SKIM

decision behavior experts

Vector Autoregressive Modeling of Longitudinal Sales Data using Simulated Populations Informed by Conjoint Experiments

Kevin Lattery VP Methodology & Innovation | Sawtooth Software Conference May 2022

Thanks for Allowing Talk

Testament to Sawtooth Software's support for HB and larger analytics community, beyond their own software

- Not representative of typical topic or directly relevant for Sawtooth Software
- We show a **custom HB model** with **conjoint** component
  - Sawtooth Software uses a different fixed HB model
  - Our model is a choice model with similarities to a conjoint simulator with softmax (or logit rule) $\frac{e^U}{\sum e^U}$
- Estimation in Stan (open source software for general HB modeling)
  - *Hamiltonian Monte Carlo* (Stan) more robust than *Gibbs Sampling* (Sawtooth Software)
    - Especially with many parameters
  - Stan faster using Linux + Arm chip + multi-threading (32 cores x 2 chains)
    - Amazon Web Services c6g.8xlarge with SKIM custom built image

SKIM

Objective : Use Longitudinal Sales Data and Conjoint for Better Predictions

Conjoint or Real World?

- Our clients (like many) trust their real world data more than survey data.
  - Want us to use real world data to develop predictive model.
- They also believe (along with academics) consumers in **real** world behave somewhat differently than the **experimental** (conjoint survey) world.
  - Adjusting scale factor is one step (tuning conjoint results)

> *However, there is no reason to expect the alternative specific constants and the scale factor to be the same for stated preference data [conjoint] as for revealed-preference data [sales data]. These parameters reflect the impact of unobserved factors, which are necessarily different in real choice situations than hypothetical survey situations.*
>
> Kenneth Train
> *Discrete Choice Methods with Simulation (March 2002)*
> *Pg 177, section 7.2*

SKIM

Vector Autoregression Often Helps Predictions

Our clients (like many) have their own forecasting models

- Models based on **aggregate** sales data for their products and competitive products over time
  - 2 – 5 years of data (monthly and some weekly)
  - Also have data like price, distribution, etc.

- Client models use previous sales to predict future sales, Vector Autoregression (VAR)
  - Previous sales are a great predictor of future sales
    - Difficult for non-VAR model to perform as well as VAR model
    - Correlations of .9 or higher, even for 12 months
    - Mean Absolute Error for family rolled-up SKUs about .15% - .30%

Vector Autoregression (VAR): Use Shares to Predict Shares

- **Vector**: At any specific time we have sales data for multiple items (100-200 in our case)
- **Autoregression**: use previous sales to forecast new sales

Previous Shares

SKU	Time 1	Time 2	Time 3	Time 4	Time 5
1	20%	18%	17%	17%	16%
2	30%	31%	31%	32%	32%
3	10%	11%	12%	12%	13%
4	15%	15%	15%	16%	15%
5	25%	25%	25%	23%	24%

Future Shares

Future 1	Future 2
?	?
?	?
?	?
?	?
?	?

Time Lag

Optional Additional Exogenous Variables

Price

Distribution

…

Our Goal

- Model choices (VAR is linear)
- Framework should be a microlevel model of individual consumer dynamics/choices*
  - Respondent level conjoint is example of microlevel model
  - Model tradeoffs and detailed dynamics

Can we somehow combine conjoint experiments with real world longitudinal data into a VAR + Conjoint model?

And can we get even more accurate predictions than either alone?

