Causal Inference Libraries: What They Do and What I'd Like Them to Do

Agenda

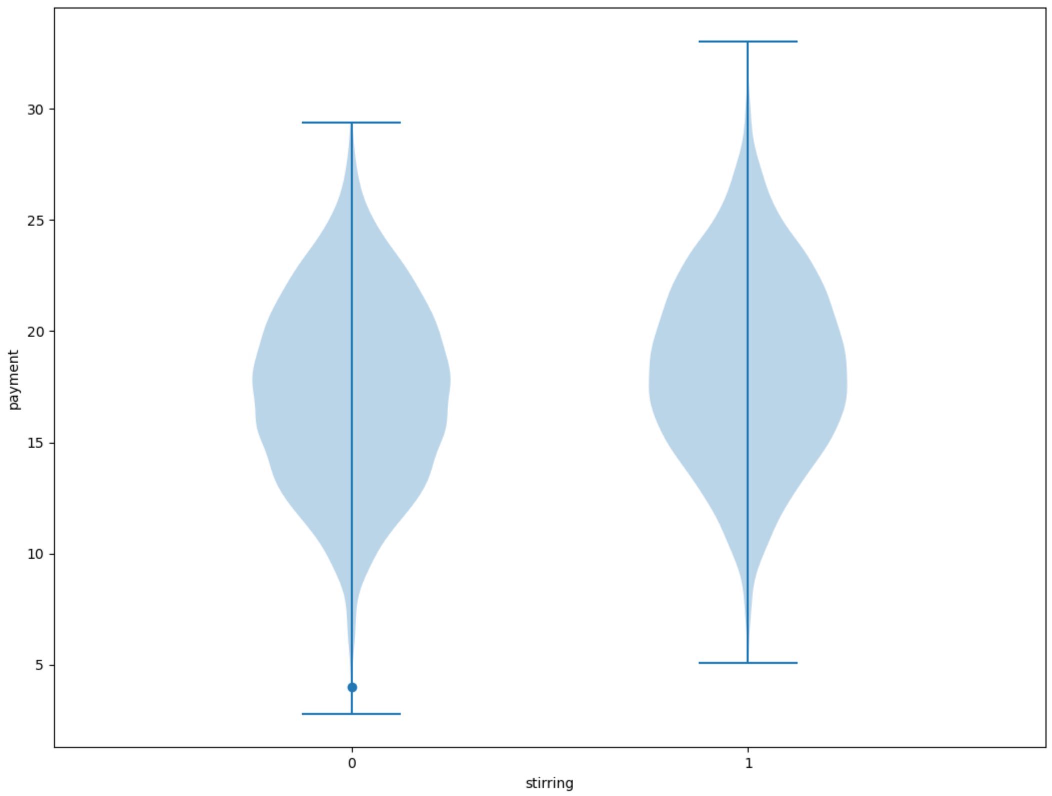
1. Why Causal Inference and why heterogeneity?
2. How can we estimate heterogeneous treatment effects on paper?
3. How can we estimate heterogeneous treatment effects in practice?
4. What are we missing from EconML and CausalML?

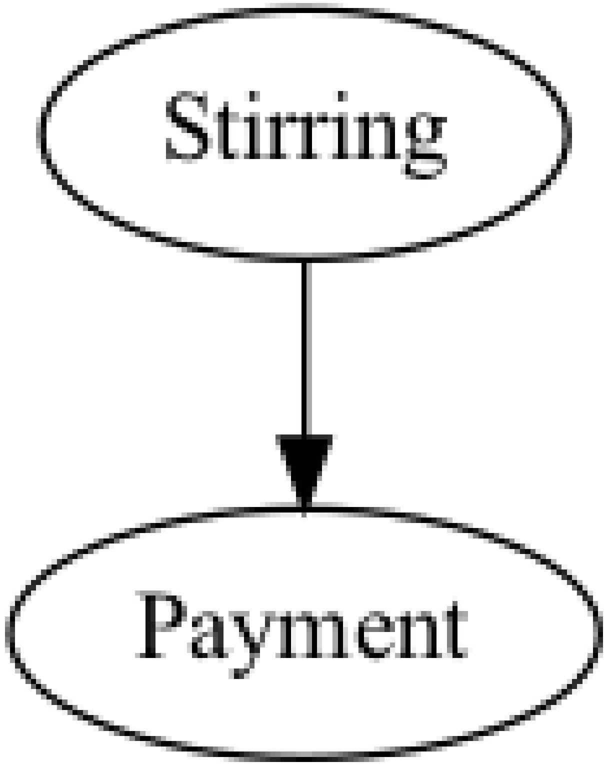
Risotto

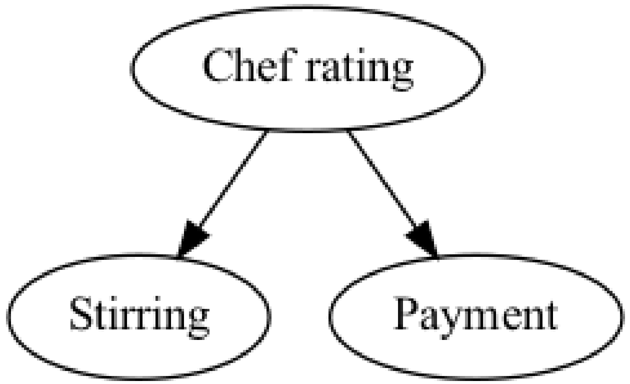
- Can either be prepared
    - in a laborous and delicate fashion, involving a lot of **stirring**
    - in a cut-throat, cantine style fashion, **not** involving a lot of **stirring**
- Consumers of risotto are **free to decide how much they pay** for their risotto.
- Naturally we wonder: should we be stirring?

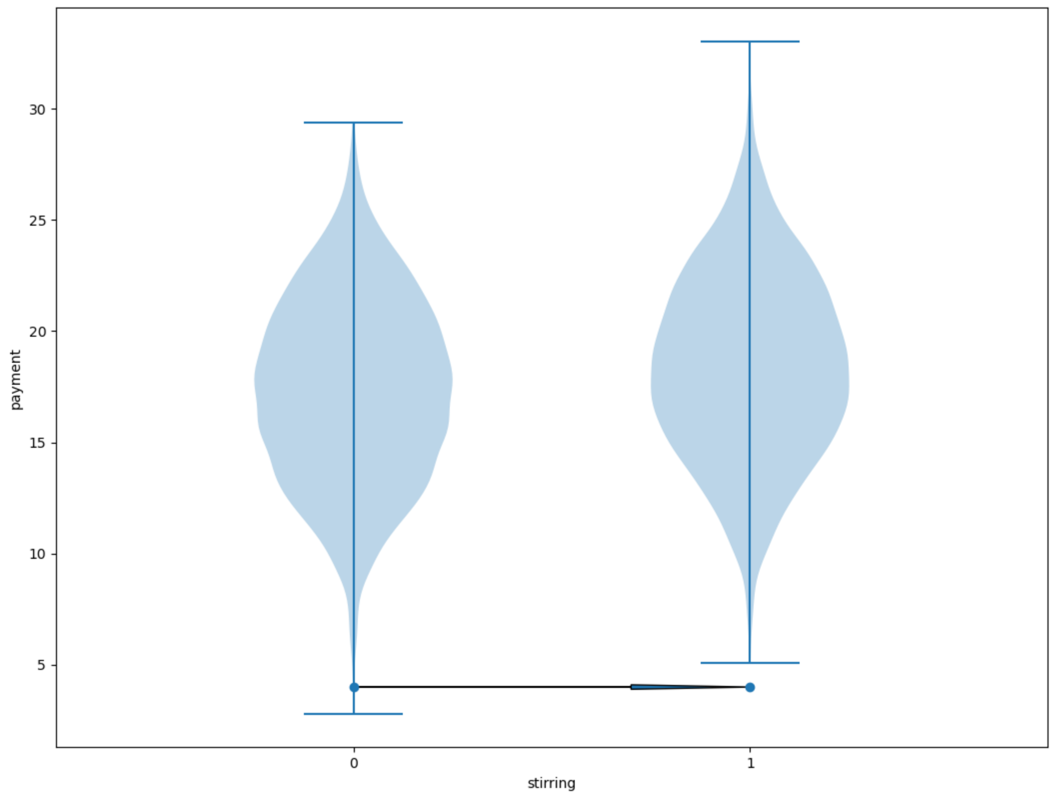
When prediction fails

What happens when we intervene on a data point from the left, i.e. `stirring = 0`, and now - keeping everything else unchanged - make sure that a gas stove is used, i.e. `stirring = 1`?

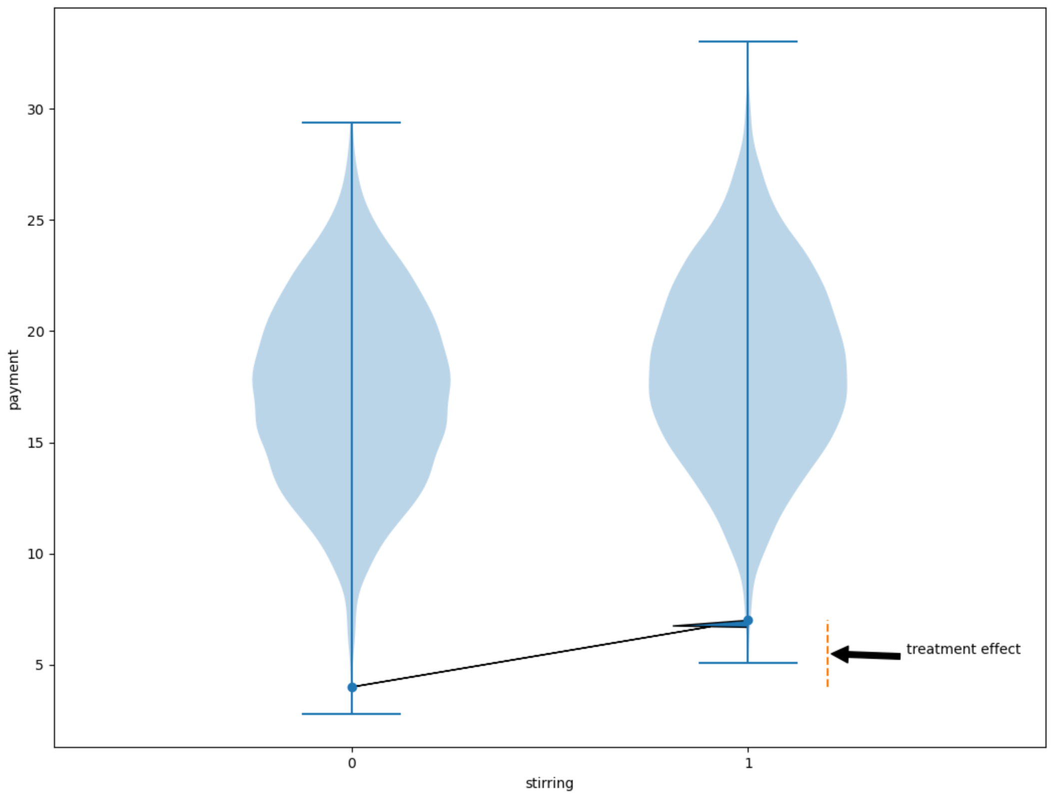



Stirring
Payment


Chef rating
Stirring
Payment

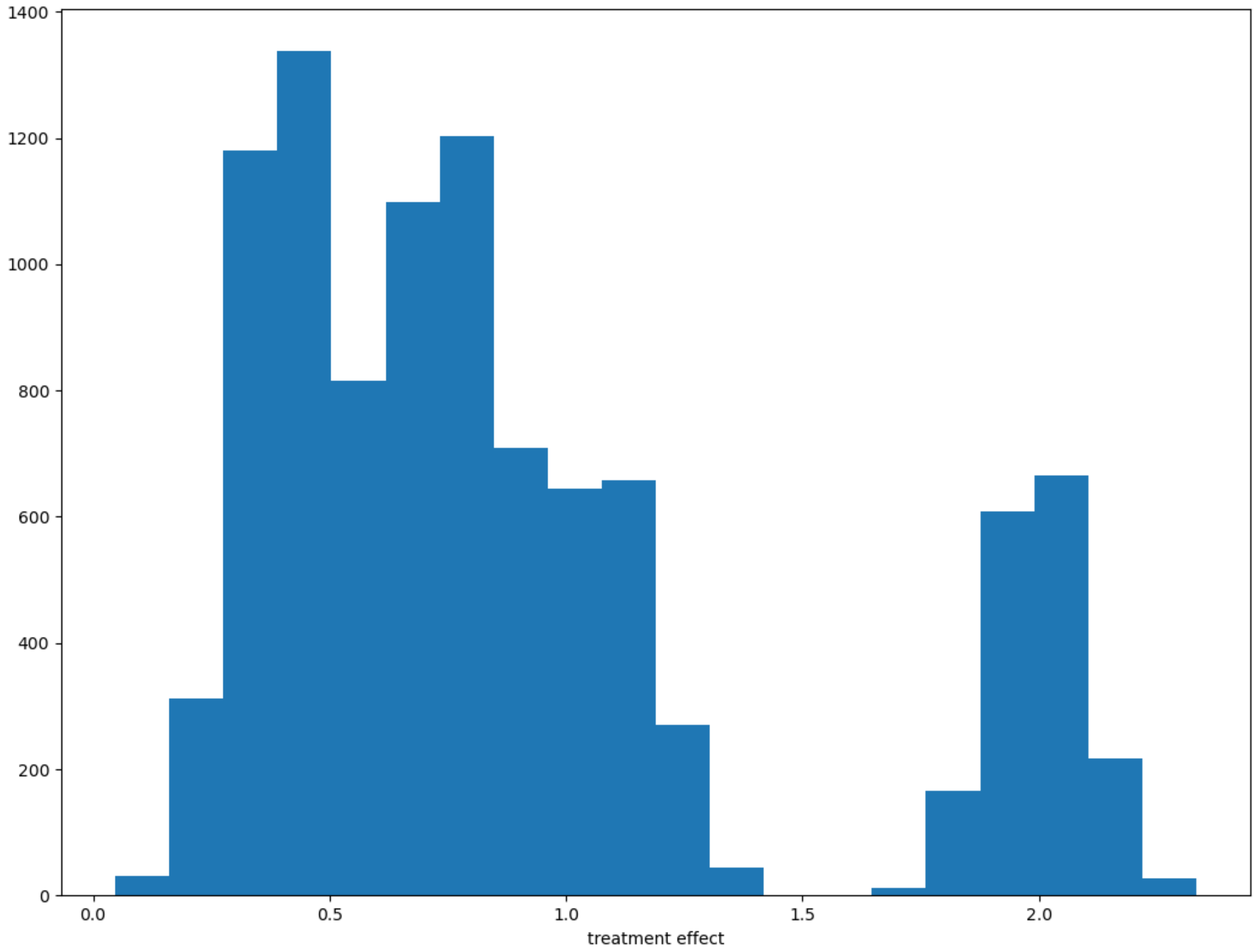


30
25
20
15
10
5
payment
0
1
stirring



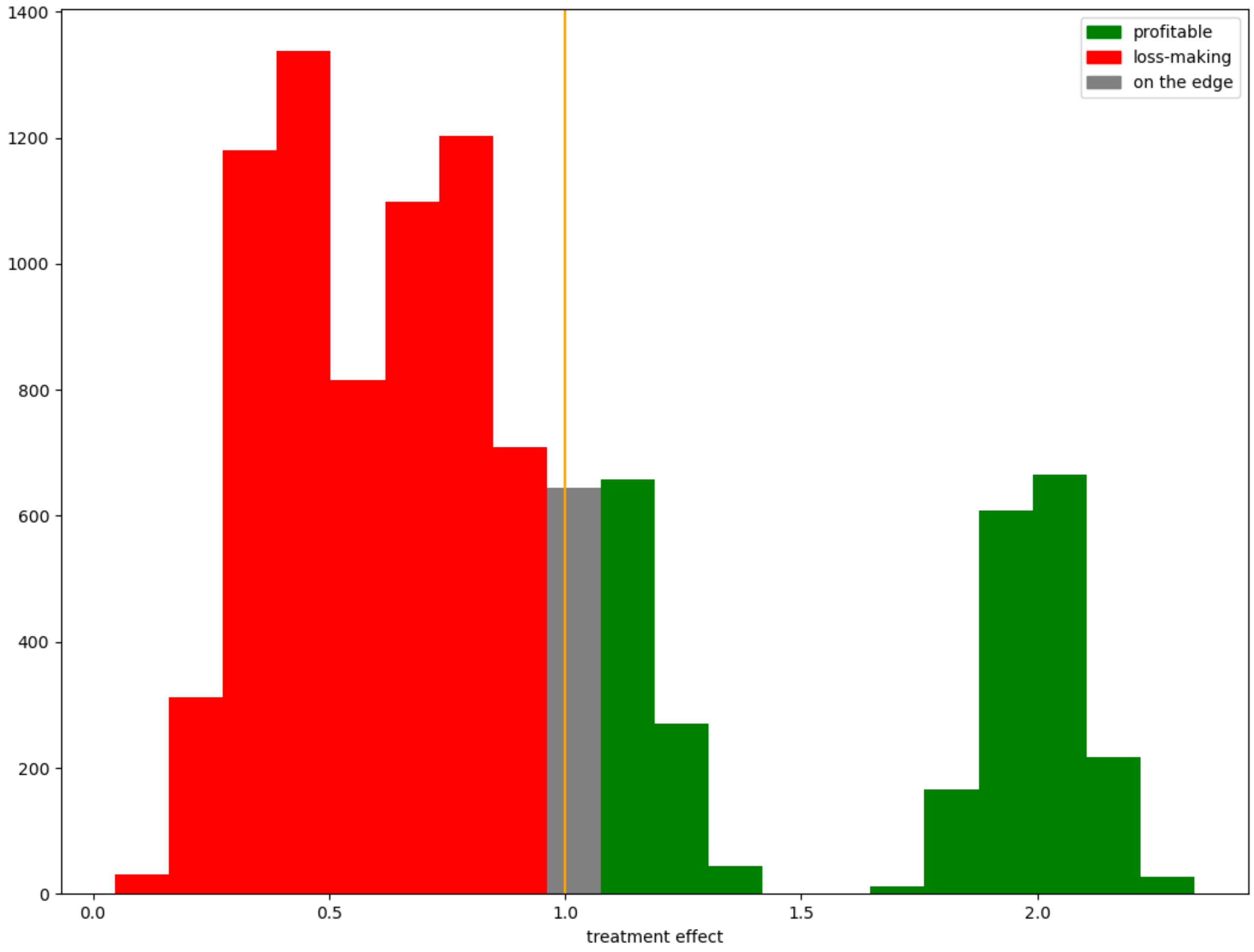
30
25
20
15
10
5
payment
0
1
stirring
treatment effect

Why heterogeneity


1400
1200
1000
800
600
400
200
0
0.0
0.5
1.0
1.5
2.0
treatment effect

To stir or not to stir, the math

- Assume that the cost of stirring amounts to 1$ per unit.
- Also assume that the revenue when never stirring is R.
- Then, the revenue when **always stirring** is $R - n \cdot 1 + \delta_1$
    - The plot from the previous slide tells us that $n \cdot 1 > \delta$.
- Revenue of **stirring when we expect it to pay off**: $R - k \cdot 1 + \delta_\pi$
    - We can condition on certain 'covariates'/features to decide for whom it pays off.
    - When doing this 'right', we get that $\delta_\pi > k \cdot 1$.


profitable
loss-making
on the edge
1400
1200
1000
800
600
400
200
0
0.0
0.5
1.0
1.5
2.0
treatment effect

Estimating heterogeneity on paper

The fundamental problem of Causal Inference

Desire

Consumer	Age of consumer	Non-stirred outcome/payment	Stirred outcome/payment	Individual treatment effect
Bob	28	21	21.8	.8
Anne	10	12	12	0

The fundamental problem of Causal Inference

Reality

Consumer	Age of consumer	Non-stirred outcome/payment	Stirred outcome/payment	Individual treatment effect
Bob	28	21	?	?
Anne	10	?	12	?

Conventional assumptions for estimating heterogeneous treatment effects

- Positivity/overlap
- Conditional ignorability/unconfoundedness
- Consistency

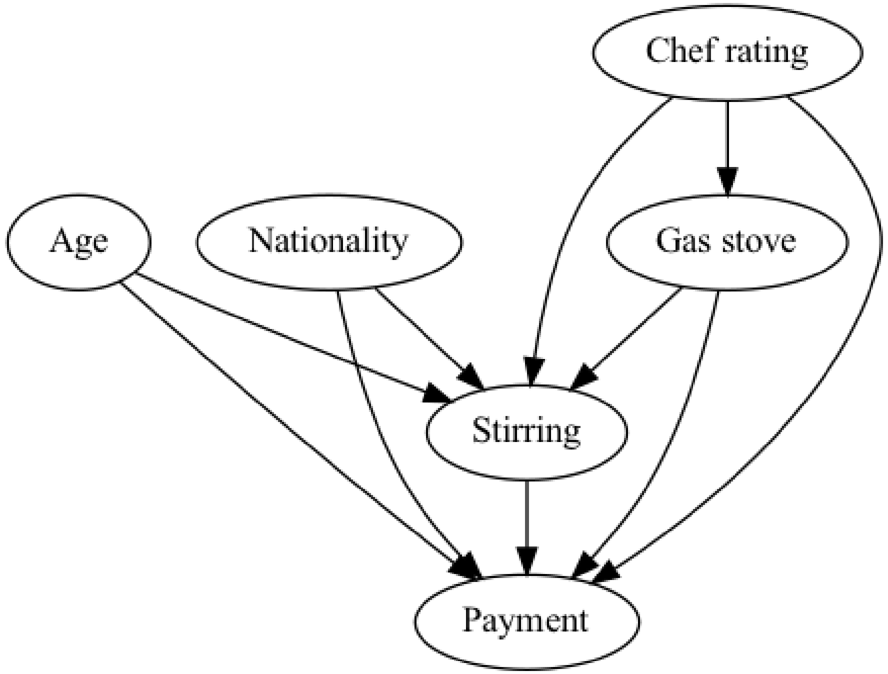
A randomized control trial usually gives us the first two for free.

What now?

- We still can't estimate the ITE.
- Yet, we can define an estimand, the Conditional Average Treatment Effect (CATE), which we can actually estimate:
  $\tau(X) := \mathbb{E}[\text{payment}|X, \text{stirring}] - \mathbb{E}[\text{payment}|X, \text{no stirring}]$

The DR-Learner

Estimating heterogeneity in practice



Risotto consumption: a simulation

Risotto consumption: a simulation

age	nationality	chef_rating	gas_stove	μ	T	τ	Y
50.77	Indonesia	0.53	1	20.73	1	0.34	21.08
59.48	Iraq	0.46	0	20.46	0	0.76	20.46
47.25	India	0.46	0	24.29	0	0.19	24.29
22.21	Italy	0.58	0	15.90	1	0.88	16.79

$Y \equiv$ the outcome, the final payment;

$T \equiv$ the treatment, whether the risotto has been stirred or not

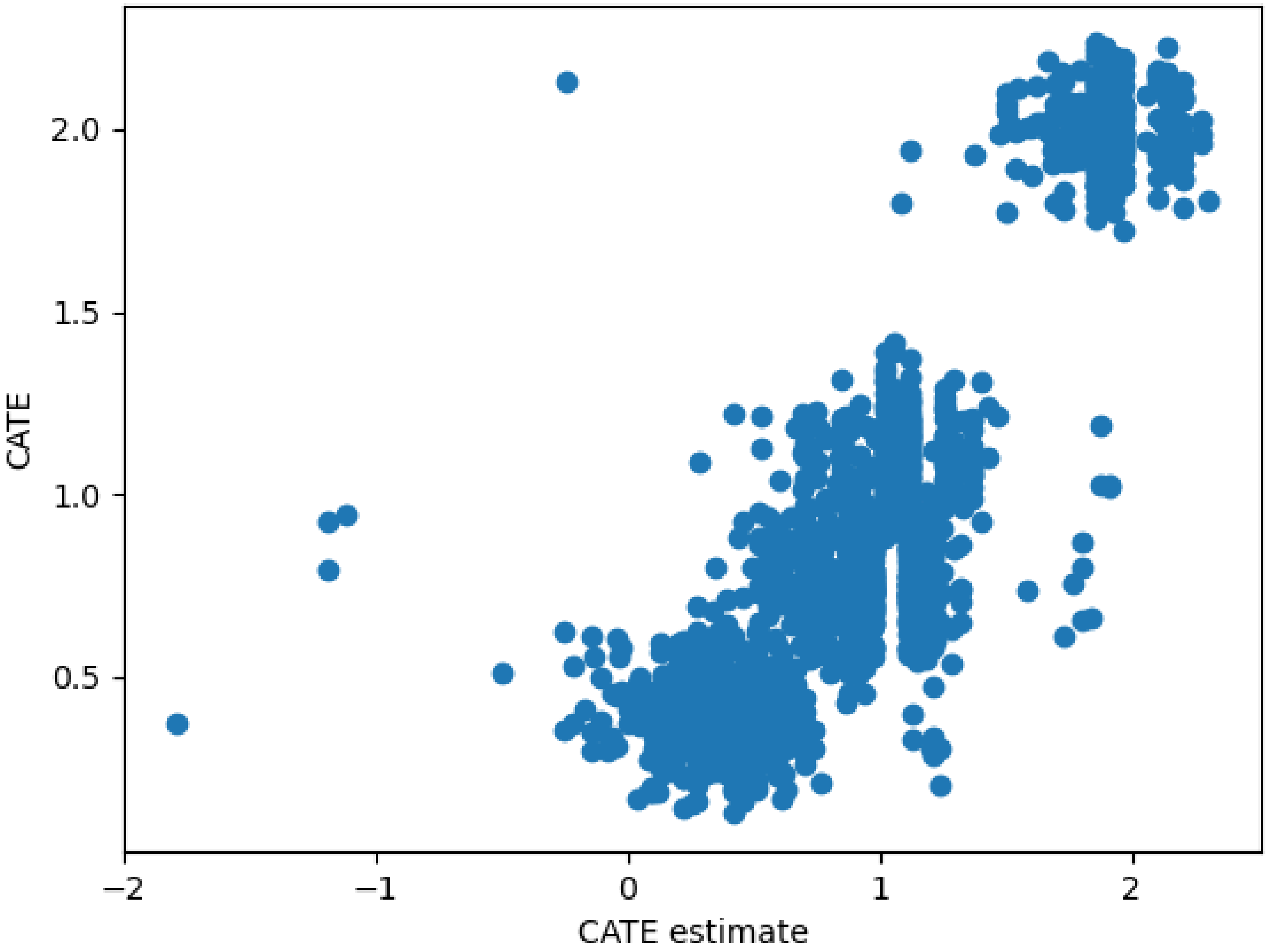
$\tau \equiv$ the treatment effect

$Y = \mu + T \cdot \tau$

```python
X = pd.concat([
        df[numerical_covariates],
        pd.get_dummies(df["nationality"])
], axis=1)

reg = lgbm.LGBMRegressor(verbosity=-1, num_leaves=4)
clf = lgbm.LGBMClassifier(verbosity=-1, num_leaves=4)
model = causalml.BaseRRegressor(
    outcome_learner=reg,
    effect_learner=reg,
    propensity_learner=clf,
)

model.fit(X=X, treatment=df[treatment], y=df[outcome])
cate_estimates = model.predict(X)
```



2.0
1.5
1.0
0.5
CATE
−2
−1
0
1
2
CATE estimate

TODO: Investigate how much the fit is better when using categorical hack