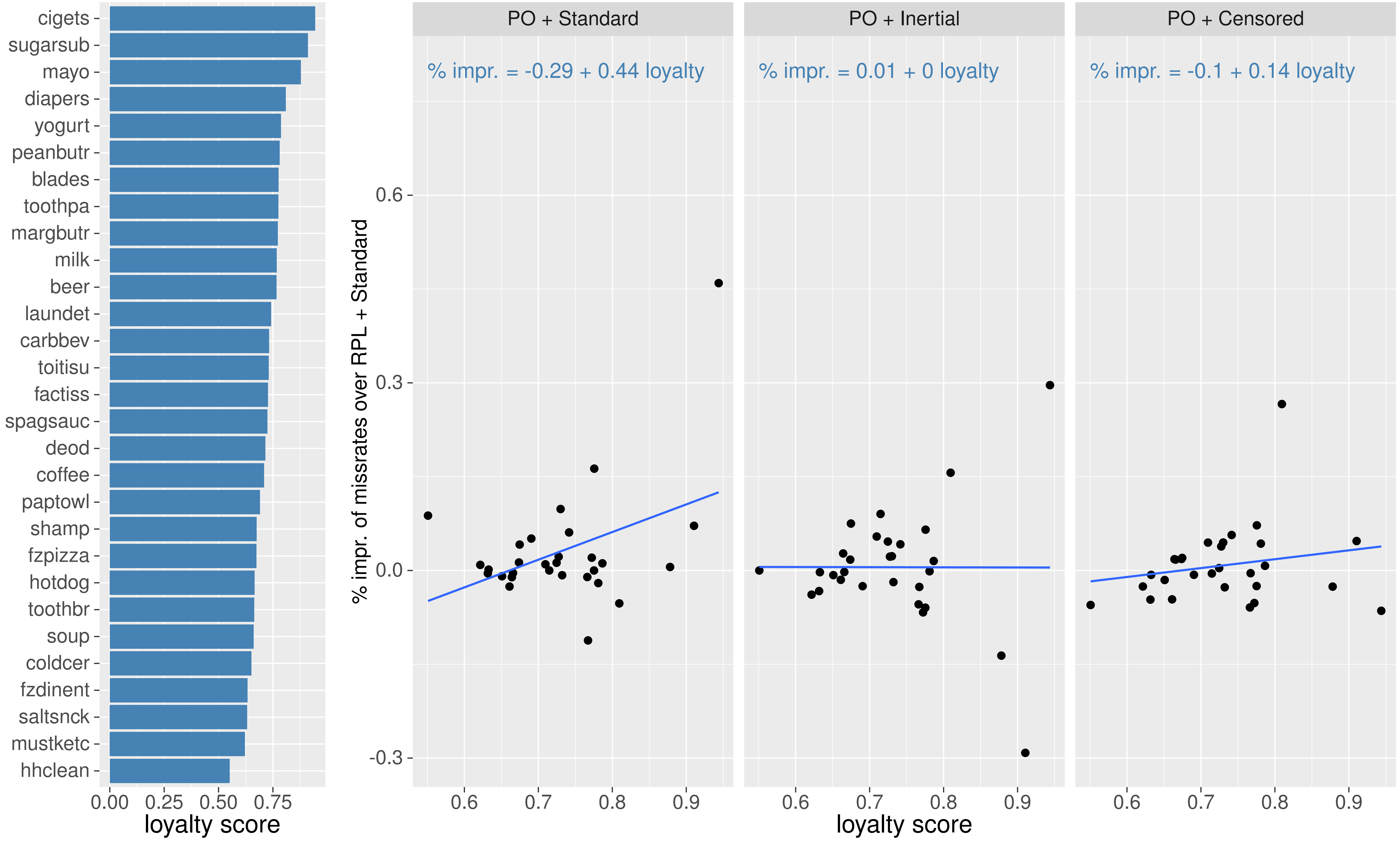
Improvements higher for categories with higher “brand loyalty scores”



