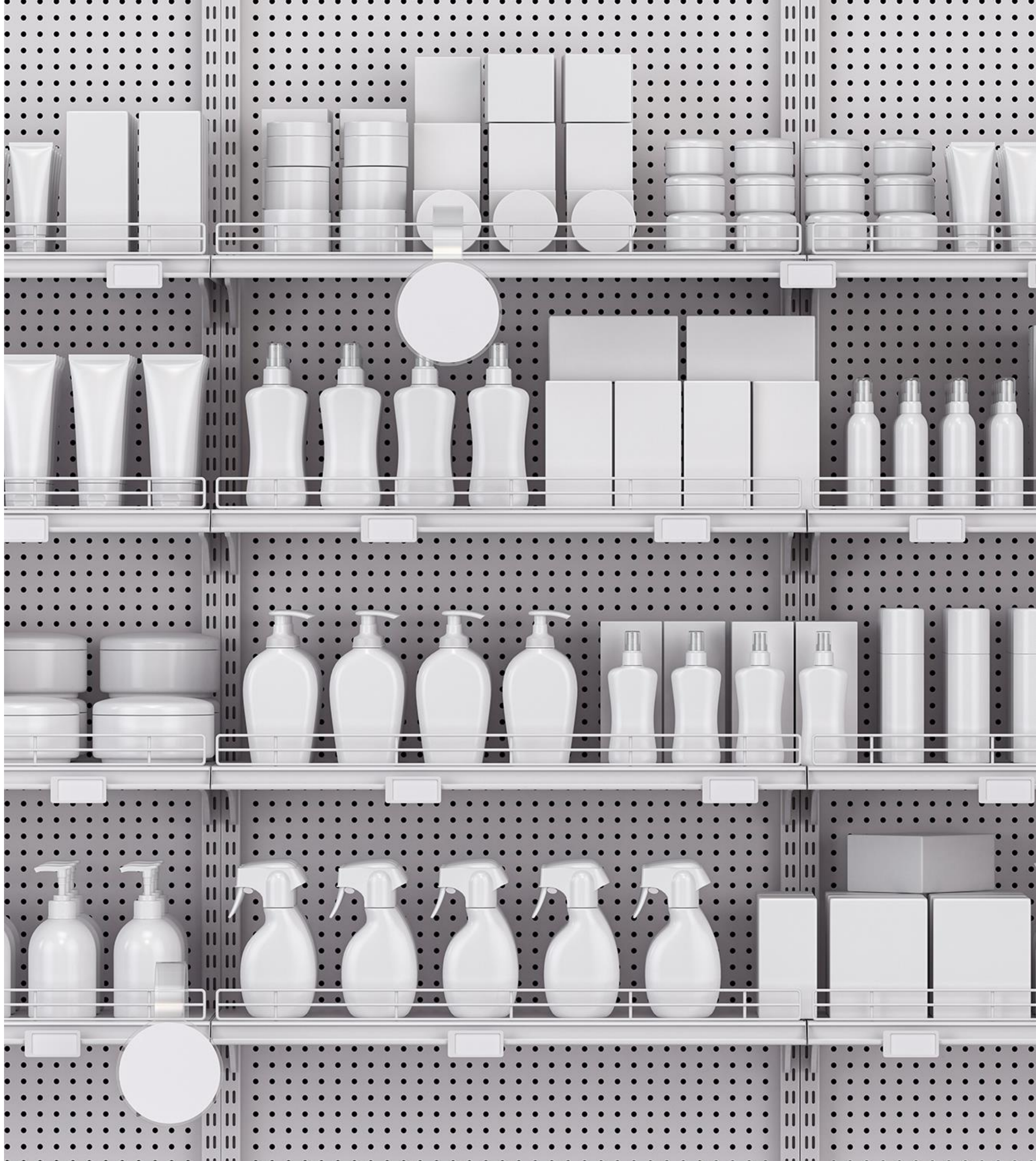
Volumetric Conjoint and the role of assortment size

Nino Hardt (SKIM)

Peter Kurz (bms marketing research + strategy)

2022 Sawtooth Software Conference

https://dx.doi.org/10.2139/ssrn.3418383



In many CPGs, demand is not discrete

- Consumers can buy **multiple units** and **multiple varieties** during any shopping trip

- There are **different package sizes** (pint, half gallon…)

