Regression and decomposition with ordinal health outcomes

Qian Wu[*] David M. Kaplan[†]

January 16, 2025

Abstract

Although ordinal health outcome values are categories like "poor" health or "moderate" depression, they are often assigned numeric values $1, 2, 3, \ldots$ for convenience. We provide results on interpretation of subsequent OLS-based analysis. For prediction, the OLS estimand's "best linear predictor" interpretation requires that these are indeed the true cardinal values. However, for description, the OLS estimand is the "best linear approximation" of a summary of the conditional survival functions, regardless of the true cardinal values. Further, for Blinder–Oaxaca-type decomposition, the OLS-based estimator is numerically equivalent to a certain counterfactual-based decomposition of the survival function, again regardless of the true cardinal values. Empirically, we decompose U.S. rural–urban differences in depression. Including a nonparametric estimator that we describe, estimators agree that 33–39% of the rural–urban difference is statistically explained by income, education, age, sex, and geographic region. The OLS-based detailed decomposition shows that almost all of this is due to income.

JEL classification: C25, I14

Keywords: Blinder–Oaxaca decomposition; Counterfactual distribution; Distribution regression; Survival function

[*]School of Statistics, Southwestern University of Finance and Economics, wuqj@swufe.edu.cn

[†]Corresponding author; Department of Economics, University of Missouri, 615 Locust Street, Columbia, MO, USA; kaplandm@missouri.edu

1 Introduction

Ordinal variables are common in health economics. Such variables take values that are not numeric but rather categorical. For example, self-reported health status often takes values "poor," "fair," "good," "very good," and "excellent," which have an order from lowest to highest, but no numeric value. Mental health variables are also often ordinal; our empirical analysis uses a measure of depression with values "none/minimal," "mild," "moderate," and "severe." Some variables are even coded with numeric values, but upon examination these values do not have a cardinal but merely ordinal meaning, like the Apgar score for newborns whose numbers are based on underlying categories like "no cyanosis" or "some flexion."

Even in raw data, ordinal outcome variables often come already coded with numeric values $1, 2, 3, \ldots$, making it easy to run OLS regression and related analyses, but this raises questions about interpretation. Ordinal variables do not even have a well-defined mean, because values like "fair" and "excellent" cannot be summed or averaged. This suggests that we must interpret the $1, 2, 3, \ldots$ coding as the cardinal values assigned to the respective categories. If those are the correct cardinal values, then we actually have a cardinal variable, and indeed OLS and such can be run and interpreted as usual. But what if those are not the true cardinal values?

This question has been addressed by several papers in health economics that take seriously the ordinal nature of such outcomes. These papers specifically consider measuring health inequality or polarization given an ordinal health outcome variable. For example, Allison and Foster (2004), Apouey (2007), Abul Naga and Yalcin (2008), and Kobus and Miłoś (2012) all agree that such measures should be "scale invariant" in the sense of not depending on whether we code the categories with cardinal values $1, 2, 3$ or $1, 2, 10$ or $1, 7, 8$, etc. The median-preserving spread of Allison and Foster (2004) and the inequality indices proposed and studied by Apouey (2007), Abul Naga and Yalcin (2008), and Kobus and Miłoś (2012) can all be interpreted without any cardinal values. Kaplan and Zhao (2023) also characterize inequality in terms of a cardinal latent variable that generates the observed ordinal variable, which allows a range of cardinal values associated with each ordinal category. To emphasize the ordinal/cardinal distinction, instead of "scale invariant," we call all the above approaches *cardinalization-robust* because their results are robust to any assignment of cardinal values to categories.

In the same spirit, we make five contributions to the cardinalization-robust interpretation of OLS regression and OLS-based Blinder–Oaxaca decomposition with an ordinal outcome. First, for prediction, we have a negative result: the best predictor of Y is sensitive to the cardinal values assigned to the categories, so there is no cardinalization-robust "best linear predictor" interpretation of OLS. That is, if we code the Y values as $1, 2, 3, \ldots$, then the interpretation of the OLS estimand as the best linear predictor crucially depends on $1, 2, 3, \ldots$ being the true cardinal values. Second, for description, the OLS estimand with Y coded $1, 2, 3, \ldots$ can be interpreted as the best linear approximation of the sum (across all Y categories) of conditional survival function values. This is explained in more detail in Section 2.3, but the main point is that unlike with prediction, the $1, 2, 3, \ldots$ coding here is innocuous: the OLS interpretation does not require those values to be the true cardinal values.

Third, as a practical contribution for ordinal decomposition, we describe how to apply the methodology of Chernozhukov, Fernández-Val, and Melly (2013) to get a cardinalization-robust decomposition of the survival function difference. This approach is based on a counterfactual distribution that combines the marginal $\boldsymbol{X}$ distribution of one group with the conditional distribution (of Y given $\boldsymbol{X}$) from the other group. 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). We also describe how to implement a specific nonparametric version of this approach that we apply in our empirical analysis.

Fourth, again with Y coded as $1, 2, 3, \ldots$, we derive a numerical equivalence between two very different estimators: the OLS-based Blinder–Oaxaca "mean" decomposition, and the cardinalization-robust counterfactual survival function decomposition based on Chernozhukov, Fernández-Val, and Melly (2013) when using OLS to estimate linear probability models for the distribution regression step. Thus, despite seeming like it relies on $1, 2, 3, \ldots$ as cardinal values, the OLS-based Blinder–Oaxaca decomposition actually has a cardinalization-robust interpretation. Among other practical implications, this means we can reinterpret previously published Blinder–Oaxaca results in terms of a robust counterfactual survival function decomposition. It also means that going forward we can more confidently use OLS-based Blinder–Oaxaca, and more appropriately interpret its results, with only the caveat that other estimators may reduce functional form misspecification. The Blinder–Oaxaca approach also readily produces a "detailed decomposition" that shows how much of the overall difference is statistically explained by each variable individually.

