How causal machine learning can leverage marketing strategies: Assessing and improving the performance of a coupon campaign

https://arxiv.org/abs/2204.10820

Henrika Langen[†] and Martin Huber[‡]

[†]University of Helsinki
[‡]University of Fribourg

The Causal Data Science Meeting
November 7-8, 2022

In a Nutshell

- **Motivation**: machine learning (ML) literature on the rise in the field of business analytics and marketing, but nearly all studies on predictive ML
- **Data**: Retailer Data on purchasing behavior of 1,582 customers during 18 coupon campaigns
- **Methodological Approach:**
  - Estimation of treatment effect heterogeneity with causal forests (Athey, Wager, 2018)
  - Estimation of ATE of coupon reception (for a certain product category) and Group Average Treatment Effects (GATE) for selected customer groups (as defined by covariates)
  - Determination of optimal coupon distribution strategy with optimal policy learning (Athey, Wager, 2021)

Motivation

How Causal ML Can Leverage Research on Coupon Campaigns

Advantages over non-ML causal inference approaches

- Identifiction of causal effect of coupon campaigns based on high-dimensional observational data
  - ⇒ all potential controls can be considered
  - ⇒ data-driven selection of relevant confounders
  - ⇒ confounders can enter into the estimation in a flexible, possibly non-linear way

Advantages over predictive ML

- informs about effectiveness of marketing campaigns
  - ⇒ allows answering “what if” questions
  - ⇒ can reveal targeting strategy that maximizes effect

Data

- Data on socio-economic characteristics, purchasing behavior and coupon reception of 1,582 retail store customers during 18 campaign periods
- Campaign periods are partly time-overlapping $\Rightarrow$ definition of "artificial" campaign periods
- Pooled Cross-Sectional Analysis:
  - **Treatment D:** Receiving coupons (of specific type)
  - **Outcome Y:** Average per-day expenditures
  - **Covariates X:** Socio-economic characteristics, period dummies, expenditures and coupon reception in pre-treatment period (and reception of other coupon types)

Identification

- **Assumption 1 (Conditional Independence of the Treatment):**
  $Y(d) \perp D | X$ for all $d \in \{0,1\}$.
- **Assumption 2 (Common Support):**
  $0 < Pr(D = 1|X) < 1$.
- **Assumption 3 (Relaxation of the SUTVA):**
  - the coupons provided to one individual in a certain campaign period do not affect the potential outcome of another individual during the same period
  - there is only one version of the control condition

Methods: Causal Forests

- Determining splitting rules that maximize the heterogeneity of treatment effects in the resulting subsamples to estimate Conditional Average treatment effects (CATEs)
- Algorithm:
  - take random set of 30 covariates and split the data in 2 subsamples
  - build causal tree in one subsample and estimate the ATE with the doubly robust approach by Robinson (1988) in all final nodes using the other subsample
  - repeat 2,000 times and average over all trees to obtain CATE estimates

$\Rightarrow$ can serve as base for estimation of ATE, GATEs by age, family size, income group and pre-period expenditures and for learning the optimal distribution strategy through optimal policy learning

Methods: Estimation of the ATE and GATEs

- **ATE:** Estimation based on modified version of Augmented Inverse Probability Weighting (AIPW), with estimates for CATE, propensity score and potential outcomes obtained from the causal forest
- **GATEs by age, family size, income group and pre-period expenditures:** Regression of CATEs on group identifiers with heteroscedasticity and cluster-robust standard errors

Methods: Optimal Policy Learning

Determination of coupon distribution scheme that maximizes the expected ATE, i.e., identification of decision rules for whether a coupon should be offered to a customer as a function of X

Algorithm:

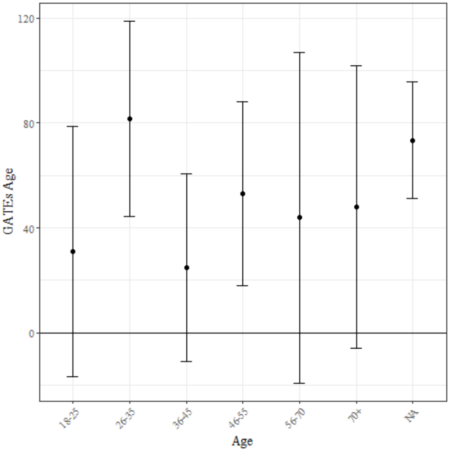
- Split sample at every value of X and estimate ATE for when either of the resulting subsamples receives coupon
- Identify sample split and coupon assignment that would maximize expected ATE
- Repeat in the resulting subsamples until certain pre-defined depth of the resulting decision tree