- Brands need to decide **which** and **how many** varieties to offer

SKIM

Discrete vs Volumetric

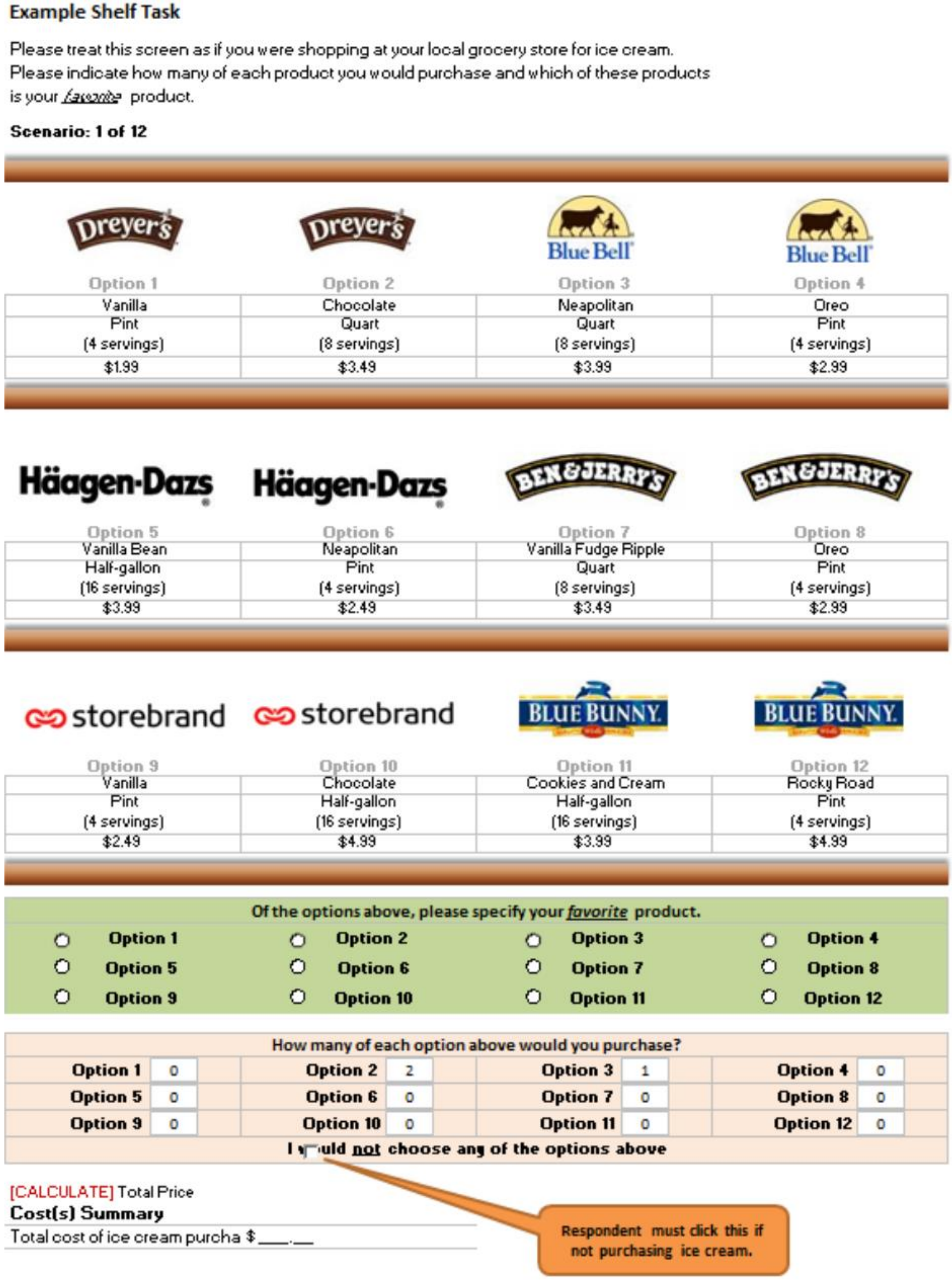
SKIM

Discrete vs Volumetric considerations

- Do you assume total **category sales constant**?
  - If yes, it may suffice to understand **market shares**
    - Are **quantity** and **brand choice** independent?
      - If yes, quantity-then-choice model may suffice
      - If no, use a volumetric demand model like the Multiple Discrete-Continuous Demand model (MDCM)
  - If no, we need to model what drives **volume**

Discrete and Quantity choice – side by side comparison

- Respondents chose the '**most preferred**' option first, **then** they selected **quantities** of any of the items
- No-Choice option last
- Does it matter if we use **discrete choice** or **quantity choice** data?


Example Shelf Task

Please treat this screen as if you were shopping at your local grocery store for ice cream. Please indicate how many of each product you would purchase and which of these products is your *favorite* product.

Scenario: 1 of 12

Dreyer's	Dreyer's	Blue Bell	Blue Bell
Option 1	Option 2	Option 3	Option 4
Vanilla	Chocolate	Neapolitan	Oreo
Pint	Quart	Quart	Pint
(4 servings)	(8 servings)	(8 servings)	(4 servings)
$1.99	$3.49	$3.99	$2.99

Häagen-Dazs	Häagen-Dazs	Ben & Jerry's	Ben & Jerry's
Option 5	Option 6	Option 7	Option 8
Vanilla Bean	Neapolitan	Vanilla Fudge Ripple	Oreo
Half-gallon	Pint	Quart	Pint
(16 servings)	(4 servings)	(8 servings)	(4 servings)
$3.99	$2.49	$3.49	$2.99

storebrand	storebrand	Blue Bunny	Blue Bunny
Option 9	Option 10	Option 11	Option 12
Vanilla	Chocolate	Cookies and Cream	Rocky Road
Pint	Half-gallon	Half-gallon	Pint
(4 servings)	(16 servings)	(16 servings)	(4 servings)
$2.49	$4.99	$3.99	$4.99

Of the options above, please specify your *favorite* product.

○ Option 1	○ Option 2	○ Option 3	○ Option 4
○ Option 5	○ Option 6	○ Option 7	○ Option 8
○ Option 9	○ Option 10	○ Option 11	○ Option 12

How many of each option above would you purchase?

Option 1	0	Option 2	2	Option 3	1	Option 4	0
Option 5	0	Option 6	0	Option 7	0	Option 8	0
Option 9	0	Option 10	0	Option 11	0	Option 12	0

