Counterfactual Analysis Based on Grouped Data: Application to Poverty and Material Deprivation.

Minghai Mao*
Liaoning University

Antonio Raiola[1]
Universidad Carlos III de Madrid

October 25, 2023

Abstract

We propose to estimate the counterfactual decomposition of distributional statistics using aggregated data. The counterfactual estimation is a two-step procedure: first, we aggregate the data, and then we evaluate a mean (or regression) within each group, re-weighting them to account for the group's composition relative to a reference population. Our approach enables a *path-independent* detailed decomposition, highlighting the contributions of each group and characteristic. We recommend using a data-driven method to create the groups. The estimator constructed in this manner can handle challenging settings characterized by numerous non-ordered characteristics and data sparsity. We apply the newly proposed methodology to decompose the Great Recession's effects on Spain's poverty indices. The results suggest that variations in composition were the primary driver of the increase in poverty rates. While labor market disruptions pushed the rates higher, demographic changes and migratory movements had the opposite effect.

Keywords: Counterfactual Decompositions; Machine Learning; Partitioning Estimation; AROPE; Poverty Analysis.

*Liaoning University, China. Email: maominghai@lnu.edu.cn
[1] Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe (Madrid), Spain. Email: araiola@eco.uc3m.es

1 Introduction

In this paper, we propose to estimate counterfactual statistics using aggregated data. The method consists of a two-step procedure: initially, we aggregate the data using data-driven approaches or apriori knowledge about natural clusters in the data; then, we estimate the counterfactual functional by evaluating a mean (or a regression) in each group and re-weighting them for the groups' composition of a reference population. The approach builds upon the standardization technique commonly used in demography and epidemiology[1]. Neison [1844] introduced it to compare mortality rates across Great Britain districts taking into account different age compositions. In this context, the standardization technique consists of re-weighting the mortality rates of given age groups for the age composition of the same groups in a different district.

To decompose the difference between the functional evaluated in two different populations, we add and subtract the counterfactual to the difference. After collecting the terms, we obtain a two-components decomposition. One term is the difference in functionals we would observe if the populations differed only in the composition of the groups (Composition component); the other is the difference in functionals we would observe if there were no differences in compositions (Structural/Residual component).

A unique feature of this framework is the possibility of obtaining a "grouped" detailed decomposition. It is common in applied works employing parametric restrictions to further decompose the decomposition's components in the contribution of each characteristic, whereas doing the same is not straightforward under non-parametric specification. The decomposition with aggregated data provides a valid alternative by naturally defining the contributions of each group and each characteristic (under additional restrictions) to the two decomposition components. This information is helpful to understand the role of each group in the application at hand.

When ex-ante information about natural clusters in the data are unavailable, we propose to group the data with the Classification and Regression Trees (CART) algorithm, proposed by Breiman et al. [1984]. The groups formed with this approach are homogeneous in their contribution to the conditional mean function (CART). The homogeneity reduces

[1]See for instance: Ahmad et al. [2001], Jemal et al. [2005], Anderson and Rosenberg [1998],Shryock and Siegel [1980], Althauser and Wigler [1972]

the information loss implied by the data aggregation and makes the detailed decomposition easier to interpret. The CART algorithm can handle all types of data, including categorical, ordinal, and continuous variables; further, it is insensitive to outliers and adaptive to non-linearity and interactions.

Aggregating with CART allows estimating counterfactual decompositions in troublesome settings where the presence of many characteristics, often of unordered nature, and sparse data make it hard to employ other non-parametric solutions. The inequality of opportunity literature provides some instances of these settings where many circumstances affecting the individuals are considered[2].

We report Montecarlo simulations comparing the partitioning estimator to other parametric and non-parametric methods used to estimate counterfactual decompositions. The results show that the partitioning estimator under CART grouping has a similar or better performance than the other non-parametric methods considered in most of the simulated environments.

In the last section, we decompose the great recession effects on the poverty level of Spain. For the identification of individuals in poverty conditions, we use the concept of At-risk-of-poverty-or-social-exclusion (AROPE). The AROPE is a composite index taking value one if an individual is in one of the following conditions: income-based poverty (AROP60), material deprivation (MD), or low work intensity (LJ). The average of individuals in one of these conditions determines the relative rate. The vector of characteristics (or poverty risk factors) consists of socio-economic variables, including working status, educational level, age, sex, nationality, country of birth, household type, household size, household's head sex and age.

Our findings show that compositional variations are the main driver of the shock effects on the aggregate index (AROPE), the income-based measure (AROP60), and the low work intensity measure (LJ). Conversely, the residuals components of these indices decomposition are mostly negative, suggesting that after-recession poverty indices would have been lower than before the recession if no variation in composition occurred. Consistently with previous studies[3], we find that the risk factors of income-based poverty and material de-

[2] See Ramos and Van de Gaer [2021], Brunori et al. [2019], Brunori [2017] among others.

[3] Previous works (See Layte et al. [2001], Berthoud et al. [2004] and Ayala et al. [2011]) have already shown that only a small percentage of individuals that exit income-based poverty can leave conditions of

privation indices strongly differ. Given our choice of risk factors, the CART algorithm does not determine any useful split, and the counterfactual MD index is numerically identical to the index itself. Finally, we analyze the detailed decomposition of the AROPE 2008-2014 differential. According to the decomposition, groups impacted by labor market disruptions worsen the crisis' consequences by increasing their relative weight in the population and having a higher after-recession within-group AROPE rate. Instead, other mechanisms, such as demographic changes and migratory movements, appear to reduce the relative weights of high-risk groups, hence alleviating recession impacts.

2 Related Literature

The literature on counterfactual decompositions starts with Kitagawa [1955], which formalized the procedure introduced by Neison [1844] and proposed to employ the counterfactual rates for decomposing the difference in rates of two populations. Kitagawa [1964] extended the approach and the decomposition to statistics different from rates. It took a decade, before Oaxaca [1973] and Blinder [1973] (hereafter OB) introduced and popularized the counterfactual decomposition in economics. The two authors used it to decompose average wage differentials among social groups such as male/female and black/white under a simple linear model for the conditional mean wage. In their model, the residual component of the decomposition identifies the degree of discrimination in the market. Moreover, the linearity assumption allows retrieving detailed decompositions evidencing the contribution of each characteristic on the two components.

In the last thirty years, several authors proposed a variety of methods extending the decomposition to general distributional statistics, other than the mean. These methods possess high flexibility due to their non-parametric or semi-parametric specification of the underlying distribution function. DiNardo et al. [1996] employ non-parametric regression mixed with a specification of the conditional density to obtain counterfactual distributions through re-weighting. Machado and Mata [2005] use semi-parametric quantile regressions and simulation methods, Rothe [2010b] estimate the distribution and any other functional of it with kernel regressions, while Chernozhukov et al. [2013] propose a variety of semi-

material deprivation.

parametric specifications.

In the poverty analysis literature, Biewen and Jenkins [2005], Bourguignon and Ferreira [2005], and Bourguignon et al. [2008] perform OB decomposition using a parametrization of the household income distribution. Ayala et al. [2011] analyze the impact of regional differences on the statistical relationship between income-based poverty measures and severe material deprivation indicators in Spain. The methodology used involves a logit specification for the conditional probabilities of being poor and generalized OB decomposition for non-linear specification (see Yun [2004]). The same methodology is also employed in Gradín [2012] and Ayala et al. [2021] where the former decomposes poverty differentials among ethnic groups in Brazil and the latter temporal evolution of poverty rates in rural and urban areas of Spain. While, Essama-Nssah and Lambert [2016] decompose different poverty measures by re-weighting their recentered influence functions, where the conditional expectation of the RIF is modeled using a linear parametrization.

3 Setting

Let Y be a non-negative random variable that models the outcome of interest with cumulative density function $F_Y(\cdot) \in \mathcal{F}$. Consider $A : \mathcal{F} \to \mathbb{R}$,

$$A(F_Y, g) = \int g(y) dF_Y(y) = \mathbb{E}\left[g(Y)\right] \tag{1}$$

With $\mathbb{E}\left[g^2(Y)\right] = \int g^2(y) dF_Y(y) < \infty$. The class of elements defined by $A(\cdot)$ includes a big variety of interesting statistics. Examples are the mean, cumulative distribution function evaluated at a point, social welfare functions such as poverty indices, etc. We can extend the decomposition techniques proposed in this article to non-linear functionals of the distribution by deriving the respective influence function, as in Firpo et al. [2009]. In this case, g(y) is given by the RIF.

Let X an $\mathbb{R}^p$-valued random vector of characteristics and assume to observe both X and Y for two populations, denoted with $j \in \{0, 1\}$. Where $F^{(j)}(\cdot)$ indicates the cumulative density function of individual characteristics conditional on belonging to population (j), and $dF^{(j)}(\cdot)$ the corresponding probability measures.

We distinguish among two different data generating process defining the nature of the relationship between $(Y^{(1)}, X^{(1)})$ and $(Y^{(0)}, X^{(0)})$. One is characterized by samples from the two populations of different size which (we assume to) grow proportionally; that is, the data consists of two samples $\{Y_i^{(1)}, X_i^{(1)}\}_{i=1}^{N^1}$ and $\{Y_i^{(0)}, X_i^{(0)}\}_{i=1}^{N^0}$, such that $N^1/N^0 \to \gamma$ for some $\gamma > 0$. This structure model cross-population analysis and is usually paired with an independence assumption between the observed vectors of the two populations. In the other structure, samples $\{Y_i^{(1)}, Y_i^{(0)}, X_i^{(1)}, X_i^{(0)}\}_{i=1}^{N}$ are drawn from $(Y^{(1)}, Y^{(0)}, X^{(1)}, X^{(0)})$. This is suitable to model dependence relationship between two populations in panel data. Classical examples are inter-temporal comparison, where the same population is observed in two different periods of time, or when $X^{(1)}$ is a deterministic function of $X^{(0)}$ such that $X^{(1)} = f(X^{(0)})$ for some known function $f : \mathbb{R}^p \to \mathbb{R}^p$. The distinction in the counterfactual framework among the two structures has already been pointed out in Chernozhukov et al. [2013] and Rothe [2010b]. Throughout this article, we adopt the notation relative to the former structure since it nests the latter as the special case $N^1 = N^0 = N$.

We restrict the relationship between the covariates' support of the two populations to identify the counterfactual functional.

Assumption 1 (Covariates Support)

The support of $X^{(0)}$ is a subset of the support of $X^{(1)}$; that is, $\mathcal{X}^{(0)} \subseteq \mathcal{X}^{(1)}$, where $\mathcal{X}^{(j)}$ denotes the support of the distribution of characteristics in population (j).

Assumption 1 is standard in this literature. It restricts the counterfactual experiment to observations in the two populations that share similar characteristics. The assumption is usually overlooked in applied works using classic OB decomposition. In these, a linear parametrization of the inner conditional expectation allows for out-of-sample extrapolations. However, it is of utmost importance in non-parametric approaches, since these do not allow to extrapolate out of the range of actual observations (apart from a small sphere around the boundaries of the supports' intersection).

Counterfactual Functional and KOB Decomposition

Let $h^{(j)}(X)$ be the function representing the conditional expected value of $g(Y)$ given the vector of characteristics and relative population j: $h^{(j)}(X) = \mathbb{E}\left[g\left(Y^{(j)}\right)|X^{(j)}\right]$. We define the counterfactual functional $A^{(1,0)}$ of population (1) with respect to population (0) as:

$$A^{(1,0)} = \int_{\mathcal{X}^{(1)}} h^{(1)}(x)dF^{(0)}(x) = \int_{\mathcal{X}^{(1)}} \mathbb{E}\left[g\left(Y^{(1)}\right)|X^{(1)} = x\right]dF^{(0)}(x) \tag{2}$$

With similar notation we indicate the functional evaluated in population (j) as $A^{(j)} = A^{(j,j)} = \mathbb{E}\left[g\left(Y^{(j)}\right)\right]$. The counterfactual experiment simply consist of re-weighting the conditional expectation of $g(Y)$ in population (1) by the probability of observing a given set of individual characteristics in population (0).

Together with the definition of a counterfactual functional, we define a Kitagawa-Oaxaca-Blinder (hereafter KOB) decomposition of the difference between $A^{(1)}$ and $A^{(0)}$ in the following way:

$$A^{(1)} - A^{(0)} = \underbrace{\left[A^{(1)} - A^{(1,0)}\right]}_{\textbf{Composition component } \Delta^C} + \underbrace{\left[A^{(1,0)} - A^{(0)}\right]}_{\textbf{Residual (or structural) component } \Delta^R}$$

This decomposition is of great value in understanding the intrinsic nature of the difference. It splits the differential into two components, one attributable to different characteristics' composition, the other to everything else other than characteristics' composition. When assumption 1 does not hold, $\mathcal{X}^{(0)} \not\subset \mathcal{X}^{(1)}$, we can still identify the counterfactual functional over the common support and decompose the difference in functionals using a four-components decomposition, as in Ñopo [2008].

