Big Value from Big Data

Sanjog Misra //
Charles H. Kellstadt Professor of Marketing
Chicago Booth

The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015

4.4 MILLION IT JOBS

will be created globally to support big data, with 1.9 million in the United States

Volume

SCALE OF DATA

40 ZETTABYTES

[43 TRILLION GIGABYTES]

of data will be created by 2020, an increase of 300 times from 2005

2020

2005

It's estimated that

2.5 QUINTILLION BYTES

[2.3 TRILLION GIGABYTES]

of data are created each day

6 BILLION PEOPLE

have cell phones

WORLD POPULATION: 7 BILLION

Most companies in the U.S. have at least

100 TERABYTES

[100,000 GIGABYTES]

of data stored

Velocity

ANALYSIS OF STREAMING DATA

The New York Stock Exchange captures

1 TB OF TRADE INFORMATION

during each trading session

Modern cars have close to

100 SENSORS

that monitor items such as fuel level and tire pressure

By 2016, it is projected there will be

18.9 BILLION NETWORK CONNECTIONS

– almost 2.5 connections per person on earth

Variety

DIFFERENT FORMS OF DATA

As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES

[161 BILLION GIGABYTES]

By 2014, it's anticipated there will be

420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO

are watched on YouTube each month

30 BILLION PIECES OF CONTENT

are shared on Facebook every month

400 MILLION TWEETS

are sent per day by about 200 million monthly active users

Veracity

UNCERTAINTY OF DATA

1 IN 3 BUSINESS LEADERS

don't trust the information they use to make decisions

Poor data quality costs the US economy around

$3.1 TRILLION A YEAR

27% OF RESPONDENTS

in one survey were unsure of how much of their data was inaccurate

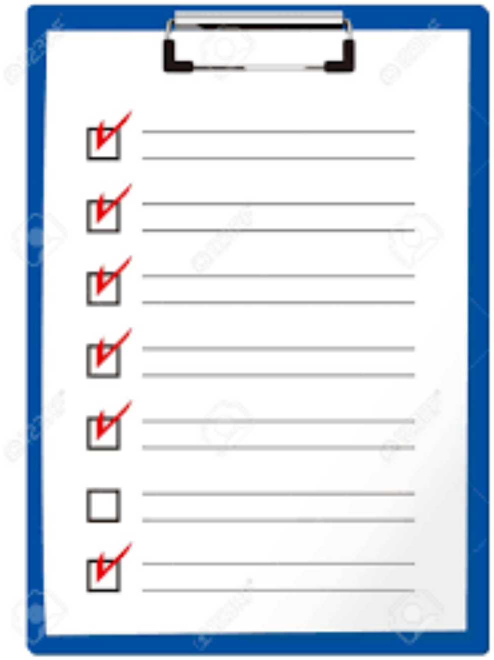
Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

IBM.

Big Data

Data is big anytime it makes you feel it is.

agenda

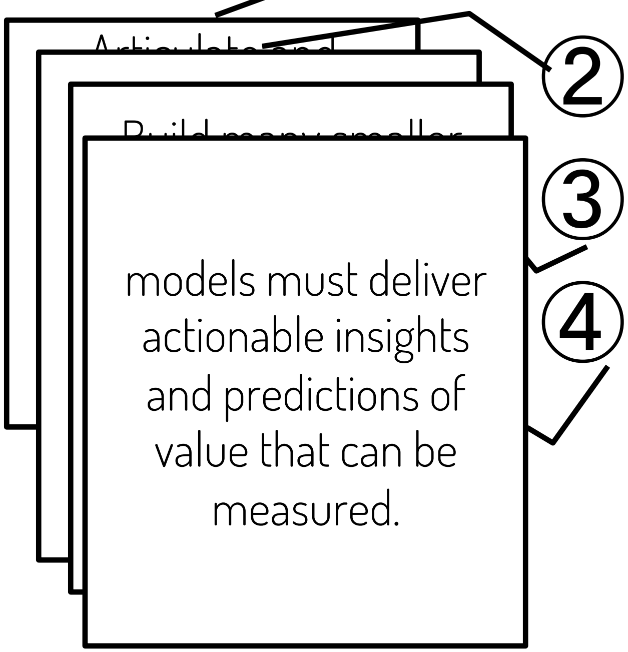


Some Ideas

Some Case Studies

Some Remarks

4 Ideas



1. Start at the end.
2. Think structurally.
3. Distribute Analytics.
4. Measure Predicted Value.

PROMOTION TARGETING

@ MGM RESORTS

Case #1

Gaming Industry

Gaming is big. Close to $50 billion.

Commercial casinos are multi-product firms providing bundles of entertainment, lodging, retail and gambling options to consumers

Like for any Marketing firm...

... Segmentation and targeting is critical

MGM Products

Resorts: Accommodations, Gaming, Spa, Retail, Entertainment, Dining

Vdara™ HOTEL & SPA

BELLAGIO®

SKYLOFTS AT MGM GRAND

MGM GRAND. LAS VEGAS

MANDALAY BAY® RESORT AND CASINO, LAS VEGAS

NEW YORK NEW YORK LAS VEGAS HOTEL & CASINO

Aria®

THE hotel AT MANDALAY BAY · LAS VEGAS

Excalibur HOTEL•CASINO•LAS VEGAS

MGM GRAND. DETROIT

Mirage® LAS VEGAS

GOLD STRIKE HOTEL & GAMBLING HALL

CITYCENTER®

THE SIGNATURE AT MGM GRAND®

LUXOR® LAS VEGAS

MGM 澳門 美高梅

Monte Carlo LAS VEGAS RESORT AND CASINO

Beau Rivage® RESORT & CASINO

CIRCUS CIRCUS. RENO

CIRCUS CIRCUS. LAS VEGAS

19 RESORTS

90% AVERAGE OCCUPANCY
(3.5 DAYS AVERAGE STAY)

$9 BILLION
ANNUAL REVENUE

62 THOUSAND
EMPLOYEES

$1.7 BILLION IN
ANNUAL TAXES

3 MILLION SQ FT OF
CONVENTION SPACE

65+ MILLION GUESTS
IN OUR DB

250+ FOOD & BEVERAGE
VENUES

47+ THOUSAND ROOMS
(41K IN VEGAS, 6K IN OTHER)

350+ RETAIL OUTLETS
(OVER 1M SQ FT OF SPACE)

End Goal

① Start at the end

MGM Resorts offers a variety of promotions to customers…

… and would like to avoid "**fleas**" …

i.e. Customers who avail promotions without generating value to the firm.

Reverse Engineer Decisions

② Think structurally.

Q: How is the data generated?

A: By consumer decisions.

A Model of Behavior

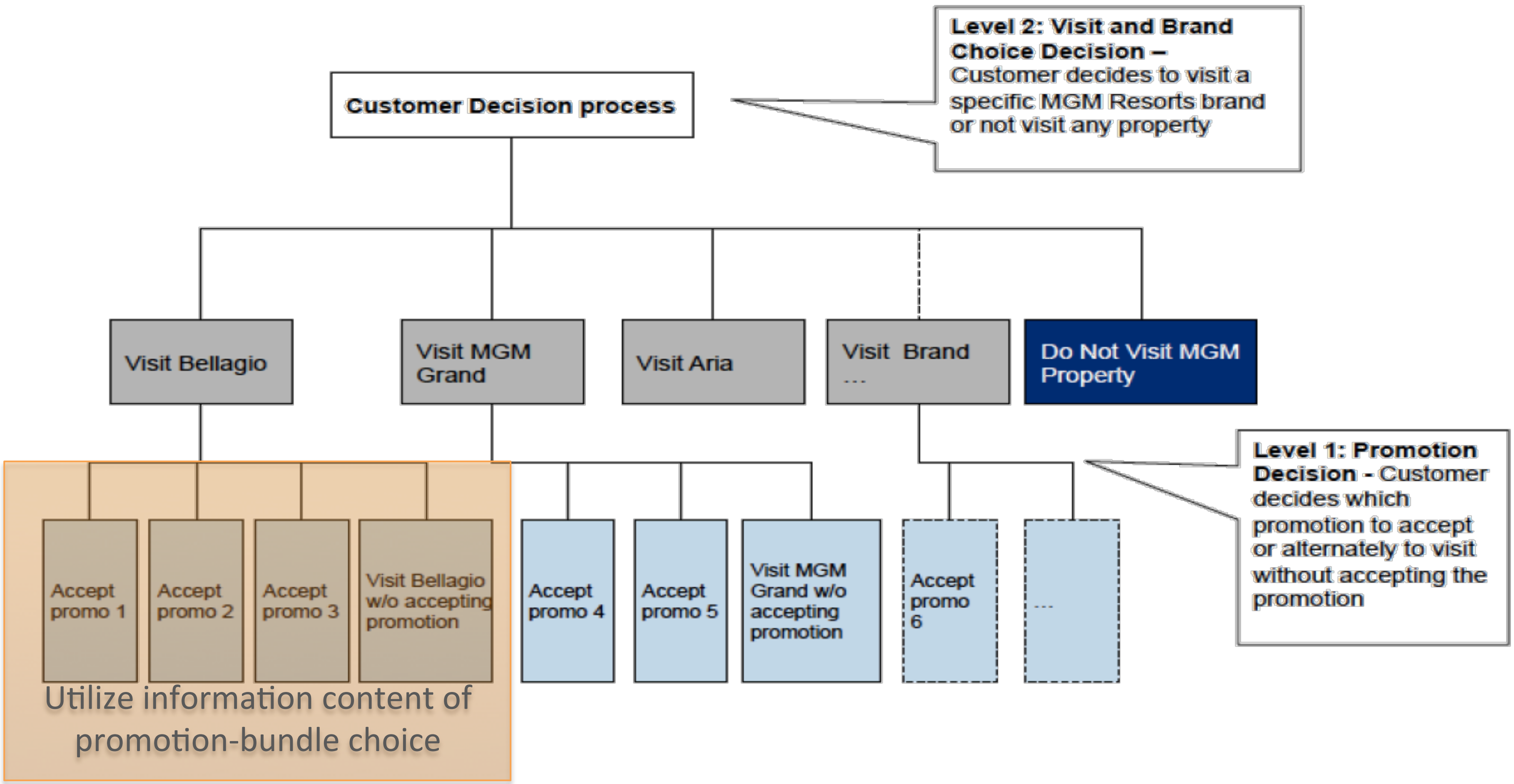
② Think structurally.

- Model the guest
  - Why visit?
  - Which promotion?
  - Which property?
  - How much to spend?

Y

Targeted promotions
(**X**)

Modeling Visitation and Spend



Reverse Engineer **Decisions**

② Think structurally.

Q: Is that it?