I would <u>not</u> choose any of the options above

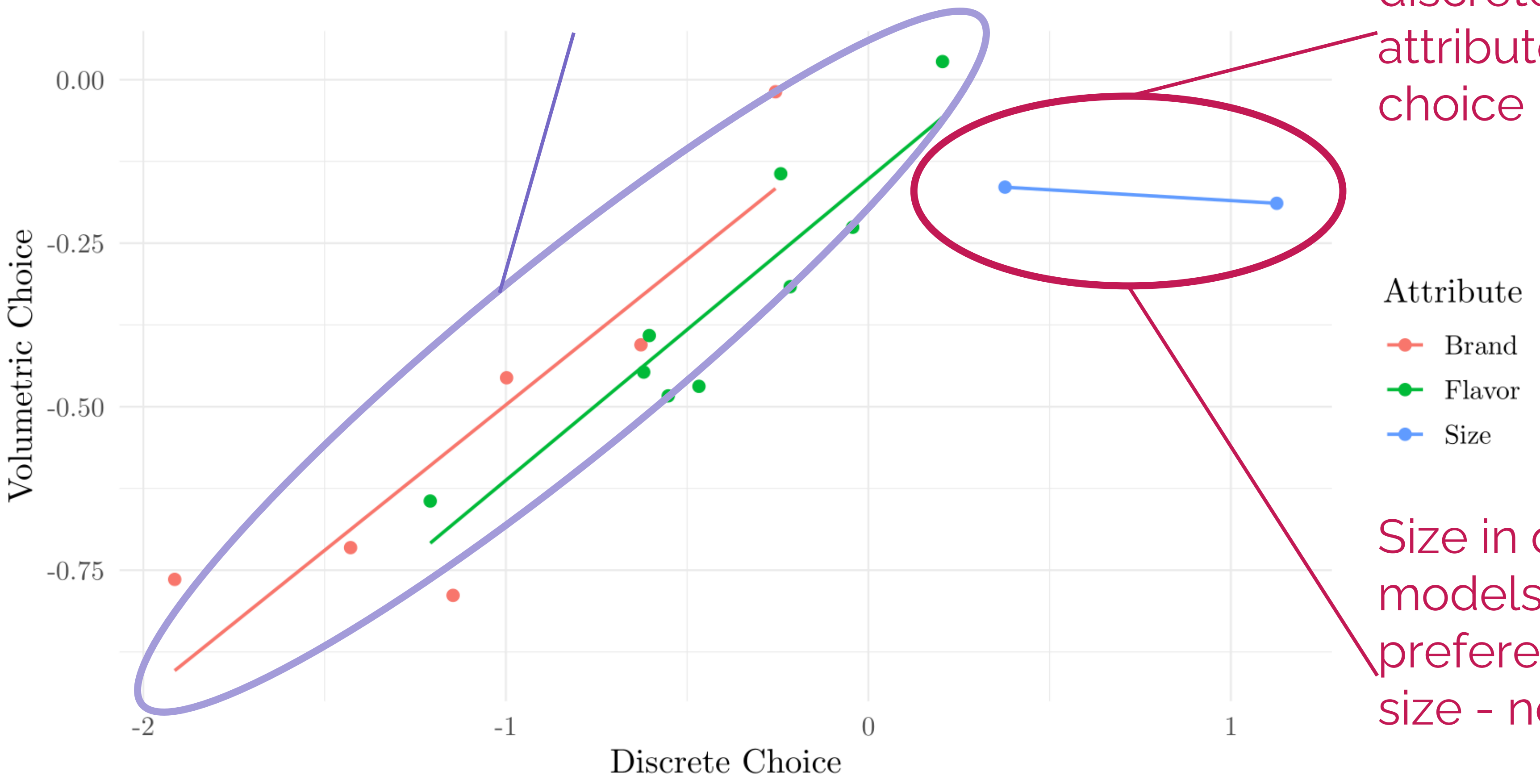
[CALCULATE] Total Price
Cost(s) Summary
Total cost of ice cream purcha $___.__

Source: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2770025 (forthcoming IJRM)

Where discrete choice and volumetric choice data agree

Consistent for most attributes/levels

Size attribute in discrete choice <> Size attribute in quantity choice



Size in quantity models captures a preference for pack size - not quantity

Source: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2770025 (forthcoming IJRM)

SKIM

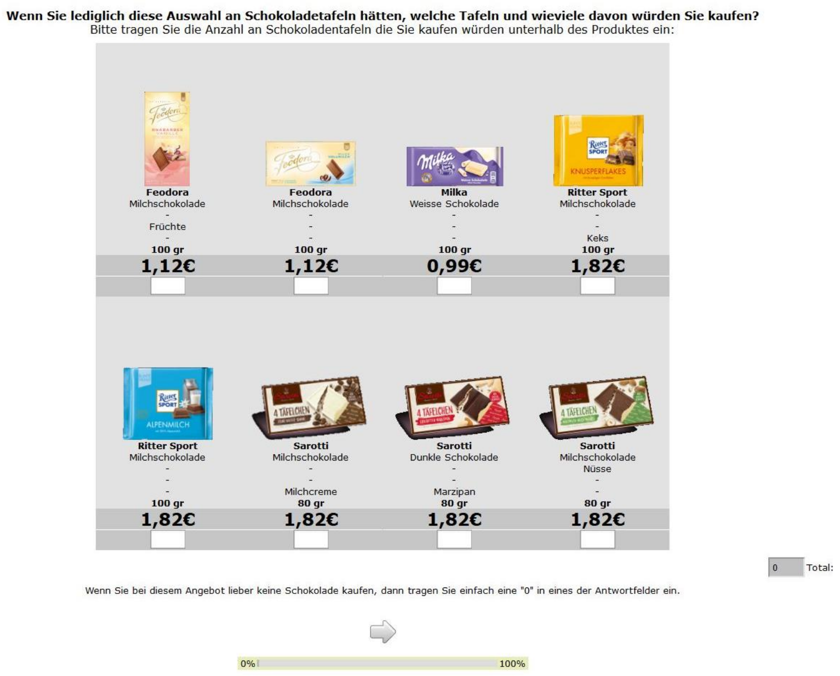
Where discrete vs volumetric matters

- MDCM (volumetric) models comprise **part-worths, budget** and **rate of satiation**, which is important to account for variety seeking
- MDCM can help distinguish between **incidence** (buying >0 units) from **volume** (quantity given >0 units) (Kim, Hardt, Kim, Allenby 2022)
- Both **composition** and **size** of assortments drive overall volume in CPGs where variety seeking exists
  - Large grocers hope to generate more sales – across all brands on the shelves - by offering a larger assortment
  - I buy more ice cream when there so many varieties to try….