4 Aggregated KOB Decomposition

In this section, we define the counterfactual functional and the relative KOB decomposition for aggregated data. There are two main advantages in grouping the data and focusing on (one of) the standardized version of the decomposition. First, by grouping observations we get rid of estimation issues linked with the non-parametric specification of the conditional expectation and data sparsity. Second, the aggregated KOB decomposition allows to retrieve a detailed decomposition with the contribution of each group on the two components. This last one mimics the Endowment/Coefficient decomposition originally introduced by OB and is a unique feature of the partitioned framework. Furthermore, under the additional restriction of linear specification within each cell, we distinguish the contributions of each covariate while still keeping a higher generality of classic OB decomposition.

Let $\mu^{(j)}(A) = \int\limits_{\{x \in A\}} dF^{(j)}(x)$ denote the probability measure of a set A under the distribution $dF^{(j)}(\cdot)$. Given a partition of the covariates' support $\mathbb{C}_L = \{C_l\}_{l=1}^{L}$, with $\mu^{(1)}(C_l) > a \;\forall\; l,\; a > 0,\; \bigcup\limits_{l=1}^{L} C_l = \mathcal{X}$ and $C_l \bigcap C_j = \emptyset \;\forall\; l \neq j$, we define the class of (aggregated) counterfactual functionals $A^{(1,0)}_{\mathbb{C}_L}$ as:

$$A^{(1,0)}_{\mathbb{C}_L} = \sum_{l=1}^{L} \underbrace{\left(\int\limits_{\{x \in C_l\}} \frac{h^{(1)}(x)}{\mu^{(1)}(C_l)} dF^{(1)}(x) \right)}_{\mathbb{E}\left[g(Y)^{(1)}|X^{(1)} \in C_l\right]} \mu^{(0)}(C_l) = \sum_{l=1}^{L} h^{(1)}(C_l)\mu^{(0)}(C_l) \tag{3}$$

With $h^{(j)}(C_l)$ denoting $\mathbb{E}\left[g\left(Y^{(j)}\right)|X^{(j)} \in C_l\right]$.

We might have a different counterfactual $A^{(1,0)}_{\mathbb{C}_L}$ for any $\mathbb{C}_L$. In general $A^{(1,0)}_{\mathbb{C}_L} \neq A^{(1,0)}$ but the difference $A^{(1,0)}_{\mathbb{C}_L} - A^{(1,0)}$ goes to zero as the size of each C_l shrinks to zero. In the case of discrete covariates, the equality holds whenever we evaluate $A^{(1,0)}_{\mathbb{C}_L}$ using the finest partition of the data.

For any of these objects we define the relative KOB decomposition of the difference in functionals:

$$A^{(1)} - A^{(0)} = \underbrace{[A^{(1)} - A^{(1,0)}_{\mathbb{C}_L}]}_{\textbf{Composition component } \Delta^{C}_{\mathbb{C}_L}} + \underbrace{[A^{(1,0)}_{\mathbb{C}_{\mathbb{L}}} - A^{(0)}]}_{\textbf{Residual (or structural) component } \Delta^{R}_{\mathbb{C}_L}}$$

Detailed decomposition

Counterfactual decompositions using linear specification of the conditional expectation (OB decompositions) allow to further decompose the contribution of each covariate on the composition and the residual components in the following way:

$$
\begin{aligned}
\Delta^C &= \int \mathbb{E}[Y^{(1)}|X^{(1)}=x]\left(dF^{(1)}(x)-dF^{(0)}(x)\right) \\
&= \sum_{p=1}^{P} \underbrace{\beta_p^{(1)}\left(\mathbb{E}\left[X_p^{(1)}\right]-\mathbb{E}\left[X_p^{(0)}\right]\right)}_{\textbf{Endowments effect of the } p \textbf{ covariate}} \\
\Delta^R &= \int \left(\mathbb{E}[Y^{(1)}|X^{(1)}=x]-\mathbb{E}[Y^{(0)}|X^{(0)}=x]\right) dF^{(0)}(x) \\
&= \sum_{p=1}^{P} \underbrace{\left(\beta_p^{(1)}-\beta_p^{(0)}\right)\mathbb{E}[X_p^{(0)}]}_{\textbf{Coefficient effect of the } p \textbf{ covariate}}
\end{aligned}
$$

Recent approaches extend the detailed decomposition for distributional statistics other than the mean. Machado and Mata [2005] propose a procedure based on quantile regression and simulation methods to decompose the contribution of each characteristic on the residual component. For the composition component, DiNardo et al. [1996] suggests a reweighting procedure to compute the contribution of a dummy covariate to the aggregate composition effect. Altonji et al. [2012] extended this last one to either continuous or categorical covariates. The shortcoming of these methods is that they are *path-dependent*; that is, the elements of the detailed decompositions depend on the path chosen to decompose the perturbation in the characteristics' distribution. Alternatively, as shown in Fortin et al. [2011], using RIF regression joint with linear specification produces path-independent decomposition of general functional.

Without parametric restrictions on the conditional expectation (or on the conditional expectation of the RIF), the standardized KOB decomposition naturally defines the contributions each group of individuals has on the decomposition's components:

$$
\begin{aligned}
A^{(1,1)} - A^{(1,0)}_{\mathbb{C}_L} = \Delta^C_{\mathbb{C}_L} &= \sum_{l=1}^{L} \underbrace{\mathbb{E}\left[Y^{(1)}|X^{(1)}\in C_l\right]\left(\mathbb{P}(X^{(1)}\in C_l)-\mathbb{P}(X^{(0)}\in C_l)\right)}_{\textbf{Composition contribution of the } l \textbf{ group}} \\
A^{(1,0)} - A^{(0,0)}_{\mathbb{C}_L} = \Delta^R_{\mathbb{C}_L} &= \sum_{l=1}^{L} \underbrace{\left(\mathbb{E}\left[Y^{(1)}|X^{(1)}\in C_l\right]-\mathbb{E}\left[Y^{(0)}|X^{(0)}\in C_l\right]\right)\mathbb{P}(X^{(1)}\in C_l)}_{\textbf{Residual contribution of the } l \textbf{ group}}
\end{aligned}
$$

This detailed decomposition works for general linear functional (or non-linear if we can get estimates of the RIF) without requiring the specification of a perturbation's pattern or the functional form of the conditional expectation. However, the decomposition depends on the grouping choice. Additionally, we obtain the contribution of each characteristic to the two decomposition components by further restricting the conditional expectation function to be stepwise linear:

Assumption 2 (Stepwise Linearity)
The CEF is assumed to be stepwise linear

$$h_l^{(j)}(x) = \sum_{l=1}^{L} x'\beta_l^{(j)}\mathbb{I}(x \in C_l)$$

For $j \in \{0, 1\}$.

Under this restriction, we obtain the following detailed decomposition

$$\begin{aligned}
\Delta^C_{\mathbb{C}_L} &= \sum_{k=1}^{p} \underbrace{\sum_{l=1}^{L} \mathbb{E}(X_k^{(1)}|X^{(1)} \in C_l)\beta_{kl}^{(1)} \left(\mu^{(1)}(C_l) - \mu^{(0)}(C_l)\right)}_{\textbf{Grouped composition contribution of the } k\textbf{th covariate (C)}} \\
\Delta^S_{\mathbb{C}_L} &= \sum_{k=1}^{p} \underbrace{\sum_{l=1}^{L} \mathbb{E}(X_k^{(0)}|X^{(0)} \in C_l)(\beta_{kl}^{(1)} - \beta_{kl}^{(0)})\mu^{(1)}(C_l)}_{\textbf{Structural contribution of the } k\textbf{th covariate (R1)}} \\
&+ \sum_{k=1}^{p} \underbrace{\sum_{l=1}^{L} \left(\mathbb{E}(X_k^{(1)}|X^{(1)} \in C_l) - \mathbb{E}(X_k^{(0)}|X^{(0)} \in C_l)\right) \beta_{kl}^{(1)}\mu^{(1)}(C_l)}_{\textbf{Composition contribution of the } k\textbf{th covariate: Mean (R2)}}
\end{aligned} \tag{4}$$

Where both components can be written as sums of the singular characteristics contributions. Notice that this specification is more general than the usual linearity assumption, which is a particular case of assumption 2, when $\beta_l^{(j)} = \beta_f^{(j)}$ for $l \neq f$ and $j \in \{0, 1\}$ or $L = 1$.

When $L = 1$, the term (C) disappears, and the terms (R1) and (R2) become equal to the coefficient and endowment effects, respectively, of the OB decomposition. Accordingly, the term (C) may be thought of as an extra term coming from the partitioning, indicating the contribution of each covariate to the difference among groups; whereas (R1) and (R2) are an extension of the endowment and composition effects of standard OB decomposition. (R1) is a weighted average of the differences in coefficients within each group, where the weights are given by the conditional mean of each covariate in the population (0); whereas (R2) reflects variation in the within-group compositions, weighted for the coefficient of each characteristic and partition of the population (1).

Estimation

Given a partition $\mathbb{C}_L$ of the data, we estimate the standardized functional by simply substituting the sample analog of all the expectations:

$$\widehat{A}^{(1,0)}_{\mathbb{C}_L} = \sum_{l=1}^{L} \widehat{h}^{(1)}(C_l)\widehat{\mu}^{(0)}_l(C_l) = \sum_{l=1}^{L} \frac{1}{N^1} \sum_{i=1}^{N^1} \frac{g(Y_i^{(1)})\mathbb{I}\{X_i^{(1)} \in C_l\}}{\widehat{\mu}^{(1)}_l(C_l)} \widehat{\mu}^{(0)}_l(C_l) \tag{5}$$

Where: $\widehat{\mu}^{(j)}_l(C_l) = (N^j)^{-1} \sum_{i=1}^{N^j} \mathbb{I}\{X_i^{(j)} \in C_l\}$ and $\mathbb{I}\{A\}$ is the indicator function, which assumes value 1 if condition A holds and 0 otherwise.

Alternatively, one can use the partitioning estimator of Cattaneo and Farrell [2013] where within each cell the unknown regression function is approximated by linear least squares using a fixed order polynomial basis (our estimator is a particular case of it with $K = 1$). Such a solution might increase the finite sample performances of the counterfactual estimator. Notice that, in general, partitioning estimators are more computationally efficient than other non-parametric methods, such as Kernel or Nearest-Neighbor regressions, as they only require evaluating the mean (or a regression) in each set rather than at each point. However, the partitioning procedure might be computationally intensive (see next section).

Inference

For finite partition, the asymptotic distribution is determined with a simple application of the delta method.

Assumption 3 (Fixed partition)

$\lim_{N\to\infty} \mathbb{C}_{L_N} = \mathbb{C}_L$ *for some finite grouping of the data* $\mathbb{C}_L$.

As the asymptotic variance of the estimator depends on the degree of dependence between the two populations, we assume the more general structure

Assumption 4 (Correlated Populations)
The data $\{Y_i^{(1)}, X_i^{(1)}, Y_i^{(0)}, X_i^{(0)}\}_{i=1}^{N}$ *are jointly i.i.d.*

Notice that under assumption 4, $N^1 = N^0$ and $\gamma = 1$.
A simple application of the law of large numbers and Slutsky's theorem shows the consistency of the counterfactual with aggregated data. We use the delta method to derive the asymptotic distribution of the counterfactual. The result is an extension of the distribution derived in Ñopo [2008] for different matching windows and interdependent data.