* Estimating Disaggregate Data using Aggregate Data Through Augmentation of Individual Choice JMR Nov 2007

Modeling Real World Changes with Simulated Shoppers

Multinomial Probit (MNP): Importance of ε_i for Sourcing Dynamics

$$
\begin{aligned}
U_1 &= \text{Int}_1 + \beta\text{price}_1 * \log(\text{price}_1) + \beta\text{dist}_1 * \log(\text{dist}_1) + \varepsilon_1 \\
U_2 &= \text{Int}_2 + \beta\text{price}_2 * \log(\text{price}_2) + \beta\text{dist}_2 * \log(\text{dist}_2) + \varepsilon_2 \\
&\dots. \\
U_n &= \underbrace{\text{Int}_n + \beta\text{price}_n * \log(\text{price}_n) + \beta\text{dist}_n * \log(\text{dist}_n)}_{\mu_i} + \varepsilon_n
\end{aligned}
$$

Where $\{\varepsilon_i\} \sim \text{MVN}(0, \Sigma)$

Not IID, covariance Σ controls sourcing

$+ \varepsilon_i$ $\quad \{\mu_i + \varepsilon_i\} \sim \text{MVN}(\mu_i, \Sigma)$

The alternative is chosen with the highest total utility among $U_i = \{\mu_i + \varepsilon_i\}$ →

Prob of choosing has
No closed form
Must simulate to get probabilities from (aggregate) utilities

SKIM

MNP with Smoothed Accept-Reject (A-R) Simulator

1) Simulate a population with utilities representing unknown error **$\{\varepsilon_i\} \sim$ MVN (μ=0, Σ)**
2) For each simulated shopper in 1)
   a) add the known/global part μ_i to their **ε_i**
   b) **compute predictions for simulated shopper**
3) Average the predictions

Kenneth Train:
Discrete Choice Methods with Simulation

2b) Can compute predictions w/many methods.
First choice is most simple.

Smoothed A-R simulator can use scaled softmax (logit rule) with λ in (0,1] = 1:

$$\frac{e^{U/\lambda}}{\sum e^{U/\lambda}}$$

Requires less sim population than first choice
Avoids small items getting 0 choice

SKIM

A-R MNP Simulator ≈ Conjoint Simulator

In a respondent level conjoint simulator we can think of utilities as a global μ plus respondent level deviations ε_i

- Compute the mean vector μ across typical conjoint respondent utilities
- Subtract μ from each respondent's utilities
    - Result is conjoint respondent's deviations from the mean $\varepsilon_i \sim \text{MVN}(\mu=0, \Sigma)$

To simulate, add μ to each row of ε_i (call this U)

- Prediction for each respondent is $\frac{e^{Ux}}{\sum e^{Ux}}$

Smoothed Accept-Reject simulator behaves like conjoint simulator and enables us to build microlevel models of choice

Our simulator differs from conjoint because it varies μ over time based on a specific time period

SKIM

Multinomial Probit For Any Two Periods

What do Simulated Shoppers buy in this Period (Lag)?

What do Simulated Shoppers Buy in this Period (Forecast)?

	SKU1	SKU2	…	SKUn

Simulated Shopper Utilities

ε_i ~MVN (μ=0, Σ)

For each row in ε_i add

$\mu_{i,lag}$ = **$int_i@lag$** + $\beta price_i$ * log($price_i@lag$) + $\beta dist_i$ * log($dist_i@lag$)

For each row in ε_i add

$\mu_{i,new}$ = **$int_i@new_i$** + $\beta price_i$ * log($price_i@new$) + $\beta dist_i$ * log($dist_i@new$)

Reminder:
Exact same errors ε_i (simulated shoppers)
Lag and forecast have different mean vectors $\boldsymbol{\mu}_i$ that we add to each simulated shopper in ε_i

Difficulty: Single **intercept** term for each sku does not work well, as it likely varies over time. We will model intercept autoregressively

SKIM

Use Known Share of Lag Period to Drop Intercept and Simplify

What do Simulated Shoppers buy in **Lag** Period?

$$\mu_{i,lag} = int_i@lag + \beta price_i * \log(price_i@lag) + \beta dist_i * \log(dist_i@lag)$$

$$\approx \log(lag\_share_i)$$

Implies

$$int_i@lag \approx \log(lag\_share_i) - \beta price_i * \log(price_i@lag) - \beta dist_i * \log(dist_i@lag)$$

	SKU1	SKU2	…	SKUn
	Simulated Shopper Utilities			
• •	$\varepsilon_i \sim MVN\ (\mu=0, \Sigma)$			