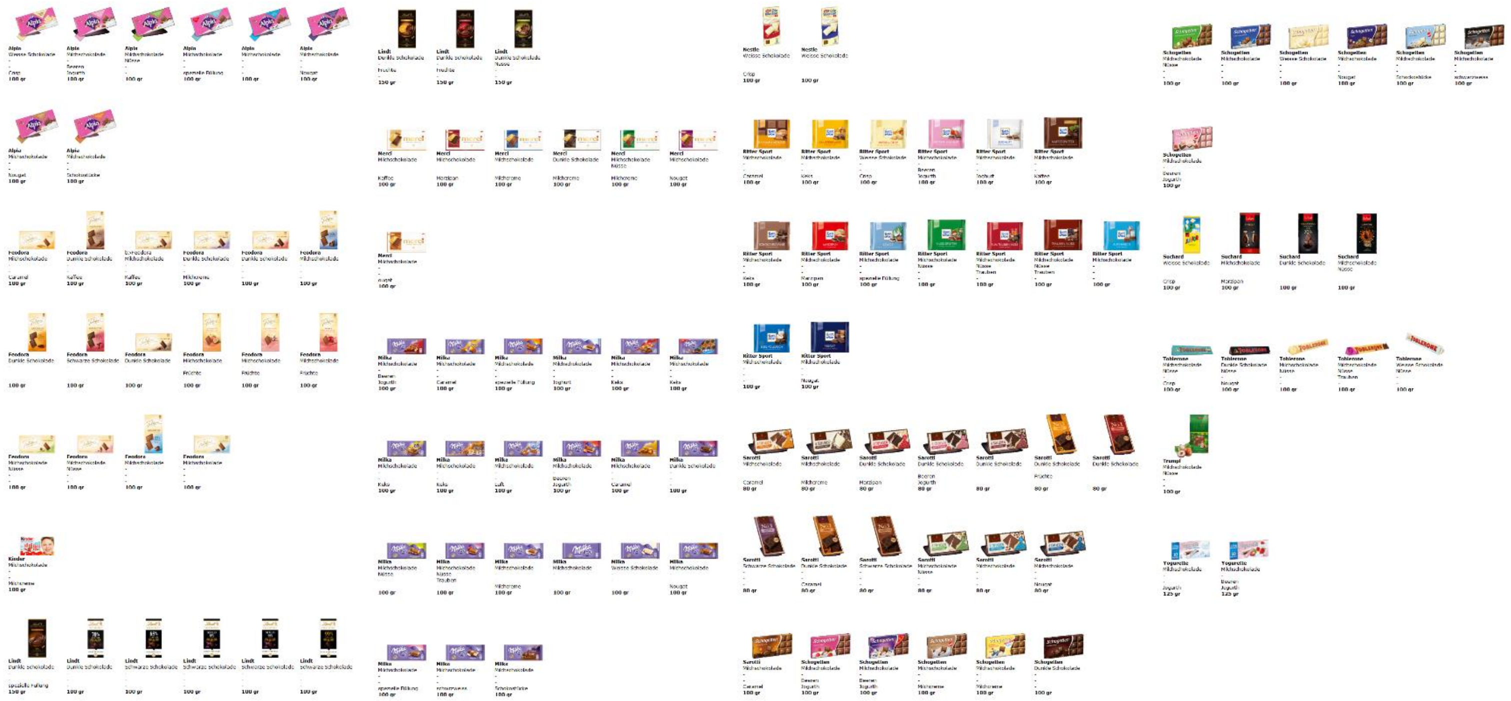
Assortment size variation

The obvious problem: Conjoint choice sets are much smaller than assortments in stores

Conjoint Survey



Top sellers in typical stores



- Can we extrapolate demand from a large assortment if we only observe demand from smaller choice sets?

SKIM

What do we know about the relationship between assortment size and demand?

- Mixed results from **empirical research** looking into the effect of temporary changes in assortment size on overall sales (Boatwright and Nunes, 2001; Sloot et al., 2006; Borle et al., 2005).
- **Behavioral research** suggests there might be 'too many' varieties and 'choice overload' (Meissner et al., 2019; Scheibehenne et al., 2010)
- The relationship between assortment size and total demand may be **inverse U-shaped** (Herzenstein et al., 2019)

SKIM

Demand models accounting for assortment size

- Choice models have been developed with fixed size of choice-set in mind
  - Discrete Choice models may result in biased estimates if variation in number of products remain unaccounted for; and adjustment term may be needed (Ackerberg & Rysman 2005)
- Multiple Discrete-Continuous Demand Models (MDCM) have also been developed for fixed choice-set sizes
  - We show that ignoring set size variation results in seriously biased predictions of demand
  - A parametric adjustment for set size allows to even extrapolate for larger set sizes

Developing a model

Modeling quantity/volumetric demand

- The most common approach is **MDCM** – Multiple Discrete Continuous Model
- This means we model so-called **corners** (stuff you don't buy) and **interior** solutions (stuff you do buy), where the quantity of interior solutions is continuous
- **Budget** constraints (and/or space constraints) prevent people from buying unlimited quantities. Budgets determine the nature of cross-effects
- In reality, we only buy full bottles, but integer constraints would make the model intractable

Standard MDCM

Baseline util inside good

Inside good

Baseline util outside good

Outside good

Budget

$$\text{Max } u(\mathbf{x}, z) = \sum_{j=1}^{N} \frac{\psi_j}{\gamma} \ln(\gamma x_j + 1) + \psi_z \ln(z) \qquad \text{s.t. } \sum_{j=1}^{N} p_j x_j + z = E$$

Satiation rate inside goods

$$\text{where } \psi_j = \exp\left(a_j^T \beta + \varepsilon_j\right)$$

- Consumers **maximize utility** from choosing quantities (x_j) of N inside goods and `leftover money' z, given available **budget** E.
- **Diminishing marginal returns** of inside (x_j) and outside (z) goods
- γ governs **rate of satiation** of inside goods
- $\psi_z = 1$ fixed for identification

Incorporating the effect of assortment size

$$u\left(\mathbf{x_t}, z_t\right) = \sum_{j=1}^{N_t} \frac{\psi_{jt}}{\gamma} \ln\left(\gamma \mathbf{x}_{jt} + 1\right) + f\left(N_t; \xi\right) \ln(z_t)$$

