The University Of Sheffield. | Department Of Economics

Decompositions: Accounting for Discrimination

Gurleen Popli

Sheffield Economic Research Paper Series

SERPS no. 2022009

ISSN 1749-8368

28 June 2022

Decompositions: Accounting for Discrimination

Gurleen Popli
Department of Economics
University of Sheffield, UK
g.popli@shef.ac.uk

June 2022

Abstract
This chapter summarises the different regression based decomposition methods used in the empirical literature to evaluate discrimination. Starting with the decomposition at the mean using the methods made popular in the 1970s by Oaxaca and Blinder, we discuss how the method has evolved over time to look beyond the means, taking into account the entire distribution of the outcomes of interest. We present the formal identification assumptions underlying the decomposition method and discuss cautions that should be exercised in interpreting them and their limitations. We also explain how the 'unexplained gap' in the decomposition, often used as a measure of discrimination, relates to the treatment effect literature.

Keywords: Decomposition, Counterfactual regressions, Distributions, Discrimination
JEL codes: J31, J71, C21

Forthcoming in Handbook of Economics of Discrimination and Affirmative Action, Springer, 2022

Acknowledgements: I would like to thank Okan Yilmaz and Meng Le Zhang for their helpful comments on an earlier draft of the chapter.

1. Introduction

The fact that discrimination exists is not a doubt; it exists and impacts every aspect of our lives, including, but not limited to, economic outcomes (in educational opportunities, and outcomes in labour and credit markets), social outcomes (network formations, and residential location), health outcomes (mortality rates, and access to health services), and interactions with the criminal justice system (Arrow 1998; Lang and Kahn-Lang Spitzer 2020; Small and Pager 2020). Within the discipline of economics, much of the theoretical discussion around discrimination started with the work of Becker (1957), while much of the empirical work has come from within the field of labour economics with the seminal works of Oaxaca (1973) and Blinder (1973), who used wage regressions to quantitatively assess the role of discrimination in explaining the observed wage gaps between men and women (in case of Oaxaca) and between whites and blacks (in case of Blinder, who also looked at the gender wage gaps).

Since Oaxaca and Blinder's work, the literature on regression based decompositions has seen substantial growth in methodological advancements and their empirical applications. This chapter gives a summary of the different regression based decomposition methods. Starting with Oaxaca and Blinder's work, we discuss how the method has evolved over time. Two key limitations of the regression based decompositions should be stated at the very beginning. First, all the methods discussed here follow the partial equilibrium approach. The wage gaps are decomposed into a part explained by the differences in endowments of the two groups, holding the returns to these endowments constant; and the differences in the returns to endowments across the two groups, holding their endowments constant. This assumes that we can change the endowments without impacting returns to them and vice versa. This is a strong assumption and often not valid.

Second, regression based decomposition methods are an accounting exercise. While we can know various factors contributing to the existing difference between wages (or any outcome of interest), decompositions do not tell us the underlying mechanisms. They can, however, help us confirm an existing hypothesis or form new ones. When we do regression based decompositions, we are not trying to detect discrimination but are attempting to quantify it. While we have to be cautious in our interpretations that not all the observed wage gaps are discrimination, nothing stops us from concluding that discrimination exists and is buried in it. The way we use regression based decompositions and the way we interpret them is critical. If done correctly and used correctly, we can give some accounting of the existing discriminations.

Section 2 of the chapter discusses the most famous decomposition method – the Oaxaca-Blinder (OB, henceforth) method. This section will also highlight issues around interpretations, including the treatment effect interpretation, and formal assumptions for identification. Section 3 will discuss the different decomposition methods that have been proposed in the literature since the OB work. Section 4 provides concluding comments.

2. Decomposing the mean

2.1 Oaxaca-Blinder method

What we refer to as the OB decomposition in economics was first employed in demography by Kitagawa (1955), and made popular in sociology by Althauser and Wigler (1972). This method has been used extensively in the empirical economics literature to study mean wage gaps between different groups in the labour market, defined over gender, race and ethnicity, age, disability, and over immigration status, among others. The decomposition has also been used to look at outcomes beyond the labour market, like consumption and expenditure differences, and inequalities over time and space (i.e. across different regions). Below we lay out the basics of this decomposition.