Fifth, we empirically examine the mental health disparity between urban and rural groups in the U.S., decomposing the distributional difference in a measure of depression. Depression is important to study due to its large aggregate effects both personally and economically, with an annual total economic burden in the U.S. estimated in the hundreds of billions of dollars (Greenberg, Fournier, Sisitsky, Simes, Berman, Koenigsberg, and Kessler, 2021). The explanatory variables are education, age, sex, income, and geographic region. Our various estimators attribute 33–39% of the depression difference to these variables. That is, these variables explain a substantial amount, but still leave a major of the urban–rural difference unexplained. We include a nonparametric estimator that performs model selection among millions of candidate models, as we describe in detail. Using the Blinder–Oaxaca approach, we report a detailed decomposition showing that income explains much more than any other covariate. Even though we do not believe that the depression categories correspond to cardinal values $1, 2, 3, 4$, our equivalence result implies that this detailed decomposition is still meaningful.

We include results for Blinder–Oaxaca-type decomposition because of its widespread use and importance. It 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. Some examples in health include decomposing differences in various biomarkers by gender (Carrieri and Jones, 2017), self-reported health by age (Idler and Cartwright, 2018), various health outcomes by education or income (Kino and Kawachi, 2020), diabetes by Latinx identity (Cartwright, 2021), and obesity/BMI by race (Sen, 2014).

Despite the importance of both decomposition and ordinal variables, there is a limited literature on decomposition with ordinal outcomes. The extensive *Handbook of Labor Economics* 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).) Some 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). Although not "wrong," such simplification loses information and precision. Bauer and Sinning (2008) propose an ordered probit/logit decomposition, but it is used only to introduce nonlinearity while still treating the ordinal outcome as if had cardinal values $1, 2, 3, \ldots$, as seen in their equations on page 200. The same is true of Demoussis and Giannakopoulos (2007).[1] Similarly, empirical work often takes $1, 2, 3, \ldots$ 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,

[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.

§§2.1–2.2), Idler and Cartwright (2018), and Pilipiec, Groot, and Pavlova (2020, §2.2). Madden (2010, §2) acknowledges the cardinalization is not fully appropriate, yet his robustness check's ordered probit decomposition still uses the same cardinalization (p. 111). However, recall that our new results say that the $1, 2, 3, \ldots$ coding in all the above-cited work actually has a cardinalization-robust interpretation in terms of survival functions. That is, all the above methods and results are still valid, just with a somewhat different interpretation that we detail in our results.

Paper structure Section 2 characterizes the interpretation of OLS regression with an ordinal outcome, both in terms of prediction and description. Section 3 describes a framework for ordinal decomposition based on the counterfactual approach of Chernozhukov, Fernández-Val, and Melly (2013), as well as our new equivalence between "mean" and survival function decomposition. Section 4 describes estimation and inference, as well as our second equivalence result that provides a meaningful, robust interpretation for the naive OLS-based Blinder–Oaxaca decomposition. Section 5 contains our empirical contributions on rural–urban mental health disparities in the U.S. Appendix A contains an additional theoretical proof, and Appendix B discusses quantile decomposition for ordinal variables.

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), National Health Interview Survey (NHIS), and ordinary least squares (OLS).

2 Ordinal regression

We consider OLS regression with an ordinal outcome variable Y whose J categories have been assigned the numeric values $1, 2, \ldots, J$, respectively. In particular, we are interested in the OLS estimand's interpretation when these are not the actual cardinal values associated with each category.

An equivalence for the "mean" is given first, followed by OLS results for both prediction and description. In later sections, we extend these results to Blinder–Oaxaca decomposition.

2.1 Mean

Although an ordinal random variable Y does not have a mean, if we assign the numeric values $1, \ldots, J$ to its J categories, then we can compute a "mean." Because we do not consider any other numeric assignments, throughout the paper we write this "mean" as $\mathrm{E}(Y)$. Lemma 1 shows that this "mean" has a meaningful, cardinalization-robust interpretation.

Lemma 1. *Given ordinal random variable Y, if we assign the numeric values $1, \ldots, J$ to its J categories, then its "mean" is $1 + \sum_{j=1}^{J-1} \mathrm{P}(Y > j)$.*

Proof. The "mean" is

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

$$= 1 + \sum_{j=1}^{J-1} \mathrm{P}(Y > j). \qquad \square$$

Lemma 1 shows the "mean" can be interpreted in terms of the survival function, which does not depend on any particular cardinal value assignment. The summand $\mathrm{P}(Y > j)$ is the survival function of Y evaluated at category j. Alternatively, the final expression could be rewritten in terms of the CDF as $J - \sum_{j=1}^{J-1} \mathrm{P}(Y \leq j)$. However, the survival function expression makes it more directly clear that higher values correspond to higher probabilities of higher-valued categories. For example, if one ordinal distribution first-order stochastically dominates another, then it has a higher survival function at all j, and thus has higher "mean" of $1 + \sum_{j=1}^{J-1} \mathrm{P}(Y > j)$.