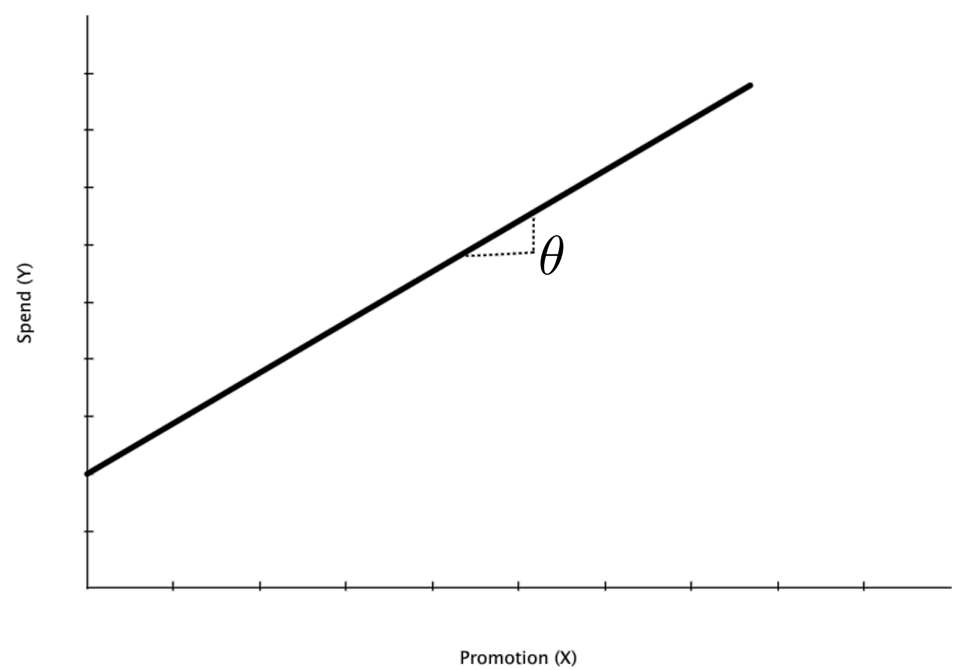
A: Nope.

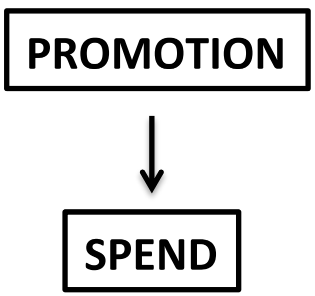
(Promotions are generated by history-dependent targeting)

Inferring Effects

② Think structurally.

Segmentation and ROI assessment requires understanding *causal* effect of promotions for each customer

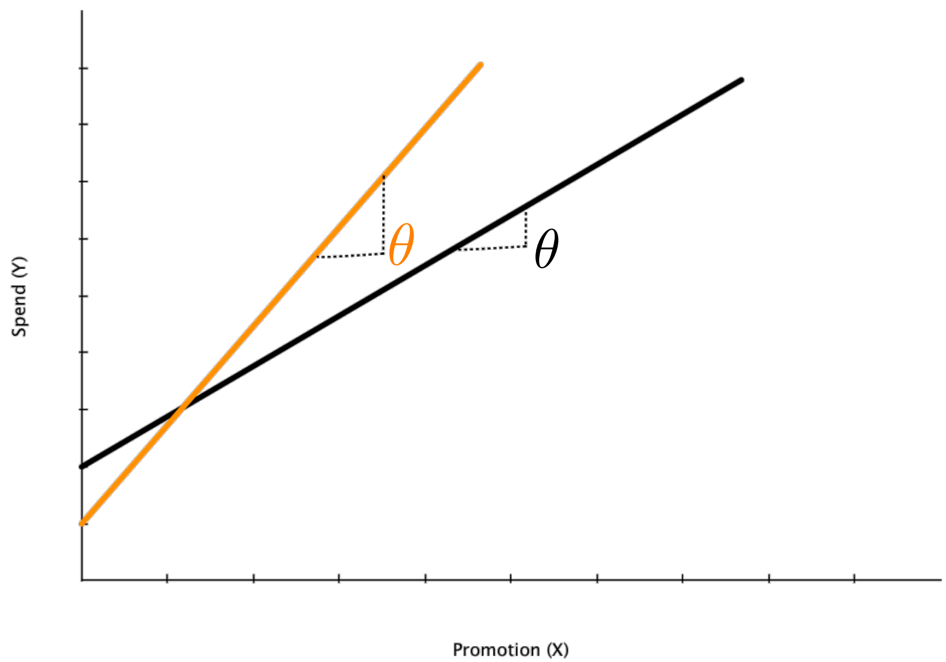


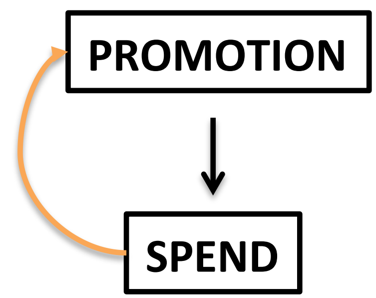


Endogeneity Concerns

② Think structurally.

Those who are likely to play more are systematically targeted more promotions. If uncorrected, *overstates* promotion effect





The Solution

② Think structurally.

Model MGM's Targeting rule

$$\Pr(play, promo) = \underbrace{\Pr(play|promo)}_{\text{Causal Effect of promotion on play}} \times \underbrace{\Pr(promo)}_{\text{Targeting Rule}}$$

- Propose a new method to exploit discontinuities in targeting rule in a likelihood framework
- Scalable, easy to use

Implementation Details

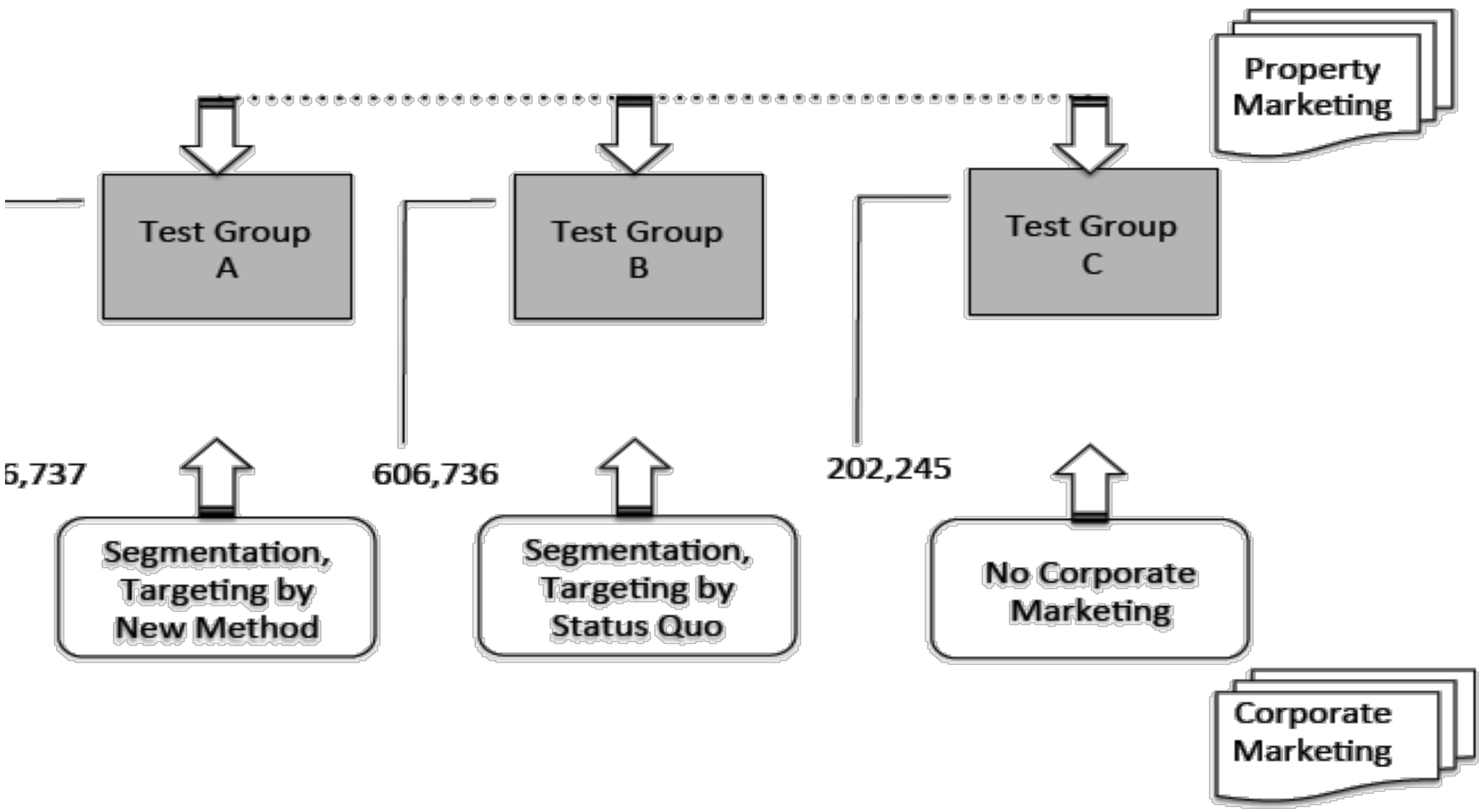
- Over **120+** models implemented
- Average of **180** variables in each model
- **20,000+** parameters
- New methods to account for measuring causal effects in this setting
- End-result:
  - property, promotion and person-specific predictions of promotion lift
  - correlation ≠ causality critique particularly acute

4) Measure Predicted Value.

Implementation Dashboards

4) Measure Predicted Value.

Experimental Test



4) Measure Predicted Value.

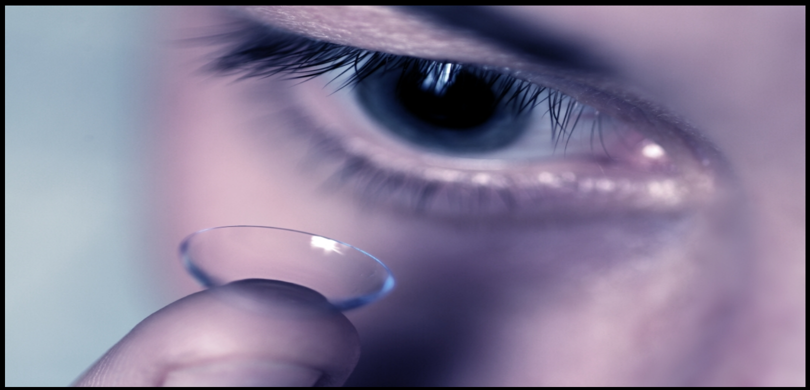
Test Results

	New	Status-Quo	Control
Adjusted Revenues	\$39.69M	\$38.97M	\$12.79M
Costs	\$14.42M	\$15.28M	\$5.05M
Margin	63.68%	60.79%	60.49%
Profit	\$25.28M	\$23.69M	\$7.74M
ΔProfit (A - B)	\$1.59M	—	—
Return on Investment (ROI)	\$2.75	\$2.55	\$2.53

ADJUSTED REV = ALL GAMING/NON-GAMING REVENUE – (*FREE PLAY* AND M LIFE POINT REDEMPTION)
COSTS = ALL COMP UTILIZATION

Overall **Impact**

- If spending same per campaign ($14.42M), over 4 campaigns per year, *incremental* revenues of **$11.7M** per year under new model
  - Interpret effect as lower bound, as further improvements possible if properties also shifted to model (ongoing)
- Other benefits
  - All properties agree on metric of value
  - Better cannibalization assessment
  - Consistency in targeting
- Changed focus of MGM to be more analytics driven.



SALESFORCE COMPENSATION
DYNAMICS

Case #2

Some Context

① Start at the end

This is a large medical devices company that sells primarily through a salesforce.

This is typical for most pharma companies (and others)

Data is aplenty.

