Not all treated are alike: estimating the proportionality of interventions with clustered dose-response functions

A. Cerqua[1], R. Di Stefano[2], R. Mattera[1]

[1]*Dep. of Social Sciences and Economics - Sapienza University of Rome*
[2]*Dep. of Methods and Models for Economics, Territory and Finance- Sapienza University of Rome*

MEMOTEF giornata della ricerca

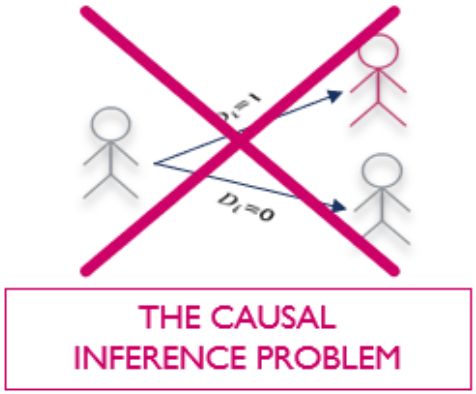
28 June 2023

An overview

- In the policy evaluation context, most applications that estimate causality from a counterfactual standpoint have focused on binary treatment scenarios (treat vs not treat).
- Evaluating causal effect when treatment is a **continuous variable**, thus assessing the proportionality of intervention, is a helpful tool in the hand of policymakers:
  - to evaluate **how the effect changes at different levels of treatment intensity**;
  - to estimate an **"optimal" level of intensity** that maximizes the impact.
- Despite this, the complexity of the continuous treatment framework makes the estimation process more challenging and limits the spread in empirical works.

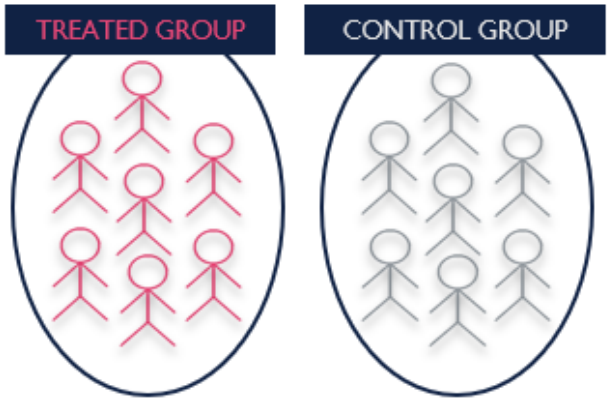
The binary case

" The causal effect of a treatment on a single individual or unit of observation is the comparison between the value of the outcome if the unit is treated and the value of the outcome if the unit is not treated" (Rubin, 1974)



missing data problem

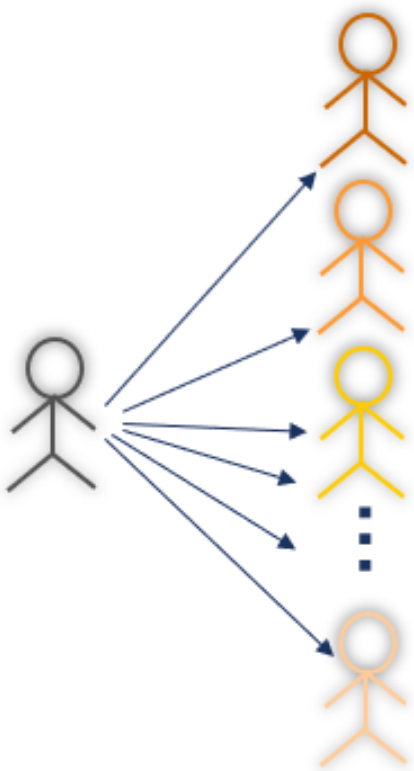
Solution



Estimate the **average effect of the treatment** for the population: compare the difference in means between the **treated group** and a **control group** that has a similar distribution of observed and unobserved characteristics.

When randomized experiments are not possible, quasi-experimental methods can be used to correct selection bias.

The continuous case



- The framework of potential outcomes becomes more complicated: potentially each unit could have been exposed to any treatment level.
- But we only observe one level for each unit.

Framework

- Units: $i = 1, ..., N$
- Treatment: $t \in \mathsf{T} = [t_0, t_1]$
- Potential outcome: $Y_i(t)$
- Estimand: Average dose-response function

$$\mu(t) = E[(Y_i(t)]$$

That function relates the potential average outcomes to the treatment intensity variable.

- Observed treatment: $T_i \in [t_0, t_1]$
- Vector of observed pre-treatment covariates: X_i
- Potential outcome corresponding to the level of observed treatment: $Y_i(T_i)$

The average dose response function

Causal effects are comparisons among potential outcomes in the same population of units, but the regression curve may represent different populations at different values of treatment.
For this, we re-balance the sample over the observed confounders and then estimate the DRF.

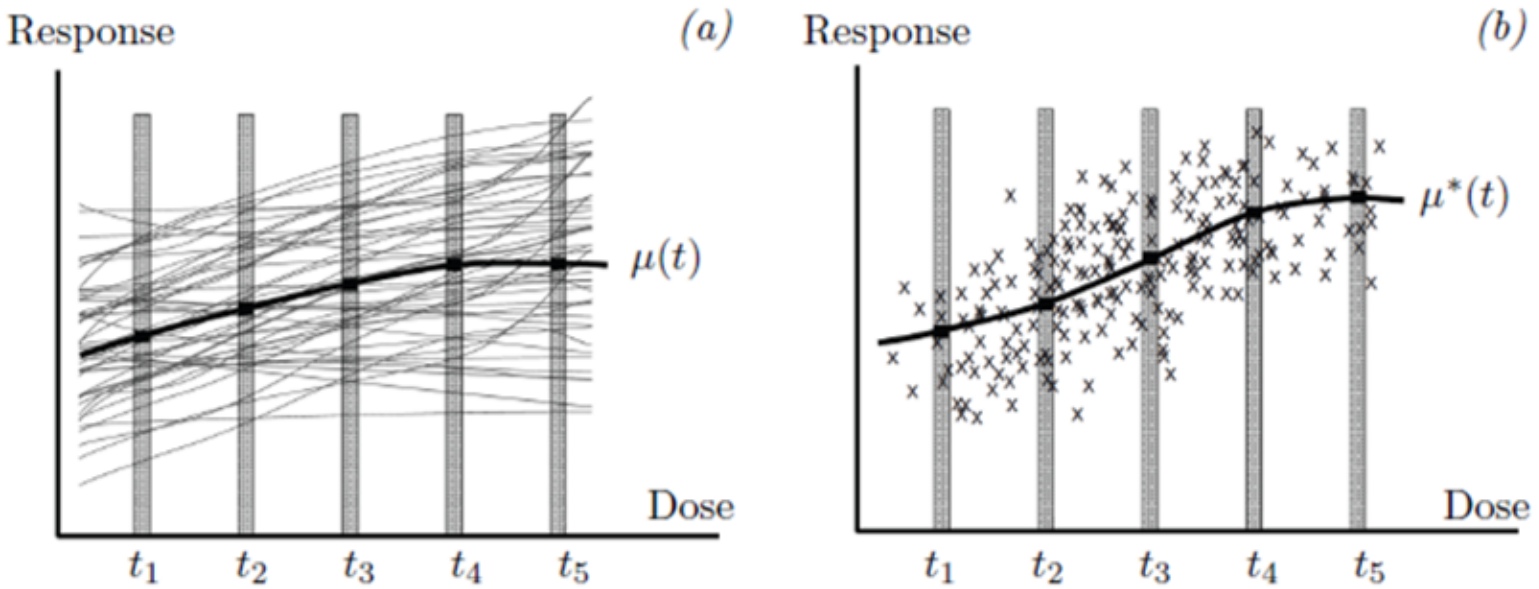


Figure 1.4: (a) Average dose-response function, $\mu(t) = E(Y_i(t))$, and (b) regression relationship between treatment and observed outcomes, $\mu^*(t) = E(Y_i(t) \mid T_i = t)$.

The generalized propensity score

- Hirano and Imbens, 2004 propose the generalized propensity score (GPS). The **idea** is to correct the selection bias into different treatment levels by comparing units that are similar in their observable characteristics.
  The GPS is a score estimated to correct for treatment endogeneity in the outcome model.

1. **TREATMENT MODEL**
   - Estimate the GPS, i.e., the conditional density of the treatment given the covariates $(r(t,X))$.
2. **OUTCOME MODEL**
   - Estimate the conditional expectation of Y on T and R, i.e.,

$$\beta(t,r) = E[Y \mid T = t, R = r] \ \forall t$$

   - Estimate the average dose response function, i.e.,

$$\mu(t) = E[\beta(t, r(t,X)) \ \forall t]$$

Assumptions

1. **Stable Unit Treatment Value Assumption (SUTVA)**
   no interference between units, no multiple versions of the same treatment.
2. **Overlap**
   $$r(t,x) = f_{T|X}(t\ x) > 0,\ \forall t \in T$$
3. **Weak unconfoundedness**
   $$Y(t) \perp T|X,\ \forall t \in T$$
   It requires conditional independence for each potential outcome $Y_i(t)$ rather than joint independence of all potential outcomes.

Theorem: If the assignment mechanism is weakly unconfounded given X, then it is also weakly unconfounded given the GPS.

$$Y(t) \perp T|r(t,X),\ \forall t \in T$$

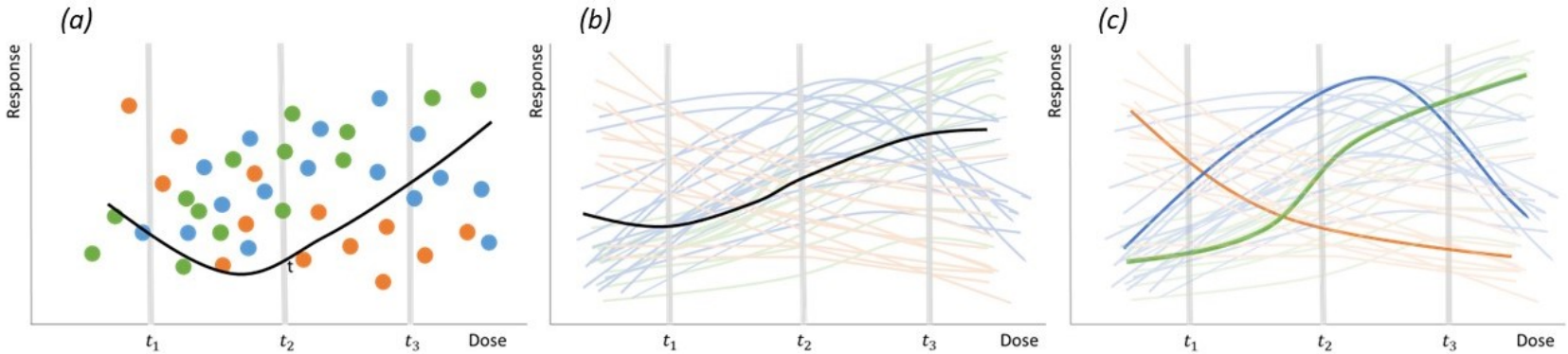
In most empirical applications:

- the uncounfoundedness assumption rarely holds;
- the observed covariates could hardly capture the heterogeneity of units;
- it relies on the correct specification of the parametric models of GPS and outcome.

(Kennedy et al. 2017, Fong et al. 2018, Yiu and Su, 2018, Tübbicke 2022, Vegetabile et al. 2021, relax parametric assumptions by directly optimising certain features of the weights rather than explicitly modelling the GPS.

Clustered GPS: the intuition

If unit heterogeneity is not highly captured by observed covariates, putting together all units and estimating a unique DRF could take imprecise and irrelevant policy implications.
We propose to incorporate clustering techniques into the GPS model to identify more interpretable treatment groups and provide a more robust and reliable framework.



$\mu(t) = E(Y_i(t)|T_i = t)$ $\mu(t) = E(Y_i(t))$ $\mu_g(t) = E(Y_i^g(t))$

Clustered GPS

Considering

$$T_i|X_i \sim N(h(\theta^T, X_i), \sigma^2),$$

by maximum likelihood or OLS, we can estimate the parameters θ^T and σ^2 and then estimate the GPS.

If θ^T changes over units, there is **(spatial) heterogeneity**. Without any structures for θ^T, we cannot identify these parameters since a repeated measurement on the same unit is not available in a cross-sectional framework.

Hence, we assume that n units are divided into G groups, and units in the same group share the same parameter values of θ^T.

Clustered GPS: how to estimate

1. **TREATMENT MODEL**
   Our procedure builds on those proposed by Sugasawa and Murakami (2021).
   Let's assume that N units are possibly clustered into G groups.
   1. An initial suitable partition is considered (covariates similarity or spatial contiguity).
   2. G clusterwise regressions of the covariates on the treatment levels are adopted.
      Considering the set of parameters within each $g = 1, ..., G$ cluster, we estimate GPS for each i-th unit i = 1,..., N.
   3. The procedure iteratively re-assigns each i-th unit to the g-th cluster if the likelihood obtained with the GPS is maximum.
   4. The procedure repeats until convergence or if the clusters' composition does not change in the subsequent iteration.

2. **OUTCOME MODEL**
   Estimate DRFs within clusters.

Clustered GPS

This has three main advantages:

- reduces p-hacking since the clusters are found in accordance with a data-driven procedure;
- identifies transparent groups allowing for a clear examination of the units within the group and facilitates a more accessible interpretation of dose-response functions;
- enhances the assurance of unconfoundedness within clusters.

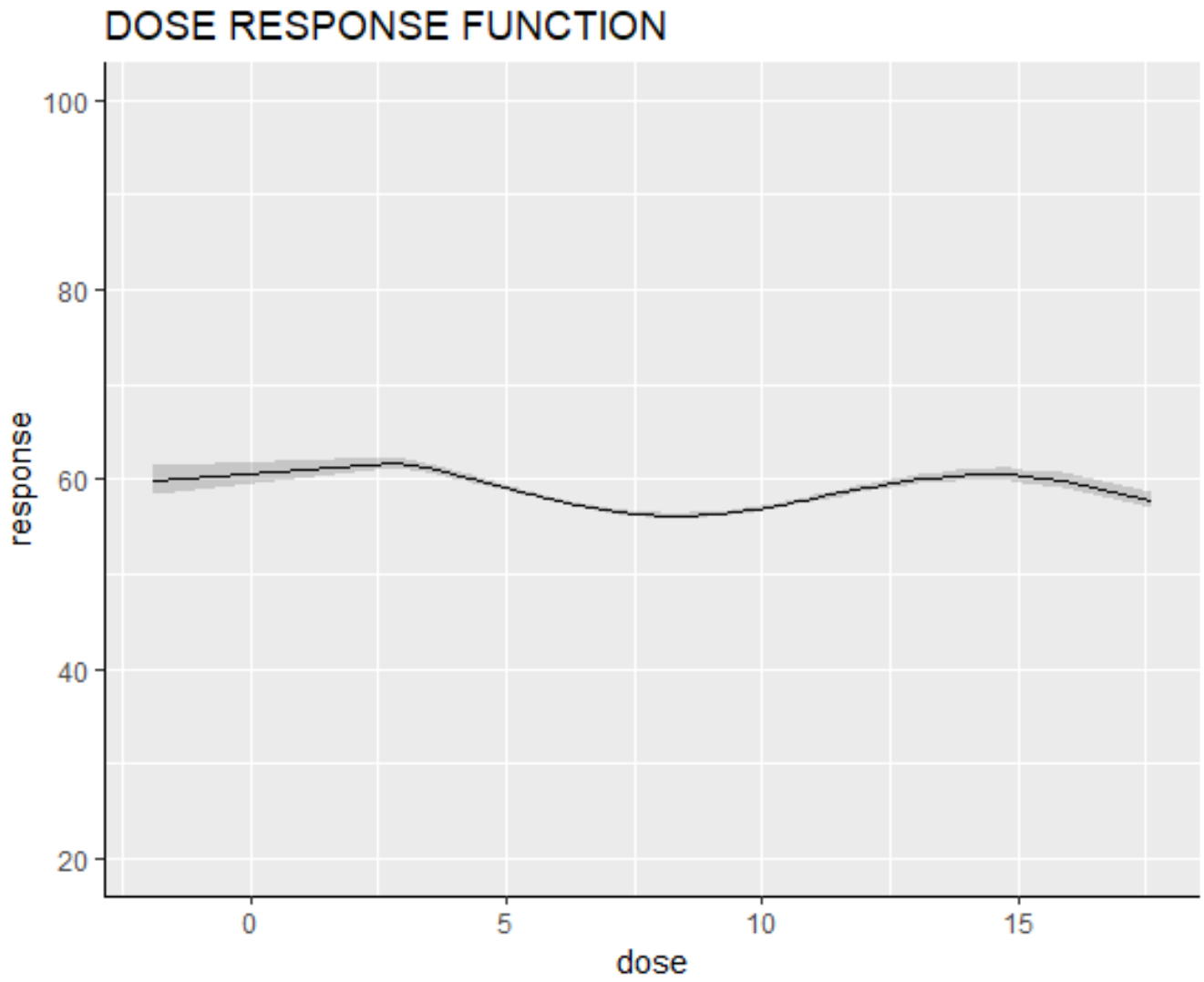
The impact of NPRR on voting

In 2023, 595 municipal elections were held in the ordinary statute regions in Italy. Of these, in 344 cases the outgoing mayor ran again (of which in 44 it was the only candidate).

AIM $\Rightarrow$ Estimate the impact of the National Plan of Recovery and Resilience funds on municipality elections voting.

OUTCOME: Percentage of voting received
TREATMENT: Per capita share of the NPRR funds assigned to the municipality
COVARIATES: region of belonging, population and population 65+, per capita income, Gini index, % waste sorting, turnout (%) in the 2022 national elections.

The impact of NPRR on voting



Obtained clusters: map

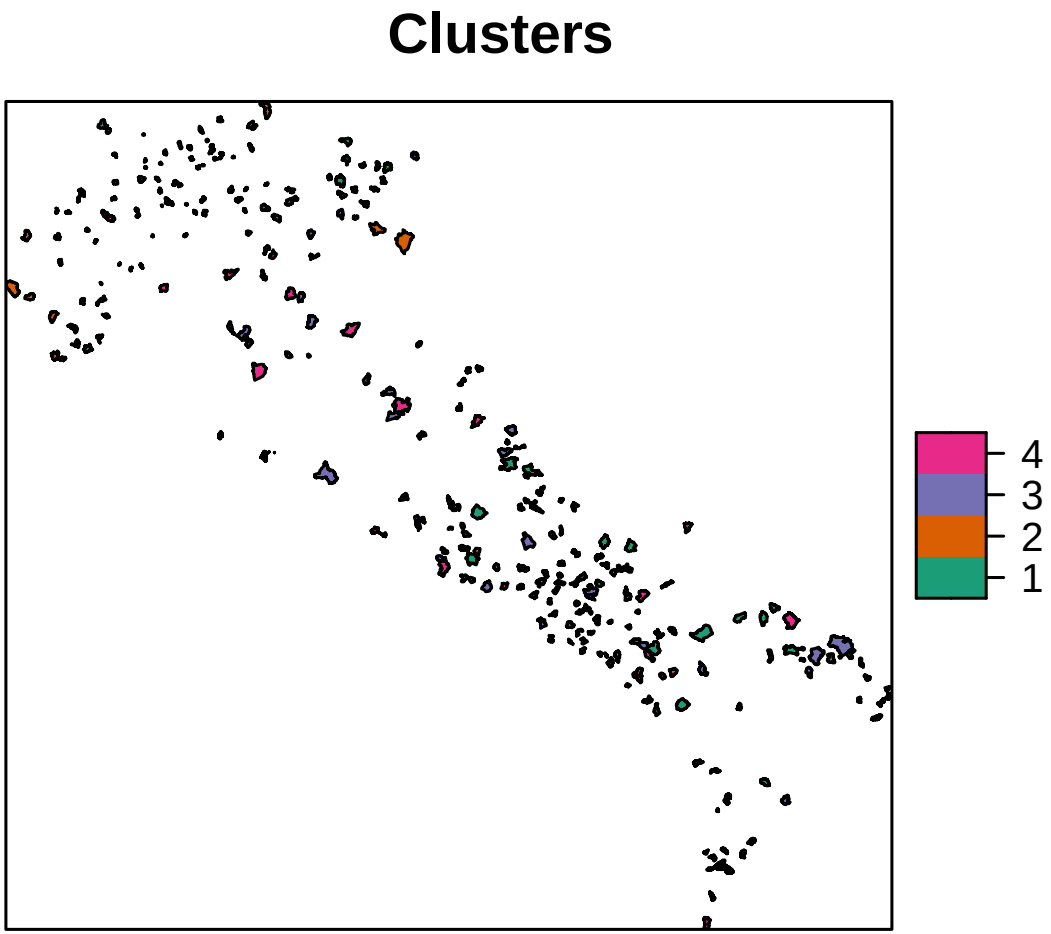
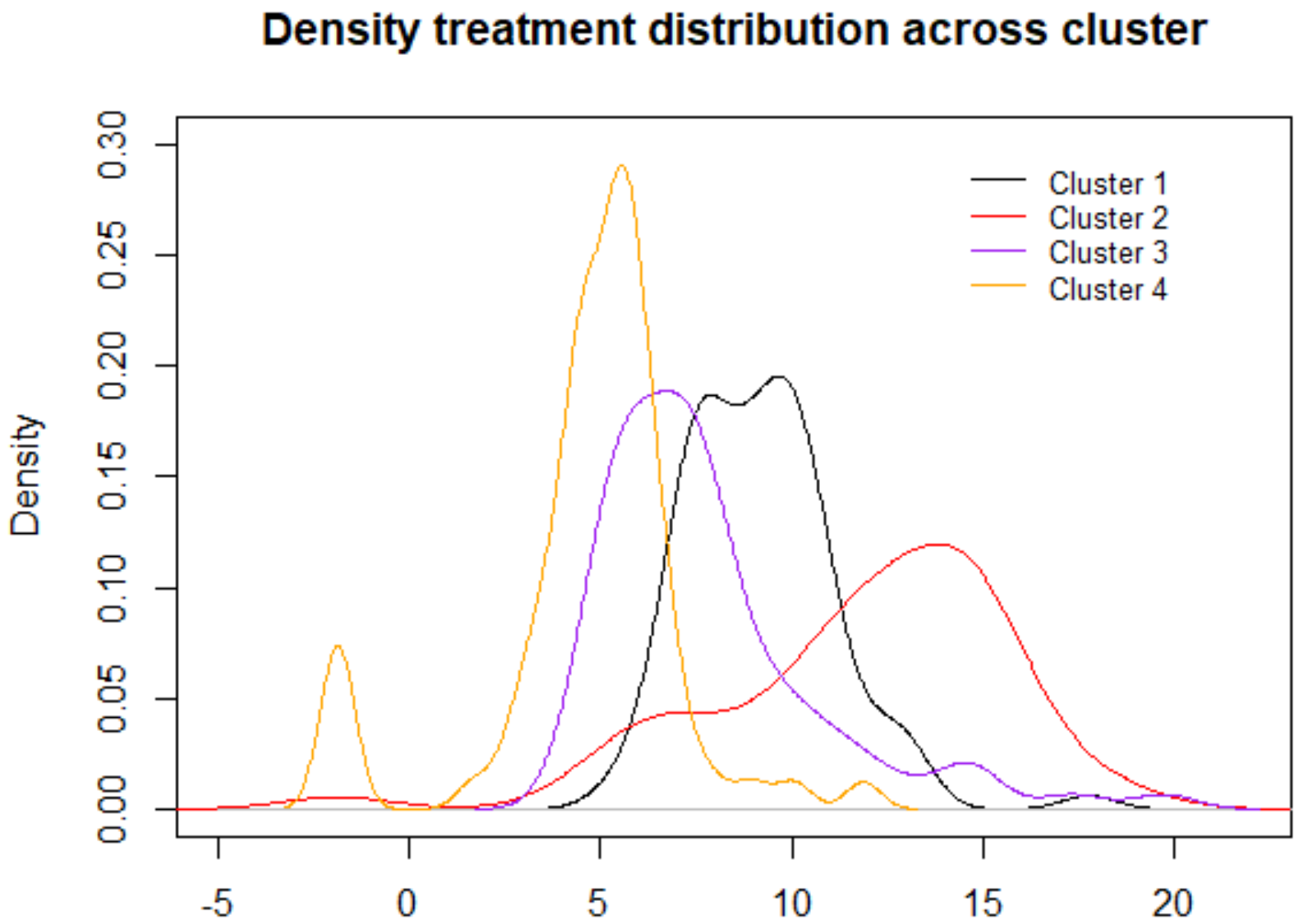


Figure 1: Clutering implied by the clustered DRF

The impact of NPRR on voting



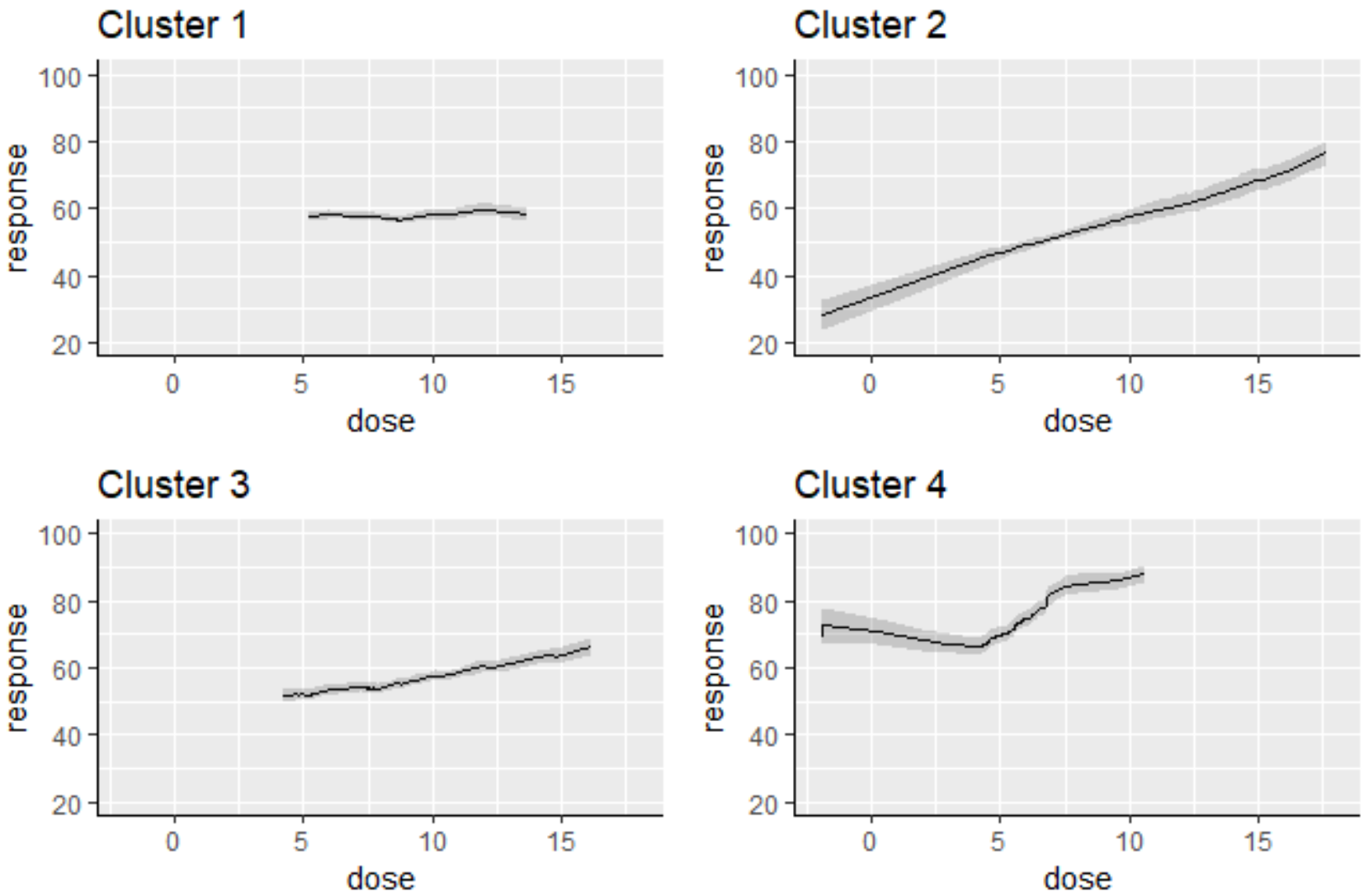
Clustering results: average values at group level

Variable	$G=1$	$G=2$	$G=3$	$G=4$
Population	8674.37	7053.80	9185.41	7648.38
Income	17861.06	17631.36	19359.84	19797.79
Gini	0.40	0.40	0.39	0.41
percentualerd2021	62.83	62.14	69.80	66.74
turnout_2022	60.41	62.50	63.64	64.36
$\text{pop}_1gen_65plus_2022$	2061.67	1631.52	2155.32	1918.31
Treatment	9.18	11.80	7.89	4.72
Outcome	55.36	55.01	57.59	60.90

GPS estimation: results from clustered regression

Variables	$G=1$	$G=2$	$G=3$	$G=4$
Constant	0.0004	29.8044***	-5.7819**	-7.7657*
Population	-0.0004	-0.0007***	-0.0000**	-0.0007**
Income	-0.0003***	-0.0005***	0.0001*	0.0003**
Gini	31.0866***	-42.4836***	45.3983***	-20.1056***
percentualerd2021	0.0356***	0.0249	-0.1683***	0.0079
turnout$_2$022	0.0293*	0.0615	0.1045***	0.1710***
pop$_1$*gen*$_6$5*plus*$_2$022	0.0018***	0.0028**	-0.0001	0.0029***
adj-R^2	0.79	0.61	0.86	0.44

The impact of NPRR on voting



Conclusions

- We present a novel procedure to estimate the causal effect of a continuous treatment framework that takes into account the unit heterogeneity.

- The CGPS has different advantages:
  - it relaxes the unconfoundedness assumption of the unit within clusters;
  - it increases the interpretability of dose response function because it identifies estimators policy relevant;
  - it reduces p-hacking since the clusters are found in accordance with a data-driven procedure.

Thank you for your attention!