Proposition 1 ***(Asymptotic Distribution)***
Let $\mu_l^{(j)}$ *be the probability of an individual in population* (j) *to show characteristics* C_l *and* $\sigma_l^{(j)} = \mathbb{V}(g(Y^{(j)})|X^{(j)} \in C_l)/\mu_l^{(j)}$ *denote the population variance of* $g(Y^{(j)})$ *for individual with characteristics* C_l*. Under assumption 1, 2 and 4, the estimated counterfactual converges to a normal distribution:*

$$\sqrt{N}\left(\widehat{A}^{(1,0)}_{\mathbb{C}_{L_N}} - A^{(1,0)}_{\mathbb{C}_L}\right) \xrightarrow{d} N\left(0, V_A\right)$$

$$V_A = \sum_{l=1}^{L}\left((\mu_l^{(0)})^2\sigma_l^{(1)} + (h_l^{(1)})^2\mu_l^{(0)}(1-\mu_l^{(0)})\right) - \sum_{l\neq f} h_l^{(1)}h_f^{(1)}\mu_l^{(0)}\mu_f^{(0)} + 2\sum_{l=1}^{L}\sum_{f=1}^{L} h_l^{(1)}c_{(l,f)}\mu_f^{(0)}$$

Where:

$$c_{(l,f)} = \frac{\mu_f^{(1)}q_{(l,f)}^{(0,1)} - q_{(f,f)}^{(1,1)}\mu_{(l,f)}^{(0,1)}}{\left(\mu_f^{(1)}\right)^2}$$

And :

$$q_{(l,f)}^{(0,1)} = \mathbb{E}\left[g(Y^{(1)})\mathbb{I}\{X^{(0)} \in C_l \wedge X^{(1)} \in C_f\}\right] \quad \mu_{(l,f)}^{(0,1)} = \mathbb{E}\left[\mathbb{I}\{X^{(0)} \in C_l \wedge X^{(1)} \in C_f\}\right]$$

If the samples of the two populations are independent, with $N^1 \neq N^0$ *and* $N^1/N^0 \to \gamma$*, the asymptotic variance is given by:*

$$V_A = \sum_{l=1}^{L}\left((\mu_l^{(0)})^2\sigma_l^{(1)} + \frac{1}{\gamma}(h_l^{(1)})^2\mu_l^{(0)}(1-\mu_l^{(0)})\right) - \frac{1}{\gamma}\sum_{l\neq f} h_l^{(1)}h_f^{(1)}\mu_l^{(0)}\mu_f^{(0)}$$

Proof. *Appendix.* ■

Further, in the appendix, we provide the asymptotic distribution for the decomposition components, both under dependent and independent data structure.

Which Grouping?

The grouping choice depends on the application context. In some, the economic theory provides natural clusters. For instance, Lemieux [2006] before decomposing wages between different years, analyzes wage variance and workforce share for 20 education/experience classes. In these settings, the groups are interpreted as *structural*, and the grouping choice is unambiguous. When ex-ante information on the groups is not available (or incomplete), determining the groups is analogous to choosing a bandwidth in Kernel regressions. Ideally, we want to adopt grouping criteria that generate homogeneous groups (based on some likeliness measure). The homogeneity reduces the information loss implicit in the grouping procedure and makes the detailed decomposition easier to interpret.

A sensible solution is to split the sample into statistically equivalent blocks (see Gessaman [1970], for instance); however, the procedure is problematic for a large dimension of the covariates' vector or when the data includes non-numeric variables.

In this article, we propose to group the data using **Classification-and-regression-tree** (CART), by Breiman et al. [1984]. The procedure is explained in detail in the next section. One advantage of the CART method is that it allows grouping with non-ordered categorical variables. As the algorithm groups the data based on the contribution of each characteristics' combination to the conditional expectation, it does not require a measure of distance in the characteristics' space.

We provide Montecarlo simulations comparing the finite sample performances of $\widehat{A}_{\mathbb{C}_L}^{(1,0)}$, under CART grouping, to other estimation methods. The results show that the partitioning estimator under CART partitioning systematically outperforms the other non-parametric methods.

5 CART

The CART algorithm is widely used in machine learning and statistics. The criterion for aggregating the data is how well a piece-wise constant function, based on the grouping, fits the conditional mean function. The groups have a meaningful interpretation even with non-ordered (and non-numerical) characteristics. We can think of them as the set of nearest neighbors for a given target observation in the group. However, deep interpretation of these groups are not possible as "Particular covariates that have strong associations with the outcome may not show up in splits because the tree splits on covariates highly correlated with those covariates" (Athey and Imbens [2019], Page 697). The algorithm consists of two parts, initial tree building and cross-validation phase. In the initial tree building phase, the algorithm recursively partitions the data according to the in-sample goodness-of-fit and a tuning parameter, called the penalty term. As the depth of the tree is regulated by the penalty term, for different choices of the parameter we obtain different partitions. In the cross-validation phase, the algorithm determines the optimal penalty term by repeatedly splitting the sample into a training sample and a cross-validation sample. The training sample is used to generate a partition and to estimate the conditional mean function; the cross-validation sample is used to evaluate the estimates based on the partition generated by the training sample. The optimal penalty term maximizes a goodness-of-fit criterion in cross-validation samples. In all the grouping procedure we use the sample of population (1) to generate the splitting rules, which are then applied to split the sample of population (0).

Initial Tree-Building Phase

Let denote a partition as $\mathbb{C}^{CART}_{L_t}$, where the sub-script L_t indicates the dependency of the number of set in each partition from the t-th iteration of the algorithm. Fix a given iteration t and partition $\mathbb{C}^{CART}_{L_t}$, then for any $C_l \in \mathbb{C}^{CART}_{L_t}$ the algorithm selects a threshold c and one of the covariates X_{ik} to split C_l into two smaller sets by minimizing the following objective function:

$$Q_{C_l}(k,c) \text{ subject to } Q_{C_l} \geq \lambda L_t + Q_{C_l}(k,c)$$

Where Q_{C_l} and $Q_{C_l}(k,c)$ denote the in-sample goodness-of-fit before and after the split respectively, such that:

$$Q_{C_l} = \sum_{i:X_i^{(1)} \in C_l} \left(Y_i^{(1)} - \widehat{h}^{(1)}(C_l) \right)^2$$

And

$$Q_{C_l}(k,c) = \sum_{i:X_i^{(1)} \in C_l, X_{ik}^{(1)} \leq c} \left(Y_i^{(1)} - \widehat{h}_l^{(1)}(C_l,c) \right)^2 + \sum_{i:X_i^{(1)} \in C_l, X_{ik}^{(1)} > c} \left(Y_i^{(1)} - \widehat{h}_r^{(1)}(C_l,c) \right)^2$$

Where $\widehat{h}_r^{(1)}(C_l,c) = \frac{\sum_{i=1}^{N^1} g(Y_i^{(1)}) \mathbb{I}\{X_i^{(1)} \in C_l\}}{\sum_{i=1}^{N^1} \mathbb{I}\{X_i^{(j)} \in C_l : X_{ik}^{(1)} \leq c\}}$ and $\widehat{h}_r^{(1)}(C_l,c) = \frac{\sum_{i=1}^{N^1} g(Y_i^{(1)}) \mathbb{I}\{X_i^{(1)} \in C_l\}}{\sum_{i=1}^{N^1} \mathbb{I}\{X_i^{(j)} \in C_l : X_{ik}^{(1)} > c\}}$.

The penalty term λL_t punishes the partitions with too many classes. For a split to be done, the splitting point c on the covariate X_{ik} should improve the in-sample goodness-of-fit criterion by at least λL_t. When X_{ik} is a factor variable (non-ordered or non-numerical), the split is done by selecting two subsets of categories of X_{ik}, rather than dividing above and below the threshold c.

Cross-Validation Phase

Let $S^{(1)}$ be the sample of population (1), and denote the partition obtained in the initial tree building phase as $\mathbb{C}^{\text{CART}}(S^{(1)}, \lambda)$. The next step is to search over a grid of values $\Lambda_M = \{\lambda_1, ..., \lambda_M\}$ the λ that maximizes the goodness-of-fit criterion in cross-validation samples. In this step, we drop the dependence of L_t from the number of iteration since it is not relevant for the discussion and consider it as a function of the sample and the penalty term; that is, $L = L(S^{(1)}, \lambda)$. Common cross-validation practices include the k-fold cross-validation and the leave-one-out cross-validation. Without loss of generality, we discuss the latter. For each $\lambda \in \Lambda_M$ and $i \in \{1, 2, ..., N^1\}$ we split $S^{(1)}$ into a training sample $S_{-i}^{(1)}$ including all the observations in $S^{(1)}$ but i and a cross-validation sample $\{i\}$. For each training sample $S_{-i}^{(1)}$ we generate a partition $\mathbb{C}^{\text{CART}}(S_{-i}^{(1)}, \lambda) = \{C_l(S_{-i}^{(1)}, \lambda)\}_{l=1}^{L(S_{-i}^{(1)}, \lambda)}$ and estimate the conditional mean in the leaf (set) of $\mathbb{C}^{\text{CART}}(S_{-i}^{(1)}, \lambda)$ where the cross-validation sample lies. That is:

$$\widehat{h}^{(1)}\left(X_i^{(1)}, \mathbb{C}(S_{-i}^{(1)}, \lambda)\right) = \sum_{l=1}^{L(S_{-i}^{(1)}, \lambda)} \frac{\sum_{j \neq i} Y_i^{(1)} \mathbb{I}\{X_j^{(1)} \in C_l(S_{-i}^{(1)}, \lambda)\}}{\sum_{j \neq i} \mathbb{I}\{X_j^{(1)} \in C_l(S_{-i}^{(1)}, \lambda)\}} \mathbb{I}\{X_i^{(1)} \in C_l(S_{-i}^{(1)}, \lambda)\}$$

We calculate the error on observation i: $e_i^{(1)}(\lambda) = Y_i^{(1)} - \widehat{h}^{(1)}\left(X_i^{(1)}, \mathbb{C}^{\scriptscriptstyle CART}(S_{-i}^{(1)}, \lambda)\right)$ and choose the λ^* that minimize the average mean square error over all N^1 observations:

$$\lambda^* = \underset{\lambda \in \Lambda_M}{\operatorname{argmin}} \sum_{i=1}^{N^1} \left[e_i^{(1)}(\lambda) \right]^2$$

Finally, we get our chosen partition $\mathbb{C}_L^{\scriptscriptstyle CART} = \mathbb{C}^{\scriptscriptstyle CART}(S^{(1)}, \lambda^*)$ and use (5) to estimate the counterfactual mean. All the above procedure using CART can be easily implemented using the R-package *Rpart*.

6 Analysis of AROPE Index in Spain

We apply the aggregated framework to investigate the effects of the great recession on AROPE (At-risk-of-poverty-or-social-exclusion) rates in Spain. The goal is to analyze how perturbations into the distribution of poverty risk factors induced by the shock affected the rise in poverty rates. We do that by decomposing the growth of AROPE rates across the 2008-2014 period (corresponding to the beginning and the peak of the recession, respectively) into two components: the increase due to variations in the distribution of poverty risk factors among groups of individuals (composition component); the increase due to variations in the probability of being poor in those groups (residual component). By aggregating the data, we determine the contribution of each group to the rise of poverty rates through the two channels.

This information might be helpful for the targeting of pro-poorness policies. A policymaker, for instance, can actuate a policy to avoid rise in the unemployment level for individuals between a given age range; thus, reducing the compositional contribution of that group[4]. At the same time, he can raise the social net for unemployed individuals between other age ranges, reducing the rise in the probability of being poor in the group. Since the policymaker has to consider other factors, such as the cost of implementing policies for different age ranges, the classification of the group contributions might be helpful for the analysis of possibly arising trade-offs.

AROPE

We identify (and aggregate) individuals in poverty condition[5] using the AROPE (At-risk-of-poverty-or-social-exclusion) measure. The European Commission adopted the measure to monitor the convergence toward the goal set up in the strategy for smart, sustainable, and inclusive growth of the European Union to reduce by at least 20 million the number of citizens living in conditions of poverty or social exclusion. The measure is a multi-dimensional index that classifies an individual as *at risk of poverty or social exclusion* if he faces at least one of the following situations:

- At risk of Poverty (AROP60): The individual lives in a household with an equivalized disposable income[6] below 60% of the national median equivalized disposable income.
- Severe material deprivation (MD): The individual cannot afford at least 4 out of 9 predefined material items considered by most people to be desirable or even necessary to live an adequate life.
- Low work intensity (LJ): The individual lives in a household where the adults worked a working time equal to or less than 20% of their total combined work-time potential during the previous year.

The proportion of individuals falling into the AROPE classification determines the AROPE rate.

[4] An example of such policies is the direct financing of the firms' labor cost (ERTE - Expedientes de Regulación Temporal del Empleo) by the Spanish government, which avoided massive layoffs during the coronavirus pandemic.

[5] See Sen [1976], Atkinson [1987], Foster et al. [1984], Zheng [1997], Sen et al. [1997], Foster [2006], Foster et al. [2013] among others for a thorough review of identification and aggregation methods of poverty conditions.

[6] The equivalized disposable income is the main welfare measure adopted by Eurostat. This is equal to the total household income divided by the OECD scale of family size.

Data and Characteristics Choice

We use data from the Survey on Living Condition (*Encuesta de condiciones de vida* or ECV) elaborated by the Spanish Statistical National Institute (INE). It is a yearly survey collecting harmonized cross-sectional and longitudinal microdata on income, poverty, social exclusion, and living conditions. The survey is part of the EU Statistics on Living and Income condition (EU-SILC), used for monitoring poverty and social exclusion within the framework for the coordination of economic policies across the European Union. It collects data on individuals' labor market status, relative earning conditions, and information about social exclusion and housing conditions. Social exclusion and housing conditions data are collected at the household level, while income data at the individual level. The analysis is performed at the individual level but accounts for the correlation within each household. We consider only individuals older than 16 years old, thus excluding child poverty from the analysis. The dataset already provides indicators that classify an observation in AROPE or one of the three components (AROP60, MD, LJ). As the LJ index is not defined for individuals older than 60 years, we set the indicator to 0 for them. It follows that our LJ index is a lower bound of the proportion of people facing low work intensity conditions. Both indicators and characteristics come from annual cross-sectional observations in the reference periods.