End Goal

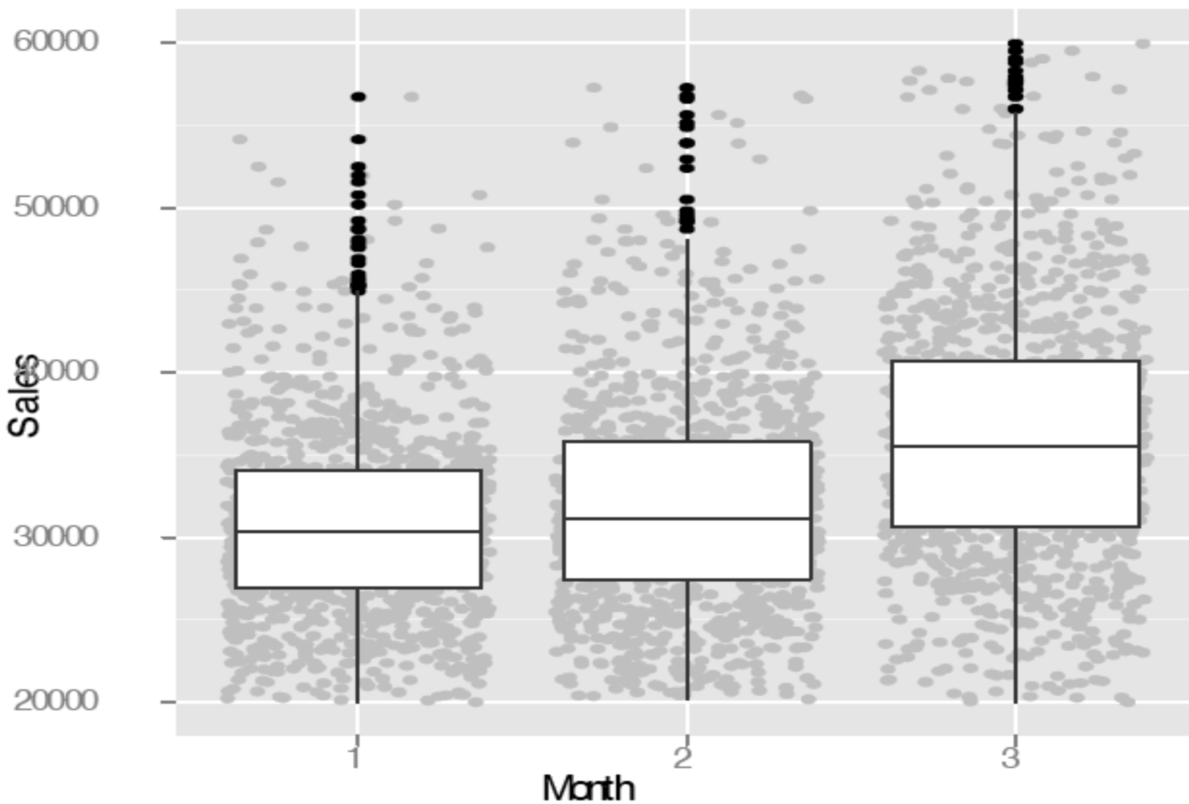
1 Start at the end

The firm's salesforce is the primary driver of sales…

… and they thought there was a problem.