- Variation in assortment size means ψ_z is identifiable if it's a function of assortment size N_t
- We find that a simple linear function of N_t is sufficient

$$f\left(N_t; \xi_1\right) = \xi_1 N_t + 1$$

SKIM

Log-Likelihood & Estimation

$$\Pr(x_t) = |J_{R_t}| \left\{ \prod_{j=1}^{R_t} \frac{\exp(-g_{jt}/\sigma)}{\sigma} \exp\left(-e^{-g_{jt}/\sigma}\right) \right\} \left\{ \prod_{i=R_t+1}^{N_t} \exp\left(-e^{-g_{it}/\sigma}\right) \right\}$$

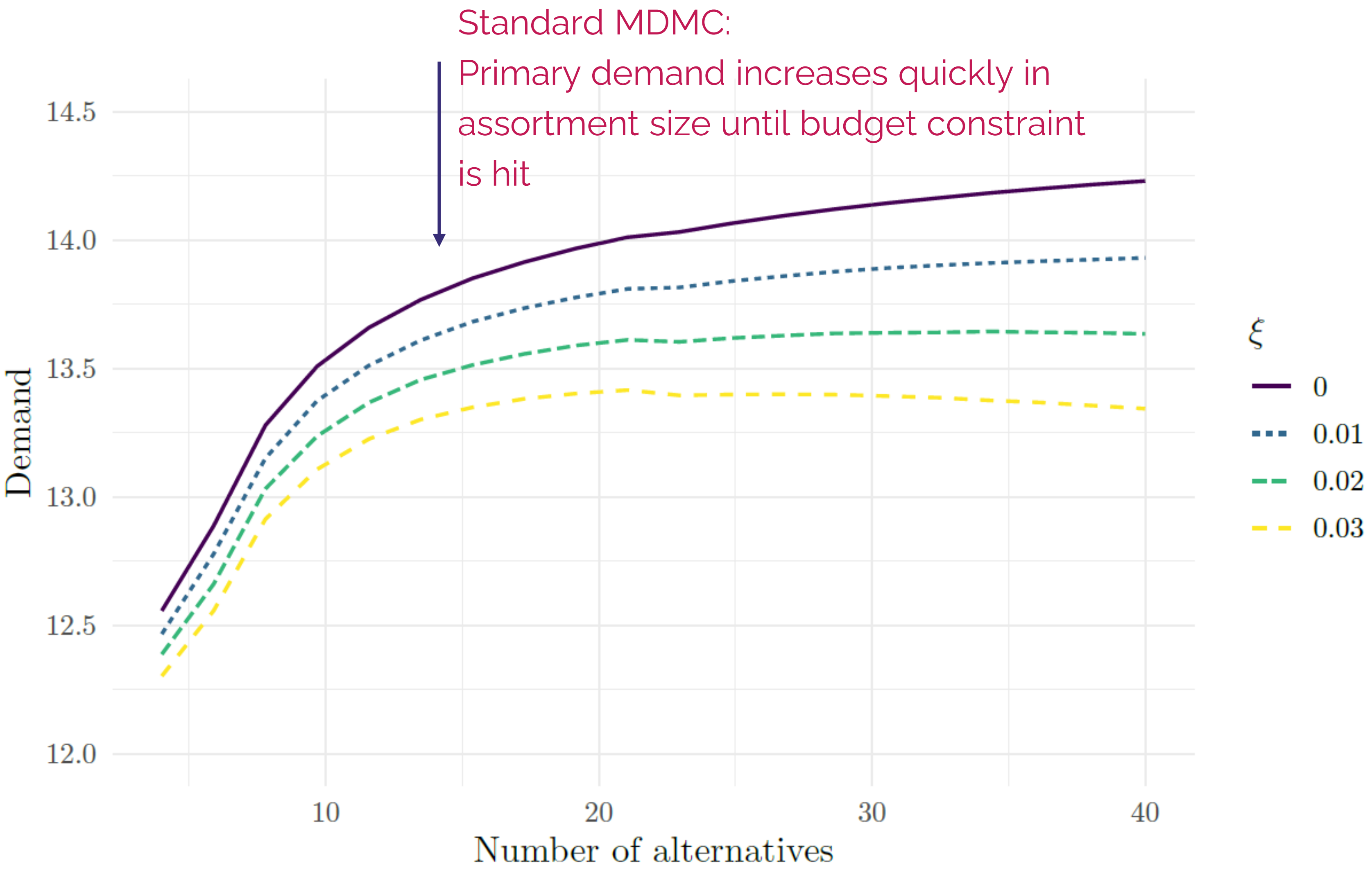
where

$$g_{kt} = -\mathbf{a}_{kt}\beta + \ln(\xi N_t + 1) + \ln(\mathbf{p}_{kt}) + \ln(\gamma \mathbf{x}_{kt} + 1) - \ln(z_t)$$

$$|J_{R_t}| = \prod_{i=1}^{R_t} \left(\frac{\gamma}{\gamma x_{i,t} + 1} \right) \left\{ \sum_{i=1}^{R_t} \frac{\gamma x_{i,t} + 1}{\gamma} \cdot \frac{p_i}{z_t} + 1 \right\}$$

- **Prior**: Standard MVN Heterogeneity and weakly-informative priors
- **Estimation**: Standard Hybrid MCMC
- **Prediction**: Numeric integration over error and parameter space

Proposed model captures flexible relationship between set size and primary demand



SKIM

Results from 2 studies

Study#1: Chocolate bars

Attributes	Levels
Brand	Alpia, Feodora, Kinder, Lindt, Merci, Milka, Nestle, Ritter, Sarotti, Schogetten, Suchard, Tobler, Trumpf, Ferrero/Yogurette
Chocolate	Milk, Dark, Black, White
Nut	Nut, No Nut
Fruit	Fruit, Berry, Grape, No Fruit
Filling	None, Yogurt, Choc Chunk, Coffee, Cookie, Black and White, Crisp, Nougat, Caramel, Milk Ccreme, Special, Marzipan