Let Y_{gi} be any outcome of interest, for individual i ($i = 1, \dots, n$) belonging to group $g \in (M, W)$. For ease of exposition, let the two groups be men (M) and women (W), and the outcome of interest be wages. Let $X_{gi} = (X_{gi1}, \dots, X_{giK})$ be a vector of K covariates which are associated with the outcome of interest. We assume that the outcome of interest is continuous and linearly related to the covariates as:

$$Y_{gi} = \beta_{g0} + \Sigma_{k=1}^{K} X_{gik}\beta_{gk} + \upsilon_{gi} \tag{1}$$

where β_{g0} and β_{gk} are parameters to be estimated; and υ_{gi} is the error term which is conditionally independent of X_{gi}, such that $E(\upsilon_{gi}|X_{gi}) = 0.$ The difference in the means of the outcomes between the two groups, $\hat{\Delta}_O^{\mu} = \bar{Y}_M - \bar{Y}_W$, where $\bar{Y}_g$ is the mean outcome for group g, is given as:

$$\hat{\Delta}_O^{\mu} = \underbrace{(\hat{\beta}_{M0} - \hat{\beta}_{W0}) + \sum_k \bar{X}_{Wk}(\hat{\beta}_{Mk} - \hat{\beta}_{Wk})}_{\hat{\Delta}_S^{\mu}\ (unexplained)} + \underbrace{\sum_k (\bar{X}_{Mk} - \bar{X}_{Wk})\hat{\beta}_{Mk}}_{\hat{\Delta}_X^{\mu}\ (explained)} \tag{2}$$

where $\bar{X}_{gk}$ is the mean of covariate X_{gik} for group g, and $\hat{\beta}_{g0}$ and $\hat{\beta}_{gk}$ are the estimated parameters, intercept and slope coefficients, from the regression models estimated separately for the two groups. The mean difference is decomposed into two parts. The first term, $\hat{\Delta}_S^{\mu}$ is the unexplained component, if the outcome of interest is wages, this is also referred to as the *wage structure effect*. The second term, $\hat{\Delta}_X^{\mu}$ is the explained component, this is also referred to as the *composition effect*.

The composition (explained) effect is the difference in wages due to differences in the observed covariates (endowments) of the individuals across the two groups. For example, the composition effect is the part of the mean gender wage gap that is explained due to observable differences in mean characteristics, $\bar{X}_{Mk} - \bar{X}_{Wk}$, among the men and women, like education and labour market experience. The wage structure (unexplained) effect is the difference in mean wages due to the difference in returns to individual characteristics, $\hat{\beta}_{Mk} - \hat{\beta}_{Wk}$. The difference in the intercepts, $\hat{\beta}_{M0} - \hat{\beta}_{W0}$, is interpreted as the part of the unexplained gap attributed to group membership. The unexplained component is the gender wage gap that is often associated with labour market discrimination, omitted variables, and unobserved heterogeneity.

Using the expression given in equation (2), we can investigate both the aggregate and detailed decompositions. *Aggregate decomposition* is where we are only interested in the overall wage structure and composition effects, i.e. $\hat{\Delta}_X^{\mu}$ and $\hat{\Delta}_S^{\mu}$, which helps us understand how much of the observed gap is due to differences in endowments (composition effect) and how much of the gap is due to differences in returns to those endowments (wage structure effect). If we are interested only in aggregate decompositions, we do not even need to separately estimate the wage regressions for women. The unexplained component can be simply computed as, $\hat{\Delta}_S^{\mu} = \hat{\Delta}_O^{\mu} - \sum_k (\bar{X}_{Mk} - \bar{X}_{Wk})\hat{\beta}_{Mk}$.

There is, however, often an interest in *detailed decomposition*, where we wish to know the contribution of each individual covariate, X_{giK}, to both the wage structure and composition effect. For example, to understand the gender wage gap, it can be of interest to know how much of the composition gap is due to differences in education between men and women and how much is due to differences in the labour market experience between the two groups; as the policy implications from the two can be different. Similarly, we might also be interested in differences in returns to education versus differences in returns to labour market experience when looking at the wage structure effect.

