Ordinal Decomposition

Qian Wu* David M. Kaplan†

May 30, 2024

Abstract

The famous Blinder–Oaxaca decomposition estimates the statistically "explained" proportion of a between-group difference in means, but ordinal variables have no mean. A common approach assigns cardinal values $1, 2, 3, \ldots$ to the ordinal categories and runs the conventional OLS-based decomposition. Surprisingly, we show such results are numerically identical to a decomposition of the survival function when estimating the counterfactual using OLS-based distribution regression, even if the cardinalization is wrong. Still, reporting the counterfactual helps transparency and wide-sense replication, and to mitigate functional form misspecification, we describe and implement a nonparametric estimator. Empirically, we decompose U.S. rural–urban differences in mental health.

JEL classification: C25, I14

Keywords: Blinder–Oaxaca decomposition, counterfactual distribution, distribution regression, survival function

*Corresponding author; School of Statistics, Southwestern University of Finance and Economics, wuqj@swufe.edu.cn

†Department of Economics, University of Missouri, kaplandm@missouri.edu

1 Introduction

The Blinder–Oaxaca decomposition is commonly used to decompose an overall mean difference in outcome between two groups into two components: one attributed to the group difference in explanatory variable means, and the other to differences in regression coefficients. Using the same idea published by Kitagawa (1955) and used earlier in the 1940s (see her footnote 3), the papers of Blinder (1973) and Oaxaca (1973) have over 20,000 citations in Google Scholar, with over 6000 of those coming since 2019, spanning the fields of economics, public health, sociology, medicine, demography, and others. The original papers decompose wages by sex and/or race.

Unfortunately, this approach cannot be applied directly to ordinal outcome variables, which are important in many social sciences like economics. An ordinal variable does not have a well-defined mean because its values are ordered categories rather than numerical, cardinal values. For example, individuals often rate their general health using the categories poor, fair, good, very good, and excellent; such values cannot be averaged, unless we assign a cardinal value to each category. In our empirical application, we decompose mental health differences between urban and rural residents in the U.S., using a depression variable with values none/minimal, mild, moderate, and severe. Besides health, ordinal variables are also common for bond ratings, consumer confidence, political indices, and other outcomes.

Despite this importance of both decomposition and ordinal variables, there is a limited literature on decomposition with ordinal outcomes. The extensive *Handbook of Labor* chapter on “Decomposition Methods in Economics” (Fortin, Lemieux, and Firpo, 2011) includes discussion of many population functionals and estimators and causal identification, but does not include the word “ordinal” anywhere in its 102 pages. (And “ordered” only appears in the context of parametric estimation of conditional distributions for a continuous outcome after “discretizing the outcome variable” (p. 70).) Bauer and Sinning (2008) propose an ordered probit/logit decomposition in their (4), but it is used only to introduce nonlinearity, while still treating the ordinal outcome as if it were discrete with cardinal values $1, 2, 3, \ldots$;

the same is true of Demoussis and Giannakopoulos (2007).[1] Similarly, empirical work often treats the numerically coded category labels $(1, 2, 3, \ldots)$ as cardinal values and then runs the standard OLS-based Blinder–Oaxaca decomposition; for example, see Pan, Liu, and Ali (2015, §2.4), Awaworyi Churchill, Munyanyi, Prakash, and Smyth (2020, §§2.1–2.2), and Pilipiec, Groot, and Pavlova (2020, §2.2). Madden (2010, §2) acknowledges the cardinalization is not fully appropriate (but facilitates variable-by-variable decomposition) and argues that the ordered probit results in his appendix offer evidence of robustness, but his ordered probit decomposition still uses the same cardinalization (p. 111). Other empirical work simply reduces the ordinal variable to a binary variable before doing a probit-based decomposition; for example, see Zhang, Bago d'Uva, and van Doorslaer (2015, eqn. (7)) and Hauret and Williams (2017, p. 217).

In this context, we make three contributions. Practically, we recommend a unified framework based on Chernozhukov, Fernández-Val, and Melly (2013), first constructing a counterfactual distribution and then decomposing one or more summary statistics. Theoretically, we show that population decompositions using the common but dubious cardinalization $1, 2, 3, \ldots$ can actually be interpreted meaningfully in terms of survival functions even if the "mean" interpretation is wrong. Further, we show that the estimated Blinder–Oaxaca decomposition with such a cardinalization is numerically equivalent to a counterfactual-based survival function decomposition using OLS-based distribution regression. Empirically, we apply the foregoing to decompose U.S. rural–urban mental health disparities.

Practically, we suggest that ordinal decompositions follow the approach of Chernozhukov, Fernández-Val, and Melly (2013) to estimate a certain counterfactual outcome distribution, after which summary statistics can be computed to estimate the "explained" proportion. Although they consider a continuous outcome, so their quantile regression results do not apply, their results based on distribution regression still hold for an ordinal outcome. From a computational perspective, they first estimate models like their (3.1) where the dependent

[1] Their [7] and [8] have an important typo: the left-hand sides should have expectations of S rather than probabilities, as is clear from the text ("expected JS") and the right-hand sides, and equation [9] later.

variable represents whether or not the ordinal outcome is at or below a certain category, re-estimating the model for each such category, using probit or logit, or a nonparametric alternative like series logit. This is essentially a generalization of the binary outcome decomposition of Fairlie (2005) that traces back (at least) to Even and Macpherson (1990) and Farber (1987); it is also more flexible than the ordered probit/logit decomposition of Bauer and Sinning (2008, eqn. (4)), who additionally impose the $1, 2, 3, \ldots$ cardinalization. The counterfactual distribution of Chernozhukov, Fernández-Val, and Melly (2013) combines the conditional outcome distribution from one group with the marginal distribution of covariates from the other group, as in their (2.1). Qualitatively, the difference between the first group and the counterfactual is the "explained" component, whereas the difference between the counterfactual and the second group is the unexplained component. To simplify communication of results, a particular summary statistic can be chosen.

Theoretically, first we show that choosing the summary statistic as the "mean" after assigning cardinal values $1, 2, 3, \ldots$ is equivalent to a summary statistic based only on the survival function. Using such a cardinalization is temptingly convenient: often these values are coded in the raw data, and treating them as cardinal allows use of conventional statistical methods like linear regression and the original Blinder–Oaxaca decomposition. However, taken literally, this cardinalization is often inappropriate: it assigns the same value to everyone within a category that more realistically represents a range of values, and it assumes the cardinal difference between any two consecutive categories is identical. For example, in our mental health application with our depression variable, it assigns the same cardinal value to everyone in the "mild" category (which represents a range of individuals' latent depression levels), and it assumes the cardinal difference between "none" and "mild" is the same as the difference between "mild" and "moderate" and the difference between "moderate" and "severe." Despite this, we show the "mean" decomposition can be interpreted as a decomposition of the difference in survival functions averaged across categories. The survival function is well-defined for any ordinal variable, without any cardinalization.

With enough ordinal categories, the median and other quantile differences may also be decomposed. Unlike the mean, all quantiles are well-defined for ordinal variables. Ordinal quantile comparisons can also be interpreted in terms of quantiles of a latent variable under certain conditions (Kaplan and Zhao, 2023, Thm 2.2). However, computing a "percent explained" would still require cardinalization. For example, if the overall median difference on a 24-point ordinal scale is from category 10 to category 20, and the counterfactual "explained" median is category 18, then to call this "80% explained" implicitly assumes a constant difference between each pair of consecutive categories. Here, reporting each group's median along with the counterfactual median helps transparency.

Our second theoretical contribution establishes a numerical equivalence between the estimated explained proportion from the "naive" Blinder–Oaxaca decomposition (with $1, 2, 3, \ldots$ cardinalization) and a particular counterfactual-based estimate. Specifically, when using OLS (linear probability model) for the distribution regressions and decomposing the survival function difference (averaged across categories), the estimated explained proportion is identical to that from naive Blinder–Oaxaca decomposition, for any dataset. Importantly, this allows us to reinterpret many naive Blinder–Oaxaca decomposition estimates in the literature in a more sophisticated and robust way, in terms of a survival function decomposition that does not rely on any cardinalization. That is, if we took the same data but ran a survival function decomposition using the counterfactual from OLS-based distribution regression, then our explained proportion would be exactly the same as those reported. Of course, going forward, we may prefer to use a distribution regression estimator based on logit or series logit rather than OLS, but only for the purpose of reducing functional form misspecification.

In an earlier version of this work (Wu, 2023), a latent mean decomposition based on heteroskedastic ordered probit was also explored. Although this seemed to be a novel idea, it proved to be a bad idea: results were very sensitive to misspecification of both the heteroskedasticity function and the latent error distribution, whose assumed shape (Gaussian) cannot be tested consistently due to the unobserved nature of the latent model. This is closely related to the arguments of Bond and Lang (2019), who criticize latent mean analysis with ordinal data for the same reason, showing how even the signs ($+/-$) of empirical results from the happiness literature can be reversed by adding skew to the assumed latent distribution. Further, unlike the "mean" decomposition above, the latent mean decomposition lacks a meaningful interpretation under misspecification. Thus, we do not recommend latent mean decompositions.

Empirically, we examine the mental health disparity between urban and rural groups in the U.S., decomposing the distribution difference attributed to education, age, sex, income, and region. Our various model estimation methods attribute 33–39% of the mental health difference to these factors. Additionally, we verify that the naive Blinder–Oaxaca decomposition estimate is identical to the OLS-based counterfactual survival function decomposition.

Paper structure Section 2 describes the unified framework for ordinal decomposition based on the counterfactual distribution, largely following Chernozhukov, Fernández-Val, and Melly (2013). Section 3 contains our first equivalence, for the population decomposition. Section 4 describes estimation and inference, as well as our second equivalence result. Section 5 contains our empirical contributions on rural–urban mental health disparities in the U.S.

Notation and abbreviations Random and non-random vectors are respectively typeset as, e.g., $\boldsymbol{X}$ and $\boldsymbol{x}$, while random and non-random scalars are typeset as X and x, and random and non-random matrices as $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{x}}$. The indicator function is $\mathbb{1}\{\cdot\}$, with $\mathbb{1}\{A\} = 1$ if event A occurs and $\mathbb{1}\{A\} = 0$ if not. Acronyms used include those for Akaike information criterion (AIC), cumulative distribution function (CDF) linear probability model (LPM), and National Health Interview Survey (NHIS).

2 A unified framework

This section introduces the counterfactual distribution framework used for both our practical and theoretical contributions. Practically, we suggest any ordinal decomposition proceed in two steps: first, follow the approach of Chernozhukov, Fernández-Val, and Melly (2013) to construct a counterfactual distribution; second, choose a particular summary statistic to decompose. This section describes the counterfactual distribution at the population level, adapting the formulas of Chernozhukov, Fernández-Val, and Melly (2013) to ordinal outcomes.

The following are the main variables and functions. Ordinal outcome Y is a random variable with underlying categorical values like "low," "medium," and "high" that for notational convenience are labeled as $\{1, 2, \ldots, J\}$. Covariate vector $\boldsymbol{X}$ is a random vector including an intercept and other explanatory variables. Cumulative distribution functions (CDFs) have subscripts of the corresponding random variables: $F_Y(\cdot)$ for the CDF of Y, $F_{\boldsymbol{X}}(\cdot)$ for the CDF of $\boldsymbol{X}$, and $F_{Y|\boldsymbol{X}}(\cdot \mid \boldsymbol{x})$ for the conditional CDF of Y given $\boldsymbol{X} = \boldsymbol{x}$. The survival function is the complement of the CDF: $S_Y(y) \equiv \mathrm{P}(Y > y)$, or equivalently $S_Y(\cdot) = 1 - F_Y(\cdot)$. The two groups (populations) of interest are labeled A and B, generally used as superscripts. Thus, for group A: Y^A is the ordinal outcome with CDF $F_Y^A(\cdot)$ and survival function $S_Y^A(\cdot)$, $\boldsymbol{X}^A$ is the covariate vector with CDF $F_{\boldsymbol{X}}^A(\cdot)$ and support $\mathcal{X}^A$, and $F_{Y|\boldsymbol{X}}^A(\cdot \mid \cdot)$ is the conditional CDF. For group B, the A superscripts are all replaced with B superscripts. Similarly, a C superscript indicates the counterfactual distribution, introduced below.

Following Chernozhukov, Fernández-Val, and Melly (2013, §2.1), the population-level counterfactual distribution is defined as follows. The thought experiment is: starting from group B, what if we keep fixed the conditional distribution but change the covariate distribution to that of group A? Thus, we can see how much of a change in the outcome distribution is statistically explained purely from the difference in covariate distributions. Because Y is ordinal with J categories, its distribution is fully characterized by the $J - 1$ values of $F_Y(y)$

for $y \in \{1, \ldots, J-1\}$. Mathematically, as in (2.1) of Chernozhukov, Fernández-Val, and Melly (2013) or (27) of Fortin, Lemieux, and Firpo (2011), the counterfactual CDF is

$$F_Y^C(y) \equiv \int_{\mathcal{X}^A} F_{Y|\boldsymbol{X}}^B(y \mid \boldsymbol{x}) \, \mathrm{d}F_{\boldsymbol{X}}^A(\boldsymbol{x}), \quad y \in \{1, \ldots, J-1\}. \tag{1}$$

As in (2.3) of Chernozhukov, Fernández-Val, and Melly (2013), this requires $\mathcal{X}^A \subseteq \mathcal{X}^B$; if instead $\mathcal{X}^B \subseteq \mathcal{X}^A$, then the A and B labels can be switched. For intuition about (1), consider the extreme cases: if $F_{\boldsymbol{X}}^A = F_{\boldsymbol{X}}^B$, then (1) yields $F_Y^C(y) = F_Y^B(y)$, and if $F_{Y|\boldsymbol{X}}^B = F_{Y|\boldsymbol{X}}^A$, then (1) yields $F_Y^C(y) = F_Y^A(y)$.

In principle, the full distributions F_Y^A, F_Y^B, and F_Y^C can be reported (and should be in an appendix, at least), but this requires $3(J-1)$ values total, so a summary can improve communication and understanding of results. We consider scalar functionals $s(\cdot)$ in Section 3, reducing the results to the three values $s(F_Y^A) - s(F_Y^B)$, $s(F_Y^C) - s(F_Y^B)$, and $s(F_Y^A) - s(F_Y^C)$, which are the total, explained, and unexplained differences, respectively. The single value $[s(F_Y^C) - s(F_Y^B)]/[s(F_Y^A) - s(F_Y^B)]$ represents the explained proportion.

3 Summary statistic interpretations and equivalences

Continuing from the counterfactual distribution, we discuss decompositions based on different summary statistics that we show yield identical explained proportions. Quantiles are discussed in the online appendix. Everything in this section is still at the population level, to describe and understand the interpretation of different possible population objects of interest. Estimation and inference follow in Section 4.

Notationally, denote differences as Δ, with the total (subscript T), explained (E), and unexplained (U) differences respectively

$$\Delta_T, \ \Delta_E, \ \Delta_U. \tag{2}$$

3.1 Survival function

Consider a decomposition based on the survival function differences summed (or averaged) across categories. Given survival functions $S^A(\cdot)$ and $S^B(\cdot)$, we summarize their difference as

$$\sum_{j=1}^{J}[S^A(j) - S^B(j)], \tag{3}$$

and similarly for other pairs of survival functions. Summing from $j = 1$ to $J-1$ is equivalent because $S^A(J) = S^B(J) = 0$. Taking the average (instead of sum) would multiply (3) by $1/J$, but ultimately the explained proportion would remain identical because the $1/J$ would cancel out in (5) below. Given (3), using the notation of (2) and adding superscript S for "survival," the corresponding differences are

$$\Delta_T^S = \sum_{j=1}^{J}[S^A(j) - S^B(j)], \quad \Delta_E^S = \sum_{j=1}^{J}[S^C(j) - S^B(j)], \quad \Delta_U^S = \sum_{j=1}^{J}[S^A(j) - S^C(j)], \tag{4}$$

and the explained proportion is

$$\frac{\Delta_E^S}{\Delta_T^S} = \frac{\sum_{j=1}^{J}[S^C(j) - S^B(j)]}{\sum_{j=1}^{J}[S^A(j) - S^B(j)]}. \tag{5}$$

3.2 CDF

The explained proportion for a CDF-based decomposition is identical to (5). The components in (4) equal the negative of their CDF-based analogs. For example,

$$\Delta_T^S = \sum_{j=1}^{J}[S^A(j) - S^B(j)] = \sum_{j=1}^{J}\{[1 - F^A(j)] - [1 - F^B(j)]\} = -\sum_{j=1}^{J}[F^A(j) - F^B(j)],$$

and similarly for the other differences in (4). Thus, the explained proportion remains the same because $(-\Delta_E^S)/(-\Delta_T^S) = \Delta_E^S/\Delta_T^S$.

3.3 "Mean"

Consider the convenient "mean" decomposition often used in practice. That is, assume the numeric category labels $1, 2, \dots, J$ are actually cardinal values. As discussed in the introduction, if taken literally, this is a very strong assumption that is often doubtful in practice. However, it turns out to be equivalent to the survival function-based decomposition in (4) and (5), which has a meaningful interpretation without requiring any cardinalization.

To see this, using the labels as cardinal values, the "mean" of ordinal Y is

$$
\begin{aligned}
\mathbb{E}(Y) &= \sum_{j=1}^{J} j\,\mathrm{P}(Y=j) \\
&= \mathrm{P}(Y=1) + 2\,\mathrm{P}(Y=2) + \cdots + J\,\mathrm{P}(Y=J) \\
&= \quad \overbrace{[\mathrm{P}(Y=1) + \mathrm{P}(Y=2) + \cdots + \mathrm{P}(Y=J)]}^{=1} \\
&\quad + [\qquad\qquad \overbrace{\mathrm{P}(Y=2) + \cdots + \mathrm{P}(Y=J)}^{=S_Y(1)}] \\
&\quad + \cdots \\
&\quad + [\qquad\qquad\qquad\qquad \overbrace{\mathrm{P}(Y=J)}^{S_Y(J-1)}] \\
&= 1 + \sum_{j=1}^{J-1} S_Y(j) \\
&= 1 + \sum_{j=1}^{J} S_Y(j),
\end{aligned}
$$

with the final equality because $S_Y(J) \equiv \mathrm{P}(Y > J) = 0$. Taking differences, the 1s always cancel out, yielding the expressions in (4) exactly. For example, using superscript μ for "mean,"

$$
\Delta_T^{\mu} \equiv \mathbb{E}(Y^A) - \mathbb{E}(Y^B) = [1 + \sum_{j=1}^{J} S_Y^A(j)] - [1 + \sum_{j=1}^{J} S_Y^B(j)] = \sum_{j=1}^{J} [S_Y^A(j) - S_Y^B(j)] = \Delta_T^S, \quad (6)
$$

and similarly $\Delta_E^{\mu} = \Delta_E^S$ and $\Delta_U^{\mu} = \Delta_U^S$, implying the explained proportion also remains the same: $\Delta_E^{\mu}/\Delta_T^{\mu} = \Delta_E^S/\Delta_T^S$.

Theorem 1 states the core equivalence in general terms.

Theorem 1. *Let W and Z be discrete random variables with possible values $\{1, 2, \ldots, J\}$. Then, $\mathbb{E}(W) = 1 + \sum_{j=1}^{J} S_W(j)$ and $\mathbb{E}(W) - \mathbb{E}(Z) = \sum_{j=1}^{J} [S_W(j) - S_Z(j)]$, where $S_W(j) \equiv \mathrm{P}(W > j)$ and $S_Z(j) \equiv \mathrm{P}(Z > j)$ are the survival functions.*

Proof. See derivation of (6), replacing Y^A with W and Y^B with Z. □

Theorem 1 implies that, serendipitously, we can interpret a "mean" decomposition as a survival function decomposition. That is, if a paper reports results for an ordinal "mean" decomposition, then even if we disagree with the "mean" interpretation, we can still agree about the relative magnitude of explained and unexplained components.

3.4 Implications for regression-based decomposition

Theorem 1 applies to the usual regression-based decomposition, with the usual caveats about functional form misspecification. To start, consider the general nonparametric conditional "mean" model $Y^B = m^B(\boldsymbol{X}^B) + U^B$ with $m^B(\boldsymbol{x}) \equiv \mathbb{E}(Y^B \mid \boldsymbol{X}^B = \boldsymbol{x})$ so $\mathbb{E}(U^B \mid \boldsymbol{X}^B) = 0$. From Theorem 1, we can interpret $\mathbb{E}(Y^B \mid \boldsymbol{X}^B = \boldsymbol{x})$ in terms of conditional survival functions as $1 + \sum_{j=1}^{J} S^B_{Y|\boldsymbol{X}}(j \mid \boldsymbol{x})$; that is, the nonparametric regression describes how this sum of conditional survival function values varies with $\boldsymbol{x}$. The counterfactual mean is $\mathbb{E}(Y^C) = \mathbb{E}[m^B(\boldsymbol{X}^A)]$, which by Theorem 1 can be interpreted in terms of the sum of counterfactual survival function values. The decomposition is thus

$$\mathbb{E}(Y^A) - \mathbb{E}(Y^B) = \overbrace{\mathbb{E}(Y^A) - \underbrace{\mathbb{E}[m^B(\boldsymbol{X}^A)]}_{\mathbb{E}(Y^C)}}^{\text{unexplained}} + \overbrace{\underbrace{\mathbb{E}[m^B(\boldsymbol{X}^A)]}_{\mathbb{E}(Y^C)} - \mathbb{E}(Y^B)}^{\text{explained}}. \tag{7}$$

By Theorem 1, the decomposition in (7) can be interpreted in terms of survival functions. Thus, a nonparametric regression-based "mean" decomposition always has a survival function interpretation, without any assumption about cardinalization.

Theorem 1 can also be used to interpret Blinder–Oaxaca decompositions using linear regression. If we have a properly specified linear model $m^B(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}^B$ and $m^A(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}^A$, then (7) takes the familiar Blinder–Oaxaca form

$$\begin{aligned}
\mathbb{E}(Y^A) - \mathbb{E}(Y^B) &= \mathbb{E}[m^A(\boldsymbol{X}^A)] - \mathbb{E}[m^B(\boldsymbol{X}^A)] + \mathbb{E}[m^B(\boldsymbol{X}^A)] - \mathbb{E}[m^B(\boldsymbol{X}^B)] \\
&= \mathbb{E}[\boldsymbol{X}^{A\prime}\boldsymbol{\beta}^A] - \mathbb{E}[\boldsymbol{X}^{A\prime}\boldsymbol{\beta}^B] + \mathbb{E}[\boldsymbol{X}^{A\prime}\boldsymbol{\beta}^B] - \mathbb{E}[\boldsymbol{X}^{B\prime}\boldsymbol{\beta}^B)] \\
&= \overbrace{\mathbb{E}(\boldsymbol{X}^A)'(\boldsymbol{\beta}^A - \boldsymbol{\beta}^B)}^{\text{unexplained}} + \overbrace{[\mathbb{E}(\boldsymbol{X}^A) - \mathbb{E}(\boldsymbol{X}^B)]'\boldsymbol{\beta}^B}^{\text{explained}}. \qquad (8)
\end{aligned}$$

If the linear model is misspecified, then we can interpret it as the "best linear approximation" of the conditional mean function (e.g., Kaplan, 2022, §7.4), although "best" does not necessarily mean "good." In that case, the counterfactual "mean" $\mathbb{E}(Y^C)$ is approximated by $\mathbb{E}[m^B(\boldsymbol{X}^A)]$. Approximation or not, the population regression-based decomposition of "means" can be interpreted as a decomposition of survival functions.

Section 4.2 shows how this population equivalence extends to estimation.

4 Estimation and inference

Sections 4.1 and 4.3 essentially follow the estimation and inference of Chernozhukov, Fernández-Val, and Melly (2013). Theoretically, ordinal Y is simpler than continuous Y (as in their paper) because there are only $J-1$ values at which we need to estimate the counterfactual CDF, rather than a continuum of an infinite number of points. Thus, their asymptotic results all hold. Our first contribution in this section is to gather practical guidance, which we follow in our provided code.

Our second contribution is the new equivalence result in Section 4.2.

4.1 Estimation

The distribution regression model as in (3.1) of Chernozhukov, Fernández-Val, and Melly (2013) separately models the conditional CDF evaluated at each $y \in \{1, \ldots, J-1\}$ in turn. Let $\Lambda(\cdot)$ be the link function, such as the standard normal or logistic CDF. Let $\boldsymbol{P}(\boldsymbol{x})$ be a vector of transformations of the original covariate vector $\boldsymbol{x}$; for example, it can include squares, interactions, higher-degree polynomial terms, or other basis functions like B-splines. Let $\boldsymbol{\gamma}_y$ be the coefficient vector corresponding to category $y \in \{1, \ldots, J-1\}$. Then, the model is

$$F_{Y|\boldsymbol{X}}(y \mid \boldsymbol{x}) = \Lambda\bigl(\boldsymbol{P}(\boldsymbol{x})'\boldsymbol{\gamma}_y\bigr), \quad y \in \{1, \ldots, J-1\}. \tag{9}$$

Chernozhukov, Fernández-Val, and Melly (2013, p. 2217) note the link function $\Lambda(\cdot)$ is not as important as having a sufficiently flexible $\boldsymbol{P}(\boldsymbol{x})$. A popular choice of estimator is the series logit from Hirano, Imbens, and Ridder (2003, p. 1170), where $\Lambda(\cdot)$ is the logistic CDF and $\boldsymbol{P}(\boldsymbol{x})$ contains polynomials or other basis function transformations. Model selection techniques such as cross-validation can be used to select an appropriately flexible (but not too flexible) model in practice. More on basis expansions and model selection can be found in textbooks like that of Hastie, Tibshirani, and Friedman (2009, Chs. 5 and 7).

The model in (9) is estimated using (only) data from group B, yielding $\hat{\boldsymbol{\gamma}}_y^B$ for $y \in \{1, \ldots, J-1\}$. Weights can be used as appropriate. For given y and $\boldsymbol{x}$ values, the estimated conditional CDF is $\hat{F}^B_{Y|\boldsymbol{X}}(y \mid \boldsymbol{x}) = \Lambda\bigl(\boldsymbol{P}(\boldsymbol{x})'\hat{\boldsymbol{\gamma}}_y^B\bigr)$.

The estimated conditional CDF for group B is then plugged into the counterfactual distribution formula from (1) along with the estimated marginal distribution of $\boldsymbol{X}^A$. Without sampling weights, integrating against $\hat{F}^A_{\boldsymbol{X}}$ is equivalent to averaging over the sample values of $\boldsymbol{X}^A$, so the estimated counterfactual CDF is as given at the end of Remark 3.1 of Chernozhukov, Fernández-Val, and Melly (2013):

$$\hat{F}^C_Y(y) = \int_{\mathcal{X}^A} \hat{F}^B_{Y|\boldsymbol{X}}(y \mid \boldsymbol{x})\,\mathrm{d}\hat{F}^A_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{n_A}\sum_{i=1}^{n_A} \Lambda\bigl(\boldsymbol{P}(\boldsymbol{X}_i^A)'\hat{\boldsymbol{\gamma}}_y^B\bigr), \quad y \in \{1, \ldots, J-1\}, \tag{10}$$

where $\boldsymbol{X}_i^A$ are the observations in the group A sample for $i = 1, \ldots, n_A$; see also page 71 of Fortin, Lemieux, and Firpo (2011). If there are weights, then a weighted average can be taken:

$$\sum_{i=1}^{n_A} \tilde{w}_i^A \Lambda\bigl(\boldsymbol{P}(\boldsymbol{X}_i^A)'\hat{\boldsymbol{\gamma}}_y^B\bigr),$$

where $\tilde{w}_i^A \equiv w_i^A / \sum_{i=1}^{n_A} w_i^A$ normalizes the original weights w_i^A to sum to 1; the unweighted formula above is the special case with $\tilde{w}_i^A = 1/n_A$ for all i.

The actual group A and B outcome distributions can be estimated with the usual estimators. Without weights, for each $y \in \{1, \dots, J-1\}$,

$$\hat{F}_Y^A(y) = \frac{1}{n_A} \sum_{i=1}^{n_A} \mathbb{1}\{Y_i^A \le y\}, \quad \hat{F}_Y^B(y) = \frac{1}{n_B} \sum_{i=1}^{n_B} \mathbb{1}\{Y_i^B \le y\},$$

where the Y_i^A are observations from the group A sample for $i = 1, \dots, n_A$, and the Y_i^B are observations from the group B sample for $i = 1, \dots, n_B$. With weights, similar to above,

$$\hat{F}_Y^A(y) = \sum_{i=1}^{n_A} \tilde{w}_i^A \, \mathbb{1}\{Y_i^A \le y\}, \quad \hat{F}_Y^B(y) = \sum_{i=1}^{n_B} \tilde{w}_i^B \, \mathbb{1}\{Y_i^B \le y\},$$

where again $\tilde{w}_i^A \equiv w_i^A / \sum_{i=1}^{n_A} w_i^A$ and similarly $\tilde{w}_i^B \equiv w_i^B / \sum_{i=1}^{n_B} w_i^B$ normalize the raw weights to sum to one in each sample.

Given the three estimated CDFs $\hat{F}_Y^A$, $\hat{F}_Y^B$, and $\hat{F}_Y^C$, the desired summary statistic can be computed from each, and then the decomposition components. Letting $s(\cdot)$ denote the function that computes the scalar summary statistic,

$$\hat{\Delta}_T = s(\hat{F}_Y^A) - s(\hat{F}_Y^B), \quad \hat{\Delta}_E = s(\hat{F}_Y^C) - s(\hat{F}_Y^B), \quad \hat{\Delta}_U = s(\hat{F}_Y^A) - s(\hat{F}_Y^C),$$

and the estimated explained proportion is $\hat{\Delta}_E / \hat{\Delta}_T$.

4.2 Another equivalence

Here, we establish a numerical equivalence between two seemingly very different estimators of the explained proportion of a decomposition. The first estimator uses the survival function decomposition in (5), where the counterfactual distribution is estimated as in Section 4.1 using the identity link function $\Lambda(a) = a$ (i.e., by OLS with a linear probability model). The second estimator naively applies the conventional OLS-based Blinder–Oaxaca decomposition of the "mean," interpreting the coding $Y \in \{1, \dots, J\}$ as cardinal values.

This equivalence says that we can now more robustly and meaningfully reinterpret published results based on seemingly inappropriate application of Blinder–Oaxaca decomposition to ordinal outcomes. Specifically, even if the paper dubiously claims to decompose the "mean" of an ordinal outcome, we can interpret the estimated explained proportion in terms of survival functions and a counterfactual distribution that does not depend on any particular cardinalization. Although other estimators may help reduce misspecification when estimating the counterfactual, using the naive Blinder–Oaxaca estimate may still be useful for initial exploratory analysis.

Theorem 2. *Assuming both are well-defined given the data, the following two estimates of the explained proportion are numerically identical. First estimate: after coding* Y *with cardinal values* $Y \in \{1, 2, \ldots, J\}$*, estimate the conventional Blinder–Oaxaca mean decomposition in* (8)*, specifically the explained proportion*

$$\frac{(\bar{\boldsymbol{X}}^A - \bar{\boldsymbol{X}}^B)'\hat{\boldsymbol{\beta}}^B}{\bar{Y}^A - \bar{Y}^B},$$

where as usual $\hat{\boldsymbol{\beta}}^B$ *is the OLS-estimated coefficient vector from regressing* Y *on* $\boldsymbol{X}$ *in sample* B*. Second estimate: take the survival function decomposition's estimated explained proportion*

$$\frac{\sum_{y=1}^{J}[\hat{S}^C(y) - \hat{S}^B(y)]}{\sum_{y=1}^{J}[\hat{S}^A(y) - \hat{S}^B(y)]}$$

as in (5)*, and compute* $\hat{S}^C(\cdot)$ *with the counterfactual distribution estimator in* (10) *with the special case* $\Lambda(x) = x$ *and* $\boldsymbol{P}(\boldsymbol{x}) = \boldsymbol{x}$*, with* $\hat{\boldsymbol{\gamma}}_y^B$ *estimated by OLS regression of* $Z_y \equiv \mathbb{1}\{Y \le y\}$ *on* $\boldsymbol{X}$ *using data sample* B*.*

Proof. Recall from Theorem 1 that decomposing the "mean" is equivalent to decomposing the average survival function difference; below, we use the "mean" for simplicity. For example, Theorem 1 allows us to write

$$\frac{\frac{1}{J}\sum_{y=1}^{J}[\hat{S}^C(y) - \hat{S}^B(y)]}{\frac{1}{J}\sum_{y=1}^{J}[\hat{S}^A(y) - \hat{S}^B(y)]} = \frac{\bar{Y}^C - \bar{Y}^B}{\bar{Y}^A - \bar{Y}^B},$$

where $\bar{Y}^A$ and $\bar{Y}^B$ are the sample means when coding Y with cardinal values $\{1, 2, \ldots, J\}$, and $\bar{Y}^C$ is similarly the mean of the estimated counterfactual distribution $\hat{F}^C$ with the same cardinal values.

Consider the standard Blinder–Oaxaca decomposition with the following notation. Let $\bar{Y}^A$ and $\bar{Y}^B$ denote the two sample means when coding $Y \in \{1, 2, \ldots, J\}$. Let $\bar{\boldsymbol{X}}^A$ and $\bar{\boldsymbol{X}}^B$ also denote sample means. These $\boldsymbol{X}$ vectors may include transformations of an original set of variables. Let $\hat{\boldsymbol{\beta}}^B$ be the OLS coefficient vector estimate. Let $\boldsymbol{Y}^B = (Y_1^B, Y_2^B, \ldots, Y_{n_B}^B)'$ be the column vector of observations Y_i^B for $i = 1, \ldots, n_B$. Let $\underline{\boldsymbol{X}}^B$ be the matrix with row i equal to the transpose of $\boldsymbol{X}_i^B$, so $\underline{\boldsymbol{X}}^B = (\boldsymbol{X}_1^B, \boldsymbol{X}_2^B, \ldots, \boldsymbol{X}_{n_B}^B)'$. The vector of residuals and its orthogonality property are

$$\hat{\boldsymbol{U}}^B \equiv \boldsymbol{Y}^B - \underline{\boldsymbol{X}}^B \hat{\boldsymbol{\beta}}^B \quad \text{with} \quad (\underline{\boldsymbol{X}}^B)' \hat{\boldsymbol{U}}^B = \mathbf{0}. \tag{11}$$

In the conventional Blinder–Oaxaca decomposition like in (8), the explained proportions in the population and sample can respectively be written as

$$\begin{aligned} \rho_E &\equiv \frac{\Delta_E}{\Delta_T} = \frac{[\mathbb{E}(\boldsymbol{X}^A) - \mathbb{E}(\boldsymbol{X}^B)]'\boldsymbol{\beta}^B}{\mathbb{E}(Y^A) - \mathbb{E}(Y^B)} = \frac{\mathbb{E}(\boldsymbol{X}^A)'\boldsymbol{\beta}^B - \mathbb{E}(Y^B)}{\mathbb{E}(Y^A) - \mathbb{E}(Y^B)}, \\ \hat{\rho}_E &= \frac{(\bar{\boldsymbol{X}}^A)'\hat{\boldsymbol{\beta}}^B - \bar{Y}^B}{\bar{Y}^A - \bar{Y}^B}. \end{aligned} \tag{12}$$

It remains to show that the counterfactual distribution-based decomposition, when using OLS to estimate linear probability models for each category of Y, yields a term identical to $(\bar{\boldsymbol{X}}^A)'\hat{\boldsymbol{\beta}}^B$ in the numerator of (12), as seen below.

Now consider the OLS estimates of the counterfactual CDF. Define indicators $Z_j^B \equiv \mathbb{1}\{Y^B \leq j\}$ for $j \in \{1, \ldots, J-1\}$, so

$$Y^B = J - \sum_{j=1}^{J-1} Z_j^B. \tag{13}$$

Let $\hat{\boldsymbol{\gamma}}_j^B$ be the OLS coefficient vector estimate from regressing Z_j^B on $\boldsymbol{X}^B$. Analogous to $\boldsymbol{Y}^B$, let $\boldsymbol{Z}_j^B$ be the vector of n_B observations. For each $j \in \{1, \ldots, J-1\}$, the vector of residuals and its orthogonality property are

$$\hat{\boldsymbol{V}}_j^B \equiv \boldsymbol{Z}_j^B - \underline{\boldsymbol{X}}^B \hat{\boldsymbol{\gamma}}_j^B \quad \text{with} \quad (\underline{\boldsymbol{X}}^B)' \hat{\boldsymbol{V}}_j^B = \boldsymbol{0}. \tag{14}$$

Using the above and the orthogonality property of OLS residuals, we can derive the relationship between $\hat{\boldsymbol{\beta}}^B$ (from (11)) and the $\hat{\boldsymbol{\gamma}}_j^B$. Assume the constant term is the first element in vector $\boldsymbol{X}^B$; that is, $\boldsymbol{X}^B = (1, \ldots)'$, so the first column of matrix $\underline{\boldsymbol{X}}^B$ is all ones. Let $\boldsymbol{e}_1 \equiv (1, 0, \ldots, 0)'$ have the same length as the $\hat{\boldsymbol{\gamma}}_j^B$, and let $\boldsymbol{1} \equiv (1, \ldots, 1)'$ be a vector of n_B ones. Combining (13) and (14),

$$\boldsymbol{Y}^B = J\boldsymbol{1} - \sum_{j=1}^{J-1} \boldsymbol{Z}_j^B = J\boldsymbol{1} - \sum_{j=1}^{J-1} (\underline{\boldsymbol{X}}^B \hat{\boldsymbol{\gamma}}_j^B + \hat{\boldsymbol{V}}_j^B) = \underline{\boldsymbol{X}}^B \overbrace{\left(J\boldsymbol{e}_1 - \sum_{j=1}^{J-1} \hat{\boldsymbol{\gamma}}_j^B \right)}^{=\hat{\boldsymbol{\beta}}^B} + \sum_{j=1}^{J-1} (-\hat{\boldsymbol{V}}_j^B),$$

where the equality

$$J\boldsymbol{e}_1 - \sum_{j=1}^{J-1} \hat{\boldsymbol{\gamma}}_j^B = \hat{\boldsymbol{\beta}}^B \tag{15}$$

is implied by the orthogonality

$$(\underline{\boldsymbol{X}}^B)' \sum_{j=1}^{J-1} (-\hat{\boldsymbol{V}}_j^B) = - \sum_{j=1}^{J-1} \overbrace{(\underline{\boldsymbol{X}}^B)' \hat{\boldsymbol{V}}_j^B}^{=\boldsymbol{0}} = \boldsymbol{0}, \tag{16}$$

which follows from the orthogonality condition in (14).

Now consider the "mean" of the counterfactual distribution. Taking the expectation of (13) and using the linearity of the expectation operator,

$$\mathbb{E}(Y^B) = J - \sum_{j=1}^{J-1} \mathbb{E}(Z_j^B).$$

Thus, the estimated counterfactual "mean" is

$$\hat{\mathbb{E}}(Y^C) = J - \sum_{j=1}^{J-1} (\bar{\boldsymbol{X}}^A)' \hat{\boldsymbol{\gamma}}_j^B = (\bar{\boldsymbol{X}}^A)' \overbrace{\left(J\boldsymbol{e}_1 - \sum_{j=1}^{J-1} \hat{\boldsymbol{\gamma}}_j^B \right)}^{=\hat{\boldsymbol{\beta}}^B \text{ by (15)}} = (\bar{\boldsymbol{X}}^A)' \hat{\boldsymbol{\beta}}^B.$$

This final expression is identical to the term in the Blinder–Oaxaca decomposition numerator in (12), so the estimated explained proportion is identical. □

4.3 Inference

Inference for the Δ components can use the bootstrap in Algorithm 2 of Chernozhukov, Fernández-Val, and Melly (2013). Their bootstrap (and the corresponding theory) is for $s(\hat{F}_Y^C)$. For $s(\cdot)$ like the "mean," analytic confidence intervals may be readily available, or as long as the bootstrap is being run anyway, they can be bootstrapped, too. The Algorithm 2 bootstrap is a very general exchangeable weight bootstrap that includes the usual bootstrap as a special case. Per their Remark 5.1, the bootstrap weights (or resamples) should be done separately and independently for groups A and B. Given each bootstrap weight vector or sample, the full estimation procedure from Section 4.1 is run, and this is repeated many times. The many bootstrap-world estimates of the Δ components can then be used in any standard bootstrap confidence interval formula as desired.

5 Empirical results: mental health disparities

We illustrate the preceding approaches through an empirical analysis of rural–urban mental health disparity in the U.S. Specifically, we decompose the overall rural–urban difference in depression using age, sex, education, income, and region. The analysis was performed in R (R Core Team, 2022), with help from the `oaxaca` package (Hlavac, 2022). Code to replicate our results is available online.[2]

5.1 Data

We use the publicly available NHIS 2022 data (National Center for Health Statistics, 2022), chosen for its inclusion of mental health assessment and recent availability. Our analysis

[2] `https://qianjoewu.github.io/`

targets individuals aged 24–64, focusing on those of working age and surpassing the average age at college graduation in the U.S.

The following variables are used. The outcome variable Y (`PHQCAT_A`) measures the severity of depressive symptoms, summarizing the eight-item Patient Health Questionnaire into four categories from low to high: "none/minimal," "mild," "moderate," and "severe." The rural and urban groups are defined using variable `URBRRL`: group A contains individuals who live in counties categorized as nonmetropolitan, while group B is large central metro counties. The vector of explanatory variables $\boldsymbol{X}$ includes education (`EDUCP_A`), sex (`SEX_A`), age (`AGEP_A`), squared age, family income (`POVRATTC_A`), and geographic region (`REGION`). The estimation uses the sampling weight variable (`WTFA_A`). We drop 282 observations (3.4%): those for which either age or urban group is missing, and those that fit our age and urban group restrictions but have another variable value missing. This leaves 7902 observations for our analysis.

Figure 1 shows a mosaic plot to visualize the distribution of depression levels for rural and urban groups. There are $n_A = 2448$ individuals in group A (rural) and $n_B = 5454$ in group B (urban), whose relative sizes correspond to the relative widths of the rural and urban bars in the plot. Further, the area of each cell (rectangle) is proportional to the sample proportion of such individuals, which is shown on the label; for example, rural individuals with mild depression constitute 4.53% of the overall sample. Each bar indicates the sample proportion. In each group, the stacked bar height represents the conditional probability of each depression category, with splits indicating CDF values. For example, within the rural bar, from bottom to top, these splits indicate $\hat{F}_Y^A(1)$, $\hat{F}_Y^A(2)$, and $\hat{F}_Y^A(3)$, respectively. The figure shows that the rural group first-order stochastically dominates the urban group in depression level in the sample because the rural CDF is below the urban CDF at each category: $\hat{F}_Y^A(j) \leq \hat{F}_Y^B(j)$ at each $j \in \{1, 2, 3, 4\}$. This suggests that overall the rural group has higher levels of depression (worse mental health) than the urban group.

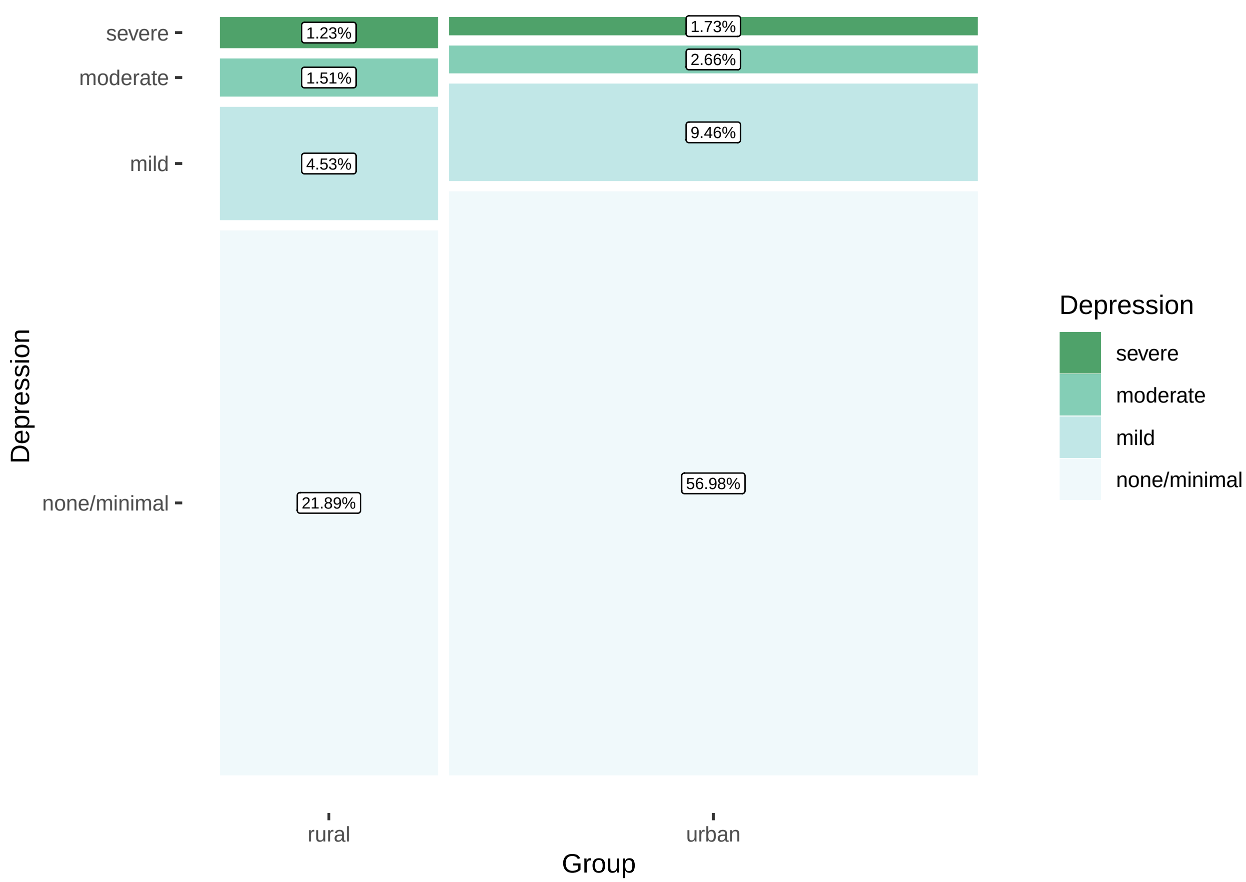

Figure 1: Urban and rural depression level distributions.

5.2 Estimation

To construct the counterfactual distribution for the rural group, we use distribution regression with three estimation methods: OLS with a linear probability model (LPM), logit, and nonparametric series logit.

For LPM/OLS and logit estimation, we include the following explanatory variables: a dummy variable for high education (equal to 1 if the individual has at least some college education), sex, age, squared age, income (family poverty ratio: family income divided by poverty threshold), a dummy variable for "high income" that equals 1 if the family poverty ratio has been top-coded (value 11), and region dummies for the Northeast, Midwest, and West (with South the base group).

For series logit estimation, we use the following model selection procedure. We run the procedure separately for each dependent variable $\mathbb{1}\{Y \leq y\}$, $y \in \{1, 2, 3\}$. The higher-order terms lack a natural ordering, so technically we do not use "series" logit because we consider many non-nested subsets of higher-order terms as candidate models. First, in every candidate model, we include a linear term for each explanatory variable described above. Second, we construct a candidate model for every possible combination of the quadratic terms, which include squared age, squared income, and interaction terms like high education*age. Third, we conduct logit estimation for each candidate model and compute the AIC (Akaike, 1974). Fourth, we select the model with the lowest AIC as the optimal model. Fifth, we consider adding certain cubic terms if the corresponding quadratic terms are included in the selected model, and consider adding quartic terms if the cubic terms are selected, etc.

In practice, due to computational constraints, in addition to linear terms we include age*income, age*high income, squared age, and squared income in every candidate model. Each of the 22 additional quadratic terms may be included or excluded, yielding $2^{22} =$ 4,194,304 candidate models. The selected model is different for each dependent variable $\mathbb{1}\{Y \leq y\}$.[3] For each $\mathbb{1}\{Y \leq y\}$, adding cubed income and/or cubed age to the selected

[3]Beyond the baseline terms (high education, sex, age, income, high income, Northeast, Midwest, West,

quadratic model resulted in worse (higher) AIC, so we use the selected quadratic models for estimation.

In our decomposition result, we include bootstrapped standard errors, which were computed using the procedure outlined by Hlavac (2022, §2.4) and described here in Method 1.

Method 1. *[bootstrapped standard errors]*

1. *Take R random samples with replacement from the relevant set of observations, separately and independently for groups A and B (per Section 4.3).*
2. *In each approach, estimate and perform the decomposition for the sample from Step 1.*
3. *Calculate the bootstrapped standard error as the standard deviation of the R decomposition estimates from Step 2.*

We use $R = 1000$.

5.3 Results

We decompose the "mean" difference between rural group A and urban group B, equivalent to decomposing the survival function per Theorem 1. We use the three estimators in Section 5.2 as well as the naive conventional Blinder–Oaxaca decomposition estimator, to verify out Theorem 2.

Table 1 displays the estimated rural, urban, and counterfactual CDFs. For the counterfactual, the estimated CDF values are similar across estimators, particularly OLS and logit. This suggests that OLS provides a reasonable approximation here, and with computation time in seconds instead of hours. At minimum, OLS seems very practical for exploratory analysis, although for the final analysis a nonparametric estimator like series logit may be preferred.

---

age*income, age*high income, squared age, and squared income), the selected model with dependent variable $\mathbb{1}\{Y \leq 1\}$ includes high education*age, high education*income, high education*high income, sex*income, sex*Northeast, age*West, income*Midwest; the selected model with $\mathbb{1}\{Y \leq 2\}$ includes high education*sex, high education*Northeast, sex*age, age*Midwest, and income*Northeast; and the selected model with $\mathbb{1}\{Y \leq 3\}$ includes sex*income, income*Midwest, high income*Northeast, and high income*West.

Table 1: Estimated actual and counterfactual CDFs.

Group	$\hat{F}(1)$	$\hat{F}(2)$	$\hat{F}(3)$
Rural	0.751	0.906	0.958
Urban	0.804	0.938	0.976
Counterfactual (OLS/LPM)	0.789	0.925	0.968
Counterfactual (logit)	0.790	0.926	0.968
Counterfactual (series logit)	0.786	0.923	0.968

Table 2: Decomposition results.

Model	Explained (%)	Unexplained (%)
Naive Blinder–Oaxaca	33.9	66.1
	(12.5)	(12.5)
OLS/LPM	33.9	66.1
	(12.5)	(12.5)
Logit	33.0	67.0
	(12.6)	(12.6)
Series logit	38.9	61.1
	(13.8)	(13.8)

Bootstrapped standard errors are in parentheses.

Table 2 displays the decomposition results. Given the similar counterfactual CDF estimates in Table 1, naturally the explained proportion estimates are also similar, all in the range of 33–39 percent. This suggests that education, sex, age, income, and region collectively account for approximately 33–39 percent of the difference in the depression distribution between the rural and urban groups. This is a substantial amount, but still leaves over half unexplained.

To verify Theorem 2, we also compute the naive Blinder–Oaxaca decomposition using the $1, 2, 3, \ldots$ cardinalization. As expected, compared to the OLS counterfactual approach, the decomposition results are identical. Thus, even if researchers reported only the conventional Blinder–Oaxaca decomposition with this data, we could still interpret the results in terms of a survival function decomposition with a counterfactual CDF estimated by distribution regression, robust to any alternative cardinalization.

6 Conclusion

For decomposition analysis with ordinal outcomes, we have shown that the naive OLS-based Blinder–Oaxaca decomposition with outcomes coded with values $1, 2, 3, \dots$ can be interpreted as a more sophisticated and robust decomposition of the survival function without any assumed cardinalization. This allows any such empirical results in the literature to be reinterpreted more meaningfully, and it suggests such "naive" OLS decomposition provides a practical tool for exploratory analysis. We also show how to implement a nonparametric estimator in our empirical application. In all, we provide a unified framework for ordinal decomposition that can be applied to a wide range of ordinal outcomes across economics, public health, medicine, sociology, and other social sciences.

References

Akaike, Hirotugu. 1974. "A new look at the statistical model identification." *IEEE Transactions on Automatic Control* 19 (6):716–723. URL https://doi.org/10.1109/TAC.1974.1100705.

Awaworyi Churchill, Sefa, Musharavati Ephraim Munyanyi, Kushneel Prakash, and Russell Smyth. 2020. "Locus of control and the gender gap in mental health." *Journal of Economic Behavior & Organization* 178:740–758. URL https://doi.org/10.1016/j.jebo.2020.08.013.

Bauer, Thomas K. and Mathias Sinning. 2008. "An extension of the Blinder–Oaxaca decomposition to nonlinear models." *AStA Advances in Statistical Analysis* 92 (2):197–206. URL https://doi.org/10.1007/s10182-008-0056-3.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8 (4):436–455. URL https://www.jstor.org/stable/144855.

Bond, Timothy N. and Kevin Lang. 2019. "The Sad Truth About Happiness Scales." *Journal of Political Economy* 127 (4):1629–1640. URL https://doi.org/10.1086/701679.

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly. 2013. "Inference on Counterfactual Distributions." *Econometrica* 81 (6):2205–2268. URL https://www.jstor.org/stable/23524318.

Demoussis, Michael and Nicholas Giannakopoulos. 2007. "Exploring Job Satisfaction in Private and Public Employment: Empirical Evidence from Greece." *Labour* 21 (2):333–359. URL https://doi.org/10.1111/j.1467-9914.2007.00370.x.

Even, William E. and David A. Macpherson. 1990. "Plant size and the decline of unionism." *Economics Letters* 32 (4):393–398. URL https://doi.org/10.1016/0165-1765(90)90035-Y.

Fairlie, Robert W. 2005. "An extension of the Blinder–Oaxaca decomposition technique to logit and probit models." *Journal of Economic and Social Measurement* 30 (4):305–316. URL https://doi.org/10.3233/JEM-2005-0259.

Farber, Henry S. 1987. "The Recent Decline of Unionization in the United States." *Science* 238 (4829):915–920. URL https://www.jstor.org/stable/1700925.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011. "Decomposition Methods in Economics." In *Handbook of Labor Economics*, vol. 4A, edited by Orley Ashenfelter and David Card, chap. 1. Elsevier, 1–102. URL https://doi.org/10.1016/S0169-7218(11)00407-2.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, 2nd ed. URL https://web.stanford.edu/~hastie/ElemStatLearn. Corrected 12th printing, January 13, 2017.

Hauret, Laetitia and Donald R. Williams. 2017. "Cross-National Analysis of Gender Differences in Job Satisfaction." *Industrial Relations* 56 (2):203–235.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica* 71 (4):1161–1189. URL https://www.jstor.org/stable/1555493.

Hlavac, Marek. 2022. *oaxaca: Blinder–Oaxaca Decomposition in R*. Social Policy Institute, Bratislava, Slovakia. URL `https://CRAN.R-project.org/package=oaxaca`. R package version 0.1.5.

Kaplan, David M. 2022. *Introductory Econometrics: Description, Prediction, and Causality*. Columbia, MO: Mizzou Publishing, 3rd ed. URL `https://www.themizzoustore.com/p-236916-introductory-econometrics-description-prediction-and-causality.aspx`.

Kaplan, David M. and Wei Zhao. 2023. "Comparing latent inequality with ordinal data." *Econometrics Journal* 26 (2):189–214. URL `https://doi.org/10.1093/ectj/utac030`.

Kitagawa, Evelyn M. 1955. "Components of a Difference Between Two Rates." *Journal of the American Statistical Association* 50 (272):1168–1194. URL `https://www.jstor.org/stable/2281213`.

Madden, David. 2010. "Gender Differences in Mental Well-Being: a Decomposition Analysis." *Social Indicators Research* 99 (1):101–114. URL `https://www.jstor.org/stable/40800995`.

National Center for Health Statistics. 2022. "National Health Interview Survey." `https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm`. Public-use data file and documentation, accessed 2024.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14 (3):693–709. URL `https://www.jstor.org/stable/2525981`.

Pan, Jay, Dan Liu, and Shehzad Ali. 2015. "Patient dissatisfaction in China: What matters." *Social Science & Medicine* 143:145–153. URL `https://doi.org/10.1016/j.socscimed.2015.08.051`.

Pilipiec, Patrick, Wim Groot, and Milena Pavlova. 2020. "A Longitudinal Analysis of Job Satisfaction During a Recession in the Netherlands." *Social Indicators Research* 149 (1):239–269. URL `https://doi.org/10.1007/s11205-019-02233-6`.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org`.

Wu, Qian. 2023. *Essays on the econometrics of ordinal data*. PhD dissertation, University of Missouri. URL `https://doi.org/10.32469/10355/97079`.

Zhang, Hao, Teresa Bago d'Uva, and Eddy van Doorslaer. 2015. "The gender health gap in China: A decomposition analysis." *Economics & Human Biology* 18:13–26. URL `https://doi.org/10.1016/j.ehb.2015.03.001`.

Online appendix for "Ordinal Decomposition"

Qian Wu David M. Kaplan

May 30, 2024

A Quantiles

Ordinal variables have a well-defined τ-quantile for any $0 \le \tau \le 1$. For any type of variable Y, including ordinal, the τ-quantile is generally defined as

$$\mathrm{Q}_\tau(Y) \equiv \inf\{y : F_Y(y) \ge \tau\}.$$

For example, if an ordinal variable has probability 40% of value "low," 20% "medium," and 40% "high," then "medium" is the smallest value such that the CDF is at least 0.5, so "medium" is the 0.5-quantile (median). Similarly, "low" is the smallest value such that the CDF is at least 0.25, so "low" is the 0.25-quantile.

Given that, the actual and counterfactual τ-quantiles are all well-defined: $\mathrm{Q}_\tau(Y^A)$, $\mathrm{Q}_\tau(Y^B)$, and $\mathrm{Q}_\tau(Y^C)$. Reporting these can provide a sense of how much of the overall difference is statistically explained by the covariates. Specifically, the closer $\mathrm{Q}_\tau(Y^C)$ is to $\mathrm{Q}_\tau(Y^A)$ than to $\mathrm{Q}_\tau(Y^B)$, the more is explained. These ordinal comparisons can also be interpreted in terms of quantiles of latent distributions under certain conditions as shown by Kaplan and Zhao (2023).

However, without imposing cardinal values, the Δ differences are not well-defined, nor are their relative magnitudes like Δ_E/Δ_T. For example, imagine $\mathrm{Q}_\tau(Y^A) = 10$, $\mathrm{Q}_\tau(Y^C) = 8$, and $\mathrm{Q}_\tau(Y^B) = 2$. It is tempting to say that $\Delta_E^\tau = 8 - 2 = 6$ and $\Delta_T^\tau = 10 - 2 = 8$, so $6/8 = 75\%$ is explained, meaning the difference in the covariate distributions can statistically explain most of the overall difference. However, imagine the categories labeled 2 through 8 have values "infinitesimal," "negligible," "minuscule," "tiny," "extremely small," "very small," and "small," followed by category 9 "medium" and category 10 "large." Thus, the difference between "infinitesimal" and "small" is explained, while the difference between "small" and "large" is unexplained, suggesting that actually most of the difference is unexplained. That is, the "75%" assumes the categories are all evenly spaced, but here clearly categories 2 through 8 are much closer in value than 8 through 10. In general, comparing the values of $\mathrm{Q}_\tau(Y^A)$, $\mathrm{Q}_\tau(Y^B)$, and $\mathrm{Q}_\tau(Y^C)$ can be insightful, but quantifying the differences generally requires a subjective cardinalization.

Another limitation of quantile decomposition is that often $\mathrm{Q}_\tau(Y^A) = \mathrm{Q}_\tau(Y^B)$ when J is small. For example, in our depression data, the median is "None/minimal" in both urban and rural groups, even though the proportion of each group in that category is significantly different. Because of this discreteness of quantiles with ordinal data, the median (or other typical quantiles) may not reveal any group difference in the first place, let alone provide a precise decomposition. However, if the number of categories J is sufficiently large, then quantile decomposition can be insightful.