As potential drivers of poverty, we choose a set of characteristics considered relevant factors of risk for being poor or materially deprived[7]. These include information both at the household and individual levels. At the household level, we report the household's type (single person, couple with or without children, single-parent households, etc.) and the household head's gender and age. At the individual level, we choose a set of variables that characterizes age, education, employment, sex, the country of birth, and the nationality of each individual in the sample.

[7] In the appendix, we report a detailed breakdown of the variables used in this analysis.

Table 1: **Descriptive Statistics**

	Sex	**Age**	**Nat**	**CountryB**	**Educ**	**WorkS**	**H-type**	**HH-sex**	**HH-age**	**H-syze**
2008										
mean	1.524	61.747	1.090	1.124	2.465	3.514	9.764	1.358	67.900	3.192
median	2	61	1	1	2	3	9	1	67	3
q-25	1	47	1	1	1	1	8	1	57	2
q-75	2	76	1	1	3	5	12	2	79	4
std.dev	0.499	18.715	0.396	0.463	1.584	2.756	2.787	0.479	14.693	1.312
2014										
Mean	1.522	57.359	1.081	1.139	2.590	3.988	9.558	1.363	63.661	3.067
Median	2	57	1	1	2	5	9	1	63	3
p-25	1	43	1	1	1	1	8	1	53	2
p-75	2	71	1	1	5	6	12	2	74	4
St.dev	0.499	18.643	0.374	0.487	1.662	2.577	2.858	0.480	14.776	1.299

Descriptive statistics of all the characteristics considered in the analysis in 2008 and 2014. The table reports mean, median, 25th, 75th percentiles, and standard deviation.

In Spain, there are 17 regions plus the autonomous cities of Ceuta and Melilla. We exclude these last two from the analysis due to their small size, geographic position, and particular economic environment of the two cities, which make their statistics not comparable with those of the other 17 regions. We drop from the sample observations with missing values. The drop affects a percentage of individuals smaller than 3% of the whole sample. The sample size goes from a minimum of twenty-six thousand in 2014 to a maximum of thirty thousand observations in 2008. The survey furnishes population weight for the adult population (16+) that we employ in the estimation of the statistics and relative counterfactuals.

The Evolution of Poverty in Spain

The impact of the great recession on the poverty level in Spain has been dramatic. The proportion of individuals in AROPE rose from 22.6% of the total (adult) population in 2008 to a peak of 28% in 2014. Even after 2014, despite the recovery of the Spanish economy, we still observe AROPE rates ranging above 26% of the total population.

Table 2: **National Poverty Rates**

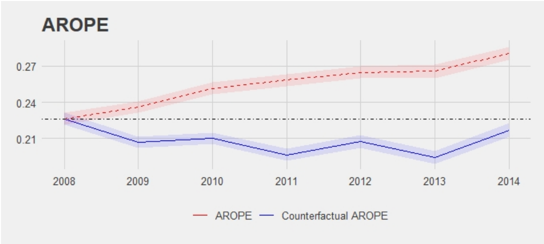
Year	AROPE	AROP60	MD	LJ
2008	22.6	18.5	3.1	5.3
2009	23.6	19.0	4.0	5.9
2010	25.1	19.4	4.4	8.4
2011	25.8	19.4	4.3	10.2
2012	26.4	19.7	5.5	10.9
2013	26.5	19.2	5.6	12.0
2014	28.0	20.7	6.5	13.0
2015	27.7	20.9	5.8	11.8
2016	27.1	21.1	5.5	11.3
2017	25.7	20.3	4.8	9.6
2018	25.5	20.6	5.1	8.1
2019	24.3	19.3	4.4	8.0
2020	25.4	19.6	6.4	7.4

The table reports the evolution of the AROPE rates and its three components in Spain from 2008 to 2020. The values indicate percentage of the population.

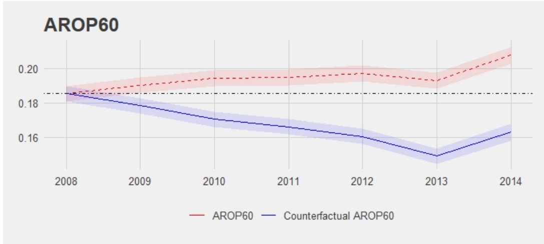
The AROPE components followed similar trends. Variations in the income-based measure (AROP60) mainly reflect the ones observed in the aggregate index, for most individuals in AROPE are also in AROP60. Conversely, the MD and LJ indicators show weaker correlations with the AROPE trend. Nonetheless, the effects of the recession on these two indicators have been proportionally much higher than for the aggregate and income-based measures. The proportion of individuals in low work intensity and material deprivation conditions more than doubled in the 2008-2014 period; LJ index increases are even more dramatic considering that our LJ measure is a low bound of the true portion of the population in low-work intensity conditions.

Decomposition of National Rates

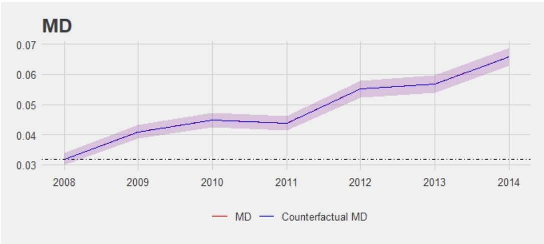
We decompose the AROPE rate and its three components along the 2008-2014 period. At first, we derive for each index and each period $t \in \{2008, ..., 2014\}$ the counterfactual rate, using the pre-recession population as a reference ($A_{\mathbb{C}_{\mathbb{L}}}^{(t,2008)}$ in the notation of the previous sections). Figure 1 illustrates the decomposition of the four rates, the indices time series are depicted in red, while the counterfactual rates are in blue. Each graph reports a dashed black line corresponding to the indices level in 2008, the differences between the red and blue line and the blue and black line determine the compositional and residual components, respectively.



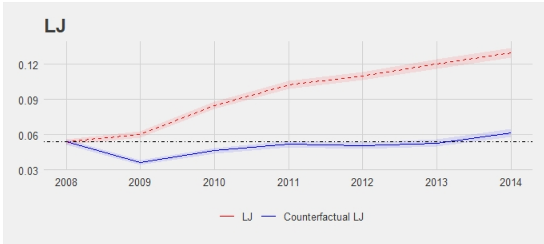
(a) AROPE



(b) AROP60



(c) MD



(d) LJ

The graphs illustrate the time series of the four rates, AROPE, AROP60, MD, LJ, in red together with the relative counterfactual rate, $A_{\mathbb{C}_{\mathbb{L}}}^{(t,2008)}$, in blue for $t \in \{2008, ..., 2014\}$.

While in Table 3, we report the decompositions of the four indices 2008-2014 differentials, together with the number of groups determined by the CART algorithm.

Compositional variations appear as the main driver of the shock effects on the aggregate index. Similar patterns show in the decomposition of the LJ and AROP60 rates, which perturbations seem to drive the raise in AROPE rates. As unemployment is a primary risk factor for poverty, it is not surprising that variations in characteristics' composition had a massive impact on the increase of poverty rates. The national shock drastically reduced the demand for labor, particularly in sectors like construction, thus, moving a big portion of the population from low-risk factors (full-time employment) to high-risk factors (part-time employment or unemployment).

The interpretation of residual components is more difficult than that of composition components, as they depend on unobserved factors. Loosely speaking, positive (negative) residual components implies that the poverty conditions of groups of individuals worsened (improved) (keeping constant the population composition). Regional policies, for instance, might affect the relative poverty conditions of these groups and consequently determine the sign and magnitude of the residual components.

In Figure 1 (a-b), we observe negative residual components across most of the recession period, which suggests that if the population kept the same characteristics composition of 2008, the poverty ratio would have actually decreased[8]. However, our model cannot capture the channel through which these adjustments occurs.

Table 3: **Decomposition of National Poverty Rates**

Index	Δ	**Comp.**	**Res.**	**Groups**
AROPE	5.3	6.1 (27.19)	-0.7 (-2.05)	24
AROP60	2.2	4.4 (27.8)	-2.2 (-6.47)	18
MD	3.3	0 (NaN)	3.3 (31.01)	1
LJ	7.6	6.8 (34.45)	0.7 (3.76)	21

Decomposition of 2008-2014 national poverty rates differentials. Δ reports the difference of the index in the two period; **Comp.** and **Res.** report the composition and residual components of the decomposition; **Groups** are the number of groups generated by the CART algorithm. T-ratio with 0 null hypothesis in parenthesis.

The decomposition of the MD index shows a zero composition effect throughout the period considered. The reason is that the CART algorithm does not find useful splits for this index and consequently the counterfactual index is numerically identical to the index itself. These evidence, suggesting that risk factors of material deprivation and income-based measure strongly differ, are consistent with previous studies showing that only a small percentage of individuals exiting income-based poverty also leave conditions of material deprivation[9].

Detailed decomposition

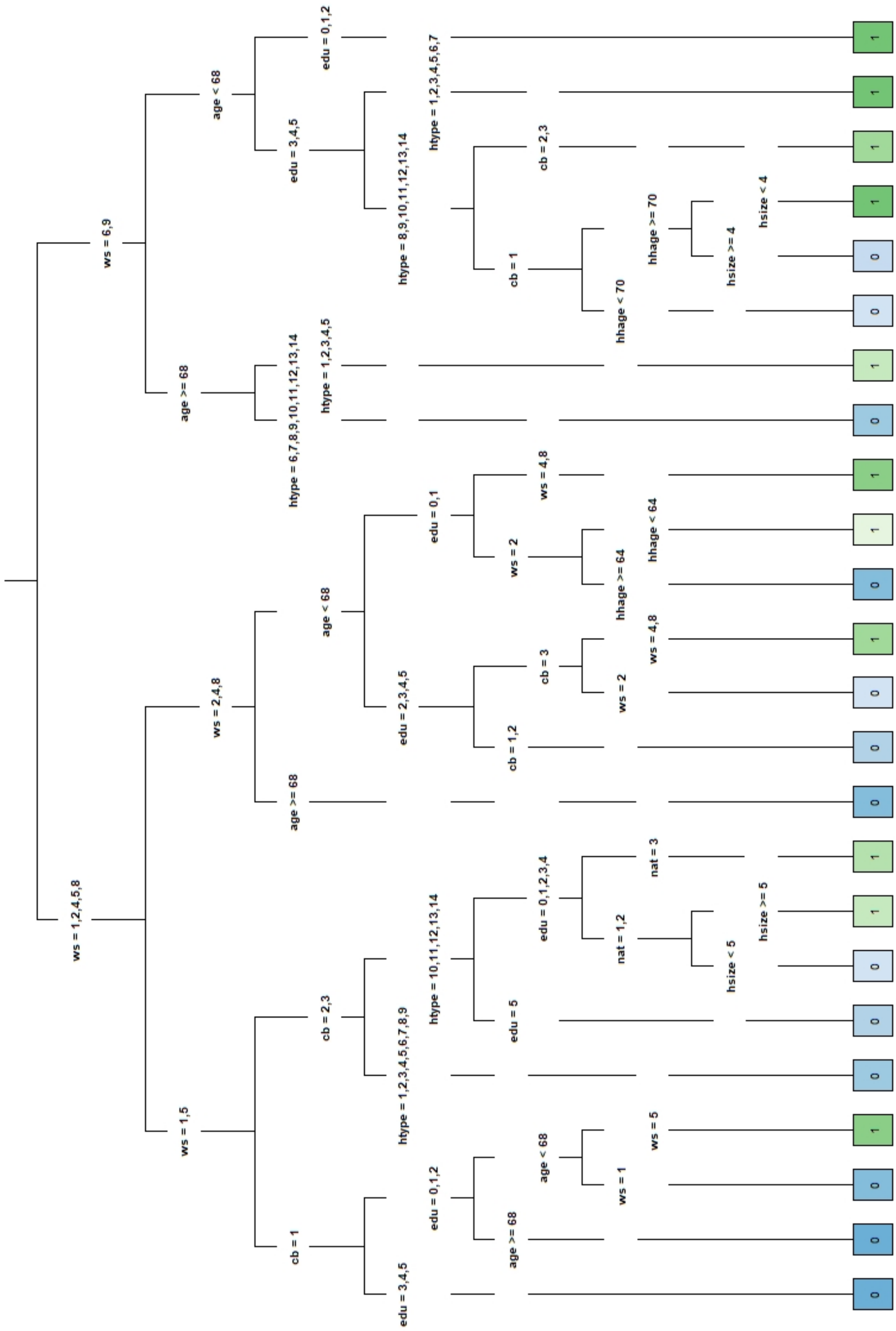
In this sub-section, we extrapolate information about groups of individuals and how they contributed to the rapid rise of poverty rates. For the sake of brevity, we report and comment only on the detailed decomposition of the 2008-2014 AROPE differential, and we relegate tables of analogous analysis on the other components to the appendix. The tree generated by the CART algorithm, represented in Figure 2, furnishes information about the importance of each characteristic in the splitting rules and returns an immediate overview of the generated groups. CART calculates the importance[10] of each variable

[8] Such considerations assume causal character only when the relationship between conditional and marginal distribution is restricted (See Firpo et al. [2009]). These assumptions are unrealistic when we analyze a global shock with general equilibrium effects.