The OB decomposition has a counterfactual interpretation, which has been exploited in later methodological innovations. Consider,

$$Y_i^C = \hat{\beta}_{M0} + \Sigma_{k=1}^{K} X_{Wik}\hat{\beta}_{Mk} \tag{3}$$

where Y_i^C is the counterfactual wage for women obtained by using their own characteristics, but the estimated parameters are from regression for men. The counterfactual wages tell us what women's wages would have been if they had their own characteristics but their characteristics were rewarded as men's are. Using the notion of counterfactual wages, we can write the decomposition in equation (2) as:

$$\underbrace{\bar{Y}_M - \bar{Y}_W}_{\hat{\Delta}_O^{\mu}} = \underbrace{\bar{Y}_M - \bar{Y}^C}_{\hat{\Delta}_X^{\mu}} + \underbrace{\bar{Y}^C - \bar{Y}_W}_{\hat{\Delta}_S^{\mu}} \tag{4}$$

where $\bar{Y}^C$ is the mean counterfactual wage distribution. The first term of equation (4) is the explained component and the second term is the unexplained component. The unexplained component is the difference in wages that women would have received if they had their own characteristics but were treated as men are in the labour market and the actual wages of women.

2.2 Issues with the decompositions

This section discusses three main cautions that should be exercised or issues that should be kept in mind when interpreting the findings from regression based decompositions. These cautions apply to the OB decomposition of the mean and all the extensions and methods proposed since then, including those discussed in section 3.

First is the issue of the reference group. To understand this, let's consider the counterfactual wages generated for women, given by equation (3). In this format, it is clear that men are assumed to be the reference group, and male returns to characteristics (the estimated coefficients) are assumed to be non-discriminatory, i.e. we assume that these are the returns that would prevail in the market for both men and women in the absence of discrimination. This is not an innocuous assumption; choice of reference group can alter how the observed wage gap is attributed to the effect of wage structure and composition.

The choice of reference group has been discussed in the literature, with various alternatives being proposed. Cotton (1988) gives an excellent graphical illustration of the inherent assumption made when choosing one group as a reference relative to the other group, and proposes using a weighted average of the estimated coefficients for the two groups as the non-discriminatory coefficients. Neumark (1988), on the other hand, proposed using estimated coefficients from a pooled regression for the two groups. Jones and Kelly (1984) propose what is known as the three-fold decomposition:

$$\hat{\Delta}_O^{\mu} = \underbrace{(\hat{\beta}_{M0} - \hat{\beta}_{W0}) + \sum\nolimits_k \bar{X}_{Wk}(\hat{\beta}_{Mk} - \hat{\beta}_{Wk})}_{wage\ structure\ effect} + \underbrace{\sum\nolimits_k (\bar{X}_{Mk} - \bar{X}_{Wk})\hat{\beta}_{Mk}}_{composition\ effect} \quad (5)$$

$$+ \underbrace{\sum\nolimits_k (\bar{X}_{Mk} - \bar{X}_{Wk})(\hat{\beta}_{Mk} - \hat{\beta}_{Wk})}_{interaction\ term}$$

The first term, in equation (5), is the *pure* wage structure effect, as before this reflects how much of the gender wage gap results from differences in how women's characteristics are actually valued in the market and how they would be valued if they had the same rates of return as of men. The second term is the *pure* composition effect, as before, this reflects how much more women would earn if they had the same characteristics as men, but nothing else has changed. The third term is the interaction term, this is the amount that women would gain if they had the characteristics as men and if these characteristics had returns similar to men. In the empirical analysis, this term is often small and is hard to interpret when we have multiple covariates. For a further discussion of the alternative measures proposed in the literature and a unified framework to compare them see Oaxaca and Ransom (1994).

The second issue concerns the base group or the 'omitted group' problem. The omitted group problem is discussed in detail by Oaxaca and Ransom (1999) and Oaxaca (2007). This is a concern if we want to do a detailed decomposition of the wage structure (unexplained) component and have categorical covariates. If we are interested only in aggregate decompositions, then this is not a concern. To illustrate the problem, let's assume the only covariate we have are sectors of employment: primary, manufacturing, and service. In the regression framework, we include two dummies for two sectors, and one sector is omitted as the base category. Say we arbitrarily set the primary sector ($sec1$) as the omitted category and include dummies for manufacturing ($sec2$) and service ($sec3$) sector in the regression equation. This gives us the wage structure effect:

$$\hat{\Delta}_S^{\mu} = (\hat{\beta}'_{M0} - \hat{\beta}'_{W0}) + \sum\nolimits_{k=2,3} \bar{X}_{Wseck}(\hat{\beta}_{Mseck} - \hat{\beta}_{Wseck}) \quad (6)$$

In the presence of categorical covariates, the difference in the intercepts of the two wage regressions for the two groups can be written as:

$$\widehat{\beta'}_{M0} - \widehat{\beta'}_{W0} = (\hat{\beta}_{M0} + \hat{\beta}_{Msec1}) - (\hat{\beta}_{W0} + \hat{\beta}_{Wsec1}) = (\hat{\beta}_{M0} - \hat{\beta}_{W0}) + (\hat{\beta}_{Msec1} - \hat{\beta}_{Wsec1}) \quad (7)$$

When we have categorical covariates difference in intercepts, $\widehat{\beta'}_{M0} - \widehat{\beta'}_{W0}$, includes not only the gap attributed to the group membership, $\hat{\beta}_{M0} - \hat{\beta}_{W0}$, but also the gap attributed to belonging to the omitted sector, $\hat{\beta}_{Msec1} - \hat{\beta}_{Wsec1}$. The latter will change depending on the sector that is omitted.

The omitted group's issue arises only in the wage structure effect and does not impact the composition effect; neither does this problem arise if there is only one binary dummy variable as a covariate (Oaxaca 2007). However, the problem gets complicated if we have more than one categorical covariate. Further, as Jones (1983) pointed out, the base group problem also exists in continuous covariates that do not have a natural scale, like test scores. Some solutions to the base group problem have been proposed in the literature, see Gardeazabal and Ugidos (2004) and Yun (2005), which involve the normalisation of coefficients of the categorical variables. However, there is no agreement on this, these normalisations tend to be sample specific and different researchers can use different sets of normalisations. Base groups should be chosen based on context and economic meaning.

The third issue is of self-selection *within* groups and *between* groups. The most common example of self-selection *within* groups is of differential selection into the labour force by gender. Self-selection *between* groups arises when there is an element of choice over group membership, e.g. union versus non-union workers. In the presence of selection (whether between groups or within groups), the estimated coefficients of the wage regression are biased. There are unobserved variables that are correlated with both selection, whether it is the choice of participation in the labour force or the choice of joining unions, and wages. A solution to this problem is to estimate a selection corrected wage regression, similar to that proposed by Hekman (1979). The selection-corrected wage regression then yields the unbiased estimates of the wage regressions and allows for subsequent decomposition.

Let the selection corrected wage regression be given as:

$$Y_{gi} = \beta_{g0} + \Sigma_{k=1}^{K} X_{gik}\beta_{gk} + \sigma_g \lambda_{gi} + \upsilon_{gi} \quad (8)$$

where λ_{gi} is the control variable to correct for selection, if the Heckman selection model is used this will be the inverse mills ratio estimated from the first step of the selection model; and σ_g is the estimated coefficient for the control variable. This now changes the decomposition, as the selection corrected wage regressions have an additional term; the decomposition is now given as:

$$\hat{\Delta}_O^{\mu} = \underbrace{(\hat{\beta}_{M0} - \hat{\beta}_{W0}) + \sum_k \bar{X}_{Fk}(\hat{\beta}_{Mk} - \hat{\beta}_{Wk})}_{\hat{\Delta}_S^{\mu}\ (unexplained)} + \underbrace{\sum_k (\bar{X}_{Mk} - \bar{X}_{Wk})\hat{\beta}_{Mk}}_{\hat{\Delta}_X^{\mu}\ (explained)} + (\bar{\lambda}_M \hat{\sigma}_M - \hat{\lambda}_W \hat{\sigma}_W) \quad (9)$$

where the first two terms are as before and the last term $(\bar{\lambda}_M \hat{\sigma}_M - \hat{\lambda}_W \hat{\sigma}_W)$ is the difference in the wage gap attributed to differences in selection bias. How the selection term (control variable and the parameter associated with it) is attributed to the different parts of the wage decomposition has implications for interpreting discrimination; for a discussion see Neuman and Oaxaca (2005).