What do Simulated Shoppers Buy in **Forecast** Period?

$$\mu_{i,new} = int_i@lag + \beta trend_i * [\text{time diff}] + \beta price_i * \log(price_i@new) + \beta dist_i * \log(dist_i@new)$$

Sub for $int_i@lag$

$$\mu_{i,new} = \log(lag\_share_i) - \beta price_i * \log(price_i@lag) - \beta dist_i * \log(dist_i@lag) + \beta trend_i * [\text{gap periods}] + \beta price_i * \log(price_i@new) + \beta dist_i * \log(dist_i@new)$$

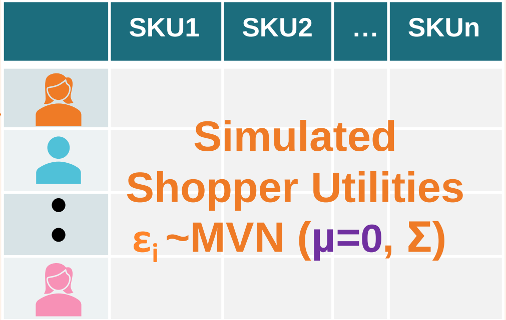
Final Model: Paired MNP and VAR of Differences

1

What do Simulated Shoppers buy in **Lag** Period?

For each row in ε_i add

$\mu[sku_i]_{lag}$ = $\log(share[sku_i]_{lag})$



What do Simulated Shoppers Buy in **Forecast** Period?

For each row in ε_i add

$\mu[sku_i]_{new}$ = $\log(share[sku_i]_{lag}) + \beta trend_i * [\text{time diff}] + \beta price_i * [\log(price_i@new) - \log(price_i@lag)] + \beta dist_i * [\log(dist_i@new) - \log(dist_i@lag)]$

2

Fit difference in log(shares) for simulated vs observed:

$\log(Sim_{Fore}) - \log(Sim_{Lag}) \sim \log(Obs_{Fore}) - \log(Obs_{Lag})$

$\log(Sim_{Fore}) - \log(Sim_{Lag}) + \log(Obs_{Lag}) \sim \log(Obs_{Fore})$

$\text{Softmax}(\log(Sim_{Fore}) - \log(Sim_{Lag}) + \log(Obs_{Lag})) \sim Obs_{Fore}$

Final Model: Conceptual Explanation

1

What do Simulated Shoppers buy in **Lag** Period?

Calibrate Shoppers Approximately to Shares in Lag Period

	SKU1	SKU2	…	SKUn
• •				

Simulated Shopper Utilities

$\varepsilon_i \sim MVN(\mu=0, \Sigma)$

What do Simulated Shoppers Buy in **Forecast** Period?

Lag Calibrated Shoppers + Changes from Lag Period

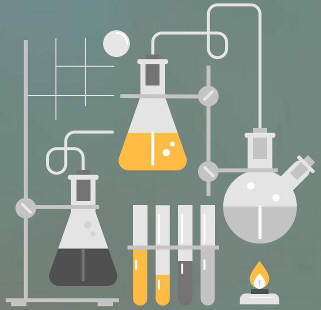
2

Adjust Forecast Simulations for Error in Calibration to Lag Period

$$\log(Sim_{Fore}) + \{\log(Obs_{Lag}) - \log(Sim_{Lag})\} \sim \log(Obs_{Fore})$$

	SKU1	SKU2	…	SKUn
	Simulated Shopper Utilities $\varepsilon_i \sim MVN\ (\mu=0,\ \Sigma)$			

ε_i as Microlevel of Consumer Dynamics: Fusion with Conjoint Experiments



Why Conjoint can Complement Real World Data

Stated-preference data [conjoint] complement revealed-preference data [sales data].
…
The advantage of stated-preference data is that the experiments can be designed to contain as much variation in each attribute as the researcher thinks is appropriate. While there might be little price variation over suppliers in the real world, the suppliers that are described in the experiments can be given sufficiently different prices to allow precise estimation. Attributes can be varied over respondents and over experiments for each respondent. This degree of variation contrasts with market data, where often the same products are available to all customers, such that there is no variation over customers in the attributes of products. Importantly, for products that have never been offered before, or for new attributes of old products, stated-preference data allow estimation of choice models when revealed-preference data do not exist.