2.2 Prediction

With a cardinal-valued Y, it is well known that the mean provides the best predictor given a quadratic loss function (e.g., Kaplan, 2022, §2.5.2):

$$\mathrm{E}(Y) = \arg\min_{g} \mathrm{E}[(Y - g)^2].$$

Continuing to maintain quadratic loss, this result then implies that the conditional mean is the best predictor of Y given vector $\boldsymbol{X}$ (e.g., Kaplan, 2022, §6.3.5). It also implies that the OLS estimand $\boldsymbol{\beta}$ is the best linear predictor (e.g., Kaplan, 2022, §7.5) in the sense of

$$\boldsymbol{\beta} = \arg\min_{\boldsymbol{b}} \mathrm{E}[(Y - \boldsymbol{X}'\boldsymbol{b})^2].$$

Despite Lemma 1, without committing to a particular assignment of cardinal value to each category, we cannot derive any such best predictor results with an ordinal Y. This is true even for the simplest case of the unconditional mean, as seen in the following counterexample. First, let $\mathrm{P}(Y = y) = 1/5$ given cardinal values $y = 1, 2, 3, 4, 5$. The true mean is thus $\mathrm{E}(Y) = 3$. If the cardinal values are really $1, \ldots, 5$, then this is indeed the best predictor of Y. However, imagine the cardinal values are instead $1, 2, 3, 4, 10$. In that case, the true mean is $(1+2+3+4+10)/5 = 4$, so the best predictor is 4. That is, we cannot simply code the categories as $1, \ldots, J$ to get a "mean" that is the best predictor regardless of the true cardinal values. Consequently, we cannot generally interpret a nonparametric regression's conditional "mean" estimand as the best predictor of Y given $\boldsymbol{X}$, and we cannot interpret the OLS estimand as the best linear predictor.

Theorem 2. *If Y is an ordinal random variable, then given a quadratic loss function, the best predictor of Y, the best predictor of Y given $\boldsymbol{X}$, and the best linear predictor of Y given $\boldsymbol{X}$ all depend on the true cardinal values corresponding to the categories of Y.*

Proof. The counterexample in the text preceding Theorem 2 shows the best predictor of Y is not invariant to the cardinal values of the categories. Given scalar $X = 1$, this is also a special case of the best predictor of Y given $\boldsymbol{X}$ and the best linear predictor of Y given $\boldsymbol{X}$. □

2.3 Description

Although prediction is sensitive to the true cardinal values, the OLS estimand can be interpreted descriptively without any cardinalization. Lemma 1 immediately generalizes to the conditional "mean" function

$$m(\boldsymbol{x}) \equiv \mathrm{E}(Y \mid \boldsymbol{X} = \boldsymbol{x}) = 1 + \sum_{j=1}^{J-1} \mathrm{P}(Y > j \mid \boldsymbol{X} = \boldsymbol{x}). \tag{1}$$

That is, when coding Y categories as $1, \ldots, J$, we can interpret the conditional "mean" as a sum of conditional survival functions, analogous to the unconditional case. This conditional "mean" function is the estimand of a nonparametric regression of Y (coded $1, \ldots, J$) on $\boldsymbol{X}$.

When Y is cardinal-valued, it is well known that the OLS estimand $\boldsymbol{\beta}$ (the vector of linear projection coefficients) can be interpreted as the best linear approximation of the conditional mean function, with "best" again in terms of quadratic loss (e.g., Kaplan, 2022, §7.4):

$$\boldsymbol{\beta} = \arg\min_{\boldsymbol{b}} \mathrm{E}\{[m(\boldsymbol{X}) - \boldsymbol{X}'\boldsymbol{b}]^2\}. \tag{2}$$

When Y is ordinal, the conditional "mean" $m(\boldsymbol{X})$ in (2) has the cardinalization-robust interpretation given in (1). Thus, the OLS estimand $\boldsymbol{X}'\boldsymbol{\beta}$ can be interpreted as the best linear approximation (in the mean squared error sense) of the sum of conditional survival functions in (1).

Theorem 3 shows that running OLS with an ordinal Y coded with values $1, \ldots, J$ yields a meaningful interpretation even if we do not believe the $1, \ldots, J$ represent cardinal values.

Theorem 3. *Let Y be an ordinal random variable whose J categories are assigned numeric values $1, \ldots, J$. Then, regardless of the true cardinal values of the categories: a) the estimand of a nonparametric regression of Y on $\boldsymbol{X}$ can be written in terms of the conditional survival function as in* (1)*; b) for OLS regression of Y on $\boldsymbol{X}$, the population estimand $\boldsymbol{X}'\boldsymbol{\beta}$ is the best linear approximation in the sense of* (2)*.*

Proof. Combine (1) and (2). □

3 Ordinal decomposition: framework and estimands

Turning attention to decomposition, this section introduces the counterfactual distribution framework used for both our practical and theoretical contributions. We use the framework of Chernozhukov, Fernández-Val, and Melly (2013), adapting their formulas to ordinal outcomes. Then, building on Lemma 1, we show how a naive "mean" decomposition is equivalent to a cardinalization-robust survival function decomposition.

3.1 Counterfactual distribution framework

First, we introduce notation for 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

$$F_Y^C(y) \equiv \int_{\mathcal{X}^A} F_{Y|\boldsymbol{X}}^B(y \mid \boldsymbol{x}) \, \mathrm{d}F_{\boldsymbol{X}}^A(\boldsymbol{x}), \quad y \in \{1, \ldots, J-1\}. \tag{3}$$

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 (3), consider the extreme cases: if $F_{\boldsymbol{X}}^A = F_{\boldsymbol{X}}^B$, then (3) yields $F_Y^C(y) = F_Y^B(y)$, and if $F_{Y|\boldsymbol{X}}^B = F_{Y|\boldsymbol{X}}^A$, then (3) yields $F_Y^C(y) = F_Y^A(y)$.