[9] See Layte et al. [2001], Berthoud et al. [2004], Ayala et al. [2011].

[10] Notice that this is the importance of the characteristics for the splitting rule in the 2014 sample.

Figure 2: AROPE Tree: The terminal leafs report the predicted AROPE of each class in the 2014 population; green colors denote high within-group AROPE rate, while blue denotes the opposite.



as the sum of a goodness of split measures for each split as the main variable plus the goodness for each split as a secondary variable. The algorithm returns the following list, ordered in terms of relevance, of the covariates: working status, age, education, country of birth, HH type, HH head age, nationality, HH size, and sex; the variable HH head sex does not provide any split, and so it has zero importance. The first one, working status, has a disproportionate weight in the splitting rules, confirming her role as primary driver of the AROPE condition.

Table 4: **Groups Details**

Cluster	**Ws**	**Age**	**Edu**	**Cb**	**Nat**	**Htype**	**HHage**	**Hsize**
1	1,5	≥ 67.5	0,1,2	1	-	-	-	-
2	6,9	< 67.5	3,4,5	1	-	8,9,10,11,12,13,14	≥ 69.5	≥ 3.5
3	1,5	-	3,4,5	1	-	-	-	-
4	2,4,8	≥ 67.5	-	-	-	-	-	-
5	6,9	< 67.5	0, 1, 2	-	-	-	-	-
6	6,9	< 67.5	3,4,5	1	-	8,9,10,11,12,13,14	< 69.5	-
7	1	< 67.5	0,1,2	1	-	-	-	-
8	6,9	< 67.5	3,4,5	1	-	8,9,10,11,12,13,14	≥ 69.5	< 3.5
9	2,4,8	< 67.5	2,3,4,5	1,2	-	-	-	-
10	1,5	-	5	2,3	-	10,11,12,13,14	-	-
11	6,9	≥ 67.5	-	-	-	6,7,8,9,10,11,12,13,14	-	-
12	1,5	-	-	2,3	-	1,2,3,4,5,6,7,8,9	-	-
13	6,9	< 67.5	3,4,5	2,3	-	8,9,10,11,12,13,14	-	-
14	4,8	< 67.5	0,1	-	-	-	-	-
15	5	< 67.5	0,1,2	1	-	-	-	-
16	1,5	-	0,1,2,3,4	2,3	3	10,11,12,13,14	-	-
17	2	< 67.5	2,3,4,5	3	-	-	-	-
18	6,9	< 67.5	3,4,5		-	1,2,3,4,5,6,7	-	-
19	2	< 67.5	0,1	-	-	-	≥ 63.5	-
20	2	< 67.5	0,1	-	-	-	< 63.5	-
21	1,5	-	0,1,2,3,4	2,3	1,2	10,11,12,13,14	-	< 4.5
22	6,9	≥ 67.5	-	-	-	1,2,3,4,5	-	-
23	4,8	< 67.5	2,3,4,5	3	-	-	-	-
24	1,5	-	0,1,2,3,4	2,3	1,2	10,11,12,13,14	-	≥ 4.5

The table provides the split rule for each group determined by the CART algorithm. Variables for which there were no useful splits have been excluded.

Table 4 and Table 5 give a thorough overview of the groups and their contribution to the AROPE rise; the former reports the splitting rule determining each group, the latter reports the relative size of the groups and the within-group AROPE rate in the 2008 and 2014 populations. Increases (decreases) in the probability of being in these groups, $\mu_l^{(14)} - \mu_l^{(08)}$, produce positive (negative) composition contributions; increases (decreases) of the within-group AROPE rate, $h_l^{(14)} - h_l^{(08)}$, determine positive (negative) residual contribution. The magnitude of the contribution on the two term is determined by the interaction between these variations and the probability of being in AROPE in 2014 (composition) or in the group in 2008 (residual).

For instance, the third group is the largest one in both periods; it consists of Spanish-born individuals employed in full-time jobs or retired with a high level of education (above or equal to secondary education). Between 2008 and 2014, the likelihood of being in this group and of being in AROPE in this group both declined, implying that the group had a negative contribution to residual and composition components. As both $h_l^{(14)}$ and $h_l^{(14)} - h_l^{(08)}$ are very small, the group contributions are of modest entity.

Table 5: **Groups Contribution**

Cluster	$\boldsymbol{h_l^{(14)}}$	$\boldsymbol{\mu_l^{(14)}}$	$\boldsymbol{h_l^{(08)}}$	$\boldsymbol{\mu_l^{(08)}}$	**ContrC**	**ContrR**
1	12.19	12.27	22.70	15.38	-37.9109	-161.6438
2	38.32	0.34	39.80	0.10	9.1968	-0.1480
3	6.85	25.36	7.03	29.36	-27.4000	-5.2848
4	18.01	7.90	32.05	7.94	-0.7204	-111.4776
5	71.61	10.74	55.86	2.28	605.8206	35.9100
6	37.13	4.73	21.00	0.62	152.6043	10.0006
7	22.79	9.11	20.33	15.49	-145.4002	38.1054
8	71.54	0.35	83.85	0.09	18.6004	-1.1079
9	29.63	14.47	23.23	11.50	88.0011	73.6000
10	27.55	0.71	30.32	1.04	-9.0915	-2.8808
11	26.15	2.30	35.22	3.54	-32.4260	-32.1078
12	28.63	2.68	30.16	3.98	-37.2190	-6.0894
13	64.18	1.43	56.66	0.18	80.2250	1.3536
14	69.63	1.64	50.79	2.08	-30.6372	39.1872
15	77.59	0.26	55.47	0.14	9.3108	3.0968
16	67.06	0.77	50.81	2.22	-97.2370	36.0750
17	48.89	0.48	49.98	0.66	-8.8002	-0.7194
18	79.77	0.95	83.06	0.07	70.1976	-0.2303
19	22.70	0.24	29.63	0.24	0.0000	-1.6632
20	51.67	0.43	36.97	0.48	-2.5835	7.0560
21	46.56	0.98	45.90	0.74	11.1744	0.4884
22	63.46	0.28	64.55	0.23	3.1730	-0.2507
23	61.15	1.19	65.32	0.99	12.2300	-4.1283
24	57.48	0.30	47.30	0.52	-12.6456	5.2936

The table shows the estimated frequencies and AROPE probabilities for each group. The last two columns shows the contribution of each group to composition (ContrC) and residual (ContrR) components. All the statistics have been derived using working-age population survey weights.

Consider now the fifth group, which consists of working-age individuals permanently unable to work (or in another class of economic activity) with a low level of education. After the recession, the incidence of this group in the whole population rose from 2.28% in 2008 to 10.74% in 2014, and the within-group AROPE in 2014 stood at 71.61%. The interaction between these two factors resulted in the highest contribution to the composition component among all groups. The mechanism driving these group-frequency variations are the labor market perturbations consequent to the shock. Moreover, the group analysis reveals patterns in the data that cannot be seen by looking solely at the aggregate statistics. For instance, consider the first and sixteenth groups. The former consists of Spanish elderly individuals, employed full-time or retired, with a low educational level. The latter, instead, consists of foreign-born individuals, in full-time employment or retired, without higher education, living in a household with dependent children. Both groups reduced their incidence in the 2008-2014 period resulting in a negative contribution to the composition component. While the mechanism behind the weight reduction of groups three and five is identifiable in the loss of jobs experienced across the period, it is not clear what pushed the reduction in the weight of groups one and sixteen. For group one, two mechanisms might be thought of: the first is through an increase in the education level, resulting in a lower proportion of elderly individuals with low education; the second, through demographic changes in the 2014 population, resulting in an overall smaller proportion of elderly individuals. For group sixteen, instead, we can think of a mix of job loss and migratory movements of foreign-born individuals toward more prosperous countries.

Characteristics Contribution

In the final sub-section of the application, we assume a linear specification for the data within each cell. The restriction allows distinguishing the contribution of each covariate on the composition and residual components, as shown in section 3, while still keeping a degree of generality larger than usual decomposition with the linear specification. The characteristics' contributions to the composition component (C) are unique, the ones to the residual component are split into two parts: one reflecting variations into returns on characteristics (R1); the other variations of within-cluster composition (R2). The interpretation of (R1) and (R2) are analogous to the Endowment/coefficients components of

Table 6: **Characteristics Contribution**

	Cons.	**Sex**	**Age**	**Nat**	**Cb**	**Edu**	**Ws**	**Hsize**	**Htype**	**HHsex**	**HHage**
AROPE											
C	0.074	-0.015	0.037	-0.009	-0.002	-0.011	-0.016	-0.024	0.019	0.000	0.013
R1	-0.049	0.046	0.043	-0.088	-0.040	0.010	-0.032	0.082	-0.082	0.000	-0.051
R2	0.000	0.001	0.003	-0.001	0.000	0.000	0.000	-0.001	0.001	0.000	0.001
AROP60											
C	0.119	-0.010	-0.011	0.002	-0.013	-0.003	0.001	-0.005	-0.027	0.002	-0.010
R1	0.160	-0.001	-0.013	-0.012	0.039	0.007	-0.039	0.066	-0.043	-0.029	-0.153
R2	0.000	0.003	0.002	-0.001	0.000	-0.005	0.000	-0.003	0.001	0.000	0.005
MD											
C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R1	0.108	-0.003	-0.032	-0.013	0.024	-0.016	0.034	0.032	-0.041	0.017	-0.077
R2	0.000	0.000	0.005	0.000	0.000	-0.002	0.005	-0.001	0.001	0.000	0.005
LJ											
C	0.028	-0.010	0.042	0.000	-0.004	-0.010	-0.002	0.002	0.008	0.002	0.008
R1	-0.110	0.001	0.191	0.012	0.000	-0.023	-0.078	0.020	-0.003	-0.001	-0.006
R2	0.000	0.000	-0.001	0.000	0.000	-0.002	0.000	0.000	0.000	0.000	-0.002

The table shows the contribution of each characteristics to the composition component (C), and to the structure (R1) and compositional (R2) part of the residual component (see section 2).

OB decomposition, while (C) can be seen as the contribution of each characteristic to the difference among groups. The most meaningful contributions to the difference between groups in the AROPE decomposition are from age, working status, household size, and type. The highest coefficient contributions are made by household size and type, as well as nationality; age is the most important for the endowment component. For the AROP60 decomposition, household's head age cover the major role. The decomposition of the MD index in table 6 is numerically equivalent to the classic OB decomposition. As the CART algorithm left the data unpartitioned for this index, the contribution of each characteristic to the difference among groups is zero. The highest coefficient contribution is again the household age. In the LJ index, age seems the most relevant explanatory variable both in C and R1; in R1 also working status covers a large part.

Conclusion

We propose approximating counterfactual functionals using a re-weighting procedure based on aggregated data. The technique allows a path-independent detailed decomposition with the contribution of the groups on the decomposition's components; under additional restriction on the data generating process, the detailed decomposition also returns the contribution of each characteristics. We propose to form a partition of the data using the CART algorithm. Under this method, the estimator has good finite sample performances, and the groups possess meaningful interpretation. In the last section, we use the methodology to decompose rises in poverty rates in Spain after the great recession. The results show that variations in the composition of poverty risk factors, such as unemployment, explain most of the rise in AROPE rates, while the rise in material deprivation rates is left unexplained. The detailed decomposition shows that perturbations in the distribution of the group labor force raised poverty, but perturbation in demographics worked in the opposite sense.

A Appendix

A.1 Simulations

Our simulation design compares the finite sample performance of the grouped decomposition (using CART and NN clustering) to other commonly used parametric and non-parametric decomposition methods. The parametric approaches considered are the classic OB decomposition (linear parametrization) and the Machado-Mata method, under linear specification of the conditional quantiles. The non-parametric estimators include Kernel regressions (see Rothe [2010a]) and matching estimator (see Ñopo [2008]). We divide the simulation study in two parts. First, we discuss when the covariate vector $X^{(g)}$ consists of continuous variables, using three different specifications for the regression function $\gamma^{(g)}(X^{(g)})$. Second, we discuss when $X^{(g)}$ consists of both discrete and continuous variables, using four different specifications for $\gamma^{(g)}(X^{(g)})$. The simulated samples have sizes (n_1, n_0) varying from $(1000, 1000)$ to $(5000, 5000)$. For each model and estimator we perform 1000 Monte Carlo replications, reporting the mean absolute error (MAE) and root mean square error (RMSE) of the counterfactual mean estimates.