- Using this attribute space, we can cover 80% of sales in typical stores

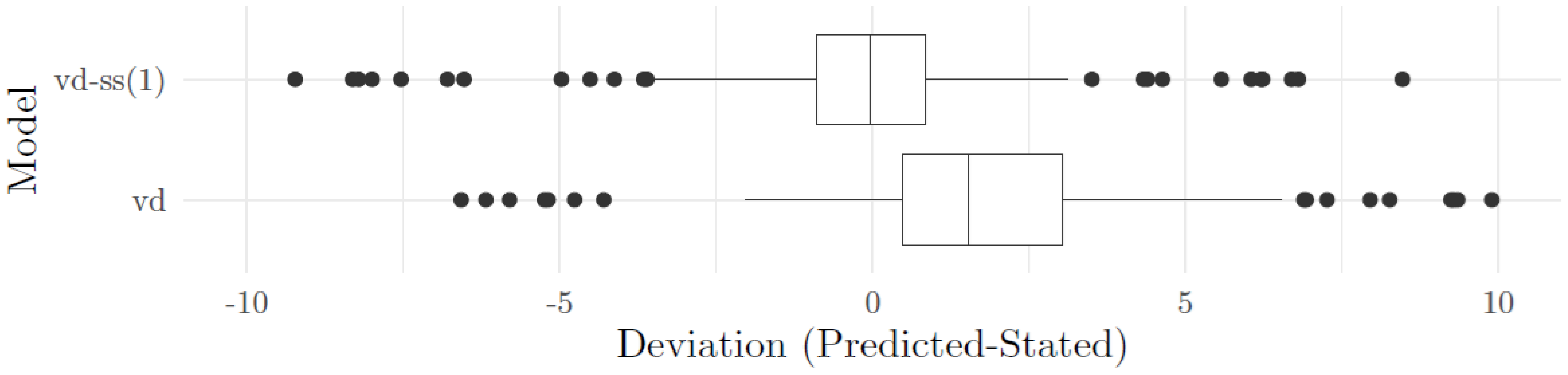
Study#1: Chocolate bars - Descriptives

Data	Units per task		Varieties per task	
# Alternatives	mean	sd	mean	sd
8	**1.31**	1.25	1.05	0.93
18	**1.95**	2.08	1.54	1.61

- No choice `overload'.
- Respondents buy more varieties and **overall higher quantities when offered larger assortment**
- How will primary demand extrapolate from 18 to 100 alternatives?

SKIM

Study#1: Predicted demand vs self-stated quantity during last trip (assuming typical assortment)



- Boxplot shows deviation of **predicted** primary demand from **self-stated last purchase quantity**
- Proposed model unbiased, while **standard MDCM severely overpredicts** quantity

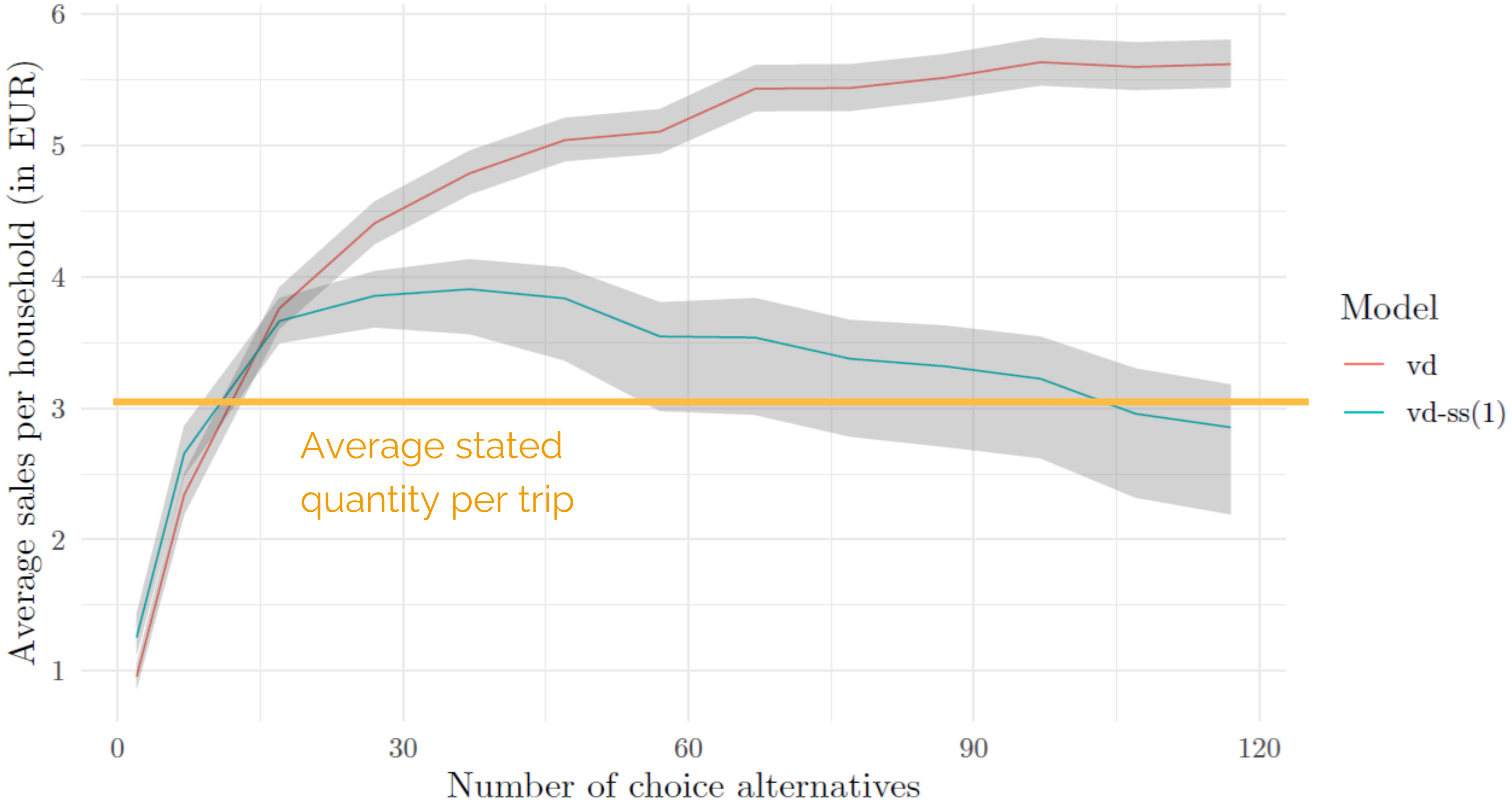
SKIM

Study#1: Chocolate bars – Market extrapolation

model	mean
vd	434,730
vd-ss(1)	218,742
stated 'last purchase'	215,422

- Assumptions: 25m households, 3 shopping trips/mo
- Predicted annual demand of **218,742** is close to the actual number: **240,000**
- **Ignoring** the set size adjustment, demand is **over-predicted by 80%**