3.2 Summary statistic interpretations and equivalences

The full distributions F_Y^A, F_Y^B, and F_Y^C can and should be reported, but this requires reporting $3(J-1)$ values, so a summary can facilitate communication and understanding of results. We show an equivalence between a naive "mean" decomposition and a robust survival function decomposition. An alternative quantile decomposition is discussed in Appendix B. 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

$$\Delta_T,\ \Delta_E,\ \Delta_U. \tag{4}$$

For an ordinal outcome, a natural decomposition compares survival function differences summed (or averaged) across categories. This does not depend on the cardinal values of the categories. Although in a different context, this shares the spirit of Theorem 1 of Kobus and Miłoś (2012), who find that any health inequality index satisfying certain axioms can be written as transformations of the category frequencies; our expression in (5) similarly depends only on category frequencies. Given survival functions $S_Y^A(\cdot)$ and $S_Y^B(\cdot)$, we summarize their difference as

$$\sum_{j=1}^{J}[S_Y^A(j) - S_Y^B(j)], \tag{5}$$

and similarly for other pairs of survival functions. Summing from $j = 1$ to $J-1$ is equivalent because $S_Y^A(J) = S_Y^B(J) = 0$. Taking the average (instead of sum) would multiply (5) by $1/J$, but ultimately the explained proportion would remain identical because the $1/J$ would cancel out in (7) below. Given (5), using the notation of (4) and adding superscript S for "survival," the corresponding differences are

$$\Delta_T^S = \sum_{j=1}^{J}[S_Y^A(j) - S_Y^B(j)], \quad \Delta_E^S = \sum_{j=1}^{J}[S_Y^C(j) - S_Y^B(j)], \quad \Delta_U^S = \sum_{j=1}^{J}[S_Y^A(j) - S_Y^C(j)], \tag{6}$$

and the explained proportion is

$$\frac{\Delta_E^S}{\Delta_T^S} = \frac{\sum_{j=1}^{J}[S_Y^C(j) - S_Y^B(j)]}{\sum_{j=1}^{J}[S_Y^A(j) - S_Y^B(j)]}. \tag{7}$$

This is equivalent to a CDF-based decomposition. The components in (6) equal the negative of their CDF-based analogs. For example,

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

and similarly for the other differences in (6). Thus, the explained proportion remains the same because $(-\Delta_E^S)/(-\Delta_T^S) = \Delta_E^S/\Delta_T^S$.

Extending Lemma 1, the survival function decomposition is also equivalent to a naive "mean" decomposition after coding the Y categories as $1, \ldots, J$. We state this as a corollary to a more general result.

Theorem 4. *Let W and Z be discrete random variables with possible values $\{1, 2, \ldots, J\}$. Then,* $\mathrm{E}(W) - \mathrm{E}(Z) = \sum_{j=1}^{J}[S_W(j) - S_Z(j)]$, *where* $S_W(j) \equiv \mathrm{P}(W > j)$ *and* $S_Z(j) \equiv \mathrm{P}(Z > j)$ *are the survival functions.*

Proof. Plugging in for $\mathrm{E}(W)$ and $\mathrm{E}(Z)$ from Lemma 1,

$$\begin{aligned}\mathrm{E}(W) - \mathrm{E}(Z) = 1 + \sum_{j=1}^{J-1} S_W(j) - \left[1 + \sum_{j=1}^{J-1} S_Z(j)\right] &= \sum_{j=1}^{J-1}[S_W(j) - S_Z(j)] \\ &= \sum_{j=1}^{J}[S_W(j) - S_Z(j)]. \qquad \square\end{aligned}$$

Corollary 5. *Given distributions of ordinal random variables Y^A, Y^B, and counterfactual Y^C, the survival function decomposition is equivalent to the "mean" decomposition after coding Y category values as $1, \ldots, J$, in the sense that the explained proportion in* (7) *equals the "mean"-based explained proportion.*

Proof. Theorem 4 implies the following for the "mean"-based decomposition components. The total difference is

$$\Delta_T^\mu \equiv \mathrm{E}(Y^A) - \mathrm{E}(Y^B) = \sum_{j=1}^{J}[S_Y^A(j) - S_Y^B(j)] = \Delta_T^S.$$

Similarly, for the explained difference,

$$\Delta_E^\mu \equiv \mathrm{E}(Y^C) - \mathrm{E}(Y^B) = \sum_{j=1}^{J} [S_Y^C(j) - S_Y^B(j)] = \Delta_T^S.$$

Thus, the explained proportions are also equal: $\Delta_E^\mu / \Delta_T^\mu = \Delta_E^S / \Delta_T^S$. □

Corollary 5 serendipitously implies that we can interpret a "mean" decomposition as a survival function decomposition. Thus, if a paper reports results for an ordinal "mean" decomposition, then even if we disagree with a literal "mean" interpretation, we can still agree about the relative magnitude of explained and unexplained components.

3.3 Implications for regression-based decomposition

Corollary 5 applies to regression-based decomposition. Coding Y as $1, \ldots, J$, let $m^B(\boldsymbol{x}) \equiv \mathrm{E}(Y^B \mid \boldsymbol{X}^B = \boldsymbol{x})$, which by (1) can be interpreted more robustly as $1 + \sum_{j=1}^{J-1} S_Y^B(j \mid \boldsymbol{X}^B = \boldsymbol{x})$. The counterfactual mean is $\mathrm{E}(Y^C) = \mathrm{E}[m^B(\boldsymbol{X}^A)]$. The decomposition is

$$\mathrm{E}(Y^A) - \mathrm{E}(Y^B) = \overbrace{\mathrm{E}(Y^A) - \underbrace{\mathrm{E}[m^B(\boldsymbol{X}^A)]}_{\mathrm{E}(Y^C)}}^{\text{unexplained}} + \overbrace{\underbrace{\mathrm{E}[m^B(\boldsymbol{X}^A)]}_{\mathrm{E}(Y^C)} - \mathrm{E}(Y^B)}^{\text{explained}}. \tag{8}$$

By Corollary 5, the decomposition in (8) 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.

In Section 4.2, we provide an even more precise equivalence result for when OLS is used to estimate the decomposition.

4 Ordinal decomposition: computation and equivalence

Sections 4.1 and 4.3 closely follow the estimation and inference of Chernozhukov, Fernández-Val, and Melly (2013). We continue the notation introduced in Section 3.1. Theoretically, ordinal Y is simpler than continuous Y (as in Chernozhukov, Fernández-Val, and Melly,

2013) 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. This shows that the naive Blinder–Oaxaca "mean" decomposition estimator that seems to assume cardinal values $1, \ldots, J$ can be interpreted as a survival function decomposition estimator that is cardinalization-robust.