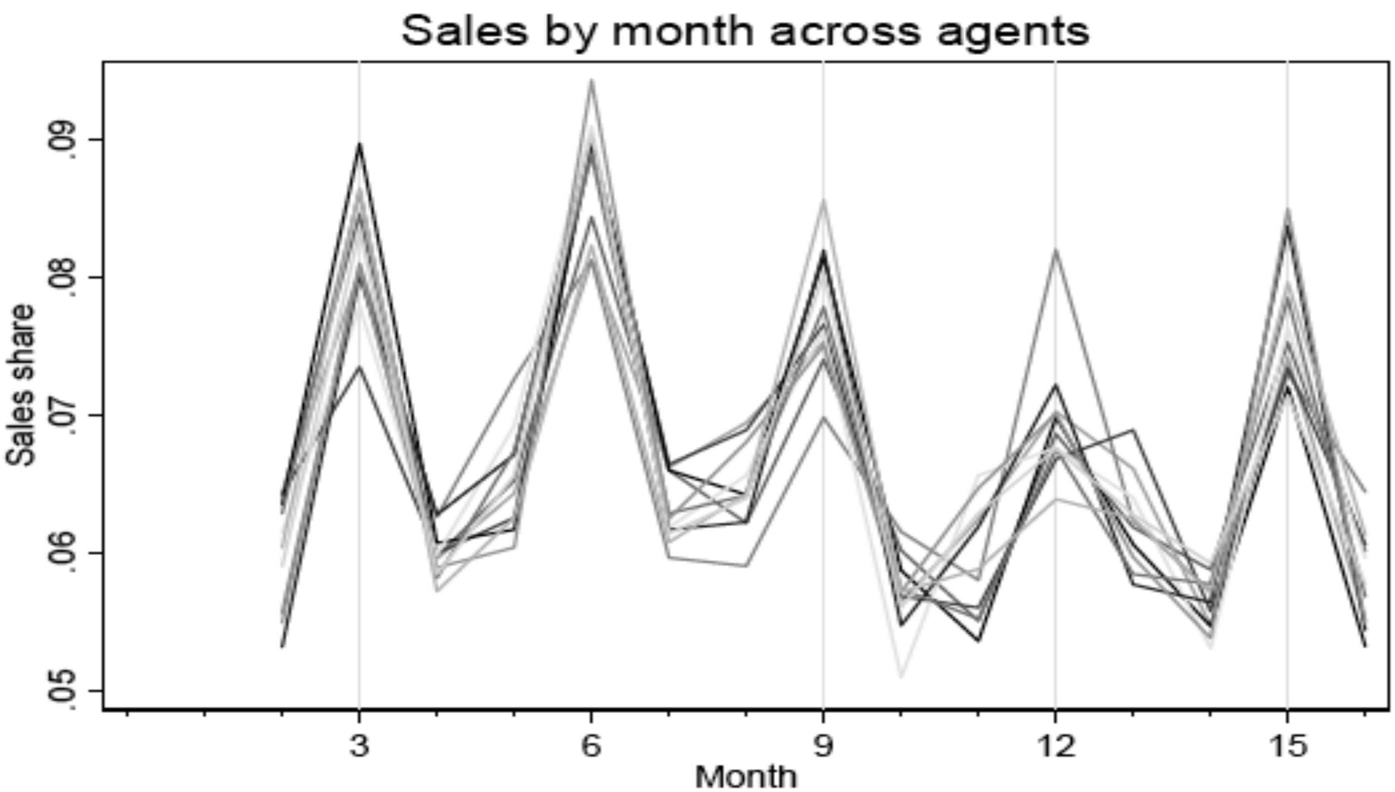
Within Quarter Sales

①Start at the end



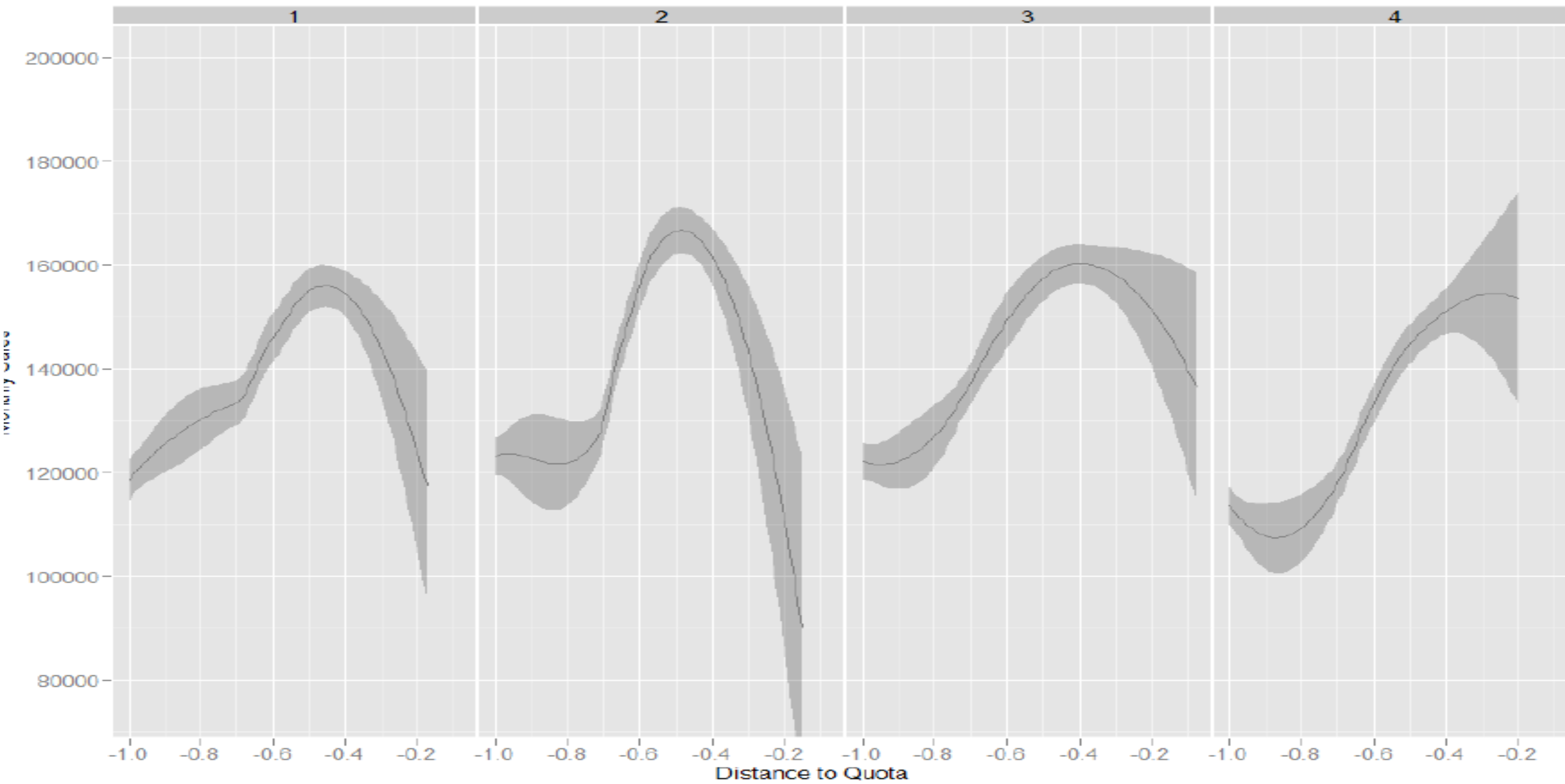
Sales Patterns

① Start at the end



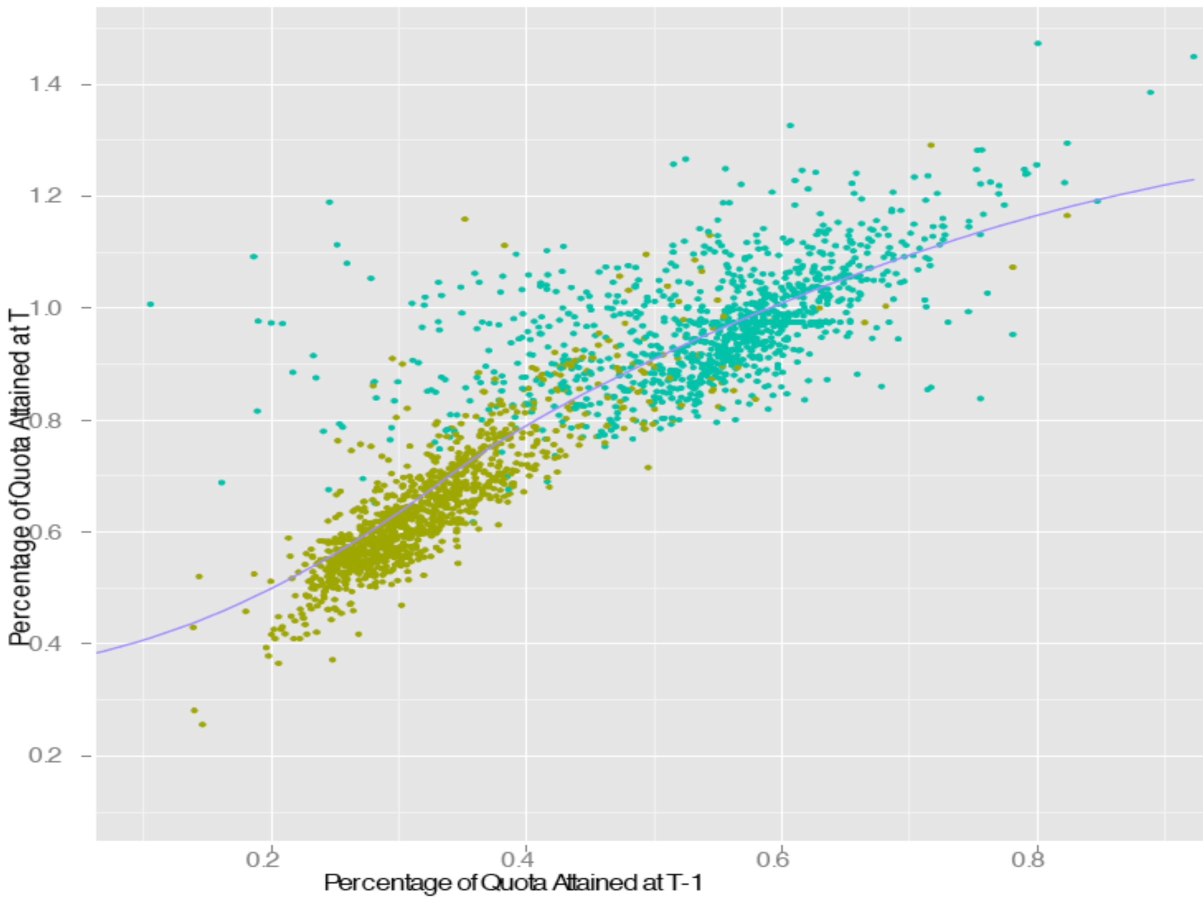
Effort Timing

① Start at the end



Effort Gaming

① Start at the end



End Goal

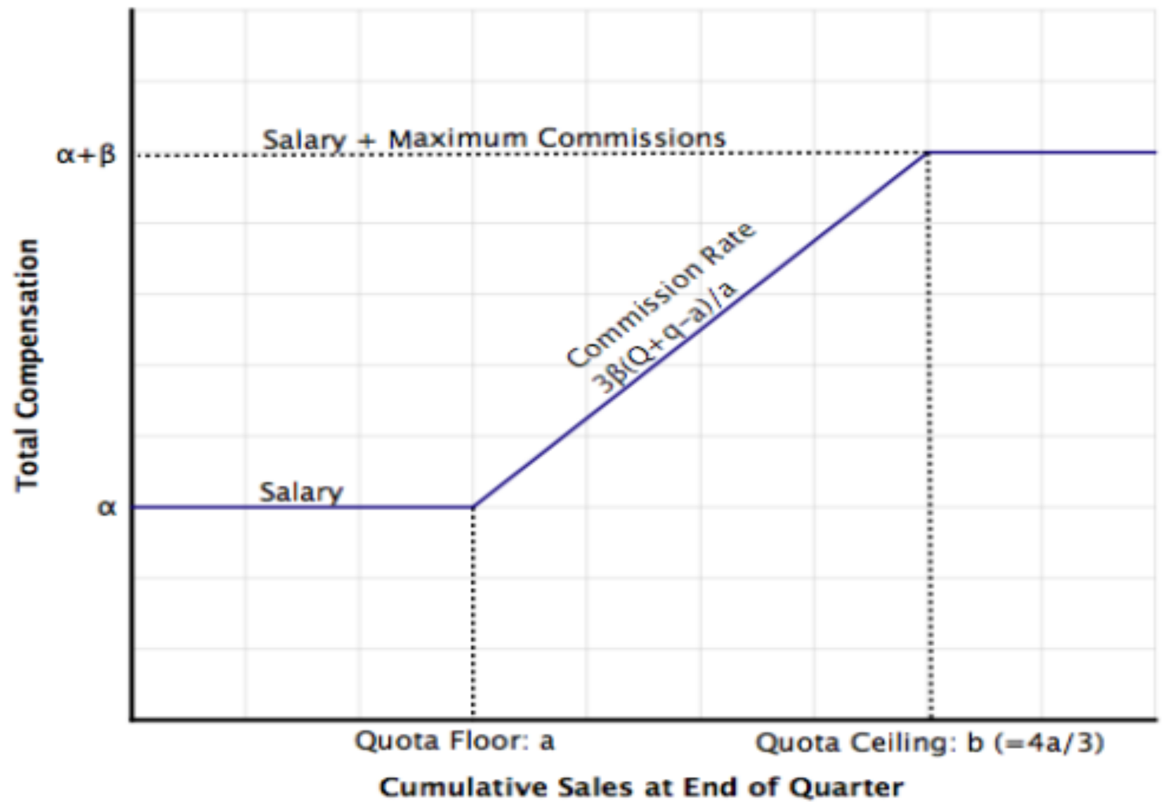
① Start at the end

We need to make salespeople put in effort in the right amount at the right time.

Easier said than done.

Compensation Plan

② Think structurally.



The Virtual Salesperson

2 Think structurally.

The Salesperson's thought process …

- I like money. I hate working.
- I like money. I hate uncertainty and risk.
- I will work if I make enough money.
- If it is not too risky to wait and see what happens. I wait.
- If I work hard you raise my quotas. I wont work hard.
- If I have no chance of being in the money. I give up.
- If I have made all the money I can make. I stop working.
- And yes. I like money.

Virtual Salesperson

2 Think structurally.

$$V\left(Q_t, a_t, N, \chi_t\right) = \max_{\chi_{t+1}, e} \left\{ \begin{array}{l} u\left(Q_t, a_t, N, \chi_t, e\right) + \\ +\rho \int_v \int_\varepsilon V\left(Q_{t+1} = 0, a_{t+1} = a\left(Q_t, q\left(\varepsilon_t, e\right), a_t, v_{t+1}\right), 1, \chi_{t+1}\right) \\ \qquad \times f\left(\varepsilon_t\right) \phi\left(v_{t+1}\right) d\varepsilon_t dv_{t+1} \end{array} \right.$$

The value from exerting effort today

=

The rewards you get today

+

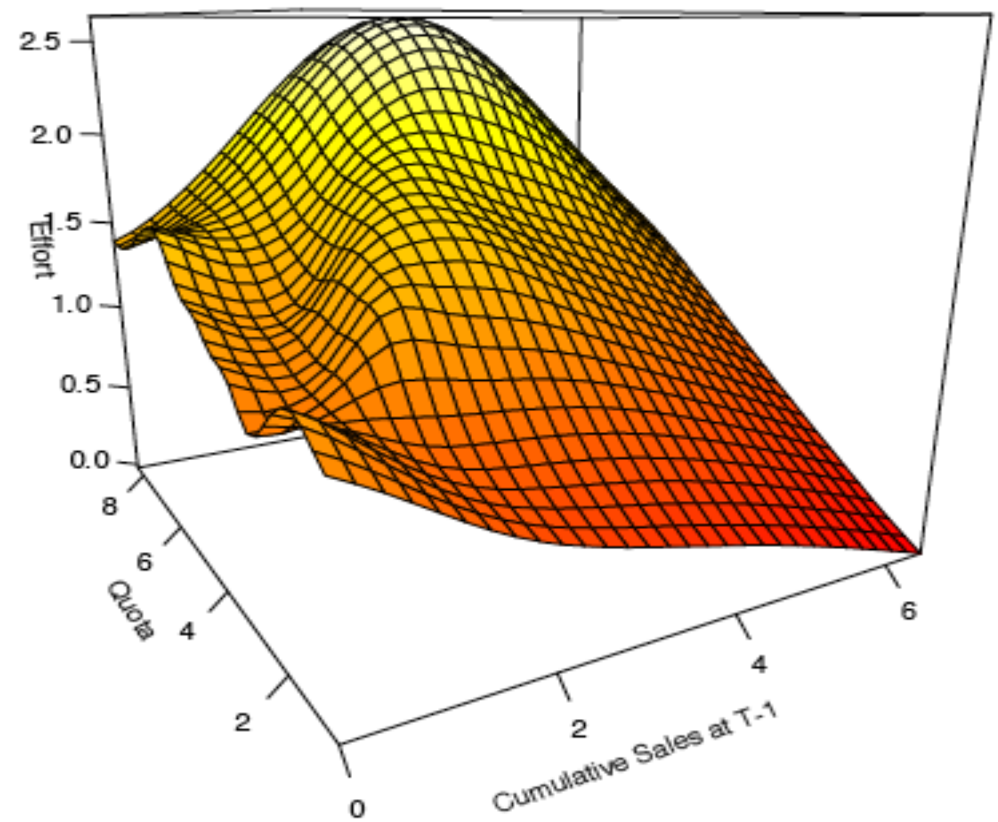
The rewards you get tomorrow
because of effort today

The Virtual Salesperson

② Think structurally.

The Virtual Salesperson

② Think structurally.



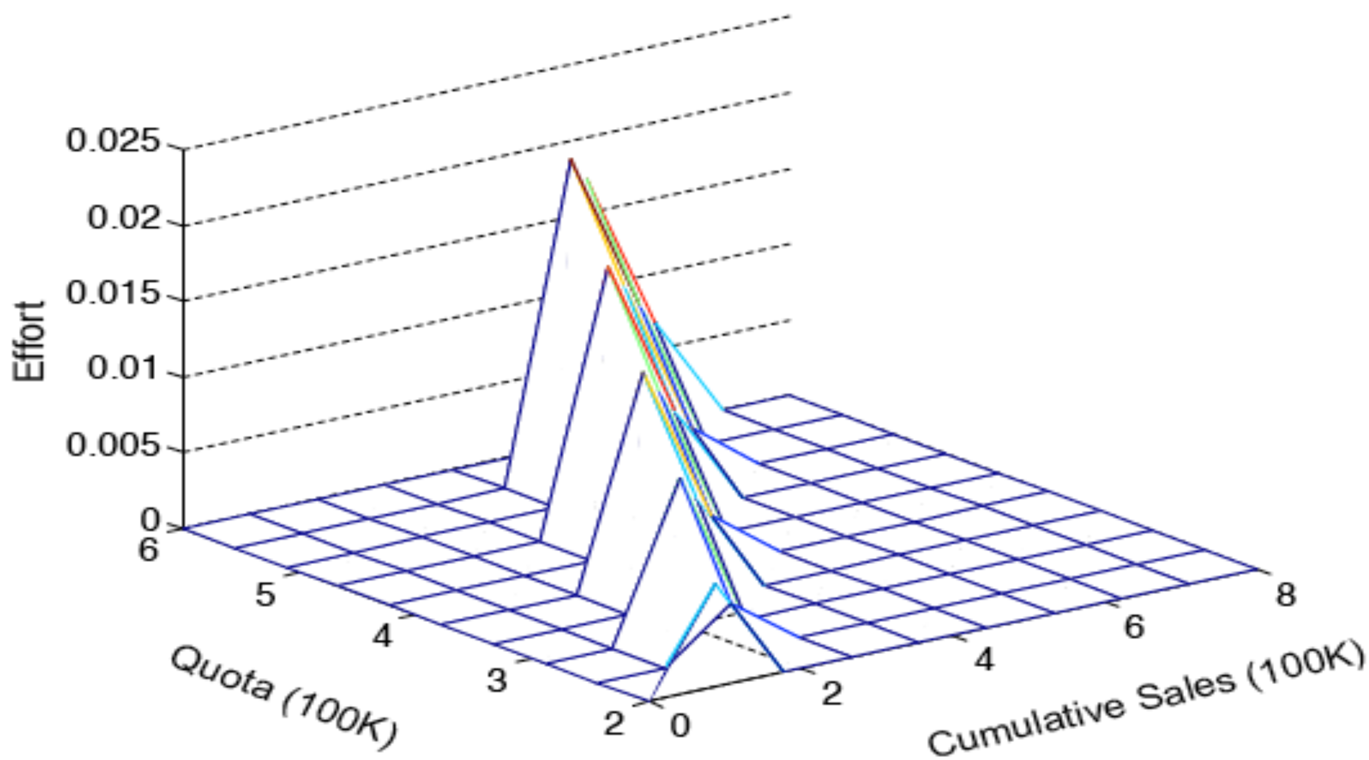
The Virtual Salesperson

② Think structurally.

$$V\left(Q_t, a_t, N, \chi_t\right) = \max_{\chi_{t+1}, e} \left\{ \begin{array}{l} u\left(Q_t, a_t, N, \chi_t, e\right) + \\ +\rho \int_v \int_\varepsilon V\left(Q_{t+1} = 0, a_{t+1} = a\left(Q_t, q\left(\varepsilon_t, e\right), a_t, v_{t+1}\right), 1, \chi_{t+1}\right) \\ \qquad\qquad \times f\left(\varepsilon_t\right) \phi\left(v_{t+1}\right) d\varepsilon_t dv_{t+1} \end{array} \right\}$$

The Virtual Salesperson

② Think structurally.