We assume the data generating process (DGP) of outcomes $Y^{(g)}$ according to $Y^{(g)} = \mu + \gamma^{(g)}(X^{(g)}) + V^{(g)}$ for $g \in \{0, 1\}$ with $\mu = 5$, where the residual $V^{(g)}$ follows the truncated normal distribution $N(0, (1 + \gamma^{(g)}(X^{(g)}))^2, -3, 3)$.

Continuous Variables

In the first simulation design, the covariates vector $X^{(g)}$ consists of K continuous variables. Each component is independently distributed as an exponential r.v. $f(x, \lambda_g) = \lambda_g e^{-\lambda_g x}$ truncated at 1 with the decay parameter $\lambda = 3$ for population 0 and $\lambda = 4.5$ for population 1. We discuss K varying from 2, 5 to 10 and three specifications that only allows the first two components in $X^{(g)}$ to affect $\gamma^{(g)}(X^{(g)})$ as following:

$$\begin{aligned}
&\text{Model 1:} \quad Y_i^{(1)} = \mu + X_{i1}^{(1)} + X_{i2}^{(1)} + V_i^{(1)} \\
&\text{Model 2:} \quad Y_i^{(1)} = \mu + \sin(\pi X_{i1}^{(1)}) + \sin(\pi X_{i2}^{(1)}) + V_i^{(1)} \qquad (6) \\
&\text{Model 3:} \quad Y_i^{(1)} = \mu + \mathbf{1}(X_{i1}^{(1)} \geq 0.5) + \mathbf{1}(X_{i2}^{(1)} \geq 0.5) + V_i^{(1)}
\end{aligned}$$

Whereas the model for population 0 is $Y_i^{(0)} = \mu + X_{i1}^{(0)} + X_{i2}^{(0)} + V_i^{(0)}$ for all DGP settings. We denote the three models as linear, nonlinear, and discontinuous, respectively.

The tuning parameters are chosen as follow: the matching estimator only uses the nearest neighbor; the bandwidth for the Kernel estimator is $h_b = 1.5\sigma_{x_k} n^{-(2K+1)}$ [Rothe, 2010a], where σ_{x_k} is the standard deviation of the respective covariates. For the grouping estimator: the minimum size of each partition, for both CART and NN, is $h_c = 2n_1^{\frac{1}{3}}$; in the CART algorithm, we set the initial penalty term to 0 cp = 0; in the k mean clustering algorithm we set the initial number of partitions as $\frac{n}{2h_c}$ and the maximum number of iteration to 100.

Table 7 reports the results of the first set of simulations. Not surprisingly, in the linear model, the OB and quantile decompositions have the smallest MAE and RMSE. But their performance drastically decrease when the model is not linear. For the smooth nonlinear model assumed in DGP 2, all the non-parametric approaches show significant advantages over parametric estimators when $K = 2$. In the case of the sparse model ($K = 5$ and $K = 10$), we find that only the CART algorithm keeps good performance, while all other non-parametric methods suffer the severe curse of dimensionality. In some cases, the non-parametric approaches perform worse than the inconsistent linear or quantile regression. In the discontinuous model of DGP 3, the CART algorithm is superior to other non-parametric methods, regardless of the dimension K. The results in DGP 2 and 3 show that the partitioning estimator with K-means clustering performs similarly to the Kernel estimator. Both Kernel and NN-partitioning have smaller MAE and RMSE than the matching estimator when $K = 2$, but bigger error when $K = 5, 10$.

To conclude: Parametric methods perform well in the linear model, but the non-parametric estimators outperform them in the other DGPs. The CART algorithm performs well in the sparse case for all the DGP settings, while the other non-parametric estimators have optimal performances only in the lower dimensional case.

Mixed Variables

Our second simulation design considers as covariates a continuous and an unordered categorical variables. The continuous variable, $X^{(g)}$, is distributed as in the first simulation design. The categorical variable $Z^{(g)}$ has support on an unordered set $\{a_1, a_2, ..., a_L\}$, where L varies from 5 to 10. We equivalently write $Z^{(g)}$ as L dummy variables $Z_t^{(g)} = \mathbf{1}(Z^{(g)} = t)$. For $t = 1, ..., L$, the probabilities $\mathbf{P}(Z_t^{(0)} = 1) = \frac{1}{L}$ and $\mathbf{P}(Z_t^{(1)} = 1) = F(F^{-1}(t/L, \lambda_0), \lambda_1) - F(F^{-1}((t-1)/L, \lambda_0), \lambda_1)$, where $F(x, \lambda)$ is the cumulative distribution function of $f(x, \lambda_g)$. We discuss four specifications:

$$\begin{aligned}
&\text{Model 1: } Y_i^{(1)} = \mu + X_i^{(1)} + \sum_{t=1}^{L} \beta_t Z_{it}^{(1)} + V_i^{(1)} \\
&\text{Model 2: } Y_i^{(1)} = \mu + \sin(2\pi X_i^{(1)}) + \sum_{t=1}^{L} \beta_t Z_{it}^{(1)} + V_i^{(1)} \\
&\text{Model 3: } Y_i^{(1)} = \mu + \sin(2\pi X_i^{(1)}) + \sum_{t=1}^{L} \beta_t Z_{it}^{(1)} + \sum_{t=1}^{L} \beta_t Z_{it}^{(1)} \sin(2\pi X_i^{(1)}) + V_i^{(1)} \\
&\text{Model 4: } Y_i^{(1)} = \mu + \mathbf{1}(X_i^{(1)} \leq \frac{1}{2}) + \sum_{t=1}^{L/2} Z_{it}^{(1)} + V_i^{(1)}.
\end{aligned} \tag{7}$$

Whereas the model for population 0 is $Y_i^{(0)} = \mu + X_i^{(0)} + \sum_{t=1}^{L-1} \beta_t Z_{it}^{(0)} + V_i^{(0)}$. For the Kernel estimator, we use the Kernel and bandwidth as in Racine and Li [2004]. Their approach computes the bandwidth via the cross-validation method. To reduce the computation involved, we first simulate each DGP setting 100 times and then adopt the average bandwidth. Our unfeasible bandwidth gives some advantage to the Kernel regression, as it is closer to the optimal bandwidth than the bandwidth available in applications. For the other non-parametric methods, we use the same tuning parameters as in the first experiment. Table 8 reports the simulation results for mixed data. In the linear model, the parametric approaches have the smallest MAE and RMSE. In DGP2 (the non-linear smooth model), the non-parametric methods dominate the parametric estimators when L = 5. However, the performance pattern among non-parametric approaches is different from the first experiment. As the nearest neighbor matched pairs often are located in the same cell given by dummy variables, the matching estimator keeps good performance even when L = 10. According to Table 8, the CART performs better than matching only if the DGP involves the interaction term between continuous and discrete variables. The Kernel method proposed by Racine and Li [2004] can not adjust to large L, and its performance is similar to the NN-Clustering partition methods. Finally, in the discontinuous model, the CART approach performs better than all the other estimators. In conclusion, the CART algorithm and the matching perform well with discrete variables and many categories, but the CART has higher performance in the discontinuous model. The other two non-parametric methods cannot suffer when dealing with many categories.

Model	K	$n_1(n_0)$	OB	Quantile	Kernel	Matching	Part	CART
				MAE				
Linear	2	1000	0.0360	0.0426	0.0398	0.0494	0.0412	0.0581
	2	5000	0.0158	0.0331	0.0176	0.0229	0.0176	0.0224
	5	1000	0.0389	0.0457	0.0723	0.0538	0.0623	0.0605
	5	5000	0.0183	0.0307	0.0578	0.0291	0.0436	0.0283
	10	1000	0.0452	0.0508	0.1008	0.0746	0.0972	0.0677
	10	5000	0.0204	0.0319	0.0924	0.0535	0.0670	0.0286
Nonlinear	2	1000	0.0501	0.0805	0.0412	0.0556	0.0416	0.0416
	2	5000	0.0333	0.0756	0.0192	0.0249	0.0188	0.0191
	5	1000	0.0503	0.0764	0.0563	0.0543	0.0697	0.0415
	5	5000	0.0333	0.0738	0.0419	0.0256	0.0455	0.0193
	10	1000	0.0589	0.0811	0.0714	0.0615	0.1041	0.0431
	10	5000	0.0354	0.0736	0.0618	0.0341	0.0696	0.0188
Discontinuous	2	1000	0.0379	0.0372	0.0396	0.0505	0.0400	0.0368
	2	5000	0.0209	0.0174	0.0193	0.0221	0.0191	0.0168
	5	1000	0.0416	0.0395	0.0956	0.0590	0.0654	0.0368
	5	5000	0.0212	0.0190	0.0727	0.0288	0.0374	0.0163
	10	1000	0.0474	0.0457	0.1466	0.0971	0.1117	0.0374
	10	5000	0.0238	0.0213	0.1297	0.0688	0.0697	0.0175
				RMSE				
Linear	2	1000	0.0447	0.0526	0.0502	0.0620	0.0510	0.0738
	2	5000	0.0199	0.0376	0.0220	0.0286	0.0222	0.0281
	5	1000	0.0493	0.0568	0.0822	0.0664	0.0738	0.0743
	5	5000	0.0230	0.0363	0.0613	0.0359	0.0484	0.0341
	10	1000	0.0558	0.0627	0.1086	0.0889	0.1057	0.0826
	10	5000	0.0257	0.0381	0.0943	0.0593	0.0707	0.0348
Nonlinear	2	1000	0.0625	0.0934	0.0515	0.0692	0.0521	0.0523
	2	5000	0.0390	0.0790	0.0240	0.0312	0.0234	0.0239
	5	1000	0.0628	0.0904	0.0679	0.0680	0.0823	0.0512
	5	5000	0.0399	0.0783	0.0468	0.0317	0.0508	0.0242
	10	1000	0.0736	0.0976	0.0836	0.0762	0.1143	0.0542
	10	5000	0.0425	0.0792	0.0654	0.0413	0.0741	0.0238
Discontinuous	2	1000	0.0479	0.0465	0.0503	0.0632	0.0505	0.0463
	2	5000	0.0260	0.0217	0.0243	0.0278	0.0241	0.0211
	5	1000	0.0520	0.0495	0.1042	0.0727	0.0769	0.0456
	5	5000	0.0262	0.0234	0.0751	0.0355	0.0426	0.0202
	10	1000	0.0590	0.0573	0.1519	0.1107	0.1193	0.0473
	10	5000	0.0297	0.0272	0.1310	0.0737	0.0735	0.0219

Table 7: **The Counterfactual Mean**

Model	L	$n_1(n_0)$	OB	Quantile	Kernel	Match	Part	CART
			MAE					
Linear	5	1000	0.0378	0.0440	0.0654	0.0552	0.0473	0.0632
		5000	0.0173	0.0338	0.0249	0.0232	0.0178	0.0241
	10	1000	0.0372	0.0419	0.0895	0.0553	0.0733	0.0657
		5000	0.0173	0.0328	0.0357	0.0247	0.0329	0.0268
Nonlinear	5	1000	0.0447	0.0644	0.0572	0.0549	0.0517	0.0546
		5000	0.0257	0.0575	0.0221	0.0218	0.0187	0.0290
	10	1000	0.0456	0.0624	0.0736	0.0567	0.0716	0.0562
		5000	0.0260	0.0568	0.0316	0.0231	0.0351	0.0290
Interaction	5	1000	0.0479	0.0690	0.0443	0.0538	0.0552	0.0440
		5000	0.0316	0.0655	0.0176	0.0212	0.0176	0.0197
	10	1000	0.0492	0.0681	0.0581	0.0540	0.0771	0.0430
		5000	0.0315	0.0633	0.0210	0.0223	0.0394	0.0197
Discontinuous	5	1000	0.0400	0.0494	0.0527	0.0567	0.0540	0.0392
		5000	0.0187	0.0384	0.0205	0.0230	0.0194	0.0169
	10	1000	0.0411	0.0501	0.0743	0.0575	0.0868	0.0408
		5000	0.0196	0.0375	0.0268	0.0253	0.0406	0.0177
			RMSE					
Linear	5	1000	0.0476	0.0548	0.0758	0.0697	0.0582	0.0766
		5000	0.0216	0.0388	0.0303	0.0288	0.0226	0.0303
	10	1000	0.0468	0.0531	0.0978	0.0705	0.0847	0.0797
		5000	0.0217	0.0382	0.0405	0.0306	0.0387	0.0329
Nonlinear	5	1000	0.0557	0.0771	0.0689	0.0695	0.0641	0.0672
		5000	0.0313	0.0619	0.0270	0.0275	0.0231	0.0348
	10	1000	0.0562	0.0751	0.0842	0.0713	0.0839	0.0697
		5000	0.0317	0.0613	0.0365	0.0292	0.0413	0.0351
Interaction	5	1000	0.0601	0.0824	0.0547	0.0677	0.0683	0.0543
		5000	0.0369	0.0691	0.0222	0.0268	0.0222	0.0246
	10	1000	0.0621	0.0824	0.0684	0.0683	0.0911	0.0532
		5000	0.0368	0.0671	0.0260	0.0280	0.0454	0.0247
Discontinuous	5	1000	0.0499	0.0604	0.0640	0.0712	0.0663	0.0489
		5000	0.0236	0.0436	0.0257	0.0286	0.0244	0.0212
	10	1000	0.0514	0.0614	0.0860	0.0717	0.1011	0.0515
		5000	0.0246	0.0434	0.0325	0.0319	0.0470	0.0221