2.3 Treatment effect interpretation

Fortin et al. (2011) show that the wage structure effect estimated from regression based decomposition has parallels with the treatment effect literature. To explain this, let us consider union workers (U) and non-union workers (N). The difference in the average wages if everyone is paid according to the wage structure of union members and the average wages of everybody if they were paid according to the wage structure of non-union members, can be conceived as the average treatment effect (ATE) where the treatment is 'union' membership, i.e. switching workers from being non-union members to union members. Assuming non-union workers as the reference group, we can construct counterfactual wages for union workers, these are the wages that the union workers with characteristics X_{Uik} would earn if they were paid as non-union workers:

$$\bar{Y}^C = \hat{\beta}_{N0} + \sum_k \bar{X}_{Uk}\hat{\beta}_{Nk} \tag{10}$$

Using this counterfactual wage the decomposition can be written as:

$$\hat{\Delta}_O^{\mu} = \bar{Y}_U - \bar{Y}_N = \underbrace{(\bar{Y}_U - \bar{Y}^C)}_{\hat{\Delta}_S^{\mu}} + \underbrace{(\bar{Y}^C - \bar{Y}_N)}_{\hat{\Delta}_X^{\mu}} \tag{11}$$

where the difference between the average wages of the union workers and their counterfactual wages, $\bar{Y}_U - \bar{Y}^C$, is the wage structure effect. Under treatment effect interpretation $\bar{Y}_U$, the average wages of the union workers, is the average wages of the treatment group where the treatment is 'union' membership; and $\bar{Y}^C$ is then the average wages of the treated (union) workers if they were not treated, i.e. they were non-unionised. Difference between the two, $\bar{Y}_U - \bar{Y}^C$, is the 'union effect' or the average treatment effect on the treated (ATT). In this framework, the composition effect, $\hat{\Delta}_X^{\mu}$, is referred to as the selection bias.

The ATT interpretation of the wage structure holds for aggregate decompositions, including extensions that go beyond the mean, under the formal identification assumptions discussed below; however, this equivalence does not always hold for detailed decompositions. The issues of choice of the reference group and the base group problem remain in this interpretation as well; however, this approach gives us a way around the selection issue. For a detailed discussion of equivalence between the treatment effect interpretation for the OB decompositions see An and Glynn (2019).

For the OB decomposition, we had assumed the error tern to be conditionally independent of covariates, $E(v_{gi}|X_{gi}) = 0$. Under the treatment effect interpretation, this assumption can be replaced by a weaker assumption of 'ignorability'. This helps us address some issues of selection bias. Under the assumption of ignorability selection bias is allowed as long as it is the same for the two groups. For example, ability, which we do not observe, can be correlated with education, an observed covariate, as long as the correlation is the same in the two groups being considered. This is the 'selection on observables' assumption used in the treatment literature. Under this interpretation, we do not need to calculate the composition (explained) component, once we have the unexplained component we can compute the explained part as: $\hat{\Delta}_X^{\mu} = \hat{\Delta}_O^{\mu} - \hat{\Delta}_S^{\mu}$, which reflects the difference in the distribution of the covariates and the error term.

In the treatment effect literature ATT has a causal interpretation. Even though the wage structure effect is derived under the same conditions as ATT, the causal interpretation often cannot be extended to the wage structure effect for two main reasons. First, in decompositions, 'treatment' is not always a choice or something that can be manipulated easily. For example, while we can conceive union/non-union membership as a choice, gender cannot be conceived as a choice or something that can be manipulated. Further, as we discussed in the selection issue, group membership is often related to unobserved variables, this means the ignorability assumption is often violated, making the causal interpretation hard. Second, as mentioned in the introduction, decompositions are a partial equilibrium approach. For example, the wage structure effects for union wage gap tells us, holding union workers characteristics (X_{Ui}) same, if the returns to their characteristics were the same as that of non-union workers (i.e. $\beta_{Uk} = \beta_{Nk}$), then the wage structure effect would be zero. Equating the two parameters is similar to 'treating' union workers as non-union workers, but this treatment is likely to change union workers' characteristics. When X_{gi} is impacted by the treatment we cannot obtain a causal interpretation. Unless great caution has been taken to include only those characteristics that are unlikely to be impacted by the treatment, for an example of this see Neal and Johnson (1996).

2.4 Formal identification

In this section, we lay out the minimum set of assumptions needed to identify the aggregate decomposition. There are further assumptions needed to identify detailed decompositions and some of the extensions discussed in section 3. For a more technical account of these assumptions refer to Fortin et al. (2011), who set out the full set of assumptions needed for identification for all regression based decomposition methods.

Assumption 1: *Mutually exclusive groups*

The population of interest can be divided into two mutually exclusive groups, $g \in (A, B)$. We are interested in comparing the outcome Y_{gi} of the two groups. In line with the treatment effect literature, Y_{Ai} and Y_{Bi} can be interpreted as potential outcomes for individual i. If the individual is in group A (B) then we only observe Y_{Ai} (Y_{Bi}).

