Quant Marketing: Hot Topics & Recent Developments

(more specifically: marketing and ML)

Stephan Seiler, Imperial College London

EMAC 2023

Two great quant marketing events ...

QME

Quantitative Marketing and Economics

CHICAGO BOOTH

Kilts Center for Marketing

2023 Quantitative Marketing and Economics Conference

Imperial College Business School

September 1 - 2, 2023

CALL FOR PAPERS

We invite paper submissions for the upcoming 21st annual QME conference, co-sponsored by Chicago Booth Kilts Center for Marketing and the Imperial College Business School, to be held at the Imperial College Business School on **Friday, September 1 and Saturday, September 2, 2023**.

Please submit your paper in PDF format to: QME2023.hotcrp.com **by the submission deadline of April 23rd**.

The conference seeks papers dealing with empirical and theoretical issues in marketing and economics. Submissions will be evaluated by the Conference Committee.

European Quant Marketing Seminar

EMAC

EUROPEAN MARKETING ACADEMY

A European Marketing Research Online Seminar by
EMAC's Marketing Research SIG

Current Organizers: Bart Bronnenberg (Tilburg), Anja Lambrecht (LBS), Thomas Otter (Frankfurt), Dominik Papies (Tuebingen), Stephan Seiler (Imperial College London)

Talks - Spring 2023

All talks are at 14:00 CET until 26 March and CSET after, except when stated otherwise below. Please join our mailing list so we can inform you about updates and changes.

Machine Learning & Targeted Marketing

- Targeting and personalization is increasing
  - Better data
  - Technological advances

Machine Learning & Targeted Marketing

- Targeting and personalization is increasing
  - Better data
  - Technological advances

- **This talk:** Combining machine learning with causal inference & managerial objectives
  - Targeting and incrementality
  - Deriving policy from an economic objective function
  - Choosing an ML model that generates the best policy

Incrementality

Example: Churn Management (Ascarza, 2019)

Marketers can also use big data to identify which customers are at highest risk of churn—and re-engage them before they defect.

—AIMIA Institute (Rogers 2013)

The challenge, of course, is to identify customers who are at the highest risk of churn before they switch to another carrier.

—*Analytics Magazine* (2016)

More sophisticated predictive analytics software use churn prediction models that predict customer churn by assessing their propensity of risk to churn. Since these models generate a small prioritized list of potential defectors, they are effective at focusing customer retention marketing programs on the subset of the customer base who are most vulnerable to churn.

—"Customer Attrition," Wikipedia

Predicting Churn Risk

- Current industry approach
  - Use historic data customers to understand risk of churn
  - Then target current customers with highest risk

Predicting Churn Risk

- Current industry approach
  - Use historic data customers to understand risk of churn
  - Then target current customers with highest risk
  - Could be done via regularized regression (or any other ML approach)

$$Retention_i = Z_i'\beta + \varepsilon_i$$

Predicting Churn Risk

- Current industry approach
  - Use historic data customers to understand risk of churn
  - Then target current customers with highest risk
  - Could be done via regularized regression (or any other ML approach)

$$Retention_i = Z_i'\beta + \varepsilon_i$$

- What does this regression tell us?
  - What customer characteristics predict whether customer is likely to churn
  - We can target high risk customer with marketing (e.g. reminder, discount)

Predicting Churn Risk

- Current industry approach
  - Use historic data customers to understand risk of churn
  - Then target current customers with highest risk
  - Could be done via regularized regression (or any other ML approach)

$$Retention_i = Z_i'\beta + \varepsilon_i$$

- What does this regression tell us?
  - What customer characteristics predict whether customer is likely to churn
  - We can target high risk customer with marketing (e.g. reminder, discount)
  - But: not clear that high risk customers are sensitive to marketing

Causal inference + Prediction

- Combine prediction model with A/B test
  - Denote $T_i = 1$ if customer receives marketing action ($T_i = 0$ otherwise)
  - Run regression that estimates effect heterogeneity