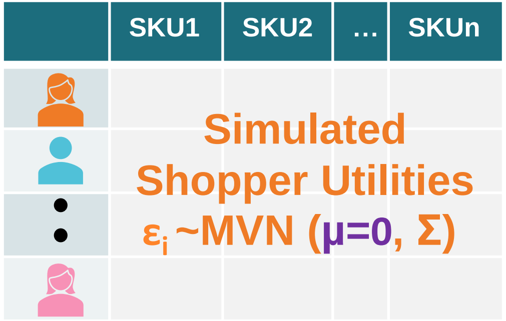
Kenneth Train
Discrete Choice Methods with Simulation
Pg 175, section 7.2

* Text in this color added (Kevin's interpretation)

SKIM

How to Estimate Σ?

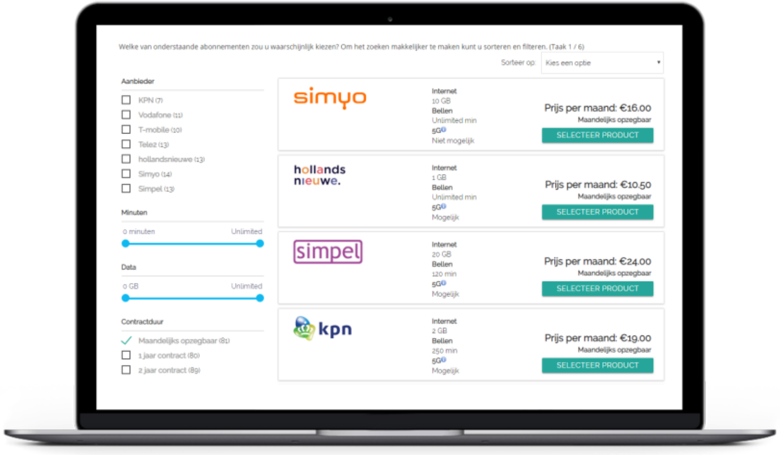
- Our microlevel choice model depends upon generating simulated shoppers from **MVN (μ=0, Σ)**
- With our specific data and hundreds of SKUs, estimating **Σ** **<u>with just sales data</u>** was problematic
  - ➢ Sourcing results were close to IIA (points or draws)
    - The posterior of correlation implied by **Σ** varies around 0 (with a range that depends on prior)
  - ➢ With many products, estimating **Σ** is one of the known difficulties for MNP



SKIM

Conjoint Experiments Can Inform Σ

- Conjoint Experiments allow us to observe changes at the respondent level to different stimuli in a designed experiment
- Respondents move from one SKU to another because of changes in price or product availability.
  - ➢ This **respondent level sourcing** information allows us to estimate the covariance of utilities **Σ** for our conjoint population.



- We conducted 5 different conjoint studies with overlapping SKUs in the same category
  - ➢ Total of 14,341 respondents

SKIM

Linking Simulated Shoppers and Conjoint Respondents

MVN (α, Σ*)

	SKU1	SKU2	…	SKUn

Conjoint Respondent Utilities

Simulated shopper utilities in real world are **similar** to our conjoint respondents

$\varepsilon_i \sim$ MVN (μ=0, Σ)

	SKU1	SKU2	…	SKUn

Simulated Shopper Utilities

Some Possible Relationships

1) Σ = k Σ*
2) Σ ~ Σ*
3) R = R* Where R is the correlation matrix
4) R ~ R* corresponding to covariance ∑ = DRD’

Academic Note:
This is an extension of the idea of building conjoint simulators that “*simulate from the upper level model*”.

SKIM

Simultaneous Estimation of Sales and Conjoint with Shared Correlation R

MVN
$(\alpha, \Sigma = D_1 R D_1')$

Resp Utilities

MNL Conjoint Tasks

Resp Predictions | **Resp Conjoint Choices**

$\underbrace{\text{Resp Predictions} \quad \text{Resp Conjoint Choices}}$

LogLikelihood Conjoint

MVN
$(0, \Sigma = D_2 R D_2')$

Sim Shopper Pop

VAR+MNP Model as Described

Sales Predictions | **Real World Share**

$\underbrace{\text{Sales Predictions} \quad \text{Real World Share}}$

wt * LogLikelihood Real World Task

$\underbrace{\text{LogLikelihood Conjoint} \quad \text{wt * LogLikelihood Real World Task}}$

Total LogLikelihood

SKIM

Use Correlation of Conjoint Utilities as Prior

$R \sim R^*$

MVN
$(\alpha, \Sigma = D_1 R^* D_1')$

Resp Utilities

MNL Conjoint Tasks

$\underbrace{\text{Resp Predictions} \quad \text{Resp Conjoint Choices}}$

LogLikelihood Conjoint

MVN
$(0, \Sigma = D_2 R D_2')$

Sim Shopper Pop

VAR+MNP Model as Described

$\underbrace{\text{Sales Predictions} \quad \text{Real World Share}}$

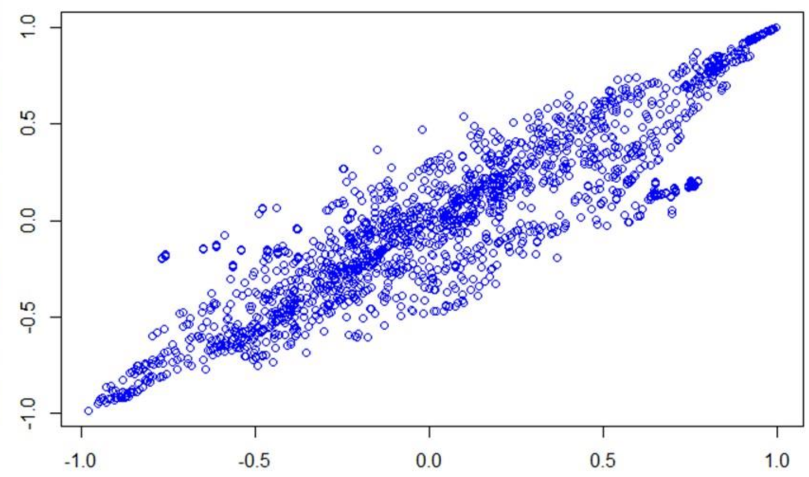
LogLikelihood Real World Task

No more combined loglikelihood or weighting
Much faster to estimate than both together

SKIM

Correlation implied by Posterior ∑ is Similar to Prior, Even when Conjoint is Vague Prior

Posterior Correlation from Estimated Covariance



Prior Correlation from Conjoint

Estimating ∑ without conjoint prior (= 0 correlation) gives posterior correlation near 0

$\Sigma \sim$ Wishart(DF = P+2 , Scale = σ_{diag} * prior_corr * t(σ_{diag})/(P+2)), where

prior_corr is fixed prior correlation from conjoint,
P is number of columns (or rows) in prior_corr,
σ_{diag} is diagonal of standard deviations (estimated parameter)
Note: in Stan we sample from Wishart using Bartlett Decomposition

SKIM

Results & Future Steps

History of Main Methods and Overall Success

Method	Covariance ∑	Sourcing	Holdout MAE SKUs Rolled to Family 20 = .2%		
			1Mo	3Mo	6mo
MNP Only	Estimated from Sales Data	IIA sourcing	18.3	26.1	40.4
MNP + Conjoint	Used Conjoint Respondents	Too much sourcing	18.0	27.8	42.2
MNP + Conjoint Joint	Sales and Conjoint Jointly	Best, but long time	18.2	23.9	31.9

Much better predictions with VAR

Simple VAR	None	Volume to Share IIA	12.7	19.9	25.5
MNP + VAR	Estimated from Sales Data	IIA sourcing	12.5	15.7	19.2
MNP + Conjoint + VAR	Used Conjoint Respondents	Too much sourcing	12.5	17.2	19.6
MNP + Conjoint Joint + VAR	Sales and Conjoint Jointly	Best, but long time	12.3	13.9	14.8
MNP + Conjoint Prior + VAR	Conjoint Prior	Similar to Above, Less Time	12.5	13.8	15.0

VAR model more accurate than Non-VAR model

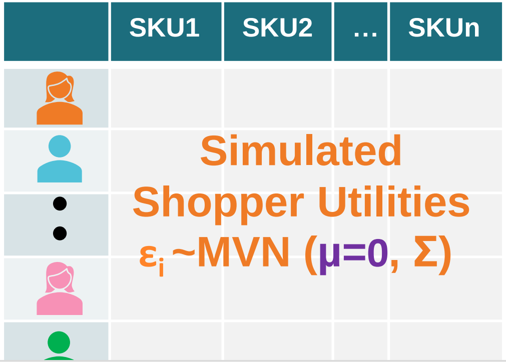
Infer intercept utility of sku from log(share[sku_i]$_{lag}$) and model differences from lag to forecast periods

SKIM

Difference of Obs ~ Difference of Simulated

What do Simulated Shoppers buy in **Lag** Period?

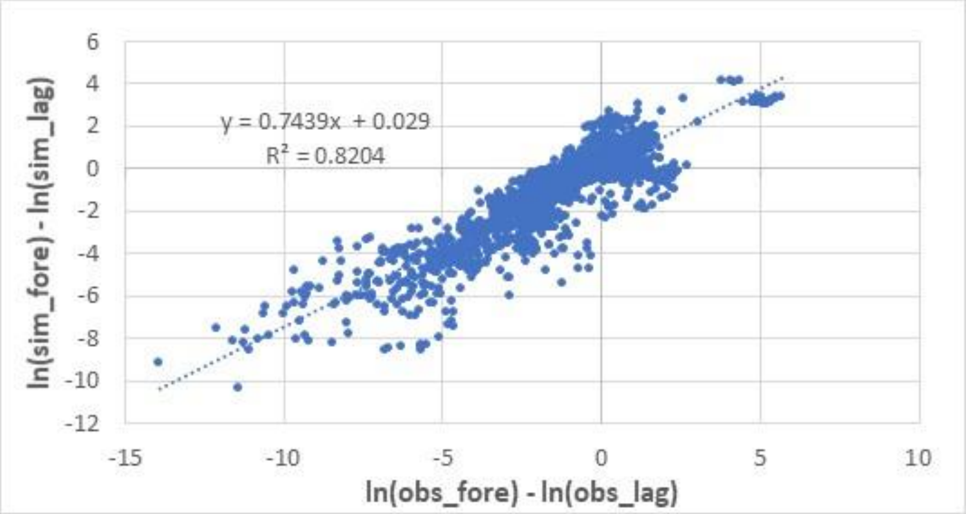
Calibrate Shoppers Approximately to Shares in Lag Period



What do Simulated Shoppers Buy in **Forecast** Period?

Lag Calibrated Shoppers + Changes from Lag Period

lo



SKIM

Difference of Obs ~ Difference of Simulated

What do Simulated Shoppers buy in **Lag** Period?

Calibrate Shoppers Approximately to Shares in Lag Period

	SKU1	SKU2	…	SKUn

Simulated Shopper Utilities
ε_i ~MVN (μ=0, Σ)

What do Simulated Shoppers Buy in **Forecast** Period?

Lag Calibrated Shoppers + Changes from Lag Period

Adjust Forecast Simulations for Error in Calibration to Lag Period

$$\log(Sim_{Fore}) + \{\log(Obs_{Lag}) - \log(Sim_{Lag})\} \sim \log(Obs_{Fore})$$

What if Sim_{Lag} is calibrated exactly to Obs_{Lag}?

SKIM

Post-Hoc Calibrate/Tune Intercept of Non-VAR

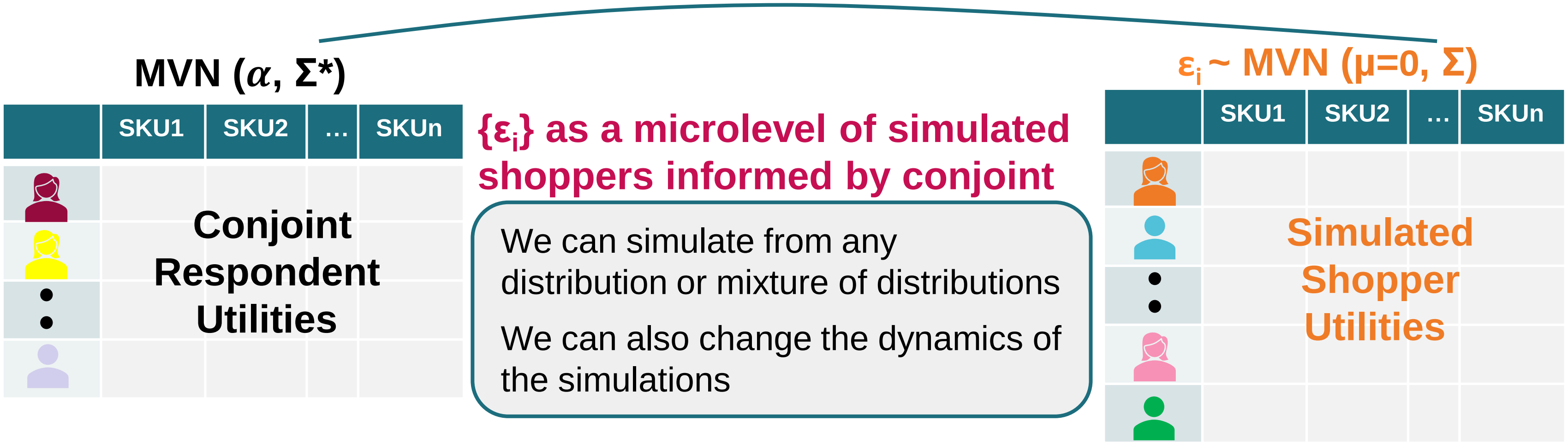
Holdout MAE
SKUs Rolled to Family
20 = .2%

Post-Hoc Tune Intercept

Method	Covariance ∑	Sourcing	1Mo	3Mo	6mo
MNP Only	Estimated from Sales Data	IIA sourcing	18.3	26.1	40.4
MNP + Conjoint	Used Conjoint Respondents	Too much sourcing	18.0	27.8	42.2
MNP + Conjoint Joint	Sales and Conjoint Jointly	Best, but long time	18.2	23.9	31.9
MNP Only + **Tune**	Estimated from Sales Data	IIA sourcing	12.7	18.9	22.2
MNP + Conjoint + **Tune**	Used Conjoint Respondents	Too much sourcing	12.9	19.2	23.1
MNP + Conjoint Joint + **Tune**	Sales and Conjoint Jointly	Best, but long time	12.6	17.5	19.9
Simple VAR	None	Volume to Share IIA	12.7	19.9	25.5
MNP + VAR	Estimated from Sales Data	IIA sourcing	12.5	15.7	19.2
MNP + Conjoint + VAR	Used Conjoint Respondents	Too much sourcing	12.5	17.2	19.6
MNP + Conjoint Joint + VAR	Sales and Conjoint Jointly	Best, but long time	12.3	13.9	14.8
MNP + Conjoint Prior + VAR	Conjoint Prior	Similar to Above, Less Time	12.5	13.8	15.0