Study#1: Extrapolating from 8/18 to >100 products



SKIM

Study#2: Air fresheners

- Respondents are first given a realistic shelf task with 57 alternatives
- Then, respondents chose among 8, 16, 24 alternatives

Please imagine these are the available air fresheners.
Which of these products would you buy?
You can buy as many units of each product as you like.

Brand	renuzit	renuzit	glade	glade	febreze	febreze	CITRUS MAGIC	CITRUS MAGIC
Fragrance	FRESH	SPICY	OUTDOOR	LAVENDER	GOURMET	CITRUS	CITRUS	FRUITY
Delivery method	Diffusor	Gel beads	Dispenser	Sticks	Candle	Gel beads	Dispenser	Fabric Flower
Type	no	yes	yes	no	no	yes	yes	yes
Price	$2.99	$4.99	$2.49	$2.49	$3.49	$3.49	$1.99	$3.99

Brand	CITRUS MAGIC	CALIFORNIA scents	CALIFORNIA scents	brightair	brightair	ARM & HAMMER	ARM & HAMMER	ARM & HAMMER
Fragrance	SPICY	TROPICAL	FLORAL	FRUITY	FRESH	FLORAL	GOURMET	OUTDOOR
Delivery method	Scent Swirl	Diffusor	Scent Swirl	Fabric Flower	Spray	Stand/Holder	Candle	Stand/Holder
Type	no	yes	no	no	no	yes	no	yes
Price	$3.99	$4.49	$4.49	$1.99	$3.99	$4.49	$4.99	$2.99

None, I wouldn't buy any of these.

0% 100%

Study#2: Demand quantities aren't constants

Number of Alternatives	Units per task mean	Units per task sd	Varieties per task mean	Varieties per task sd	$ spent per task mean	$ spent per task sd	Maximum spent mean	Maximum spent sd
8	1.05	1.27	0.79	0.80	2.80	3.84	6.30	5.46
16	1.04	1.31	0.80	0.87	2.84	4.07	6.20	5.59
24	1.43	1.76	1.11	1.17	3.88	5.27	7.59	7.17
57	4.11	6.18	2.58	2.87	6.21	10.10	6.21	10.10

- Demand increases with assortment size!
- Much higher demand in shelf task

Study#2: Extrapolation is possible

Brand	*Actual*	vd	**vd-ss(1)**	vd-ss(2)
Renuzit	1,137	1,675	1,256	1,225
Glade	479	750	559	555
Febreze	311	245	177	174
BrightAir	63	39	28	29
CitrusMagic	51	105	77	77
ArmHammer	40	14	10	10
CaliforniaScent	39	126	88	89
Total	2,120	2,953	2,196	2,162
Relative		**139%**	**104%**	**102%**

SKIM

Summary of findings

- Our model is useful when **variety seeking** exists and **primary demand** is not fixed
- It is remarkably accurate in predicting market-level sales based on volumetric conjoint and straightforward assumptions
- Ignoring set size variation results in dramatic over-prediction of total demand; the proposed model is remarkably accurate
- With data fusion becoming popular, there are many instances when assortment sizes are not fixed. Accounting for choice set-size is important!

Discussion

Volumetric choice tasks are easy to respond to

- We use "**open** sum" question and not "*constant* sum"
- This makes **volumetric tasks easier** for them to answer, because they don't need to think which portions sum up to the requested total sum.
- Respondents simply enter each number of products they like to buy
- *In some categories* we remind the **respondents how many products** they bought during the last shopping trip. This could help to make the answers mor realistic.
- Air freshener is from our experience no category where this information is needed

Brand	renuzit	renuzit	glade	glade	febreze	febreze	CITRUS MAGIC	CITRUS MAGIC
Fragrance	FRESH	SPICY	OUTDOOR	LAVENDER	GOURMET	CITRUS	CITRUS	FRUITY
Delivery method	Diffusor	Gel beads	Dispenser	Sticks	Candle	Gel beads	Dispenser	Fabric Flower
Type	no	yes	yes	no	no	yes	yes	yes
Price	$2.99	$4.99	$2.49	$2.49	$3.49	$3.49	$1.99	$3.99

Brand	CITRUS MAGIC	CALIFORNIA scents	CALIFORNIA scents	brightair	brightair	ARM & HAMMER	ARM & HAMMER	ARM & HAMMER
Fragrance	SPICY	TROPICAL	FLORAL	FRUITY	FRESH	FLORAL	GOURMET	OUTDOOR
Delivery method	Scent Swirl	Diffusor	Scent Swirl	Fabric Flower	Spray	Stand/Holder	Candle	Stand/Holder
Type	no	yes	no	no	no	yes	no	yes
Price	$3.99	$4.49	$4.49	$1.99	$3.99	$4.49	$4.99	$2.99

None, I wouldn't buy any of these.

sum =

None, clears all fields above

Boxes allow each number of products the respondent like to buy

Information about actual sum of choices

Purchase timing, Stockpiling

- Stockpiling can be an issue, i.e. larger **quantities** mean longer **purchase intervals**
- Maybe impossible to fix in a pure-conjoint environment
- Some thoughts
  - Is the good perishable?
  - Let’s say the consumer bought a big pack of cookies … isn’t it tempting to eat more? So, while large purchase volume might result in longer inter-purchase time, it might also lead to consumption acceleration! Only solution would be diary data!