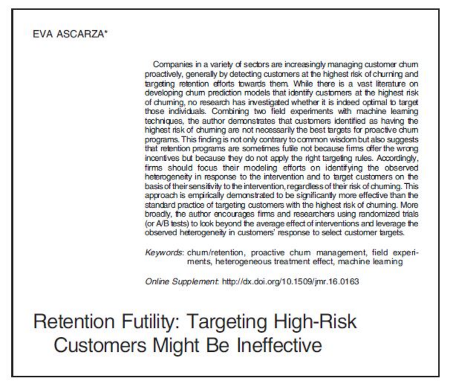
$$Retention_i = Z_i'\beta + (T_i \times Z_i)'\gamma + \varepsilon_i$$

Causal inference + Prediction

- Combine prediction model with A/B test
  - Denote $T_i = 1$ if customer receives marketing action ($T_i = 0$ otherwise)
  - Run regression that estimates effect heterogeneity
  - Then compute conditional average treatment effect (CATE) for a given set of characteristics $\tau(Z_i)$

$$
\begin{aligned}
Retention_i &= Z_i'\beta + (T_i \times Z_i)'\gamma + \varepsilon_i \\
\tau(Z_i) &= \mathbb{E}[Retention_i | Z_i, T_i = 1] - \mathbb{E}[Retention_i | Z_i, T_i = 0]
\end{aligned}
$$

Incrementality really matters



EVA ASCARZA*

Companies in a variety of sectors are increasingly managing customer churn proactively, generally by detecting customers at the highest risk of churning and targeting retention efforts towards them. While there is a vast literature on developing churn prediction models that identify customers at the highest risk of churning, no research has investigated whether it is indeed optimal to target those individuals. Combining two field experiments with machine learning techniques, the author demonstrates that customers identified as having the highest risk of churning are not necessarily the best targets for proactive churn programs. This finding is not only contrary to common wisdom but also suggests that retention programs are sometimes futile not because firms offer the wrong incentives but because they do not apply the right targeting rules. Accordingly, firms should focus their modeling efforts on identifying the observed heterogeneity in response to the intervention and to target customers on the basis of their sensitivity to the intervention, regardless of their risk of churning. This approach is empirically demonstrated to be significantly more effective than the standard practice of targeting customers with the highest risk of churning. More broadly, the author encourages firms and researchers using randomized trials (or A/B tests) to look beyond the average effect of interventions and leverage the observed heterogeneity in customers' response to select customer targets.

Keywords: churn/retention, proactive churn management, field experiments, heterogeneous treatment effect, machine learning

Online Supplement: http://dx.doi.org/10.1509/jmr.16.0163

Retention Futility: Targeting High-Risk Customers Might Be Ineffective

- Ascarza (2019) uses data from two companies and compares baseline risk targeting to the incremental approach

Incrementality really matters

EVA ASCARZA*

Companies in a variety of sectors are increasingly managing customer churn proactively, generally by detecting customers at the highest risk of churning and targeting retention efforts towards them. While there is a vast literature on developing churn prediction models that identify customers at the highest risk of churning, no research has investigated whether it is indeed optimal to target those individuals. Combining two field experiments with machine learning techniques, the author demonstrates that customers identified as having the highest risk of churning are not necessarily the best targets for proactive churn programs. This finding is not only contrary to common wisdom but also suggests that retention programs are sometimes futile not because firms offer the wrong incentives but because they do not apply the right targeting rules. Accordingly, firms should focus their modeling efforts on identifying the observed heterogeneity in response to the intervention and to target customers on the basis of their sensitivity to the intervention, regardless of their risk of churning. This approach is empirically demonstrated to be significantly more effective than the standard practice of targeting customers with the highest risk of churning. More broadly, the author encourages firms and researchers using randomized trials (or A/B tests) to look beyond the average effect of interventions and leverage the observed heterogeneity in customers' response to select customer targets.

Keywords: churn/retention, proactive churn management, field experiments, heterogeneous treatment effect, machine learning

Online Supplement: http://dx.doi.org/10.1509/jmr.16.0163

Retention Futility: Targeting High-Risk Customers Might Be Ineffective

- Ascarza (2019) uses data from two companies and compares baseline risk targeting to the incremental approach
    - Baseline Risk: target top 10% of customers with highest risk

Incrementality really matters

EVA ASCARZA*

Companies in a variety of sectors are increasingly managing customer churn proactively, generally by detecting customers at the highest risk of churning and targeting retention efforts towards them. While there is a vast literature on developing churn prediction models that identify customers at the highest risk of churning, no research has investigated whether it is indeed optimal to target those individuals. Combining two field experiments with machine learning techniques, the author demonstrates that customers identified as having the highest risk of churning are not necessarily the best targets for proactive churn programs. This finding is not only contrary to common wisdom but also suggests that retention programs are sometimes futile not because firms offer the wrong incentives but because they do not apply the right targeting rules. Accordingly, firms should focus their modeling efforts on identifying the observed heterogeneity in response to the intervention and to target customers on the basis of their sensitivity to the intervention, regardless of their risk of churning. This approach is empirically demonstrated to be significantly more effective than the standard practice of targeting customers with the highest risk of churning. More broadly, the author encourages firms and researchers using randomized trials (or A/B tests) to look beyond the average effect of interventions and leverage the observed heterogeneity in customers' response to select customer targets.

Keywords: churn/retention, proactive churn management, field experiments, heterogeneous treatment effect, machine learning

Online Supplement: http://dx.doi.org/10.1509/jmr.16.0163

Retention Futility: Targeting High-Risk Customers Might Be Ineffective

- Ascarza (2019) uses data from two companies and compares baseline risk targeting to the incremental approach
  - Baseline Risk: target top 10% of customers with highest risk
  - Incremental: target top 10% with highest treatment effect (i.e. most responsive to marketing)

Incrementality really matters

EVA ASCARZA*

Companies in a variety of sectors are increasingly managing customer churn proactively, generally by detecting customers at the highest risk of churning and targeting retention efforts towards them. While there is a vast literature on developing churn prediction models that identify customers at the highest risk of churning, no research has investigated whether it is indeed optimal to target those individuals. Combining two field experiments with machine learning techniques, the author demonstrates that customers identified as having the highest risk of churning are not necessarily the best targets for proactive churn programs. This finding is not only contrary to common wisdom but also suggests that retention programs are sometimes futile not because firms offer the wrong incentives but because they do not apply the right targeting rules. Accordingly, firms should focus their modeling efforts on identifying the observed heterogeneity in response to the intervention and to target customers on the basis of their sensitivity to the intervention, regardless of their risk of churning. This approach is empirically demonstrated to be significantly more effective than the standard practice of targeting customers with the highest risk of churning. More broadly, the author encourages firms and researchers using randomized trials (or A/B tests) to look beyond the average effect of interventions and leverage the observed heterogeneity in customers' response to select customer targets.

Keywords: churn/retention, proactive churn management, field experiments, heterogeneous treatment effect, machine learning

Online Supplement: http://dx.doi.org/10.1509/jmr.16.0163

Retention Futility: Targeting High-Risk Customers Might Be Ineffective

- Ascarza (2019) uses data from two companies and compares baseline risk targeting to the incremental approach
  - Baseline Risk: target top 10% of customers with highest risk
  - Incremental: target top 10% with highest treatment effect (i.e. most responsive to marketing)
- Incremental leads to 5-times larger reduction in churn

From Prediction to Policy

How to translate prediction into marketing policy?

- How can we derive a policy?
  - Ad hoc: target top 10% of customers with highest responsiveness (why 10%??)
  - Explicitly write down economic objective function

How to translate prediction into marketing policy?

- How can we derive a policy?
  - Ad hoc: target top 10% of customers with highest responsiveness (why 10%??)
  - Explicitly write down economic objective function

- Optimal policy for churn management setting:

$$\pi_i(Z_i, T_i = 1) \quad = \quad Pr(Retention_i = 1 | Z_i, T_i = 1) \times Fee - AdCost$$

How to translate prediction into marketing policy?

- How can we derive a policy?
  - Ad hoc: target top 10% of customers with highest responsiveness (why 10%??)
  - Explicitly write down economic objective function

- Optimal policy for churn management setting:

$$\begin{aligned}
\pi_i(Z_i, T_i = 1) &= Pr(Retention_i = 1 | Z_i, T_i = 1) \times Fee - AdCost \\
\pi_i(Z_i, T_i = 0) &= Pr(Retention_i = 1 | Z_i, T_i = 0) \times Fee
\end{aligned}$$

How to translate prediction into marketing policy?

- How can we derive a policy?
  - Ad hoc: target top 10% of customers with highest responsiveness (why 10%??)
  - Explicitly write down economic objective function

- Optimal policy for churn management setting:

$$\begin{aligned}
\pi_i(Z_i, T_i = 1) &= Pr(Retention_i = 1 | Z_i, T_i = 1) \times Fee - AdCost \\
\pi_i(Z_i, T_i = 0) &= Pr(Retention_i = 1 | Z_i, T_i = 0) \times Fee
\end{aligned}$$

$$\Delta\pi_i(Z_i) = \tau(Z_i) \times Fee - AdCost$$

Optimal Policy Derivation

$$\Delta\pi_i(Z_i) \quad = \quad \tau(Z_i) \times Fee - AdCost$$

- So what is the optimal policy?

Optimal Policy Derivation

$$\Delta\pi_i(Z_i) \quad = \quad \tau(Z_i) \times Fee - AdCost$$

- So what is the optimal policy?
  - Target all consumers with $\Delta\pi_i(Z_i) > 0$, i.e. $\tau(Z_i) > AdCost/Fee$
  - Justifies targeting consumers with highest treatment effect
  - Derives explicit threshold for how many consumer should be targeted!

Model (and Policy) Choice

How to choose the best model?

$$\Delta\pi_i(Z_i) \quad = \quad \hat{\tau}(Z_i) \times Fee - AdCost$$

- We know how to derive the optimal policy conditional on knowing the conditional average treatment effect

How to choose the best model?

$$\Delta\pi_i(Z_i) \quad = \quad \hat{\tau}(Z_i) \times Fee - AdCost$$

- We know how to derive the optimal policy conditional on knowing the conditional average treatment effect
- But we need to estimate $\hat{\tau}(Z_i)$ first
  - Which data inputs Z_i to use (demographics, past behavior, ...)
  - Choose an estimation method (lasso, random forest, neural net ...)

How to choose the best model?

$$\Delta\pi_i(Z_i) \quad = \quad \hat{\tau}(Z_i) \times Fee - AdCost$$

- We know how to derive the optimal policy conditional on knowing the conditional average treatment effect
- But we need to estimate $\hat{\tau}(Z_i)$ first
  - Which data inputs Z_i to use (demographics, past behavior, ...)
  - Choose an estimation method (lasso, random forest, neural net ...)
- How do we decide which model "works best"?
  - Typical ML approach: compare model out-of-sample fit
  - But: fit comparison has no direct relationship to policy and profits

How to evaluate different policies

- Setting
  - Optimal pricing policy (either discount or regular price)

How to evaluate different policies

- Setting
  - Optimal pricing policy (either discount or regular price)
- Typical approach
  - Estimate model $\rightarrow$ compute pricing policy $\rightarrow$ use model to calculate profits

How to evaluate different policies

- Setting
  - Optimal pricing policy (either discount or regular price)
- Typical approach
  - Estimate model → compute pricing policy → use model to calculate profits
  - Simple example: choose between two pricing policies p_a or p_b

How to evaluate different policies

- Setting
  - Optimal pricing policy (either discount or regular price)

- Typical approach
  - Estimate model → compute pricing policy → use model to calculate profits
  - Simple example: choose between two pricing policies p_a or p_b