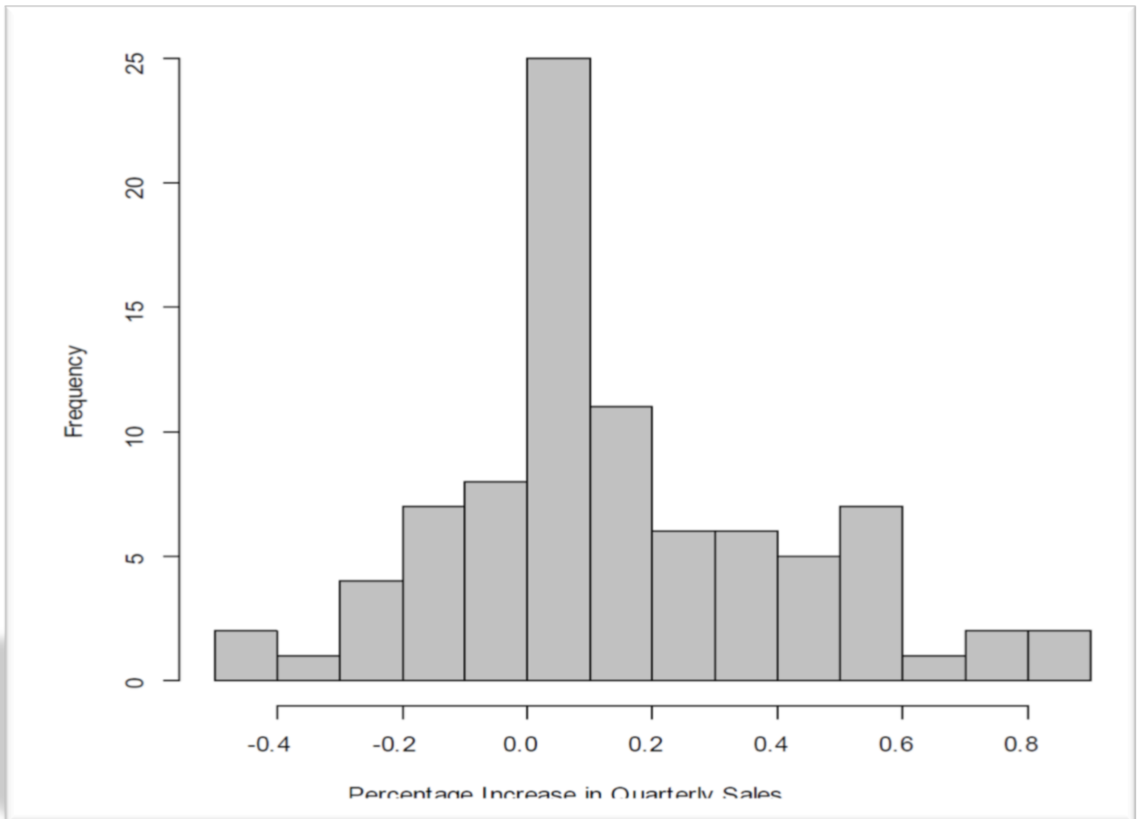


The New Plan

- A feasible set defined by cultural, legal and infrastructural constraints at firm recommended
- Firm chose one of recommended plans
- We predict sales under chosen feasible plan would increase by about 8.6% and profits by 5.2%
- New plan introduced in January, 2009
- Most changes deemed profit enhancing are incorporated (cannot reveal exact details)

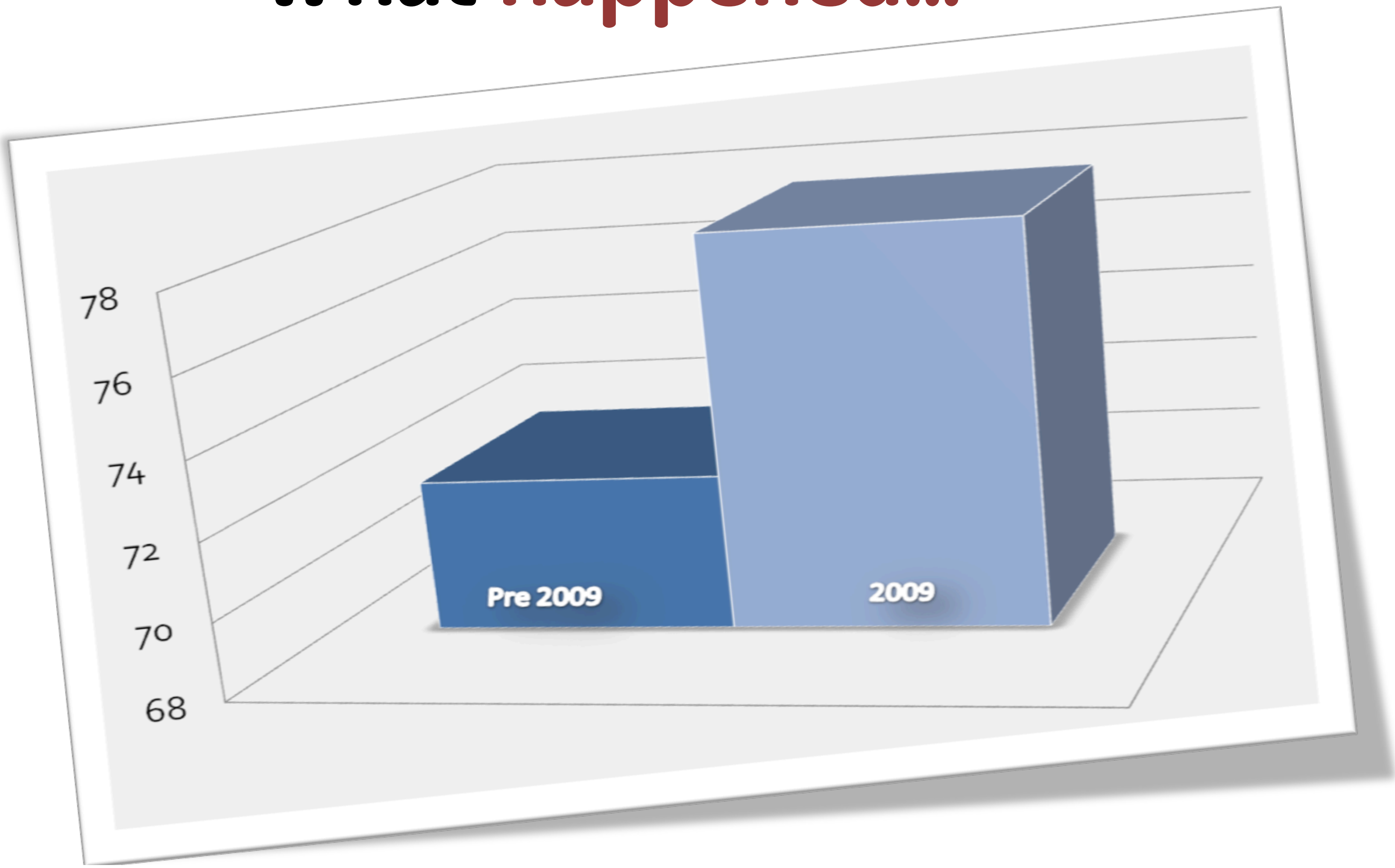
4) Measure Predicted Value.

What we expected



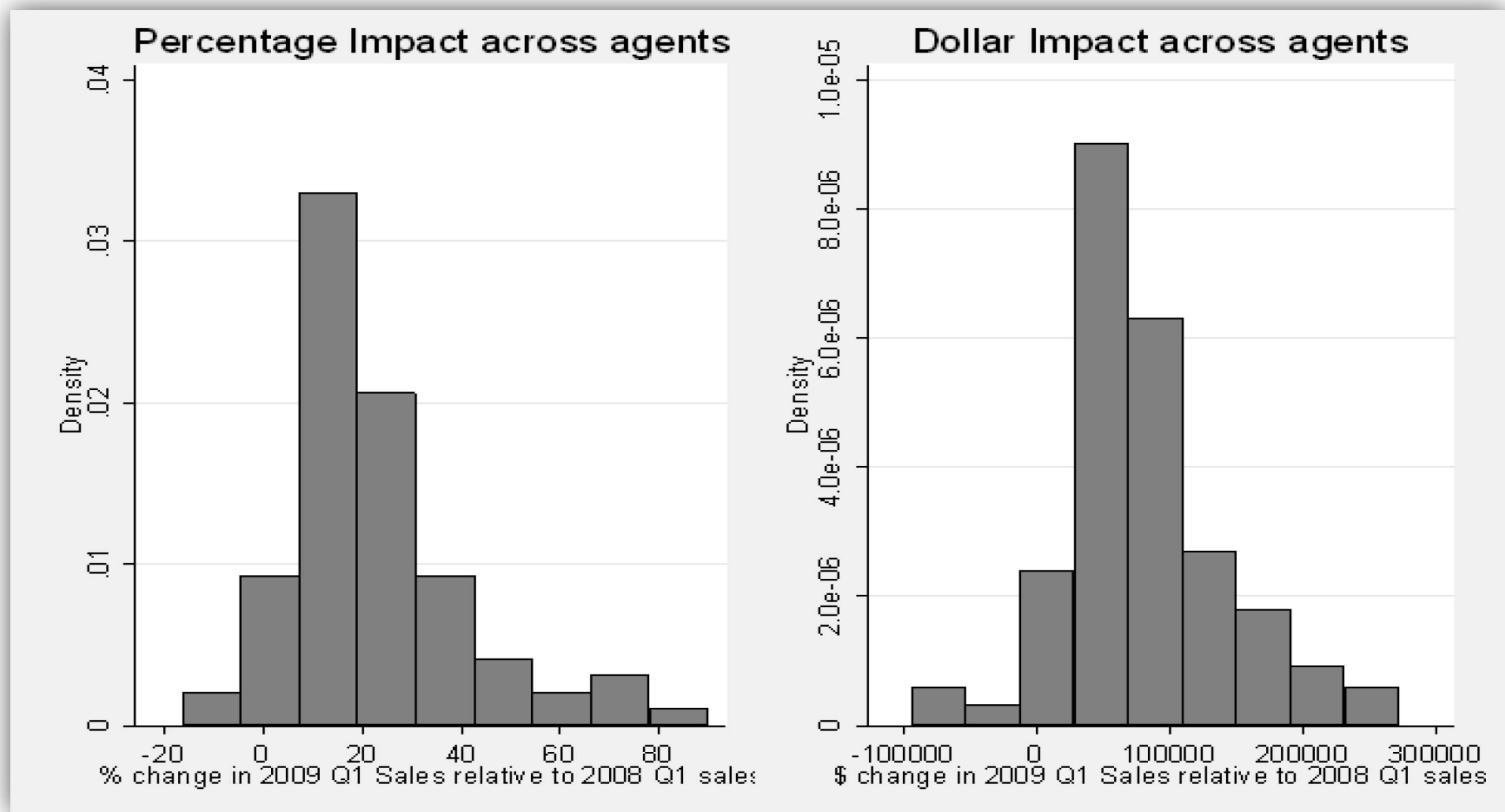
④ Measure Predicted Value.