Issue#1: Extrapolation

- Usually, we are reluctant to extrapolate, i.e. we don't like to predict beyond the convex hull of past observations
- However, we can't use shelf tasks AND show an attribute grid at the same time
- How much do we trust an extrapolation?
- A shelf-like validation task can help build confidence

SKIM

Issue#2: Single occasion

- Stick to a single purchase occasion
- Still very relevant for price promotions, generating trials (first purchases)
- Single occasion sales maximization might be enough (e.g., when there is surplus supply of a good)

Takeaways

Takeaways

- Volumetric Conjoint is fancy, but only makes sense under certain conditions:
  - **Total market sales** (i.e., primary demand) are not fixed, but can vary with size and composition of offerings
  - **Variety seeking** and simultaneous demand for multiple varieties is common
  - and/or separating **incidence** and **volume** is of interest
- In 2 applications, we have shown the approach to yield realistic market-level predictions
- Our proposed model is important for **data fusion** – when conjoint and panel data is combined
- Everything is implemented in a convenient package: https://github.com/ninohardt/echoice2

SKIM

Volumetric models using `echoice2`

- Package implements **Discrete and Volumetric demand models** with and without screening, Volumetric model with and without assortment size variation
- So far **github only**, not CRAN release yet. Compiling packages on Windows requires rtools being installed, compiling on mac requires installing xcode, gcc and configuring gcc.
- Data is supplied in 'long' (**tidy data**) format where columns are: `id`, `task` number, `alt`ernative number, quantity `x`, price, attributes
- `vd_est_vdm` runs estimates a volumetric demand model without screening and set size variation, `vd_dem_vdm` generates demand predictions
- `ec_dem_aggregate` and `ec_dem_summarise` use dplyr functions to generate summaries
- For more details, follow instructions/vignettes on github

SKIM

Volumetric models using `echoice2`

https://github.com/ninohardt/echoice2

https://github.com/ninohardt/echoice2

readme.md

Please read the vignette. It illustrates a complete workflow from estimation to market simulation. The vignette can also be found on the package website.

Functions that relate to discrete demand start in `dd_`, while functions for volumetric demand start in `vd_`. Universal functions (discrete and volumetric choice) start in `ec_`. Estimation functions continue in `est`, demand simulators in `dem`.

The package comes with a small example dataset `icecream` from a volumetric conjoint study. It contains 300 respondents.

```
  data(icecream)
  icecream %>% head
#> # A tibble: 6 x 8
#>      id  task   alt     x     p Brand     Flavor      Size
#>   <int> <int> <int> <dbl> <dbl> <fct>     <fct>       <ord>
#> 1     1     1     1     8 0.998 Store     Neapolitan  16
#> 2     1     1     2     0 0.748 Store     VanillaBean 16
#> 3     1     1     3     0 1.25  BenNJerry Oreo        16
#> 4     1     1     4     0 0.748 BenNJerry Neapolitan  16
#> 5     1     1     5     0 2.49  HaagenDa  RockyRoad   4
#> 6     1     1     6     0 1.25  HaagenDa  Oreo        16
```

Choice data data.frames or tibbles need to contain the following columns:

- `id` (integer; respondent identifier)
- `task` (integer; task number)
- `alt` (integer; alternative number within task)
- `x` (double; quantity purchased)
- `p` (double; price)
- attributes defining the choice alternatives (factor, and soon continuous as well)

- `p` (double; price)
- attributes defining the choice alternatives (factor, and soon continuous as well)

While this requires a little extra space for discrete choice data, it simplifies the workflow and makes the package versatile. It can be applied to data from choice experiments and purchase histories. It allows variance in the number of choice tasks per subject, and variance in the number of choice alternatives per task.

Estimating a simple volumetric demand model is easy. Use the `vd_est_vdm` function, and use at least 100,000 draws:

```
est_icecream <- icecream %>% vd_est_vdm(R=10000)
#> Using 16 cores
#>  MCMC in progress
#>  Total Time Elapsed: 0.13 minutes
```

Upper-level estimates can be summarized using `ec_estimates_MU`:

```
est_icecream %>% ec_estimates_MU()
#> Warning: The `x` argument of `as_tibble.matrix()` must have unique column names if `.name_repair` is omitt
#> Using compatibility `.name_repair`.
#> # A tibble: 21 x 12
#>    attribute lvl        par      mean     sd `CI-5%` `CI-95%` sig   model error
#>    <chr>     <chr>      <chr>   <dbl>  <dbl>   <dbl>    <dbl> <lgl> <chr> <chr>
#>  1 <NA>      <NA>       int    -3.24   0.519  -3.50    -2.76  TRUE  VD-c~ EV1
#>  2 Brand     BlueBell   Bran~ -0.669  0.142  -0.859   -0.496  TRUE  VD-c~ EV1
#>  3 Brand     BlueBunny  Bran~ -0.610  0.140  -0.810   -0.407  TRUE  VD-c~ EV1
#>  4 Brand     Breyers    Bran~ -0.0775 0.0936 -0.244    0.0654 FALSE VD-c~ EV1
#>  5 Brand     Dryers     Bran~ -0.528  0.120  -0.686   -0.336  TRUE  VD-c~ EV1
#>  6 Brand     HaagenDa   Bran~ -0.289  0.0863 -0.420   -0.141  TRUE  VD-c~ EV1
#>  7 Brand     Store      Bran~ -0.465  0.108  -0.625   -0.297  TRUE  VD-c~ EV1
#>  8 Flavor    ChocChip   Flav~ -0.344  0.140  -0.578   -0.109  TRUE  VD-c~ EV1
#>  9 Flavor    ChocDough  Flav~ -0.371  0.119  -0.550   -0.166  TRUE  VD-c~ EV1
#> 10 Flavor    CookieCream Flav~ -0.371 0.122  -0.563   -0.166  TRUE  VD-c~ EV1
#> # ... with 11 more rows, and 2 more variables: reference_lvl <chr>,
#> #   parameter <chr>
```

Corresponding demand predictions can be obtained using the `vd_dem_vdm` function. Here, we generate in-sample predictions: