Javier Moral Hernández
Clara Higuera-Cabañes
Álvaro Ibraín

An End-to-End Pipeline for Causal ML with Continuous Treatments: An Application to Financial Decision Making

Business Problem – Optimizing Debt Recovery

Find the optimal debt forgiveness percentage (write-off) to maximize debt recovery… for each defaulted customer → **personalization problem**

Debt to pay	Debt forgiven
70%	30%

→ What is the probability that the customer pays if I forgive 30%?

Maximize personalised debt recovery while minimizing debt loss

Why Causal ML?

1. Decision-making problem
2. Impossibility of RCT
3. Confounding bias
4. Personalization & Heterogeneous effects

Conventional causal ML framework

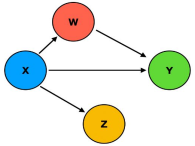
BBVA

Problem Formulation

$\mathcal{D} = (X_i, T_i, Y_i)^N$
$X \in \mathbb{R}^d$ are covariates
$T \in \mathbb{R}$ is the Treatment
$Y \in \mathbb{R}$ is the Outcome

+ causal question

Identification



backdoor criterion:

- $(T \perp\!\!\!\perp Y \mid Z)$
- $Z \cap \text{Desc}(T) = \varnothing$

Estimation

Potential Outcomes:

$m(t, z) = E[Y \mid T{=}t, Z{=}z]$

Conditional Effect:

$\square(z) = m(T{=}1, z) - m(T{=}0, z)$

Evaluation

- Refutation Tests
- Sensitivity Analysis
- QINI curve AUC
- Estimation Variance

Policy optimization

$\arg_t \max \ (1-t)\hat{y}(t)$

Personalized decisions

CausalML

DoWhy

EconML

YLearn

Conventional causal ML framework

BBVA

Problem Formulation

$\mathcal{D} = (X_i, T_i, Y_i)^N$

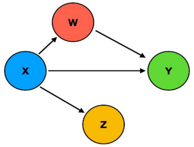
$X \in \mathbb{R}^d$ are covariates

$T \in \mathbb{R}$ is the Treatment

$Y \in \mathbb{R}$ is the Outcome

+ causal question

Identification



backdoor criterion:

- $(T \perp\!\!\!\perp Y \mid Z)$
- $Z \cap \mathrm{Desc}(T) = \varnothing$

Identification under high-dimensional data

Estimation

…tial Outcomes:

… $= E[Y \mid T{=}t, Z{=}z]$

…tional Effect:

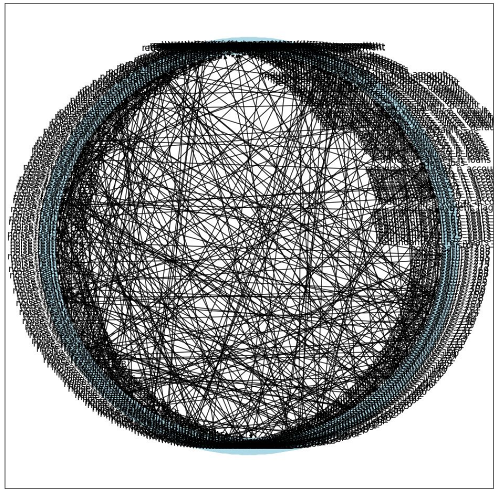
… m

Policy optimization

arg tmax $(1\text{-}t)\hat{y}(t)$

Evalu…

- Refuta…
- Sensiti…
- QINI c…
- Estima…



Personalized decisions

Conventional causal ML framework

BBVA

Positivity Assumption violation

Problem Formulation

$\mathcal{D} = (X_i, T_i, Y_i)^N$

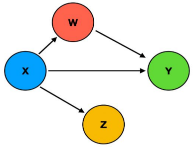
$X \in \mathbb{R}^d$ are covariates

$T \in \mathbb{R}$ is the Treatment

$Y \in \mathbb{R}$ is the Outcome

+ causal question

Identification



backdoor criterion:

- $(T \perp\!\!\!\perp Y \mid Z)$
- $Z \cap \mathrm{Desc}(T) = \varnothing$

Estimation

Potential Outcomes:

$m(t, z) = E[Y \mid T=t, Z=z]$

Conditional Effect:

$\square(z) = m(T=1, z) - m(T=0, z)$

Evaluation

- Refutation Tests
- Sensitivity Analysis
- QINI curve AUC
- Estimation Variance

Policy optimization

arg tmax $(1-t)\hat{y}(t)$

Personalized decisions

Conventional causal ML framework

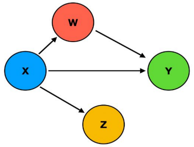
BBVA

Problem Formulation

$\mathcal{D} = (X_i, T_i, Y_i)^N$
$X \in \mathbb{R}^d$ are covariates
$T \in \mathbb{R}$ is the Treatment
$Y \in \mathbb{R}$ is the Outcome

+ causal question

Identification



backdoor criterion:

- $(T \perp\!\!\!\perp Y \mid Z)$
- $Z \cap \mathrm{Desc}(T) = \varnothing$

Estimation

Potential Outcomes:
$m(t, z) = E[Y \mid T=t, Z=z]$

Conditional Effect:
$\square(z) = m(T=1, z) - m(T=0, z)$

Continuous treatment

Evaluation

- Refutation Tests
- Sensitivity Analysis
- QINI curve AUC
- Estimation Variance

Continuous treatment

Policy optimization

arg tmax $(1-t)\hat{y}(t)$

Personalized decisions

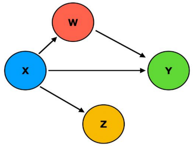
Conventional causal ML framework

Problem Formulation

$\mathcal{D} = (X_i, T_i, Y_i)^N$
$X \in \mathbb{R}^d$ are covariates
$T \in \mathbb{R}$ is the Treatment
$Y \in \mathbb{R}$ is the Outcome

+ causal question

Identification



backdoor criterion:

- $(T \perp\!\!\!\perp Y \mid Z)$
- $Z \cap \mathrm{Desc}(T) = \varnothing$

Estimation

Potential Outcomes:

$m(t, z) = E[Y \mid T=t, Z=z]$

Conditional Effect:

$\square(z) = m(T=1, z) - m(T=0, z)$

Evaluation

- Refutation Tests
- Sensitivity Analysis
- QINI curve AUC
- Estimation Variance

Policy optimization

$\arg_t \max \; (1-t)\hat{y}(t)$

Personalized decisions

Pipeline integration

Challenges - wrap up

01

Identification under high dimensionality

02

Positivity Assumption violation

03

Continuous treatment

04

Pipeline integration

Conventional causal ML framework

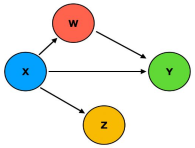
BBVA

Problem Formulation

$\mathcal{D} = (X_i, T_i, Y_i)^N$
$X \in \mathbb{R}^d$ are covariates
$T \in \mathbb{R}$ is the Treatment
$Y \in \mathbb{R}$ is the Outcome

+ causal question

Identification



backdoor criterion:

- $(T \perp\!\!\!\perp Y \mid Z)$
- $Z \cap \mathrm{Desc}(T) = \varnothing$

Personalized decisions

Policy optimization

$\arg_t\max\ (1-t)\hat{y}(t)$

Evaluation

- Refutation Tests
- Sensitivity Analysis
- QINI curve AUC
- Estimation Variance

Estimation

Potential Outcomes:
$m(t, z) = E[Y \mid T=t, Z=z]$

Conditional Effect:
$\square(z) = m(T=1, z) - m(T=0, z)$

Proposed causal ML framework

BBVA

Problem Formulation

$\mathcal{D} = (X_i, T_i, Y_i)^N$
$X \in \mathbb{R}^d$ are covariates
$T \in \mathbb{R}$ is the Treatment
$Y \in \mathbb{R}$ is the Outcome

+ causal question

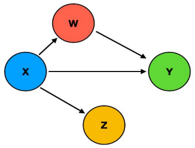
Dimensionality Reduction

- $X \in \mathbb{R}^d$ are covariates
- $d > 400$

We find controls:

- $Z \subseteq X$

Identification



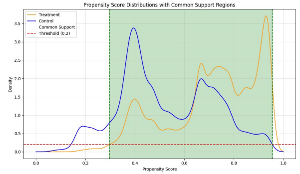
backdoor criterion:

- $(T \perp\!\!\!\perp Y \mid Z)$
- $Z \cap \mathrm{Desc}(T) = \varnothing$

Positivity Violation Handling

$P(T \in B \mid Z=z) > 0$

for every $B \subseteq \mathscr{T}$



Estimation

Potential Outcomes:

$m(t, z) = E[Y \mid T=t, Z=z]$

Conditional Effect:

$\square(z) = \partial/\partial t \; m(T=t, z)$

Evaluation

- Refutation Tests
- Sensitivity Analysis
- QINI curve AUC
- Estimation Variance

Policy optimization

$\arg_t\max \; (1-t)\hat{y}(t)$

Personalized decisions

Proposed causal ML framework

BBVA

Problem Formulation

$\mathcal{D} = (X_i, T_i, Y_i)^N$
$X \in \mathbb{R}^d$ are covariates
$T \in \mathbb{R}$ is the Treatment
$Y \in \mathbb{R}$ is the Outcome

+ causal question

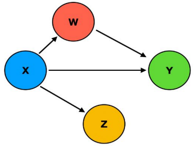
Dimensionality Reduction

- $X \in \mathbb{R}^d$ are covariates
- $d > 400$

We find controls:

- $Z \subseteq X$

Identification



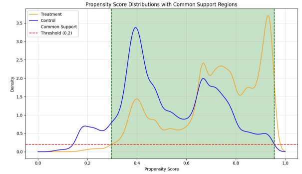
backdoor criterion:

- $(T \perp\!\!\!\perp Y \mid Z)$
- $Z \cap \mathrm{Desc}(T) = \varnothing$

Positivity Violation Handling

$P(T \in B \mid Z = z) > 0$

for every $B \subseteq \mathcal{T}$



Personalized decisions

Policy optimization

$\arg_t \max\ (1-t)\hat{y}(t)$

Evaluation

- Refutation Tests
- Sensitivity Analysis
- QINI curve AUC
- Estimation Variance

Estimation

Potential Outcomes:

$m(t, z) = E[Y \mid T=t, Z=z]$

Conditional Effect:

$\square(z) = \partial/\partial t\ m(T=t, z)$

Data Generation

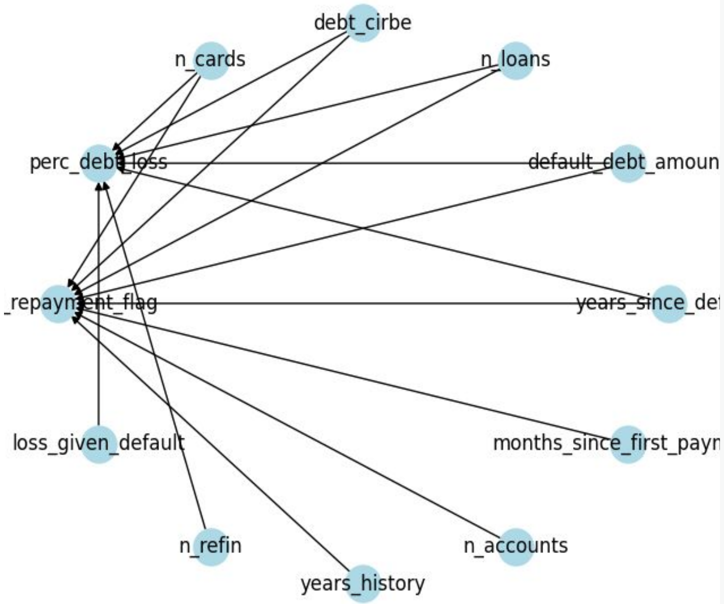
Synthetic dataset inspired in a real-world financial debt collection use case

- Real Financial variables distributions
- High-dimensional data with few controls
- Strong confounding bias
- Continuous Treatment [0, 100] and normally distributed
- Positivity Assumption violation
- Binary Outcomes sampled from conditional probabilities
- Heterogeneous Treatment effects
- Non-linearities and interactions

Data Generation

Synthetic dataset inspired in a real-world financial debt collection use case

X_{1i} = years since default
X_{2i} = default debt amount
X_{3i} = number of loans
X_{4i} = external debt
X_{5i} = number of cards
X_{6i} = loss given default
X_{7i} = number of refinances
X_{8i} = customer history length
X_{9i} = number of accounts
X_{10i} = months since first payment



C.1 Treatment assignment formula

The treatment value for each individual i is generated through:

$$T_i = \text{clip}\left(100 \cdot \frac{1}{1+e^{-\theta^\top X_i}} + \epsilon_i,\ 0,\ 100\right) \tag{3}$$

where the linear predictor $\theta^\top X_i$ is defined as:

$$\theta^\top X_i = 0.5X_{1i} + 0.4\log(1+X_{2i}) + 0.3X_{3i} + 0.3\log(1+X_{4i}) + 0.2X_{5i} + 0.3X_{6i} + 0.2X_{7i} + 0.1X_{1i}\log(1+X_{2i}) + 0.1X_{3i}^2 \tag{4}$$

and ϵ_i follows a truncated normal distribution:

$$\epsilon_i \sim \mathcal{TN}(0, \sigma^2 = 25, a = 0, b = 100) \tag{5}$$

C.2 Outcome generation formula

$$P(Y_i = 1 | T = t, X_i) = \begin{cases} 0 & \text{if } t = 0 \\ 1 & \text{if } t = 100 \\ \left(\frac{t}{100}\right)^{\exp(\eta_i)} & \text{otherwise} \end{cases} \tag{6}$$

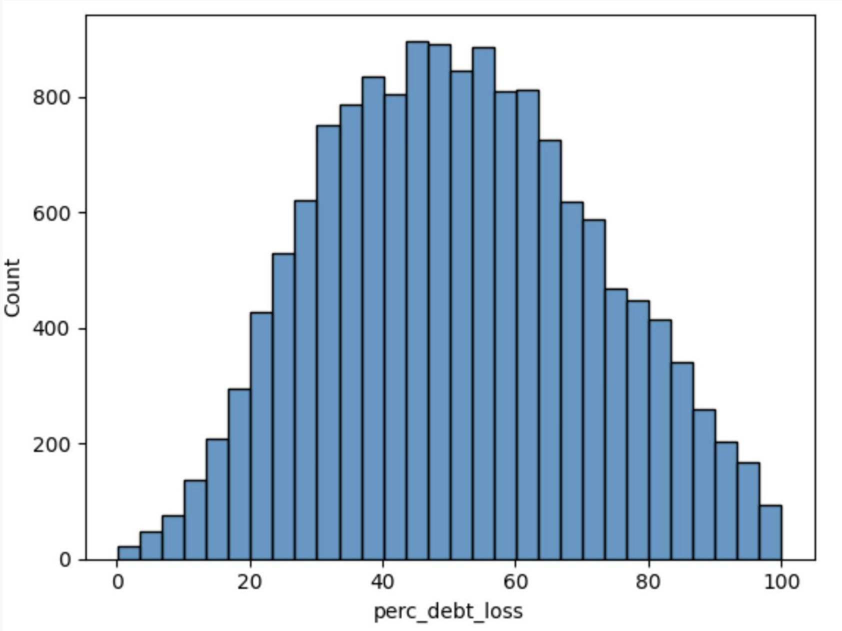
where the individual-specific coefficient η_i is computed through:

$$\eta_i = 0.6X_{1i} + 0.5\log(1+X_{2i}) + 0.5X_{3i} + 0.4\log(1+X_{4i}) + 0.3X_{5i} - 0.4X_{8i} - 0.3X_{9i} - 0.2X_{10i} + 0.1X_{1i}\log(1+X_{2i}) + 0.1X_{3i}^2 \tag{7}$$

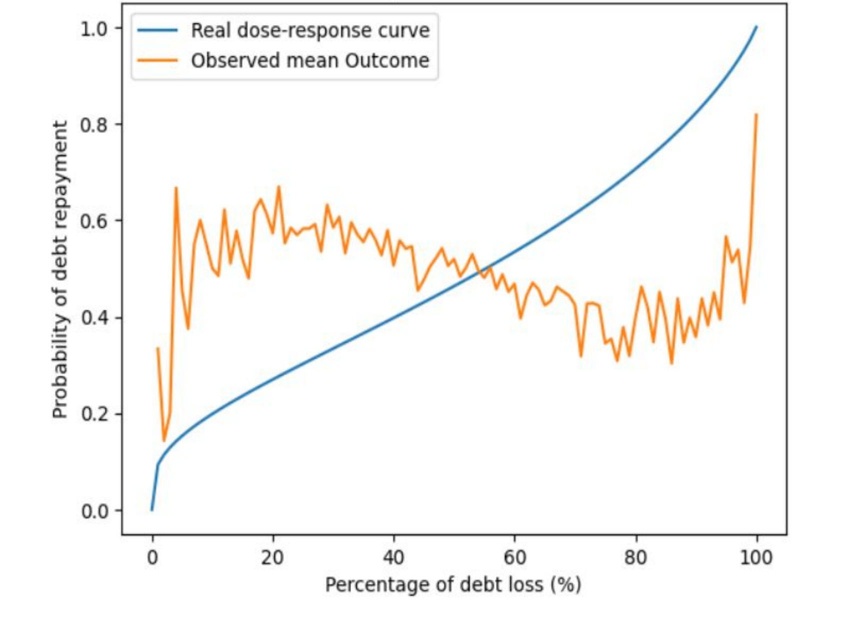
Data Generation

Synthetic dataset inspired in a real-world financial debt collection use case

Treatment Distribution



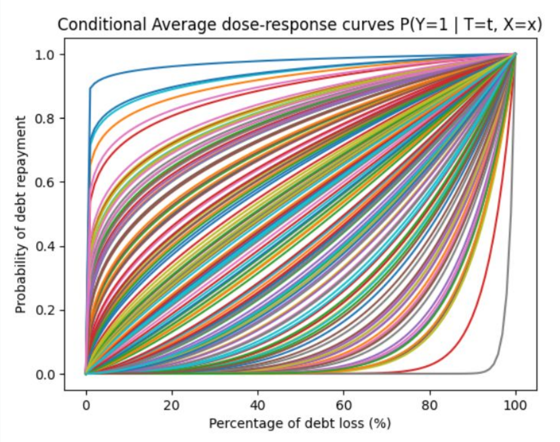
Mean observed Outcome vs Average Potential Outcome



Data Generation

Synthetic dataset inspired in a real-world financial debt collection use case

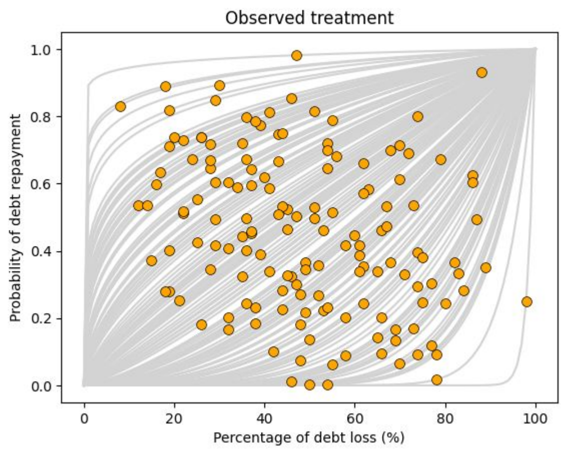
Real dose-response curves



$$P(Y_i = 1 | T = t, X_i) = \begin{cases} 0 & \text{if } t = 0 \\ 1 & \text{if } t = 100 \\ \left(\frac{t}{100}\right)^{\exp(\eta_i)} & \text{otherwise} \end{cases}$$

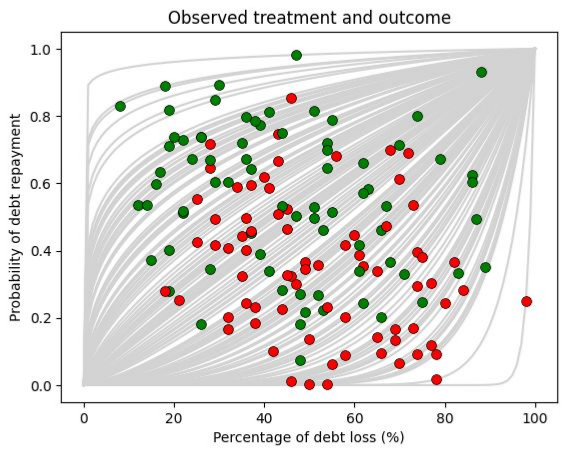
$$\tilde{p}_i = \text{clip}(P(Y_i = 1 | T = t, X_i), 0, 1)$$

Assigned observed Treatment



$$T_i = \text{clip}\left(100 \cdot \frac{1}{1 + e^{-\theta^\top X_i}} + \epsilon_i,\ 0,\ 100\right)$$

Observed Binary Outcomes



$$Y_i \sim \text{Bernoulli}(\tilde{p}_i)$$

Dimensionality Reduction & Identification

Dimensionality reduction

$\mathcal{D} = (X_i, T_i, Y_i)^N$
$X \in \mathbb{R}^d$ are covariates
$T \in \mathbb{R}$ is the Treatment
$Y \in \mathbb{R}$ is the Outcome

Dimensionality Reduction: We propose a two-stage selection framework to construct a reduced adjustment set (control candidates) $Z = Z_Y \cup Z_T$, where Z_T is a subset representing the Treatment predictors and Z_Y is a subset representing Outcome-relevant covariates.

Z_T $\cup$ Z_Y

Feature Selection for Predictive ML

$$T = f(X) + \varepsilon$$

*Methods: Recursive Feature Elimination, Sequential Forward Elimination, Permutation Feature Importance Filter, etc.

- **Potential Confounders:** $|\varrho(T, Y) - \varrho(T,Y|\, X_j)| > \epsilon$
- **Outcome-only predictors:** $\varrho(X_j, Y|\, T) > \epsilon$

*$\varrho(T,Y|\, X_j) = \varrho(\mathrm{Res}(T \sim X_j), \mathrm{Res}(Y \sim X_j))$
*$\varrho(X_j, Y|\, T) = \varrho(\mathrm{Res}(X_j \sim T), \mathrm{Res}(Y \sim T))$
where Res(·) denotes regression residuals

410 → 15 vars !

Dimensionality Reduction & Identification

Identification

$\mathcal{D} = (X_i, T_i, Y_i)^N$

$X \in \mathbb{R}^d$ are covariates

$T \in \mathbb{R}$ is the Treatment

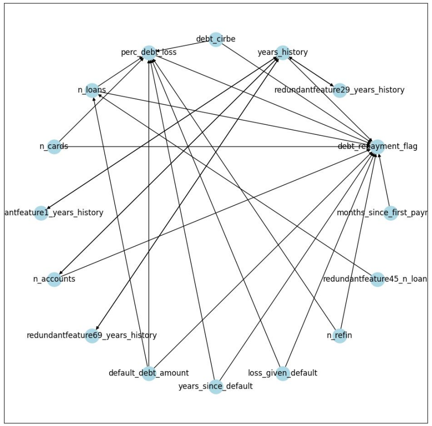
$Y \in \mathbb{R}$ is the Outcome

$Z \subseteq X$ are the controls candidates

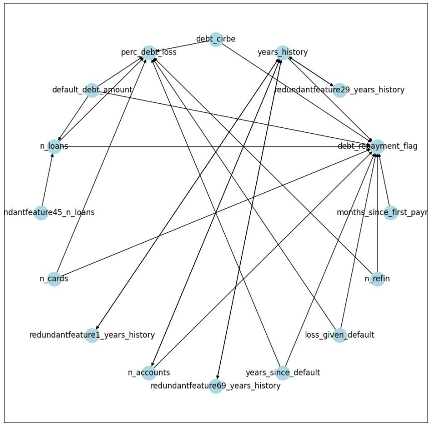
Causal Identification: ensemble of methods—Peter-Clark (PC), Fast Causal Inference (FCI), and Greedy Equivalence Search (GES)— over control candidates Z to increase confidence in consistently identified relationships.

Algorithmic outputs serve as <u>initial structural hypotheses,</u> iteratively refined through domain expertise: Edge Validation, Edge Direction, Missing Edges, Spurious Correlations, etc.