What happened...



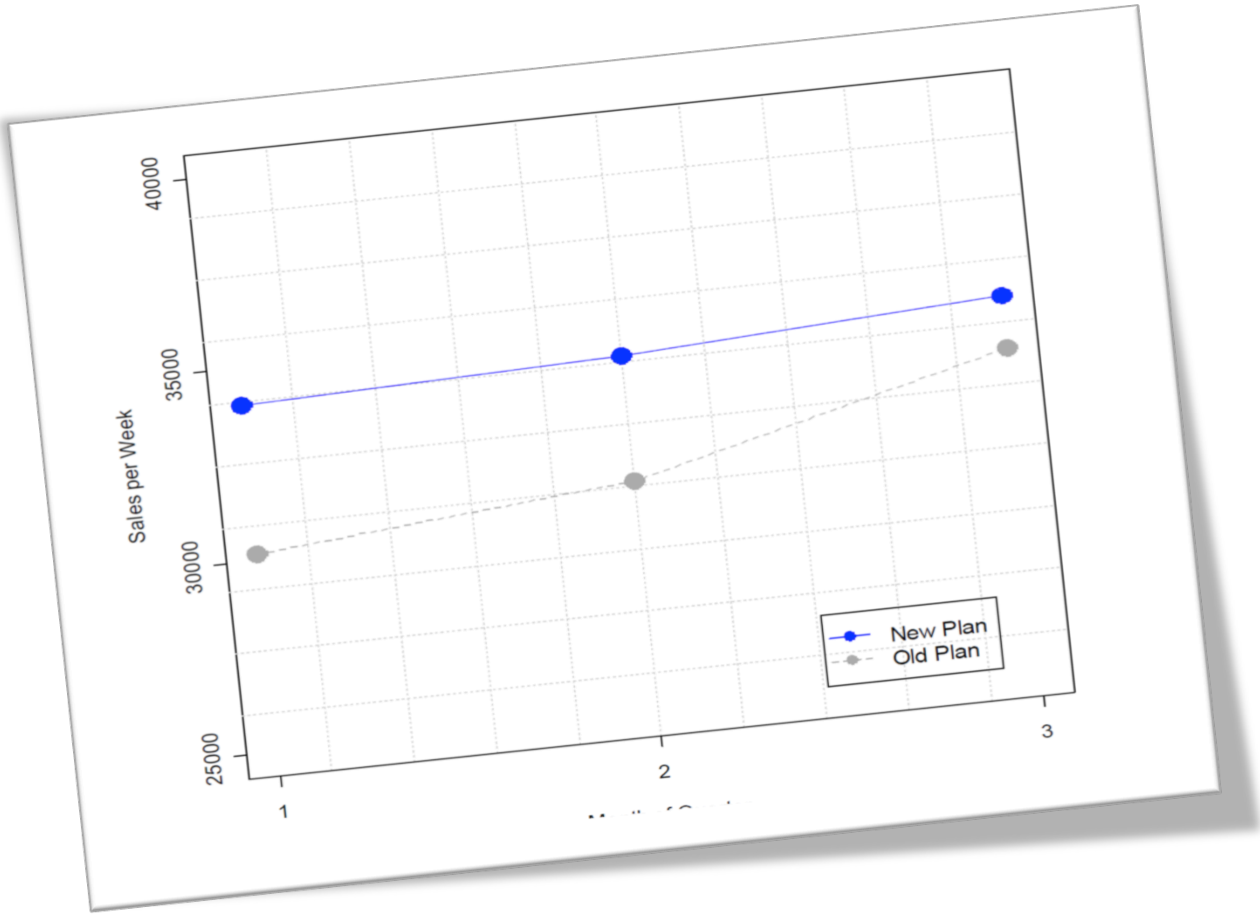
④ Measure Predicted Value.

What happened...



4 Measure Predicted Value.

What happened...



To conclude…

- Everybody's happy.
  - Firm has sustained incremental revenues of around $1MM each month.
  - As by product firm gets a nifty salesperson evaluation tool.
  - Salespeople are earning more and are more satisfied.
  - We get data. (didn't think this through!)

Case #3

Scalable Price Targeting

ZipRecruiter for enterprise

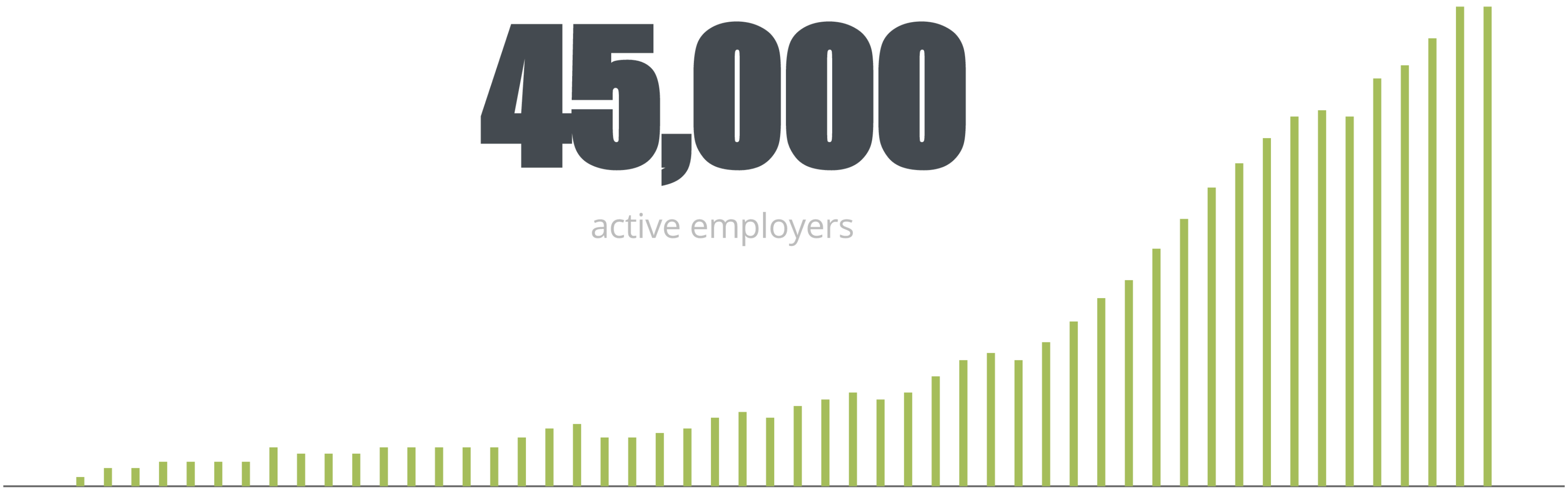
ZipRecruiter®

Find quality candidates anywhere on the internet.

Fastest growing company in HR

45,000

active employers



①Start at the end

End Goal

To construct, evaluate and implement an optimal pricing mechanism for subscribers to Ziprecruiter's service.

① Start at the end

Current Pricing

ZipRecruiter

$99

Scalable Price Targeting

① Start at the end

① Infer heterogeneity in consumer needs

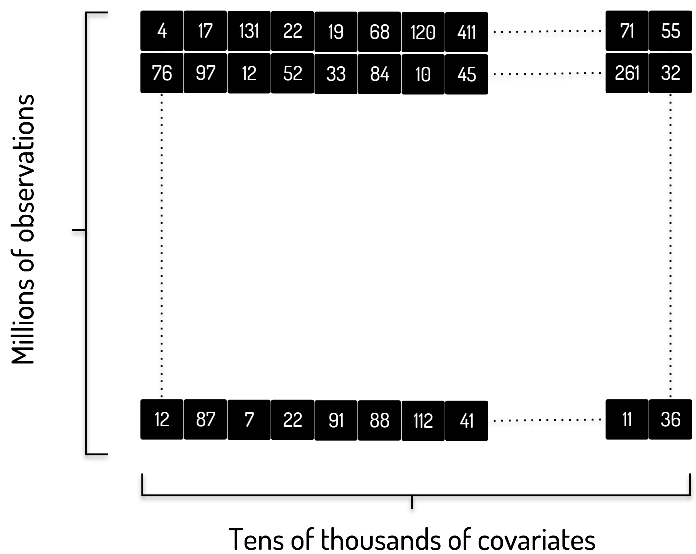
② Infer heterogeneity in consumer valuations

③ Optimize price for each customer

④ Implement & Monitor

③ Distribute Analytics.

Implementation Challenges



Dimensionality

Timeliness

③ Distribute Analytics.

Experiment #1

The experiment randomized new customers into 10 bins.

Each bin offered 3 options with varying time-commitments.

A number of job and org characteristics were collected.

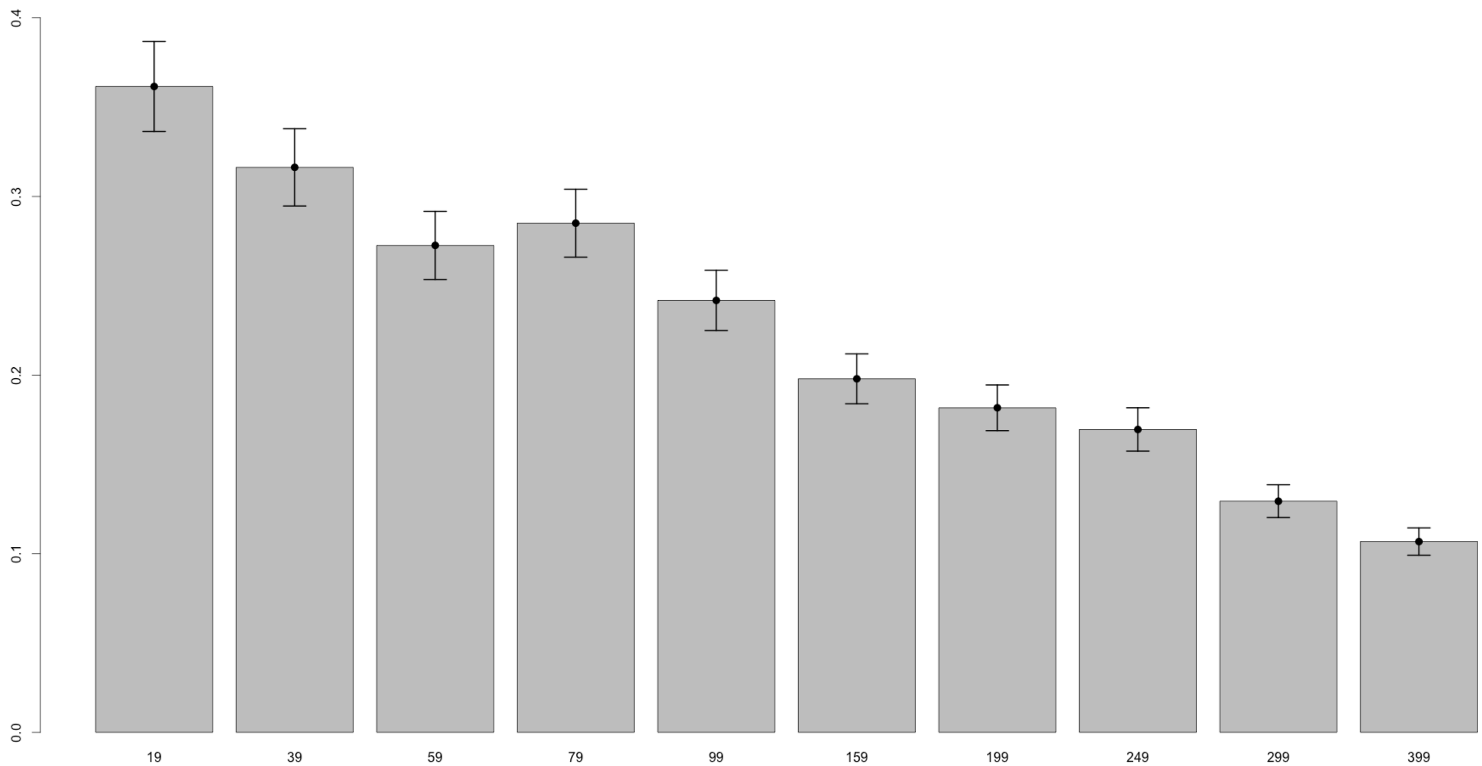
{{

Monthly	Quarterly	Annual
19	49	119
39	99	239
59	149	359
79	199	479
99	249	590
159	399	999
199	499	1199
249	629	1499
299	759	1789
399	999	2379