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. Generally, let $\Lambda(\cdot)$ be the link function, such as the standard normal or logistic CDF, and let $\boldsymbol{P}(\boldsymbol{x})$ be a column vector of transformations of the original covariate vector $\boldsymbol{x}$. For example, $\boldsymbol{P}(\boldsymbol{x})$ 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

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

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, similar to (9), the estimated conditional CDF is $\hat{F}^B_{Y|\boldsymbol{X}}(y \mid \boldsymbol{x}) = \Lambda\big(\boldsymbol{P}(\boldsymbol{x})'\hat{\boldsymbol{\gamma}}^B_y\big)$.

The estimated conditional CDF for group B is then plugged into the counterfactual distribution formula from (3) 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):

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

where $\boldsymbol{X}^A_i$ 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:

$$\sum_{i=1}^{n_A} \tilde{w}^A_i \Lambda\big(\boldsymbol{P}(\boldsymbol{X}^A_i)'\hat{\boldsymbol{\gamma}}^B_y\big),$$

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

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

$$\hat{F}^A_Y(y) = \frac{1}{n_A}\sum_{i=1}^{n_A}\mathbb{1}\{Y^A_i \le y\}, \quad \hat{F}^B_Y(y) = \frac{1}{n_B}\sum_{i=1}^{n_B}\mathbb{1}\{Y^B_i \le y\},$$

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

$$\hat{F}^A_Y(y) = \sum_{i=1}^{n_A}\tilde{w}^A_i\,\mathbb{1}\{Y^A_i \le y\}, \quad \hat{F}^B_Y(y) = \sum_{i=1}^{n_B}\tilde{w}^B_i\,\mathbb{1}\{Y^B_i \le y\},$$

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

Given the three estimated CDFs $\hat{F}^A_Y$, $\hat{F}^B_Y$, and $\hat{F}^C_Y$, the survival function decomposition and its explained proportion can be computed using (6) and (7), noting that estimated CDF

$\hat{F}(\cdot)$ implies the corresponding estimated survival function $\hat{S}(y) = 1 - \hat{F}(y)$.

4.2 Equivalence with OLS-based Blinder–Oaxaca

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 (7), 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 $Y \in \{1, \ldots, J\}$ as cardinal values. For details about the conventional Blinder–Oaxaca decomposition, see for example (15) and more generally Section 3.1 of Fortin, Lemieux, and Firpo (2011).

Theorem 6. *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, 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, and $\bar{\boldsymbol{X}}^A$ is the average of observed $\boldsymbol{X}$ values in the group A sample, with $\bar{Y}^A$ similarly the average of observed Y values in the group A sample, and with $\bar{\boldsymbol{X}}^B$ and $\bar{Y}^B$ defined similarly for group 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 (7), 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 $Z_y \equiv \mathbb{1}\{Y \leq y\}$ on $\boldsymbol{X}$ using data sample B.

Proof. See Appendix A. □

Theorem 6 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 a 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 functional form misspecification when estimating the counterfactual distribution, using the Blinder–Oaxaca estimate may still be useful for exploratory analysis. Additionally, if the functional form misspecification does not seem too large, Blinder–Oaxaca readily provides a "detailed decomposition" showing the separate contributions of each covariate; for example, see (17)–(18) and the surrounding text of Fortin, Lemieux, and Firpo (2011).

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 $s(\hat{F}_Y^C)$, where $s(\cdot)$ summarizes a distribution like $\hat{F}_Y^C$. 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. Their 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. Depression is widely studied because of its prevalence and importance. For example, Greenberg et al. (2021) estimate that in 2018 in the U.S., the aggregate economic burden of adults with major depressive disorder exceeded $300 billion. This includes the costs of medical care, suicide, and decreased work hours as well as productivity. Our analysis was performed in R (R Core Team, 2022), with help from packages `ggplot2` (Wickham, 2016), `ggmosaic` (Jeppson and Hofmann, 2023; Jeppson, Hofmann, and Cook, 2023), and `fastglm` (Huling, 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 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. We use the provided variables for education (`EDUCP_A`), sex (`SEX_A`), age (`AGEP_A`), family income (`POVRATTC_A`), and geographic region (`REGION`) to construct our explanatory vector $\boldsymbol{X}$, as described in Section 5.2. The estimation uses the sampling weight variable (`WTFA_A`). We drop 282 observations (3.4%): those for which either age or urban group is

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

missing, and those that fit our age and urban group restrictions but have another variable value missing. This leaves 7902 observations for our analysis.

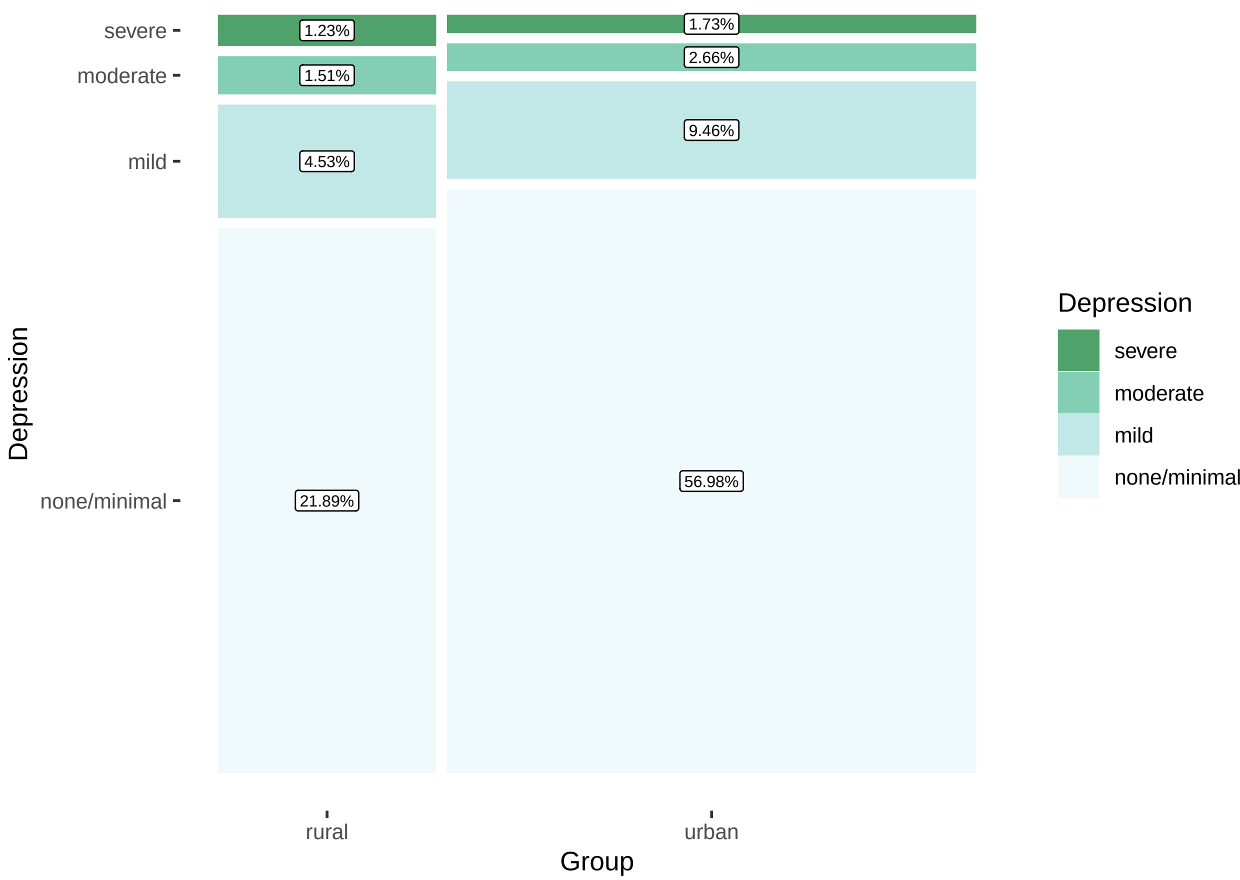

Figure 1: Urban and rural depression level distributions.

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 the rural group A and $n_B = 5454$ in the urban group B. These sample sizes are proportional to the widths of the rural and urban columns 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. Within each group's column, the stacked bar height represents the conditional probability of each depression category, with splits indicating CDF values. For example, within the rural column, 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 says that overall the rural group has higher levels of depression (worse mental health) than the urban group.

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 nonparametric 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. That is, unlike with a scalar X for which a series estimator includes X^k for $k = 0, 1, \ldots, K$, here if we include two quadratic terms (for example), we try (X_1^2, X_2^2), (X_1^2, X_3^2), (X_2^2, X_3^2), (X_1^2, X_1X_2), (X_1^2, X_1X_3), etc. 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.

Instead of AIC, cross-validation could be used for model selection, but there are some disadvantages in our setting. First, if applied to the very large number of candidate models described above, the computation time for even 5-fold cross-validation would be prohibitive (it is already nearly 20 hours with AIC). Second, although computationally faster, there are drawbacks to applying 5-fold cross-validation to select the penalization hyperparameter for lasso (Tibshirani, 1996). One drawback is that the result is sensitive to the random number generator seed used (because the observations in each fold are chosen randomly). Additionally, the conventional error measures like those available through the `glmnet` package (Friedman, Tibshirani, and Hastie, 2010; Tay, Narasimhan, and Hastie, 2023) perform poorly for the more severe depression categories that comprise a relatively small fraction of the population. For example, with outcome $\mathbb{1}\{Y_i \leq 3\}$, the intercept-only model is selected as "best." In principle, this could be addressed by coding an alternative error measure based on weighted 0–1 loss (e.g., Kaplan, 2023, §14.3.1) that makes it relatively more important to correctly predict individuals who actually have severe depression. However, because the AIC model selection works well and such details are far from our main contributions, we do not pursue this alternative further.

In practice, due to computational constraints, in addition to linear terms we always 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 quadratic model resulted in worse (higher) AIC, so we use the selected quadratic models for

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

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 Corollary 5. We use the three estimators in Section 5.2 as well as the naive conventional Blinder–Oaxaca decomposition estimator, to verify our Theorem 6.

Table 1 displays the estimated rural, urban, and counterfactual CDFs. As before, the counterfactual starts from the group B urban distribution and substitutes in the group A rural distribution of $\boldsymbol{X}$, while keeping the group B urban conditional distribution of Y given $\boldsymbol{X}$. 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 the many hours taken by our nonparametric estimator (almost 20 hours on a personal computer). At minimum, OLS seems very practical for exploratory analysis, although for the final analysis a nonparametric estimator may be preferred.

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 6, we also compute the naive Blinder–Oaxaca decomposition using the $1, \ldots, J$ 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.

Table 3 shows results of a "detailed decomposition" using the Blinder–Oaxaca estimates,

Table 3: Blinder–Oaxaca detailed decomposition results.

Variable	Rural mean	Urban mean	Explained (%)
Income (ratio)	3.274	4.715	41.6
	(0.055)	(0.050)	(12.1)
High income	0.020	0.102	-7.8
	(0.003)	(0.005)	(3.5)
Midwest	0.316	0.155	11.3
	(0.011)	(0.006)	(6.2)
West	0.152	0.342	-3.0
	(0.009)	(0.007)	(4.8)
Northeast	0.106	0.165	-2.1
	(0.009)	(0.006)	(2.1)
High education	0.528	0.685	-3.9
	(0.012)	(0.008)	(4.6)
Age	45.635	42.637	-58.0
	(0.283)	(0.184)	(28.0)
Age^2			56.3
			(28.1)
Female	0.493	0.499	-0.6
	(0.012)	(0.008)	(1.5)
Intercept	1.000	1.000	0.0
			(0.0)
Aggregate			33.9
			(12.5)

Bootstrapped standard errors are in parentheses.

which is another advantage of being able to use Blinder–Oaxaca, as our results justify. The detailed decomoposition shows how individuals variables contribute to the overall explained proportion. For each explanatory variable X_j in the vector $\boldsymbol{X}$, we show estimates of the rural mean $\mathrm{E}(X_j^A)$, urban mean $\mathrm{E}(X_j^B)$, and contribution

$$\frac{[\mathrm{E}(X_j^A) - \mathrm{E}(X_j^B)]\hat{\beta}_j^B}{\mathrm{E}(Y^A) - \mathrm{E}(Y^B)} \times 100\%$$

to the overall explained proportion. Note the sum of the estimated contributions equals the estimated explained proportion shown in Table 2. The rows are in decreasing order of absolute contributions, considering the combined contribution of the two income variables and the two age variables.

The main message is that income explains far more than any other variable. The combined contribution of income and the high-income dummy is $41.6 + (-7.8) = 33.8$, almost exactly the overall 33.9% explained. The other contributions are a mix of positive and negative values that nearly fully cancel out. As seen from the rural and urban means, urban incomes are higher, and higher incomes are associated with lower depression (negative $\hat{\beta}_j^B$ coefficient), so this partly explains the higher depression levels in rural areas. Although the contributions of Age and Age^2 initially seem even larger, their combined contribution is only -1.7. The midwest dummy has a contribution of 11.3 to the explained proportion. The means show that a much higher proportion of the rural individuals live in the midwest, and being in the midwest is associated with higher depression (positive $\hat{\beta}_j^B$ coefficient), so this also partly explains the higher depression levels in rural areas. The other contributions are all relatively small in magnitude, as well as all negative, so altogether they offset the midwest contribution. The small magnitudes are a combination of small regression coefficients and/or small differences in the rural and urban means of that X_j variable. Overall, income plays an important role in explaining the rural–urban difference in depression levels, but it still only accounts for about a third of the overall difference.

6 Conclusion

We have provided theoretical results about interpreting OLS-based analysis when an ordinal outcome Y is coded with numeric values $1, 2, 3, \ldots$. Although the "best linear predictor" interpretation of the OLS estimand requires such values to be the true cardinal values of the categories, a "best linear approximation" interpretation remains valid even when those are not cardinal values, where the approximation is of the sum of conditional survival function values. Further, the OLS-based Blinder–Oaxaca decomposition can be interpreted as a survival function decomposition the remains valid even if the $1, 2, 3, \ldots$ are not cardinal values.

This suggests such "naive" OLS-based results can be interpreted robustly and can be practically useful when dealing with the commonly used ordinal variables in health economics, epidemiology, sociology, and related areas in health, medicine, and social science.

CRediT author contribution statement

Qian Wu: Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Writing – original draft, Writing – review & editing, Visualization

David M. Kaplan: Conceptualization, Methodology, Software, Validation, Writing – original draft, Writing – review & editing

Declaration of competing interest

We (the authors) declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

For helpful feedback, we thank Zack Miller, Shawn Ni, Mike Pesko, Yuhao Yang, and especially Alyssa Carlson, as well as seminar participants from the Southwestern University of Finance and Economics and conference participants from the 2024 Midwest Econometrics Group (hosted by the University of Kentucky). This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

A Additional proof

Proof of Theorem 6. Recall from Corollary 5 that decomposing the "mean" is equivalent to decomposing the average survival function difference; below, we use the "mean" for simplicity. For example, Corollary 5 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

$$\hat{\boldsymbol{U}}^B\equiv\boldsymbol{Y}^B-\underline{\boldsymbol{X}}^B\hat{\boldsymbol{\beta}}^B\ \text{ with }\ (\underline{\boldsymbol{X}}^B)'\hat{\boldsymbol{U}}^B=\boldsymbol{0}. \tag{A.1}$$

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

$$\begin{aligned}\rho_E&\equiv\frac{\Delta_E}{\Delta_T}=\frac{[\mathrm{E}(\boldsymbol{X}^A)-\mathrm{E}(\boldsymbol{X}^B)]'\boldsymbol{\beta}^B}{\mathrm{E}(Y^A)-\mathrm{E}(Y^B)}=\frac{\mathrm{E}(\boldsymbol{X}^A)'\boldsymbol{\beta}^B-\mathrm{E}(Y^B)}{\mathrm{E}(Y^A)-\mathrm{E}(Y^B)},\\ \hat{\rho}_E&=\frac{(\bar{\boldsymbol{X}}^A)'\hat{\boldsymbol{\beta}}^B-\bar{Y}^B}{\bar{Y}^A-\bar{Y}^B}.\end{aligned} \tag{A.2}$$

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 (A.2), as seen below.

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

$$Y^B=J-\sum_{j=1}^{J-1}Z_j^B. \tag{A.3}$$

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

$$\hat{\boldsymbol{V}}_j^B \equiv \boldsymbol{Z}_j^B - \underline{\boldsymbol{X}}^B \hat{\boldsymbol{\gamma}}_j^B \quad \text{with} \quad (\underline{\boldsymbol{X}}^B)' \hat{\boldsymbol{V}}_j^B = \boldsymbol{0}. \tag{A.4}$$

Using the above and the orthogonality property of OLS residuals, we can derive the relationship between $\hat{\boldsymbol{\beta}}^B$ (from (A.1)) 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 (A.3) and (A.4),

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

where the equality

$$J\boldsymbol{e}_1 - \sum_{j=1}^{J-1} \hat{\boldsymbol{\gamma}}_j^B = \hat{\boldsymbol{\beta}}^B \tag{A.5}$$

is implied by the orthogonality

$$(\underline{\boldsymbol{X}}^B)' \sum_{j=1}^{J-1} (-\hat{\boldsymbol{V}}_j^B) = -\sum_{j=1}^{J-1} \overbrace{(\underline{\boldsymbol{X}}^B)' \hat{\boldsymbol{V}}_j^B}^{=\boldsymbol{0}} = \boldsymbol{0}, \tag{A.6}$$

which follows from the orthogonality condition in (A.4).

Now consider the "mean" of the counterfactual distribution. Taking the expectation of (A.3) and using the linearity of the expectation operator,

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

Thus, the estimated counterfactual "mean" is

$$\hat{\mathrm{E}}(Y^C) = J - \sum_{j=1}^{J-1} (\bar{\boldsymbol{X}}^A)'\hat{\boldsymbol{\gamma}}_j^B = (\bar{\boldsymbol{X}}^A)' \overbrace{\left(J\boldsymbol{e}_1 - \sum_{j=1}^{J-1} \hat{\boldsymbol{\gamma}}_j^B \right)}^{=\hat{\boldsymbol{\beta}}^B \text{ by (A.5)}} = (\bar{\boldsymbol{X}}^A)'\hat{\boldsymbol{\beta}}^B.$$

This final expression is identical to the term in the Blinder–Oaxaca decomposition numerator in (A.2), so the estimated explained proportion is identical. □

B Quantiles

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

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

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, §2.3).

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.

References

Abul Naga, Ramses H. and Tarik Yalcin. 2008. "Inequality measurement for ordered response health data." *Journal of Health Economics* 27 (6):1614–1625. URL `https://doi.org/10.1016/j.jhealeco.2008.07.015`.

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`.

Allison, R. Andrew and James E. Foster. 2004. "Measuring health inequality using qualitative data." *Journal of Health Economics* 23 (3):505–524. URL `https://doi.org/10.1016/j.jhealeco.2003.10.006`.

Apouey, Benedicte. 2007. "Measuring health polarization with self-assessed health data." *Health Economics* 16 (9):875–894. URL `https://doi.org/10.1002/hec.1284`.

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.

Carrieri, Vincenzo and Andrew M. Jones. 2017. "The Income–Health Relationship 'Beyond the Mean': New Evidence from Biomarkers." *Health Economics* 26 (7):937–956. URL https://doi.org/10.1002/hec.3372.

Cartwright, Kate. 2021. "Social determinants of the Latinx diabetes health disparity: a Oaxaca–Blinder decomposition analysis." *SSM - Population Health* 15:100869. URL https://doi.org/10.1016/j.ssmph.2021.100869.

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.

Friedman, Jerome, Robert Tibshirani, and Trevor Hastie. 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33 (1):1–22. URL https://doi.org/10.18637/jss.v033.i01.

Greenberg, Paul E., Andree-Anne Fournier, Tammy Sisitsky, Mark Simes, Richard Berman, Sarah H. Koenigsberg, and Ronald C. Kessler. 2021. "The Economic Burden of Adults with Major Depressive Disorder in the United States (2010 and 2018)." *PharmacoEconomics* 39 (6):653–665. URL https://doi.org/10.1007/s40273-021-01019-4.

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.

Huling, Jared. 2022. *fastglm: Fast and Stable Fitting of Generalized Linear Models using 'RcppEigen'*. URL https://CRAN.R-project.org/package=fastglm. R package version 0.0.3.

Idler, Ellen and Kate Cartwright. 2018. "What Do We Rate When We Rate Our Health? Decomposing Age-related Contributions to Self-rated Health." *Journal of Health and Social Behavior* 59 (1):74–93. URL https://doi.org/10.1177/0022146517750137.

Jeppson, Haley and Heike Hofmann. 2023. "Generalized Mosaic Plots in the ggplot2 Framework." *The R Journal* 14 (4):50–78. URL https://doi.org/10.32614/RJ-2023-013.

Jeppson, Haley, Heike Hofmann, and Dianne H. Cook. 2023. *ggmosaic: Mosaic Plots in the 'ggplot2' Framework*. URL https://haleyjeppson.github.io/ggmosaic/,https://github.com/haleyjeppson/ggmosaic. R package version 0.3.4.

Kaplan, David M. 2022. *Introductory Econometrics: Description, Prediction, and Causality*. Columbia, MO: Mizzou Publishing, 3rd ed. URL https://www.themizzoustore.com/product/146951?quantity=1.

———. 2023. "PhD Core Econometrics II." URL https://kaplandm.github.io/teach.html. Textbook draft.

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.

Kino, Shiho and Ichiro Kawachi. 2020. "How much do preventive health behaviors explain education- and income-related inequalities in health? Results of Oaxaca–Blinder decomposition analysis." *Annals of Epidemiology* 43:44–50. URL https://doi.org/10.1016/j.annepidem.2020.01.008.

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.

Kobus, Martyna and Piotr Miłoś. 2012. "Inequality decomposition by population subgroups for ordinal data." *Journal of Health Economics* 31 (1):15–21. URL https://doi.org/10.1016/j.jhealeco.2011.11.005.

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.

Sen, Bisakha. 2014. "Using the Oaxaca–Blinder decomposition as an empirical tool to analyze racial disparities in obesity." *Obesity* 22 (7):1750–1755. URL https://doi.org/10.1002/oby.20755.

Tay, J. Kenneth, Balasubramanian Narasimhan, and Trevor Hastie. 2023. "Elastic Net Regularization Paths for All Generalized Linear Models." *Journal of Statistical Software* 106 (1):1–31. URL https://doi.org/10.18637/jss.v106.i01.

Tibshirani, Robert J. 1996. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society: Series B* 58 (1):267–288. URL www.jstor.org/stable/2346178.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. URL https://ggplot2.tidyverse.org.

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.