<u>Probit Demand</u>

$$Y_{\text{probit}}(p)$$

How to evaluate different policies

- Setting
  - Optimal pricing policy (either discount or regular price)

- Typical approach
  - Estimate model → compute pricing policy → use model to calculate profits
  - Simple example: choose between two pricing policies p_a or p_b

Probit Demand

Pricing Policy

$$
\begin{tikzcd}
Y_{\text{probit}}(p) \arrow[r] \arrow[dr, bend right] & \mathbb{E}[\pi_{\text{probit}}(p_a)] = 0.12 \\
& \mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15
\end{tikzcd}
$$

How to evaluate different policies

- Setting
  - Optimal pricing policy (either discount or regular price)
- Typical approach
  - Estimate model $\rightarrow$ compute pricing policy $\rightarrow$ use model to calculate profits
  - Simple example: choose between two pricing policies p_a or p_b

$$
\begin{tikzcd}
\underline{\text{Probit Demand}} & \underline{\text{Pricing Policy}} & \underline{\text{Policy Evaluation}} \\
Y_{\text{probit}}(p) \arrow[r] \arrow[dr, bend right] & \mathbb{E}[\pi_{\text{probit}}(p_a)] = 0.12 & \hat{\Pi}(p_b) = \mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15 \\
 & \mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15 \arrow[ur, bend right] &
\end{tikzcd}
$$

How to evaluate different policies

- Setting
  - Optimal pricing policy (either discount or regular price)

- Typical approach
  - Estimate model → compute pricing policy → use model to calculate profits
  - Simple example: choose between two pricing policies p_a or p_b
  - Problem: model is used twice

Probit Demand | Pricing Policy | Policy Evaluation

$$
\begin{tikzcd}
Y_{\text{probit}}(p) \arrow[r] \arrow[dr] & \mathbb{E}[\pi_{\text{probit}}(p_a)] = 0.12 & \hat{\Pi}(p_b) = \mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15 \\
 & \mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15 \arrow[ur] &
\end{tikzcd}
$$

Challenge: Cross-model comparison

Probit Demand

Pricing Policy

Policy Evaluation

$$
\begin{tikzcd}[column sep=large]
 & \mathbb{E}[\pi_{\text{probit}}(p_a)] = 0.12 & \hat{\Pi}(p_b) = \mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15 \\
Y_{\text{probit}}(p) \arrow[ur] \arrow[r] & \mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15 \arrow[ur] &
\end{tikzcd}
$$

Challenge: Cross-model comparison

Probit Demand

$Y_{\text{probit}}(p)$

Pricing Policy

$\mathbb{E}[\pi_{\text{probit}}(p_a)] = 0.12$

$\mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15$

Policy Evaluation

$\hat{\Pi}(p_b) = \mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15$

Random Forest

$Y_{\text{forest}}(p)$

Challenge: Cross-model comparison

Probit Demand

$Y_{\text{probit}}(p)$

Pricing Policy

$\mathbb{E}[\pi_{\text{probit}}(p_a)] = 0.12$

$\mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15$

Policy Evaluation

$\hat{\Pi}(p_b) = \mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15$

Random Forest

$Y_{\text{forest}}(p)$

Pricing Policy

$\mathbb{E}[\pi_{\text{forest}}(p_a)] = 0.10$

$\mathbb{E}[\pi_{\text{forest}}(p_b)] = 0.08$

Challenge: Cross-model comparison

$$
\begin{tikzcd}[column sep=large]
\underline{\text{Probit Demand}} & \underline{\text{Pricing Policy}} & \underline{\text{Policy Evaluation}} \\
Y_{\text{probit}}(p) \arrow[r] \arrow[dr] & \mathbb{E}[\pi_{\text{probit}}(p_a)] = 0.12 & \hat{\Pi}(p_b) = \mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15 \\
& \mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15 \arrow[ur] &
\end{tikzcd}
$$

$$
\begin{tikzcd}[column sep=large]
\underline{\text{Random Forest}} & \underline{\text{Pricing Policy}} & \underline{\text{Policy Evaluation}} \\
Y_{\text{forest}}(p) \arrow[r] \arrow[dr] & \mathbb{E}[\pi_{\text{forest}}(p_a)] = 0.10 \arrow[r] & \hat{\Pi}(p_a) = \mathbb{E}[\pi_{\text{forest}}(p_a)] = 0.10 \\
& \mathbb{E}[\pi_{\text{forest}}(p_b)] = 0.08 &
\end{tikzcd}
$$