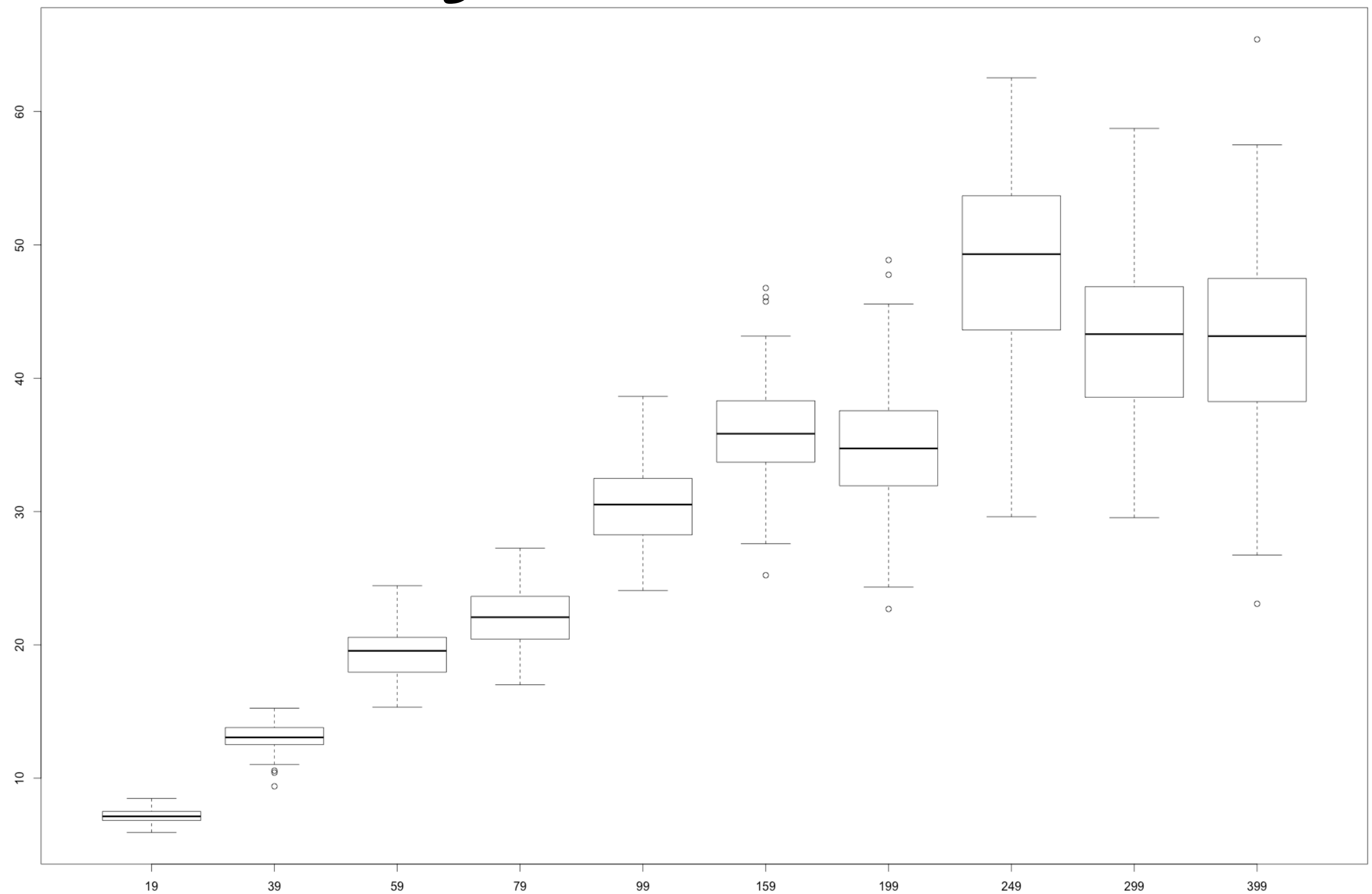
}}

Initial Results

③ Distribute Analytics.



Revenues by Price



③ Distribute Analytics.

③ Distribute Analytics.

Model & Estimation

Focused on “starters”

The utility specification for the value component and price sensitivity was linear in the covariates.

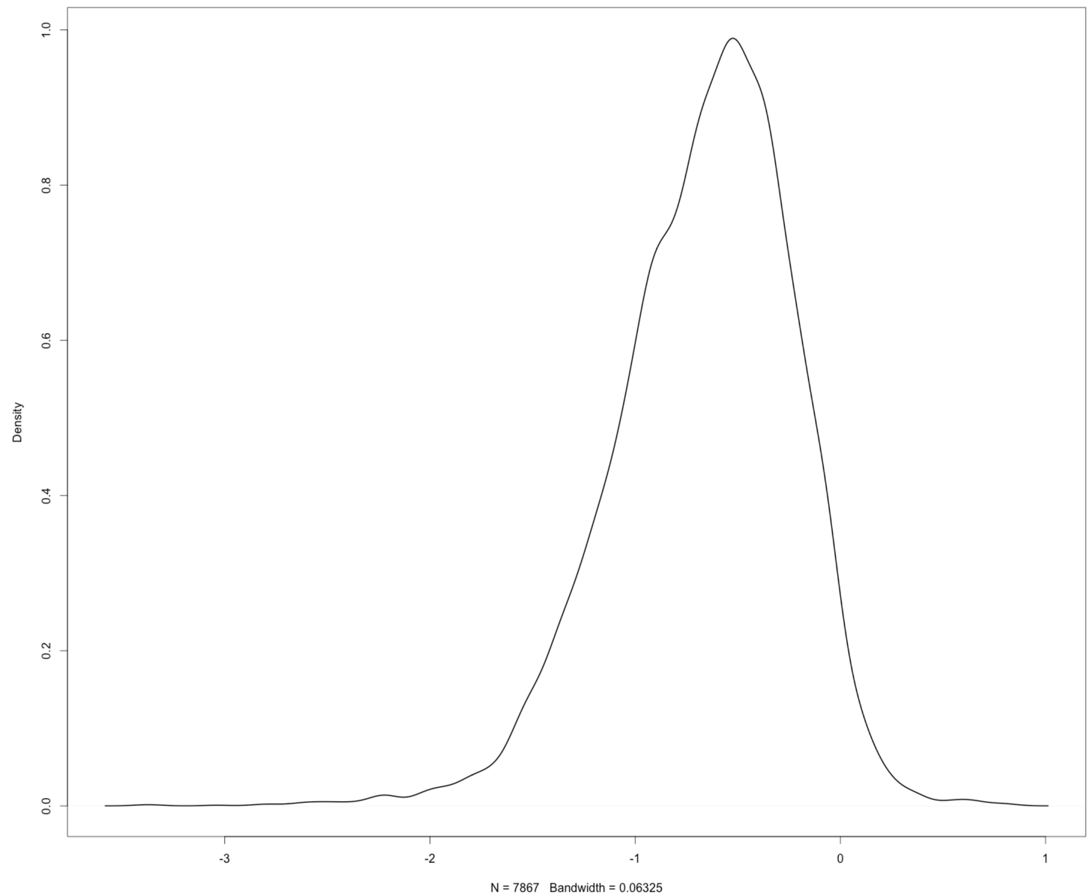
We use a weighted Likelihood Bootstrap version of a Elastic-Net Logit.

- Sophisticated machine learning approach to reduce the dimensionality of the problem and implement estimator simultaneously.
- A novel Bootstrap approach to allow for inference.

Product Value

③ Distribute Analytics.

The customer base exhibits significant variation in their valuation of the product.

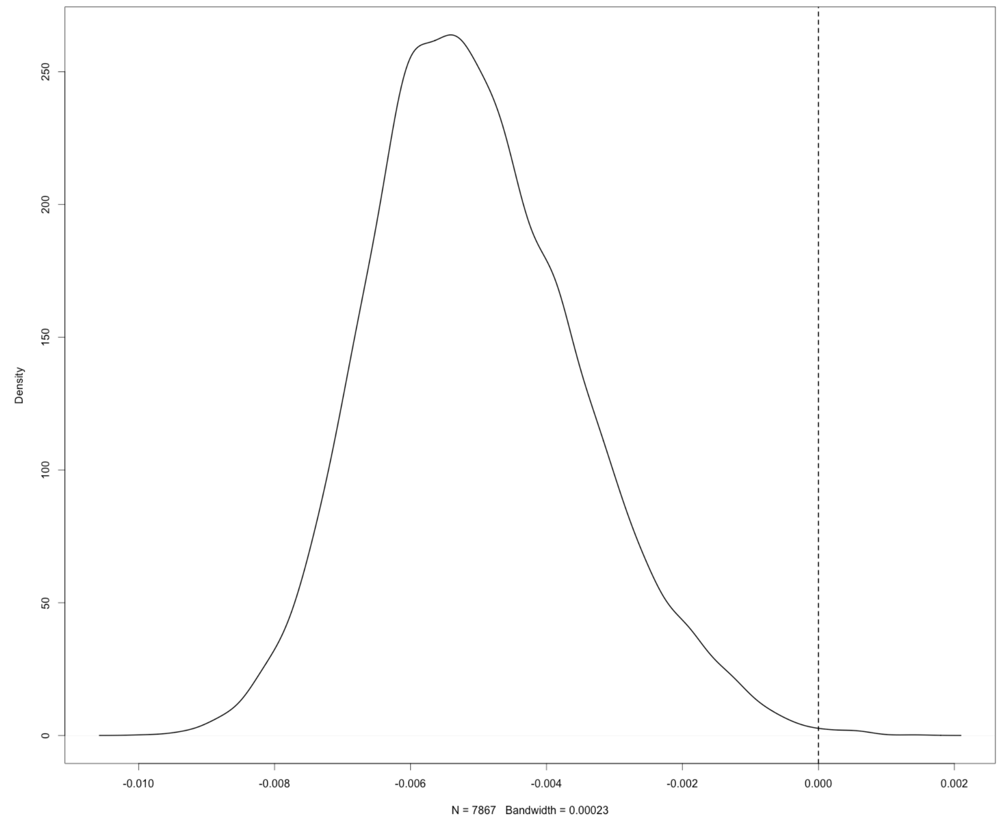


③ Distribute Analytics.