Table 8: **The Counterfactual Mean:mixed**

A.2 Asymptotic Distribution of Decomposition Components

Proposition 2 ***(Asymptotic distribution decomposition component)***
Let $\widehat{\Delta}^C_{\mathbb{C}_{L_N}} = \widehat{A}^{(1,1)} - \widehat{A}^{(1,0)}_{\mathbb{C}_{L_N}}$ *and* $\widehat{\Delta}^R_{\mathbb{C}_{L_N}} = \widehat{A}^{(1,0)}_{\mathbb{C}_{L_N}} - \widehat{A}^{(0,0)}$ *be the estimators of the composition and residual components respectively, under assumption 1,2 and 6 we have that:*

$$\sqrt{N^1}\left(\widehat{\Delta}^C_{\mathbb{C}_{L_N}} - \Delta^C_{\mathbb{C}_L}\right) \xrightarrow{d} N(0, V_{\Delta_C})$$

$$V_{\Delta_C} = \sum_{l=1}^{L}(\mu_l^{(1)} - \mu_l^{(0)})^2\sigma_l^{(1)} + \frac{1}{\gamma}\sum_{l=1}^{L}\sum_{f=1}^{L} h_l^{(1)} v_{(l,f)} h_f^{(1)}$$

Where:

$$v_{(l,f)} = \begin{cases} \mu_l^{(1)}(1-\mu_l^{(1)}) + \gamma\mu_l^{(0)}(1-\mu_l^{(0)}) & \text{if } l = f \\ -(\mu_l^{(1)}\mu_f^{(1)} + \gamma\mu_l^{(0)}\mu_f^{(0)}) & \text{if } l \neq f \end{cases}$$

And:

$$\sqrt{N^1}\left(\widehat{\Delta}^R_{\mathbb{C}_{L_N}} - \Delta^R_{\mathbb{C}_L}\right) \xrightarrow{d} N(0, V_{\Delta_R})$$

$$V_{\Delta_R} = \frac{1}{\gamma}\sum_{l=1}\left(\mu_l^{(0)}\right)^2\left(\gamma\sigma_l^{(1)} + \sigma_l^{(0)}\right) + \frac{1}{\gamma}\sum_{l=1}\sum_{f=1}\left(h_l^{(1)} - h_l^{(0)}\right)\mu^{(0)}_{(l,f)}\left(h_f^{(1)} - h_f^{(0)}\right)$$

Where:

$$\mu^{(0)}_{(l,f)} = \begin{cases} \mu_l^{(0)}(1-\mu_l^{(0)}) & \text{if } l = f \\ -(\mu_l^{(0)}\mu_f^{(0)}) & \text{if } l \neq f \end{cases}$$

Proof. *Appendix.* ■

A.3 Proofs

Asymptotic Distribution Proof.

For a given partition $\mathbb{C}_L$, let denote the conditional mean of $g(Y^{(1)})$ for a set of characteristics C_l as $\widehat{h}_l^{(1)} = \frac{1}{N_l^1}\sum_{i=1}^{N^1} g(Y_i^{(1)})\mathbb{I}\{X_i^{(1)} \in C_l\}$, where $N_l^1 = \sum_{i=1}^{N^1}\mathbb{I}\{X_i^{(1)} \in C_l\} = \widehat{\mu}_l^{(1)}N^1$. Also, let $\widehat{h}^{(1)}$ and $\widehat{\mu}^{(0)}$ be the $L-$dimensional vectors with elements $\widehat{h}_l^{(1)}$ and $\widehat{p}_l^{(0)}$ respectively, we can write the partitioning estimator as $A^{(1,0)}_{\mathbb{C}_L} = \bar{h}^{(1)} \cdot \widehat{\mu}^{(0)}$. Hence, we determine the asymptotic distribution of the estimator by applying the delta method on the joint distribution of $\widehat{h}^{(1)}$ and $\widehat{\mu}^{(0)}$.

Denote the population variance of $g(Y^{(1)})$, conditioning on a set of characteristics C_l, as $\sigma_l^{(1)}$ and the probability that an individual of population (0) shows the set of characteristics C_l as $\mu_l^{(0)}$. The asymptotic distribution of the two vectors is given by:

$$\sqrt{N^1}\left(\widehat{h}^{(1)} - h^{(1)}\right) \xrightarrow{d} N(0, V_h)$$
$$\sqrt{N^0}\left(\widehat{\mu}^{(0)} - \mu^{(0)}\right) \xrightarrow{d} N(0, V_\mu)$$

With: $V_h = \begin{bmatrix} \sigma_1^{(1)} & 0 & \cdots & 0 \\ 0 & \sigma_2^{(1)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_L^{(1)} \end{bmatrix}$ $V_\mu = \begin{bmatrix} \mu_1^{(0)}(1-\mu_1^{(0)}) & -\mu_1^{(0)}\mu_2^{(0)} & \cdots & -\mu_1^{(0)}\mu_L^{(0)} \\ -\mu_2^{(0)}\mu_1^{(0)} & \mu_2^{(0)}(1-\mu_2^{(0)}) & \cdots & -\mu_2^{(0)}\mu_L^{(0)} \\ \vdots & \vdots & \ddots & \vdots \\ -\mu_L^{(0)}\mu_1^{(0)} & -\mu_2^{(0)}\mu_L^{(0)} & \cdots & \mu_L^{(0)}(1-\mu_L^{(0)}) \end{bmatrix}$

Under Assumption 2 we have that the two populations grow proportionally, such that $\frac{N^1}{N^0} = \gamma$ for some $\gamma > 0$, and that the two samples are independent of each others. Then:

$$\sqrt{N^1}\begin{bmatrix} \widehat{h}^{(1)} - h^{(1)} \\ \widehat{\mu}^{(0)} - \mu^{(0)} \end{bmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} V_h & 0 \\ 0 & \frac{V_\mu}{\gamma} \end{bmatrix}\right) \tag{A.1}$$

Let $f(X;Y) = X \cdot Y$ then $\nabla f = (Y, X)'$, then from delta method it follows that:

$$\sqrt{N^1}\left(\widehat{A}_{\mathbb{C}_L}^{(1,0)} - A_{\mathbb{C}_L}^{(1,0)}\right) \xrightarrow{d} N\left(0, [\mu^{(0)\prime}, h^{(1)\prime}]V[\mu^{(0)\prime}, h^{(1)\prime}]'\right) = N\left(0, V_A\right)$$

Where $V_A = \sum\limits_{l=1}^{L}\left((\mu_l^{(0)})\sigma_l^{(1)} + \frac{1}{\gamma}(h_l^{(1)})^2\mu_l^{(0)}(1-\mu_l^{(0)})\right) - \frac{1}{\gamma}\sum\limits_{l\neq f} h_l^{(1)}h_f^{(1)}\mu_l^{(0)}\mu_f^{(0)}$.

The next step of the proof accounts for the different data generating process of the data, as explained in section 2. In particular we now assume $N^1 = N^0 = N$ and we relax the independence assumption between the two populations. To determine the asymptotic covariance between the two populations is necessary to decompose the conditional mean estimator. Specifically, let $\widehat{q}_l^{(1)} = \frac{1}{N}\sum\limits_{i=1}^{N} g(Y_i^{(1)})\mathbb{I}\{X_i^{(1)} \in C_l\}$ such that $\widehat{h}_l^{(1)} = \frac{\widehat{q}_l^{(1)}}{\widehat{p}_l^{(1)}}$ and let $\widehat{q}_l^{(1)}$ be the L-dimensional vector with elements $\widehat{q}_l^{(1)}$. Further, for any L-dimensional vector X define as $diag(X) = |X|$ the diagonal matrices with the main diagonal given by elements of X. Then we can write the joint vector in (A.1) as:

$$\sqrt{N}\begin{bmatrix}\widehat{h}^{(1)}-h^{(1)}\\ \widehat{\mu}^{(0)}-\mu^{(0)}\end{bmatrix}=\sqrt{N}\begin{bmatrix}|\widehat{\mu}^{(1)}| & -|\widehat{\mu}^{(1)}|^{-1}|h^{(1)}| & 0\\ 0 & 0 & I_L\end{bmatrix}\begin{bmatrix}\widehat{q}^{(1)}-q^{(1)}\\ \widehat{\mu}^{(1)}-\mu^{(1)}\\ \widehat{\mu}^{(0)}-\mu^{(0)}\end{bmatrix}$$
$$=\sqrt{N}\widehat{A}B$$

Which has the following limit distribution:

$$\sqrt{N}\widehat{A}B\xrightarrow{d}N(0,V)$$

$$V=A\mathbb{V}(B)A'=\begin{bmatrix}V_h & V_c\\ V_c' & V_\mu\end{bmatrix}.$$

Where $A=\operatorname{plim}\widehat{A}$.
Then, define $\tilde{\Lambda}=Cov\left(\sqrt{N}\widehat{\mu}^{(0)},\sqrt{N}\widehat{q}^{(1)}\right)$ with typical elements $\tilde{\lambda}_{l,f}=q^{(0,1)}_{(l,f)}-\mu^{(0)}_l q^{(1)}_f$, where $q^{(0,1)}_{(l,f)}=\mathbb{E}\left[g(Y^{(1)})\mathbb{I}\{X_i^{(1)}\in C_f\wedge X_i^{(0)}\in C_l\}\right]$. And $\Omega=Cov\left(\sqrt{N}\widehat{\mu}^{(0)},\sqrt{N}\widehat{\mu}^{(1)}\right)$ with typical elements $\omega_{l,f}=\mu^{(0,1)}_{(l,f)}-\mu^{(0)}_l\mu^{(1)}_f$, where $\mu^{(0,1)}_{(l,f)}=\mathbb{E}\left[\mathbb{I}\{X_i^{(1)}\in C_f\wedge X_i^{(0)}\in C_l\}\right]$.
After some algebra we get the following formula for the covariance term:

$$V_c=\tilde{\Lambda}|\mu^{(1)}|^{-1}-\Omega|\mu^{(1)}|^{-1}|h^{(1)}|$$

With typical elements:

$$c_{(l,f)}=\frac{\mu^{(1)}_f q^{(0,1)}_{(l,f)}-q^{(1)}_f\mu^{(0,1)}_{(l,f)}}{\left(\mu^{(1)}_f\right)^2}$$

Finally, by delta method we get the variance of the counterfactual estimator under dependent populations:

$$V_{A_D}=V_A+2\sum_{l=1}^{L}\sum_{f=1}^{L}h^{(1)}_l c_{(l,f)}\mu^{(0)}_f$$

■

Proof. Asymptotic Distribution Decomposition Components

We show the result under independent data-structure, starting from the composition component and then moving to the residual part. Following a similar reasoning to the previous proof we can write the joint distribution of $(\widehat{h}^{(1)}, \widehat{\mu}^{(0)}, \widehat{\mu}^{(1)})$ as:

$$\sqrt{N^1}\begin{bmatrix}\widehat{h}^{(1)} - h^{(1)} \\ \widehat{\mu}^{(0)} - \mu^{(0)} \\ \widehat{\mu}^{(1)} - \mu^{(1)}\end{bmatrix} = \sqrt{N^1}\begin{bmatrix}|\widehat{\mu}^{(1)}| & -|\widehat{\mu}^{(1)}|^{-1}|h^{(1)}| & 0 \\ 0 & 0 & I_L \\ 0 & I_L & 0\end{bmatrix}\begin{bmatrix}\widehat{q}^{(1)} - q^{(1)} \\ \widehat{\mu}^{(1)} - \mu^{(1)} \\ \widehat{\mu}^{(0)} - \mu^{(0)}\end{bmatrix}$$
$$= \sqrt{N}\tilde{A}B$$

Let $\Lambda = Cov\left(\sqrt{N^1}\widehat{\mu}^{(1)}, \sqrt{N^1}\widehat{q}^{(1)}\right)$ and $\bar{A} = \operatorname{plim}\tilde{A}$, then:

$$\sqrt{N^1}\tilde{A}B \xrightarrow{d} N(0, V)$$

$$V = \bar{A}\mathbb{V}(B)\bar{A}' = \begin{bmatrix}V_{h^{(1)}} & 0 & C' \\ 0 & \frac{V_{\mu^{(0)}}}{\gamma} & 0 \\ C & 0 & V_{\mu^{(1)}}\end{bmatrix}.$$

Where $C = \Lambda|\mu^{(1)}|^{-1} - V_{\mu^{(1)}}|\mu^{(1)}|^{-1}|h^{(1)}| = 0$.

Finally, we apply the delta method on the function $f(X;Y;Z) = X'(Z-Y)$ such that $\Delta f' = [(Z-Y)'; -X'; X']$ and get:

$$\sqrt{N^1}\left(\widehat{\Delta}^C_{\mathbb{C}_L} - \Delta^C_{\mathbb{C}_L}\right) \xrightarrow{d} N(0, V_{\Delta_C})$$

$$V_{\Delta_C} = \Delta f' V \Delta f = \left(\mu^{(1)} - \mu^{(0)}\right)' V_{h^{(1)}}\left(\mu^{(1)} - \mu^{(0)}\right) + \frac{1}{\gamma}h^{(1)'}\left(\gamma V_{\mu^{(1)}} + V_{\mu^{(0)}}\right)h^{(1)}$$

In the previous part we showed that $C = Cov\left(\sqrt{N^1}\widehat{\mu}^{(1)}, \sqrt{N^1}\widehat{h}^{(1)}\right) = 0$ from which it follows that $Cov\left(\sqrt{N^1}\widehat{\mu}^{(0)}, \sqrt{N^1}\widehat{h}^{(0)}\right) = 0$, hence we can write the joint asymptotic distribution of $(\widehat{h}^{(1)}, \widehat{h}^{(0)}, \widehat{\mu}^{(0)})$ as:

$$\sqrt{N^1}\begin{bmatrix}\widehat{h}^{(1)} - h^{(1)} \\ \widehat{h}^{(0)} - h^{(0)} \\ \widehat{\mu}^{(0)} - \mu^{(0)}\end{bmatrix} \xrightarrow{d} N(0, V)$$

Where:

$$V = \begin{bmatrix} V_{h^{(1)}} & 0 & 0 \\ 0 & \frac{V_{h^{(0)}}}{\gamma} & 0 \\ 0 & 0 & \frac{V_{\mu^{(0)}}}{\gamma} \end{bmatrix}.$$

Then, let $f(X;Y;Z) = (X-Y)'Z$ such that $\Delta f' = [(Z'; -Z'; (X-Y)']$, by delta method:

$$\sqrt{N^1}\left(\widehat{\Delta}^R_{\mathbb{C}_L} - \Delta^R_{\mathbb{C}_L}\right) \xrightarrow{d} N(0, V_{\Delta_R})$$

$$V_{\Delta_R} = \Delta f' V \Delta f = \frac{1}{\gamma}\left(h^{(1)} - h^{(0)}\right)' V_{\mu^{(0)}}\left(h^{(1)} - h^{(0)}\right) + \frac{1}{\gamma}\mu^{(0)\prime}\left(\gamma V_{h^{(1)}} + V_{h^{(0)}}\right)$$

■

A.4 Application Appendix

Characteristics' details

Nat & Cb

The variable *Nat* indicates the nationality of the individual, while *Cb* the country of birth. Details below:

- Spain. (1)
- Rest of European Union (EU28) (2).
- Rest of the world (3).

Educ

The variable *Educ* denotes the education level. Details below:

- Less than primary (0).
- Primary education (1).
- First-stage secondary education (2).
- Second-stage secondary education (3).
- Post-secondary education (4).
- Higher education (5).

Ws

The variable *Ws* denotes the working status of the individual. Details below

- Employed (full-time) (1).
- Employed (part-time) (2).
- Inactive (3).
- Student or in formation (4).
- Retired, early retired or have closed down a business (5).
- Permanently unable to work (6).
- Compulsory military service or substitute social service (7).
- Dedicated to housework, care of children or other persons (8).
- Other class of economic inactivity (9).

HHtype

The variable *HHtype* denotes the type of household where the individual lives. Details below:

- 1 Adult: Male < 30 years old (1).
- 1 Adult: $30 \geq$ Male ≤ 64 years old (2).
- 1 Adult: Male ≥ 65 years old (3).
- 1 Adult: Female < 30 years old (4).
- 1 Adult: $30 \leq$ Female ≤ 64 years old (5).
- 1 Adult: Female ≥ 65 years old (6).
- 2 Adults without financially dependent children[11], at least one person above 65 years of age (7).
- 2 Adults without financially dependent children, both below 65 years of age (8).
- Other type of household without financially dependent children (9).
- 1 Adult with at least a dependent child (10).
- 2 Adults with a dependent child (11).
- 2 Adults with 2 dependent children (12).
- 2 Adults with 3 or more dependent children (13).
- Other type of household with dependent children (14).

HHsyze

The variable *HHsyze* denotes the size of the household. It can take values from 1 to 14.

HHhage

The variable *HHsyze* denotes the age of the household head. It can take value from 25 to 99.

Sex

The variable *Sex* denotes the gender of the individual: Male =1; Female=2.

[11]Note: In this classification we classify as financially dependent children the following categories:

- All those under the age of 18.
- Those who are 18 and older but under 25 and economically inactive.

References

Omar B Ahmad, Cynthia Boschi-Pinto, Alan D Lopez, Christopher JL Murray, Rafael Lozano, Mie Inoue, et al. Age standardization of rates: a new who standard. *Geneva: World Health Organization*, 9(10), 2001.

Robert P Althauser and Michael Wigler. Standardization and component analysis. *Sociological Methods & Research*, 1(1):97–135, 1972.

Joseph G Altonji, Prashant Bharadwaj, and Fabian Lange. Changes in the characteristics of american youth: Implications for adult outcomes. *Journal of Labor Economics*, 30 (4):783–828, 2012.

Robert N Anderson and Harry M Rosenberg. Age standardization of death rates; implementation of the year 2000 standard. 1998.

Susan Athey and Guido W Imbens. Machine learning methods that economists should know about. *Annual Review of Economics*, 11:685–725, 2019.

Anthony Barnes Atkinson. On the measurement of poverty. *Econometrica: Journal of the Econometric Society*, pages 749–764, 1987.

Luis Ayala, Antonio Jurado, and Jesús Pérez-Mayo. Income poverty and multidimensional deprivation: Lessons from cross-regional analysis. *Review of income and wealth*, 57(1): 40–60, 2011.

Luis Ayala, Antonio Jurado, and Jesús Pérez-Mayo. Multidimensional deprivation in heterogeneous rural areas: Spain after the economic crisis. *Regional Studies*, 55(5): 883–893, 2021.

Richard Berthoud, Elena Bardasi, Mark L Bryan, and G Britain. *The dynamics of deprivation: the relationship between income and material deprivation over time.* Number 219. Corporate Document Services, 2004.

Martin Biewen and Stephen P Jenkins. A framework for the decomposition of poverty differences with an application to poverty differences between countries. *Empirical Economics*, 30(2):331–358, 2005.

Alan S Blinder. Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455, 1973.

François Bourguignon and F Ferreira. Decomposing changes in the distribution of household incomes: methodological aspects. *The microeconomics of income distribution dynamics in East Asia and Latin America*, pages 17–46, 2005.

François Bourguignon, Francisco HG Ferreira, and Phillippe G Leite. Beyond oaxaca–blinder: Accounting for differences in household income distributions. *The Journal of Economic Inequality*, 6(2):117–148, 2008.

L Breiman, JH Friedman, RA Olshen, and CJ Stone. Cart. *Classification and Regression Trees; Wadsworth and Brooks/Cole: Monterey, CA, USA*, 1984.

Paolo Brunori. The perception of inequality of opportunity in europe. *Review of Income and Wealth*, 63(3):464–491, 2017.

Paolo Brunori, Vito Peragine, and Laura Serlenga. Upward and downward bias when measuring inequality of opportunity. *Social Choice and Welfare*, 52(4):635–661, 2019.

Matias D. Cattaneo and Max H. Farrell. Optimal convergence rates, Bahadur representation, and asymptotic normality of partitioning estimators. *Journal of Econometrics*, 174(2):127–143, 2013. ISSN 03044076. doi: 10.1016/j.jeconom.2013.02.002. URL http://dx.doi.org/10.1016/j.jeconom.2013.02.002.

Victor Chernozhukov, Iván Fernández-Val, and Blaise Melly. Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268, 2013.

John DiNardo, Nicole M Fortin, and Thomas Lemieux. Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica: Journal of the Econometric Society*, pages 1001–1044, 1996.

B Essama-Nssah and Peter J Lambert. Counterfactual decomposition of pro-poorness using influence functions. *Journal of Human Development and Capabilities*, 17(1):74–92, 2016.

Sergio Firpo, Nicole M Fortin, and Thomas Lemieux. Unconditional quantile regressions. *Econometrica*, 77(3):953–973, 2009.

Nicole Fortin, Thomas Lemieux, and Sergio Firpo. Decomposition methods in economics. In *Handbook of labor economics*, volume 4, pages 1–102. Elsevier, 2011.

James Foster, Joel Greer, and Erik Thorbecke. A class of decomposable poverty measures. *Econometrica: journal of the econometric society*, pages 761–766, 1984.

James Foster, Suman Seth, and Michael Lokshin. *A Unified Approach to Measuring Poverty and Inequality*. The World Bank, 2013.

James E Foster. Poverty indices. In *Poverty, Inequality and Development*, pages 41–65. Springer, 2006.

MP Gessaman. A consistent nonparametric multivariate density estimator based on statistically equivalent blocks. *The Annals of Mathematical Statistics*, 41(4):1344–1346, 1970.

Carlos Gradín. Poverty among minorities in the united states: Explaining the racial poverty gap for blacks and latinos. *Applied Economics*, 44(29):3793–3804, 2012.

Ahmedin Jemal, Elizabeth Ward, Yongping Hao, and Michael Thun. Trends in the leading causes of death in the united states, 1970-2002. *Jama*, 294(10):1255–1259, 2005.

Evelyn M Kitagawa. Components of a difference between two rates. *Journal of the american statistical association*, 50(272):1168–1194, 1955.

Evelyn M Kitagawa. Standardized comparisons in population research. *Demography*, 1 (1):296–315, 1964.

Richard Layte, Christopher T Whelan, Bertrand Maître, and Brian Nolan. Explaining levels of deprivation in the european union. *Acta Sociologica*, 44(2):105–121, 2001.

Thomas Lemieux. Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *American Economic Review*, 96(3):461–498, 2006.

José AF Machado and José Mata. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics*, 20(4):445–465, 2005.

FGP Neison. On a method recently proposed for conducting inquiries into the comparative sanatory condition of various districts, with illustrations, derived from numerous places in great britain at the period of the last census. *Journal of the Statistical Society of London*, 7(1):40–68, 1844.

Hugo Ñopo. Matching as a tool to decompose wage gaps. *The review of economics and statistics*, 90(2):290–299, 2008.

Ronald Oaxaca. Male-female wage differentials in urban labor markets. *International economic review*, pages 693–709, 1973.

Jeff Racine and Qi Li. Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119(1):99–130, 3 2004. ISSN 0304-4076. doi: 10.1016/S0304-4076(03)00157-X.

Xavier Ramos and Dirk Van de Gaer. Is inequality of opportunity robust to the measurement approach? *Review of Income and Wealth*, 67(1):18–36, 2021.

Christoph Rothe. Nonparametric estimation of distributional policy effects. *Journal of Econometrics*, 155(1):56–70, 3 2010a. ISSN 0304-4076. doi: 10.1016/J.JECONOM.2009.09.001. URL https://www.sciencedirect.com/science/article/pii/S0304407609001924.

Christoph Rothe. Nonparametric estimation of distributional policy effects. *Journal of Econometrics*, 155(1):56–70, 2010b.

Amartya Sen. Poverty: an ordinal approach to measurement. *Econometrica: Journal of the Econometric Society*, pages 219–231, 1976.

Amartya Sen, Master Amartya Sen, Sen Amartya, James Eric Foster, James E Foster, et al. *On economic inequality*. Oxford university press, 1997.

Henry S Shryock and Jacob S Siegel. *The methods and materials of demography*, volume 2. Department of Commerce, Bureau of the Census, 1980.

Myeong-Su Yun. Decomposing differences in the first moment. *Economics letters*, 82(2): 275–280, 2004.

Buhong Zheng. Aggregate poverty measures. *Journal of economic surveys*, 11(2):123–162, 1997.