Our Approach: de-couple evaluation

Probit Demand | Pricing Policy

$$Y_{\text{probit}}(p) \to \mathbb{E}[\pi_{\text{probit}}(p_a)] = 0.12$$
$$Y_{\text{probit}}(p) \to \mathbb{E}[\pi_{\text{probit}}(p_b)] = 0.15$$

Random Forest | Pricing Policy

$$Y_{\text{forest}}(p) \to \mathbb{E}[\pi_{\text{forest}}(p_a)] = 0.10$$
$$Y_{\text{forest}}(p) \to \mathbb{E}[\pi_{\text{forest}}(p_b)] = 0.08$$

Demand Estimation & Policy Generation
(training sample)

Policy Evaluation

$$\hat{\Pi}(p_a)$$

Policy Evaluation

$$\hat{\Pi}(p_b)$$

Evaluation
(test sample)

Solution: Do Evaluation "In-Sample"

- High-level idea
  - We use only observations where (price observed in the data) = (price prescribed by the policy)
  - Then we re-weight observations to account for rate at which prices don't match
  - $\rightarrow$ Inverse probability weighted profit estimator

Simple Example

- Simple example (based on time series for one consumer)
  - Two price levels: regular and discount
  - Consumer i is observed for $T_i = 30$ trips, 20 at regular price, 10 at discounted price

Simple Example

- Simple example (based on time series for one consumer)
  - Two price levels: regular and discount
  - Consumer i is observed for $T_i = 30$ trips, 20 at regular price, 10 at discounted price
- How to use our estimator
  - If regular price is prescribed: only 20 observations are usable
  - Dividing by 2/3 re-scales them to 30 observations
  - We then divide by $T_i = 30$ to obtain consumer-specific average profits

Simple Example

- Simple example (based on time series for one consumer)
  - Two price levels: regular and discount
  - Consumer i is observed for $T_i = 30$ trips, 20 at regular price, 10 at discounted price
- How to use our estimator
  - If regular price is prescribed: only 20 observations are usable
  - Dividing by $2/3$ re-scales them to 30 observations
  - We then divide by $T_i = 30$ to obtain consumer-specific average profits
  - If prescribed price is discount: use 10 observations, divide by $1/3$

"Optimal Price Targeting" by Smith, Seiler, Aggarwal (forthcoming *Marketing Science*)

- Study price targeting using supermarket scanner data
- Estimate large set of ML models and compare out-of-sample profits

“Optimal Price Targeting” by Smith, Seiler, Aggarwal (forthcoming *Marketing Science*)

- Study price targeting using supermarket scanner data
- Estimate large set of ML models and compare out-of-sample profits

- Key result:
  - Model performance based on fit is uncorrelated with profits
  - $\rightarrow$ Important to pick model based on decision-theoretic objective function

Conclusion

Re-cap: machine learning and targeted marketing

- Need to combine ML techniques with insights from causal inference and marketing / economic theory
  - (1) Focus on differences in causal effects across customers ($\rightarrow$ incrementality)
  - (2) Derive policy from underlying profit function
  - (3) Choose ML model based on decision-theoretic objective function

Re-cap: machine learning and targeted marketing

- Need to combine ML techniques with insights from causal inference and marketing / economic theory
  - (1) Focus on differences in causal effects across customers ($\rightarrow$ incrementality)
  - (2) Derive policy from underlying profit function
  - (3) Choose ML model based on decision-theoretic objective function
- What are firms doing right now?
  - Implement none of this or possibly some version of (1) !!!

Thank You !!!