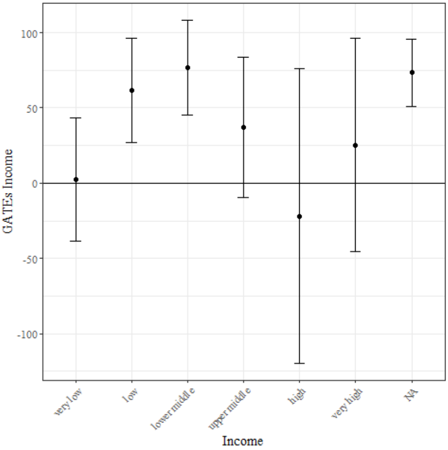
Results: ATE

	Coef.	Std. Error	Sign. Level
ATE: receiving any coupon	63.26	4.553	***
ATE: receiving coupon for ready-to-eat food	-2.90	8.118	
ATE: receiving coupon for meat/seafood	-1.42	6.045	
ATE: receiving coupon for other food	74.74	13.559	***
ATE: receiving coupon for drugstore items	60.07	6.521	***
ATE: receiving coupon for other non-food items	-26.77	6.949	***

Significance levels: . $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

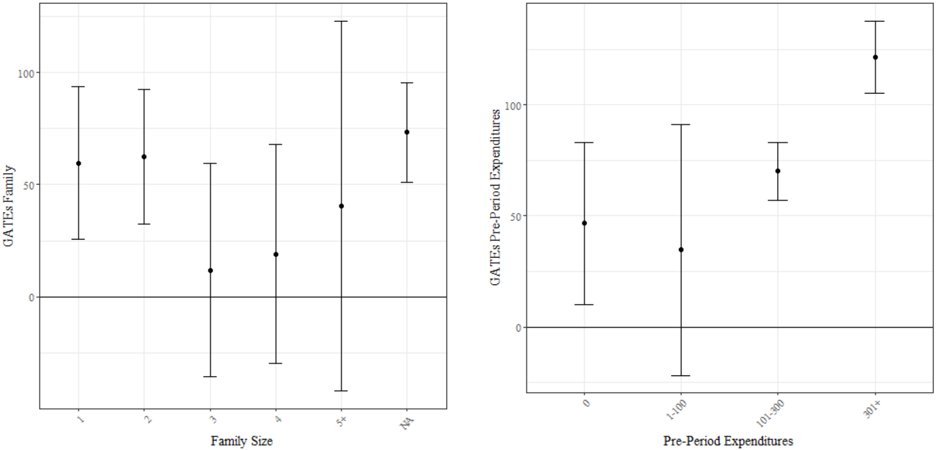
Results: GATEs





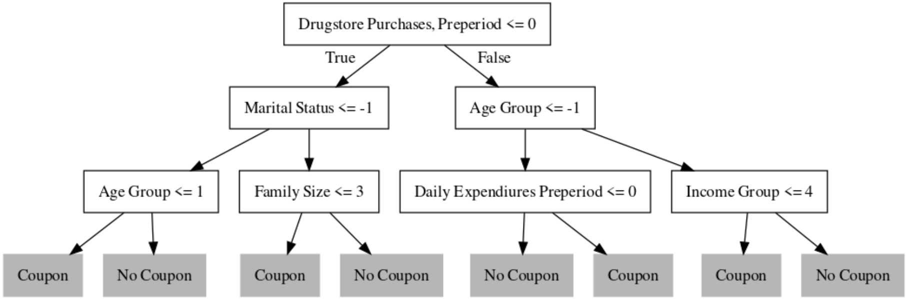
GATEs of drugstore coupons with 95% confidence interval.

Results: GATEs



GATEs of drugstore coupons with 95% confidence interval.

Results: Optimal Policy Learning



Depth-3 tree for coupons applicable to ready-to-eat food.

Results: Longer-term ATEs

	Effect in $t+1$			Effect in $t+2$		
	Coef.	Std. E.	Sign.	Coef.	Std. E.	Sign.
Any C.	37.83	4.01	***	30.59	4.456	***
Ready-to-eat Food C.	-25.81	6.226	***	4.18	8.908	
Meat/Seafood C.	9.54	5.209	.	1.13	6.244	
Other Food C.	52.50	8.779	***	2.46	7.345	
Drugstore C.	96.50	5.747	***	82.78	5.681	***
Other Non-Food C.	30.78	6.09	***	11.64	6.017	.

Significance levels: . $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Results: Robustness Checks

Investigting robustness of results: Due to the large number of observations with missing socio-economic information, the analyses are also performed on a reduced dataset containing only observations of customers whose socioeconomic background is known.

ATE		with NAs			w/o NAs	
	Coef.	Std. Error	Sign.	Coef.	Std. Error	Sign.
Any Coupon	63.26	4.553	***	53.56	8.437	***
Ready-to-Eat Food C.s	-2.90	8.118		-6.50	10.618	
Meat/Seafood C.s	-1.42	6.045		-26.94	9.374	**
Other Food C.s	74.74	13.559	***	96.72	22.897	***
Drugstore C.s	60.07	6.521	***	41.12	9.493	***
Other Non-Food C.s	-26.77	6.949	***	-18.94	9.622	*

Significance levels: . $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Conclusion

- Empirical Findings:
  - coupons belonging to two out of five categories have a positive and statistically significant overall effect on purchases, other non-food coupons have a sign. negative effect
  - effect heterogeneity between different customer subgroups can be exploited to target customers with the largest expected effect
- Causal ML methods can further be applied to evaluate and optimize a variety of other marketing and business strategies, requiring only observational data from the context of previous campaigns or business decisions