Loyalty score = fraction of purchases from most frequently purchased brand

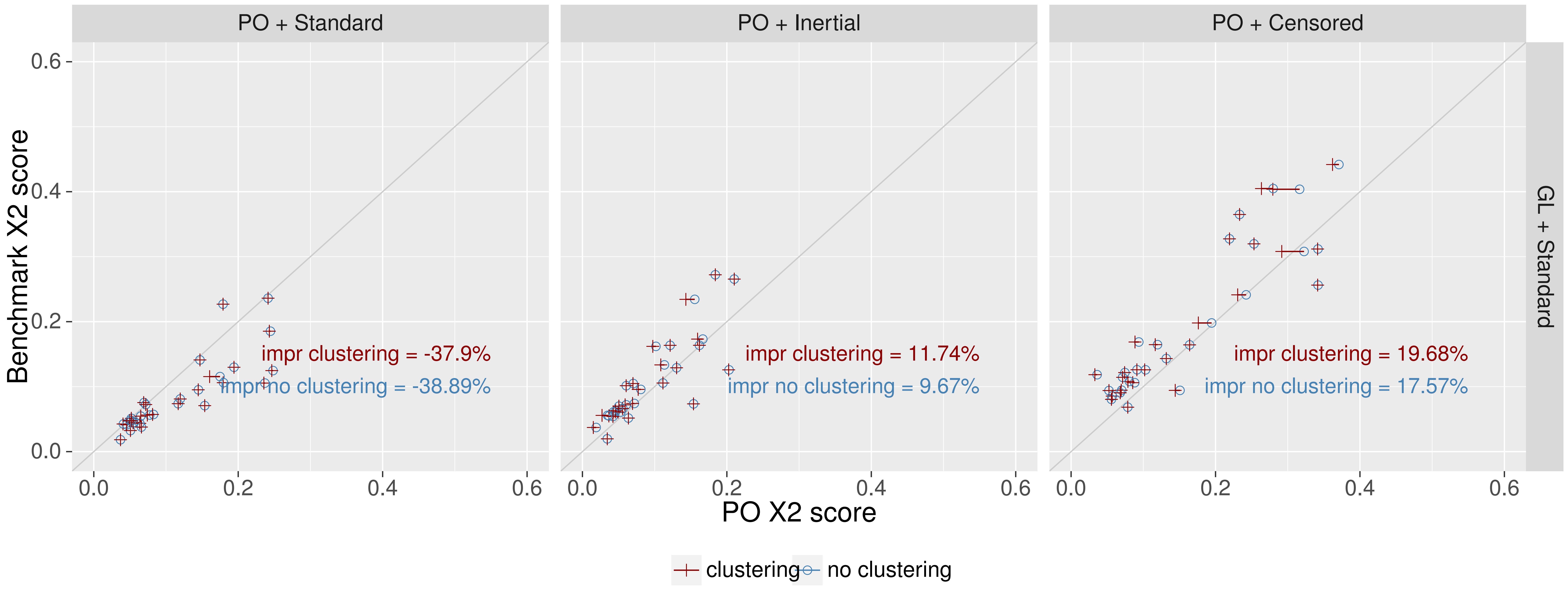


Results are similar for miss rate metric



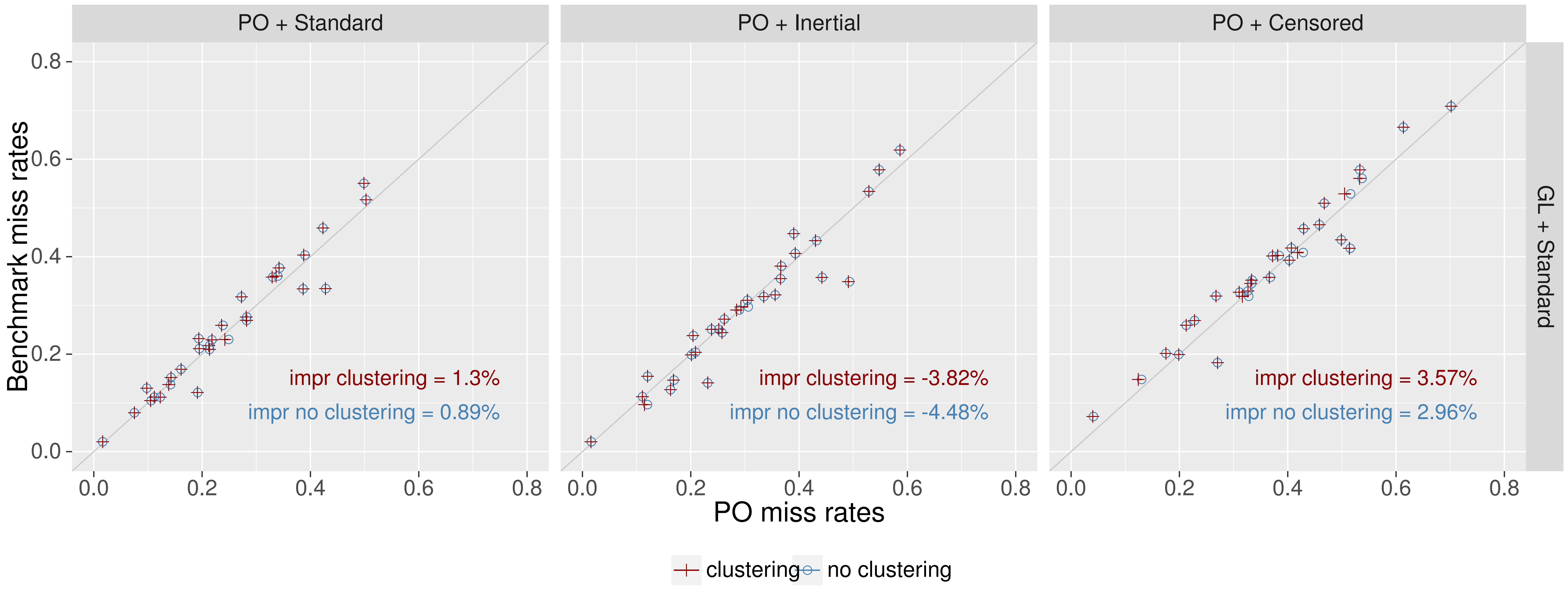
Results are similar for miss rate metric

Improvements over Gudagni-Little (GL) model for inertial and censored



Improvements obtained even after accounting for state

Improvements over Gudagni-Little (GL) model for standard and censored



GL performs well for inertial because their "loyalty" consistent with GL assumptions

A Partial-Order Based Model to Estimate Individual Preferences Using Panel Data

1. Model
   assumptions, data for estimation

2. Modular inference framework
   consideration set, clustering of DAGs, ML estimation

3. Numerical results
   results on real-world panel data from IRI dataset

4. Summary/Conclusions
   takeaway messages

Summary and key findings

KEY CONTRIBUTION

Methodology to estimate individual preferences from panel data through DAGs and choice inertia

MAIN TAKEAWAYS

DAGs and behavioral effects	significantly boost accuracy
PO+Censored	best for X2 metric comparable/best for miss-rates coverage: 71% of customer DAGs
PO+Standard	best for miss-rates