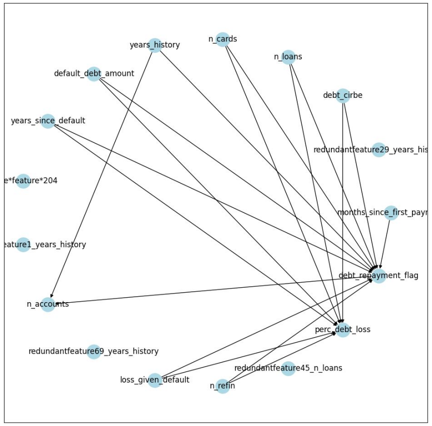
PC



FCI



GES



Addressing Positivity and Data Gaps

$\mathcal{D} = (X_i, T_i, Y_i)^N$
$X \in \mathbb{R}^d$ are covariates
$T \in \mathbb{R}$ is the Treatment
$Y \in \mathbb{R}$ is the Outcome
$Z \subseteq X$ are the controls

Detection: model-agnostic procedure to detect regions of the covariate space where lack of overlap violates the positivity assumption for a continuous Treatment based on Hirano et al. (2004) GPS framework.

01 Propensity Model

02 Conditional Density

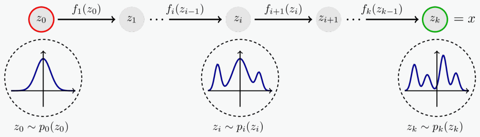
03 Overlap Diagnostics & Remedation

- **$f : Z \to \mathbb{R}$ that predicts T**
- **Out of sample residuals $\varepsilon = T - f(Z)$**

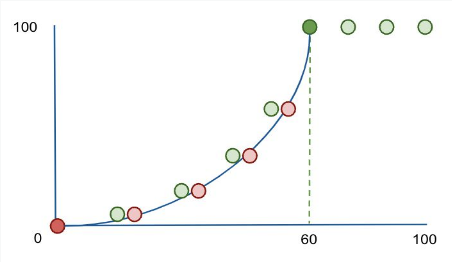
statsmodels

CatBoost

Model $g(\varepsilon \mid Z)$



$$P(T \in [t_1, t_2] \mid Z = z) = \int_{t2}^{t1} g(t - f(z) \mid z)\, dt.$$



Estimation & Evaluation with Continuous Treatment

1. **Regression adjustment with interactions:**

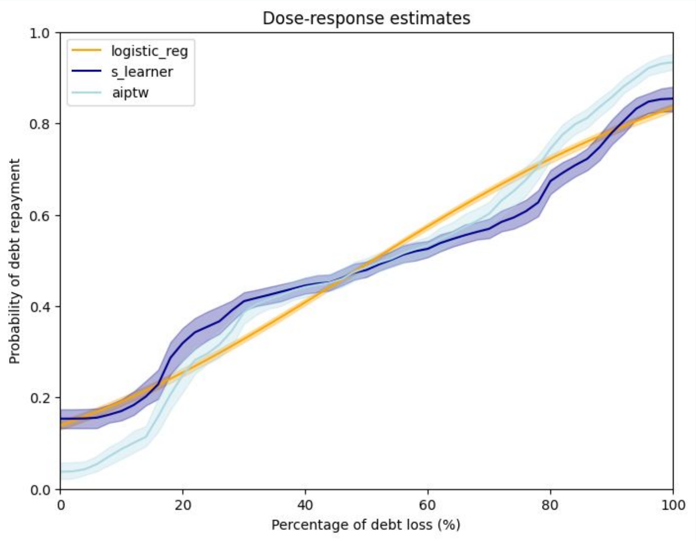
$$\hat{y}_i(t) = \beta_0 + \beta_1 t + \beta_2 Z_i + \beta_3 Z_i t$$

2. **S-learner:**

$$\hat{y}_i(t) = f(Z_i, t)$$

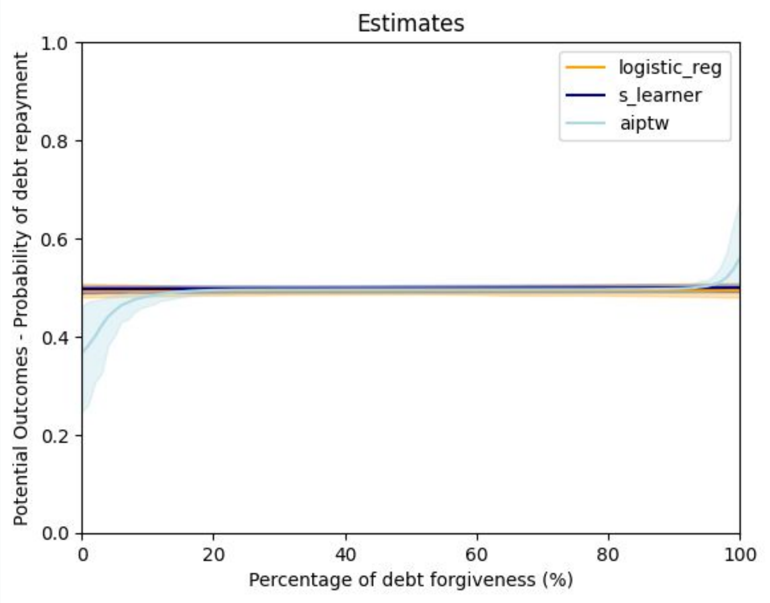
3. **Augmented IPTW:**

$$\hat{y}_i(t) = m(t, Z_i) + \frac{K(T_i - t)}{e(T_i, Z_i)}(Y_i - m(T_i, Z_i))$$

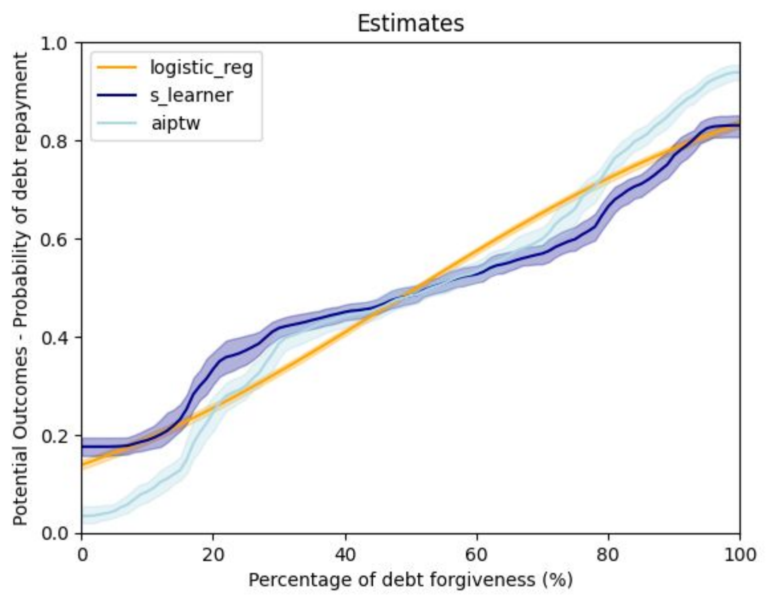


Estimation & Evaluation with Continuous Treatment

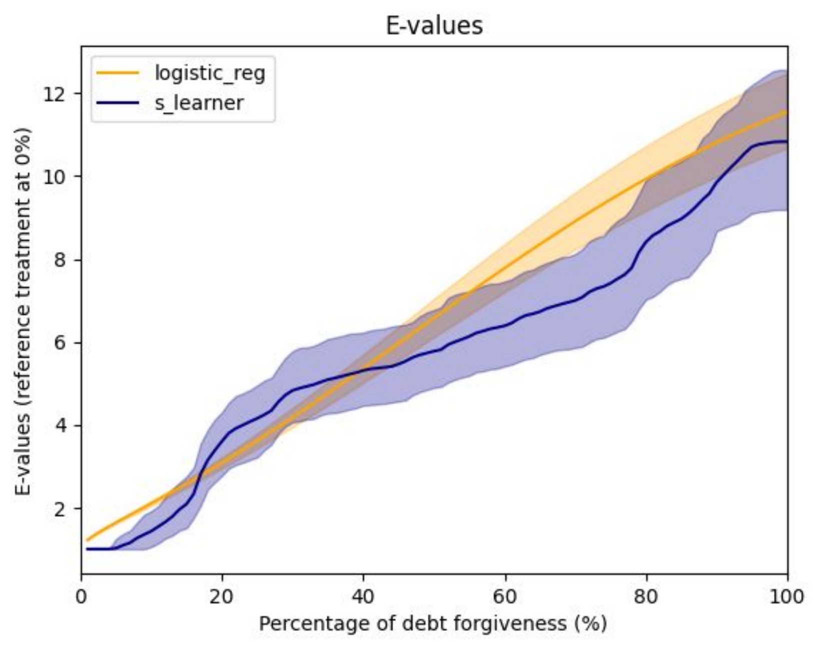
Placebo Treatment replacement

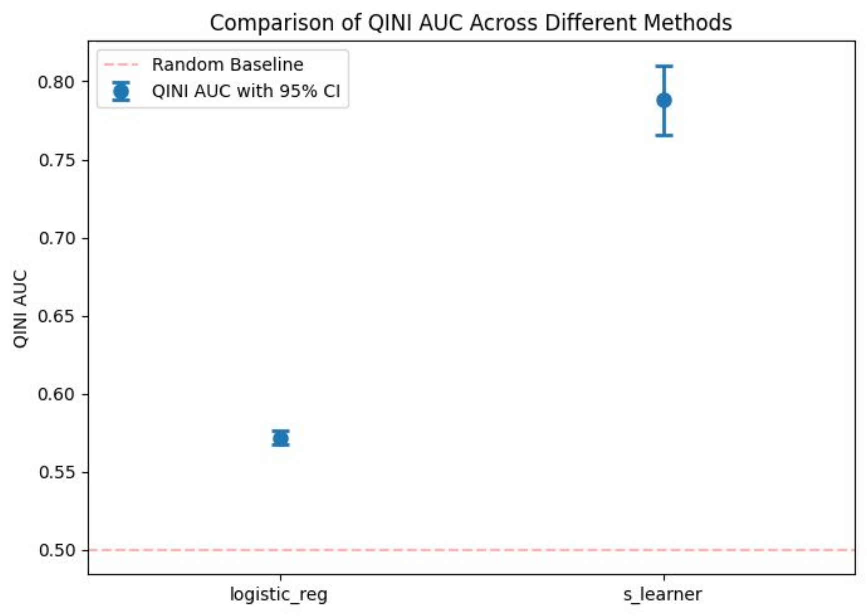


Random common causes



Estimation & Evaluation with Continuous Treatment

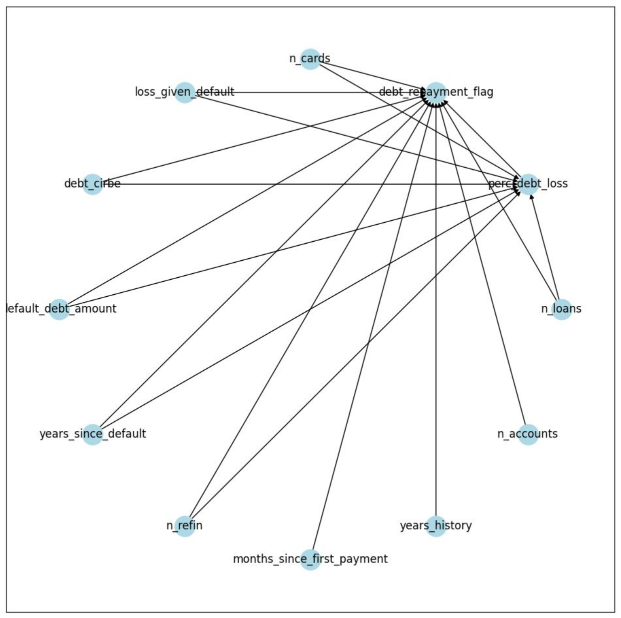




Ablation Study - Results

Table 1: Comparison of Variable Identification Performance

Method	Precision	Recall	Covariates	Runtime (min)
Baseline	0.05	1	169	>600
Dim. Reduction	0.53	1	15	3
Dim. Reduction & Identification	0.80	1	10	4



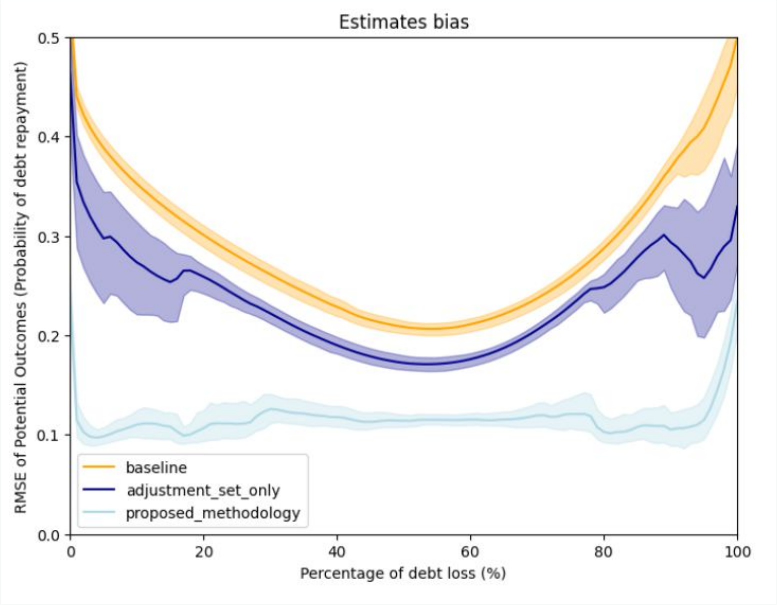
***Baseline**: ensemble of causal discovery methods (PC, FCI and GES) over the 410 covariates dataset

The proposed methodology improved precision from 0.05 to 0.80 compared to the baseline, yielding a set of 10 covariates, including all 8 true causal controls and 2 treatment-only related variables.

Ablation Study - Results

Table 2: Comparison of Estimation Bias Across Methodologies

Methodology	Mean Bias	95% Confidence Interval
Baseline	0.292	[0.279, 0.303]
Adjustment Set Only	0.236	[0.212, 0.257]
Proposed Methodology	0.117	[0.105, 0.131]



***Baseline**: S-learner adjusting for the 169 controls found in the baseline identification phase
***Adjustment Set Only**: S-learner adjusting for the 10 controls found in the dimensionality reduction and identification phase

The proposed methodology achieves superior performance with B=0.117 vs B=0.292 (baseline) representing a 60% reduction in mean bias compared to the baseline.

Conclusions & Takeaways

1. The proposed methodologies is better than standard causal ML pipelines at capturing the true controls, produces less biased estimates and archives a significant time reduction.

2. Many libraries and frameworks to address causal ML problems but gaps appear when applying it to concrete real use cases in industry

3. We propose:
   a. Dimensionality reduction + Identification for challenge 1
   b. Positivity violation diagnostics `remediation strategy for challenge 2
   c. Continuous treatment adaptation for challenge 3
   d. Pipeline for challenge 4

4. Feedback is welcome!

GitHub repo!

Questions?

BBVA