Price Sensitivity

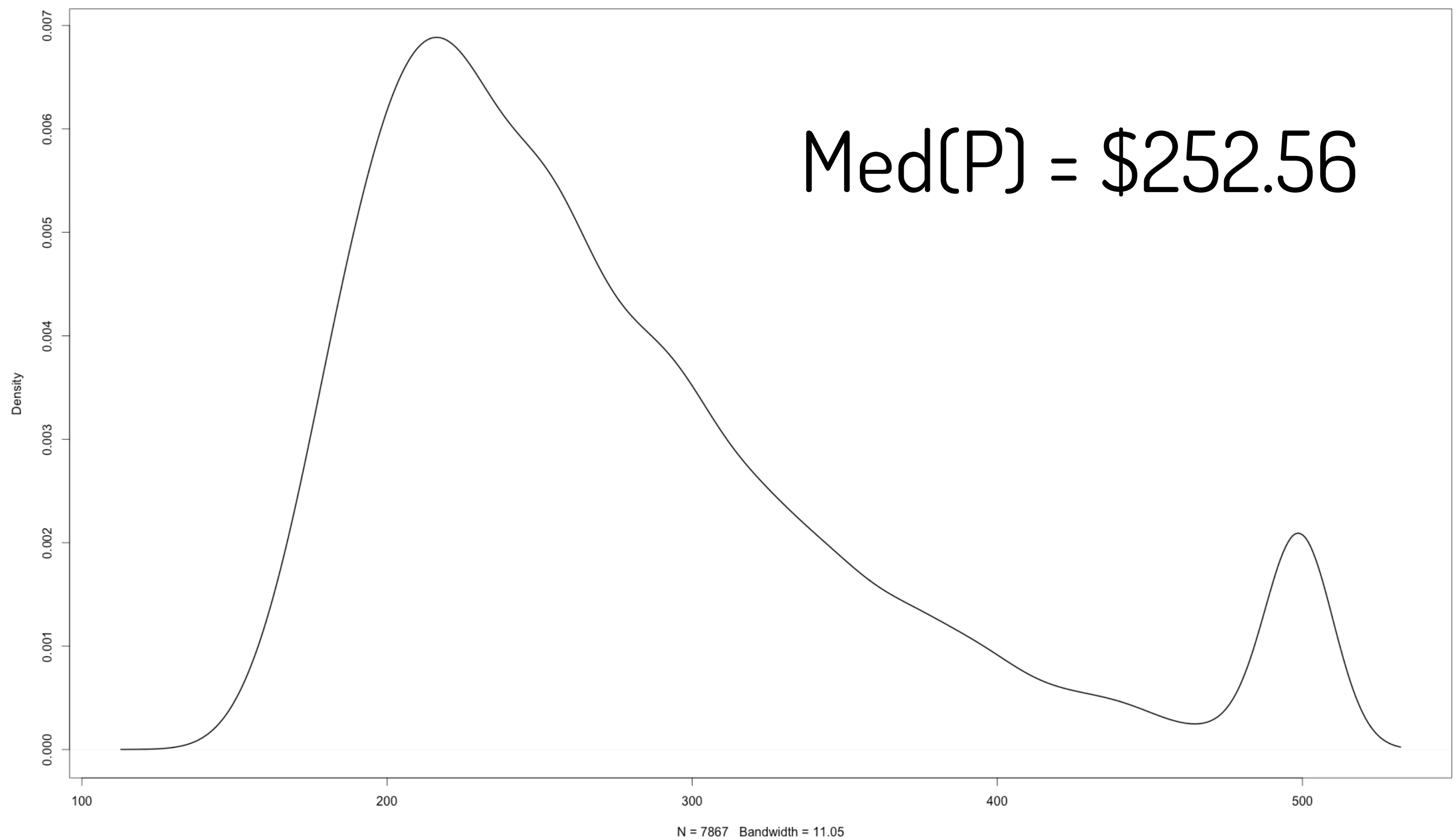
The customer base exhibits significant variation in their sensitivity to price.

*In other words price discrimination will be effective.



③ Distribute Analytics.

Optimal Prices



Who gets what price?

③ Distribute Analytics.

4) Measure Predicted Value.

Projected* Revenues

Uniform Pricing would garner about **$36 per org.**

Under the targeted pricing approach the estimated expected revenue for each starter org would be around **$43**.

- This could range from **$38-$51** because of sampling and estimation variances.

4 Measure Predicted Value.

Validation **Experiment**

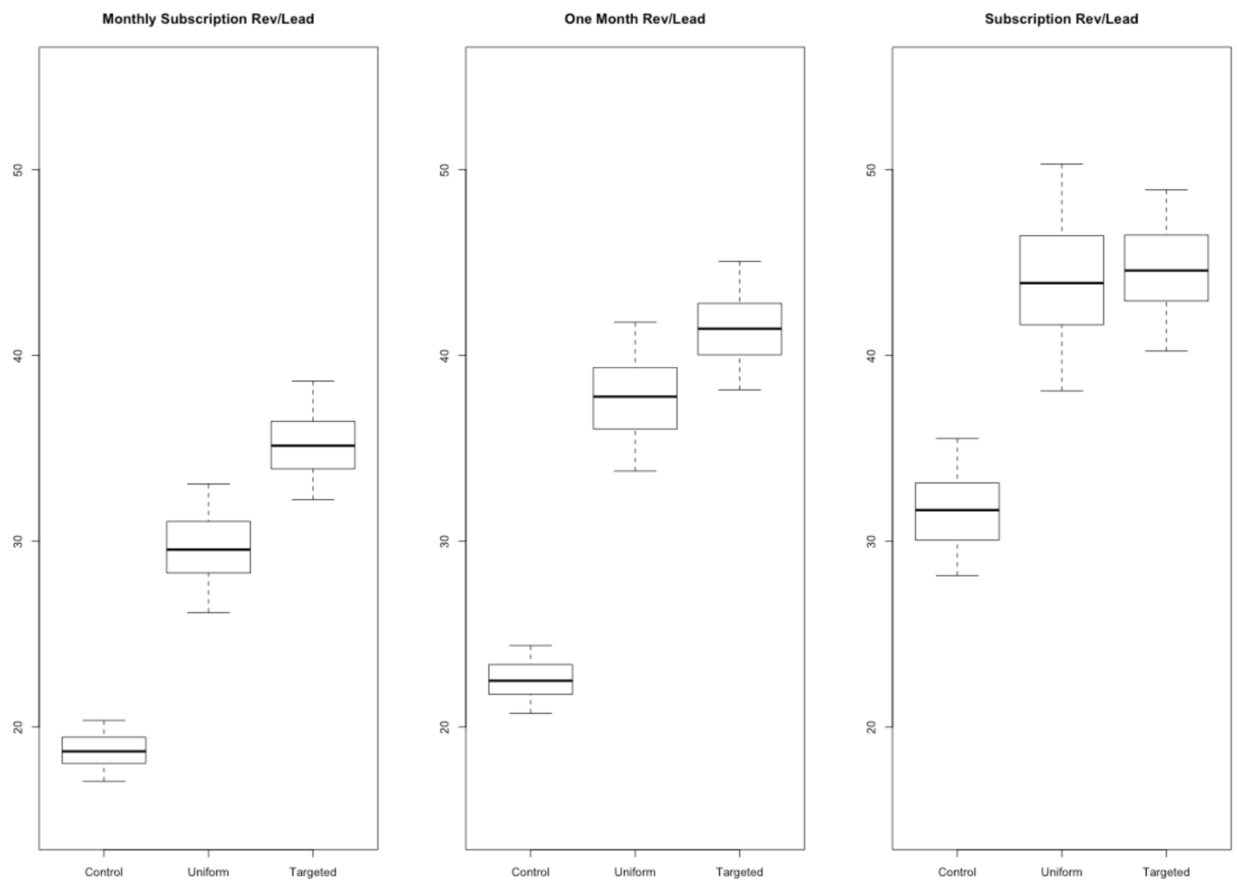
Implement a 3 bin experiment

- Status Quo
- Uniform pricing
- Targeted Pricing

Run and evaluate.

4 Measure Predicted Value.

Main Result

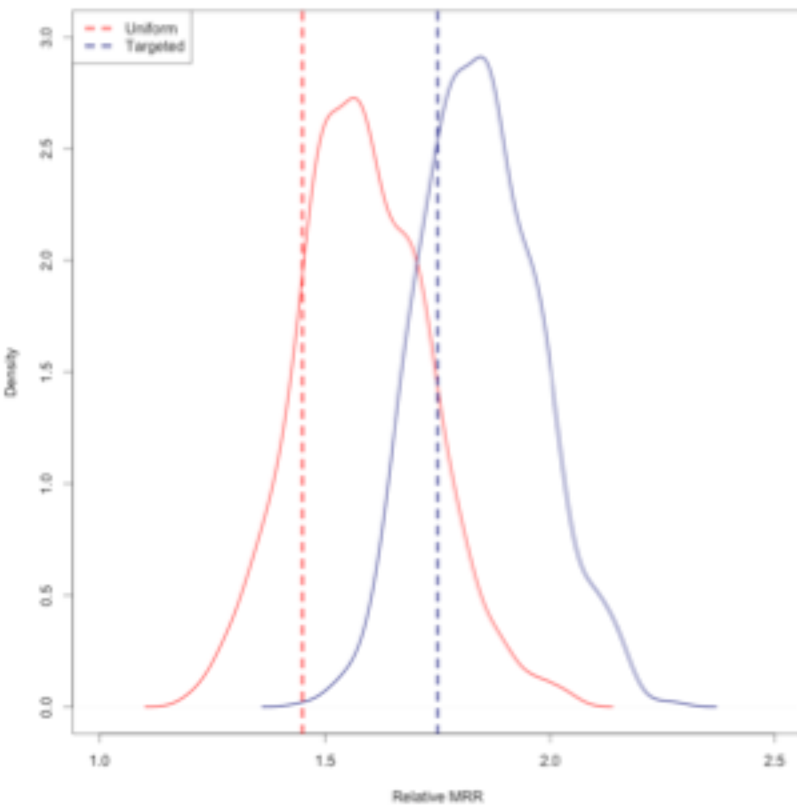


Targeted prices wins in all comparisons except for total sub rev where it is statistically tied.

Uniform (as expected) has higher variance but performs somewhat better than expected.

4 Measure Predicted Value.

Predictions Validated



Realized revenue of $41.48 is very close to prediction of $43 made before test began.

Further, the relative (to Control) MRR predictions are close to spot on.

4) Measure Predicted Value.

Results & Discussion

- Overall revenues have increased by between 30%-80%
- Conversion rates are lower and churn rates are a bit higher.
  - Implications for long run remain positive
- Firm rolling out targeted pricing for other segments of customers.

Some Takeaways

Basic economic principles coupled with sophisticated machine learning approaches can scale data driven decision making.

Ideas presented here have broader implications for advertising, direct mail and other marketing instruments.

4 Ideas

① Start at the end.

② Think structurally.

③ Distribute Analytics.

④ Measure Predicted Value.

Big Data, Analytics & Real Estate

Data Democratization & Consumer Power
Risk & Liability Assessment
Appraisals & Comps
New forms of "enhanced" data (streetscore)

Ideas:
Customized Home Recommendations
Matching Models
Computing WTP

THANK YOU

sanjog.misra@chicagobooth.edu