SKIM

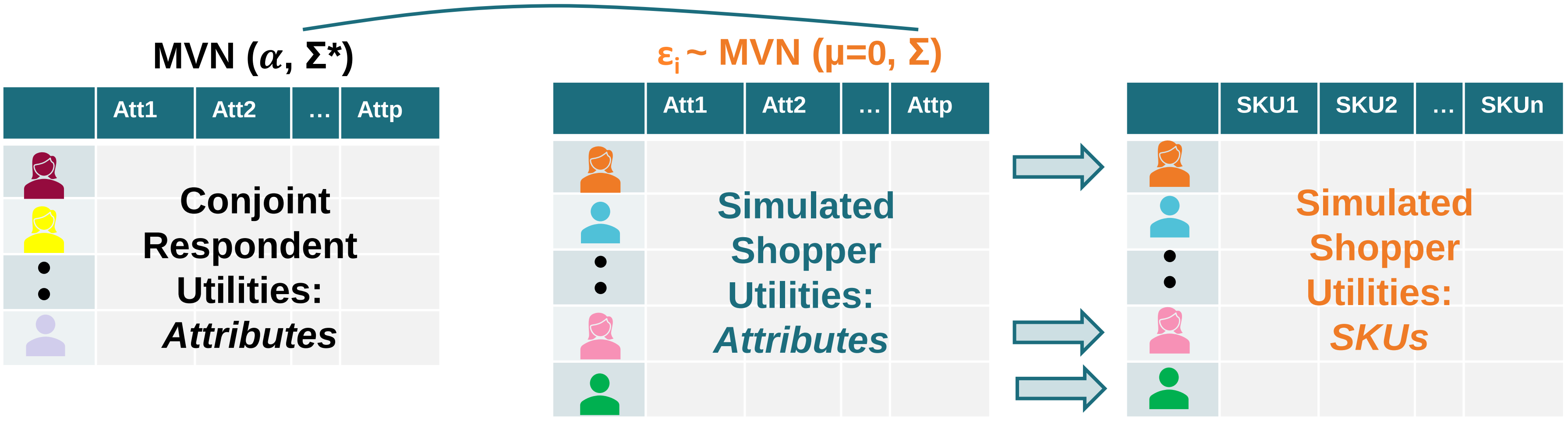
Data Fusion with Conjoint Utilities: an Extendable Framework



- **Conjoint provides information about our simulated shoppers**
- **Two-stage estimation using conjoint correlation as <u>prior</u> worked about as well as simultaneous estimation (in our studies)**

  Benefits of using conjoint as prior:
  - No need to weight sales data vs conjoint
  - Much quicker to estimate and update parameters when we get more data

Actual Extension Used Attributes and Compute SKUs



From each simulated shopper's attribute utilities $\varepsilon_{i,Att}$ we compute the corresponding utility of each SKU (based on its attribute definition), resulting in $\varepsilon_{i,SKU}$

SKIM

Conclusion: Experiments + Real World Data = Better Predictions

- **Conjoint Experiments can inform microlevel of simulated shoppers that help predict aggregate shares**

- **We described a framework that can be further expanded and modified**
- **Welcome more discussion on data fusion linking different data sources**

SKIM

Questions?

Kevin Lattery

email: k.lattery@skimgroup.com
www.skimgroup.com
@skimgroup

1) How many simulated respondents do you need?
2) What if you have real world SKUs without corresponding conjoint priors?
3) What are some of the other microlevel consumer dynamics you modeled?
4) Could we use something similar (but simpler) than this to fuse conjoint with sales data for just one time period (calibrate conjoint data to sales)?

SKIM

Appendix

Simultaneous Estimation Similar for "Static Calibration Data"

We can also set:
Sim Shopper Pop = Respondent Utilities

MVN
$(\alpha, \Sigma = D_2 R D_2')$

Resp Utilities

MNL Conjoint Tasks

Predictions **Resp Conjoint Choices**

LogLikelihood Conjoint

MVN
$(\alpha, \Sigma = D_1 R D_1)$

Sim Shopper Pop

MNL on Calibration Scenario

Mean Predictions **Agg Calibration Share**

wt * LogLikelihood CalibrationTask

Total LogLikelihood

SKIM