Assumption 2: *Structural form*

A worker i belonging to group g is paid according to the wage structure m_g, which is a function of worker's observable characteristics, X_{gi}, and unobservable characteristics, υ_{gi}:

$$Y_{Ai} = m_A(X_{Ai}, \upsilon_{Ai}) \quad \text{and} \quad Y_{Bi} = m_B(X_{Bi}, \upsilon_{Bi}) \tag{12}$$

In a more general framework, we can think of m_g as the function linking individual characteristics and their outcomes. Under linearity: $Y_{gi} = m_g(X_{gi}, \upsilon_{gi}) = \mathbf{X}_{gi}\beta_g + \upsilon_{gi}$. There are three reasons why wages can be different between groups. (1) The difference in the wage setting equations m_A and m_B. For the linear model this is the difference in the returns to the observed characteristics, β_A and β_B; or the difference in the returns to the unobservable characteristics. (2) Differences in the distribution of the observable characteristics, X_{gi}, for the two groups. (3) Differences in the distribution of the unobservable characteristics, υ_{gi}, for the two groups.

Assumption 3: *Simple counterfactual treatment*

A counterfactual wage structure, m^C, is said to correspond to a simple treatment when it can be assumed that $m^C\,(X_{Bi}, \upsilon_{Bi}) \equiv\ m_A(X_{Bi}, \upsilon_{Bi})$, or $m^C\,(X_{Ai}, \upsilon_{Ai}) \equiv\ m_B(X_{Ai}, \upsilon_{Ai})$. This assumption states, we can identify what the distribution of wages for groups A will be if the returns to their characteristics are similar to those of group B. This is the counterfactual we constructed for women, given by equation (3), and the counterfactual we constructed for union workers, given by equation (10). This assumption, however, also highlights that while we can identify what the distribution of wages for women (union workers) will be if they were paid as men (non-union workers) are, we cannot identify what the distribution of wages for women (union workers) will be if there were no labour market discrimination (no unions).

Assumption 4: *Overlapping support*

No single value of observable, X_{gi}, or unobservable, υ_{gi}, characteristics can identify group membership. This rules out a situation where the covariates in X_{gi} are different across groups, or there are values that X_{gi} can take only for one group and not for the other. The issue of overlapping support, or common support as it is also referred to, is well recognised in the treatment effect literature. This is usually not a problem if our focus is only on the decomposition of the mean but becomes an issue when looking at distributions.

An example of different covariates across groups is when let's say we want to compare wages of immigrants and natives, where the country of origin is an essential predictor of wages for the immigrants but is not relevant for natives. Similarly, when we look at gender wage gaps, often women tend to be concentrated into occupations and or industry combinations where there will be no men.

For a discussion of this and possible solutions see Nopo (2008). It is also possible that the range of values that certain covariates take differ by groups. This was discussed by Barsky et al. (2002), where the authors look at the role of income in explaining the black-white wealth gaps, and found that there are certain regions of income where no blacks are observed.

Assumption 5: *Conditional independence or ignorability*

Let $D_{gi} = \mathbf{1}\{i \in g\}$, where $\mathbf{1}\{.\}$ is the indicator function for group membership. Ignorability requires $D_{gi} \perp \upsilon | X$. Intuitively, the ignorability assumption states that the distribution of unobservable characteristics, υ, given observable characteristics, X, be the same for the two groups. In case of the decomposition of the mean and linear specification, it does not require $E\left(\upsilon_{gi} \middle| \mathrm{X}_{gi}\right) = 0$, instead all it requires is $E(\upsilon_{Ai}|\mathrm{X}_{Ai}) = E(\upsilon_{Bi}|\mathrm{X}_{Bi})$. In the treatment effects literature, this assumption is also referred to as *unconfoundedness* or *selection on observables*.

3. Going beyond the mean

There have been several extensions of the OB decomposition. In this section, we discuss some of these extensions. First, we discuss the extension proposed by Juhn et al. (1993), who give a way to account for the unobserved characteristics via a residual imputation method. Second is the reweighting method, proposed by DiNardo et al. (1996) to investigate the distribution of the unexplained differences. The third is the decomposition of the conditional distributions, based on quantile regressions, by Machado and Mata (2005). Fourth, and finally, is the decomposition of unconditional quantiles, using Recentered Influence Functions, proposed by Firpo et al. (2007, 2009). While most of the methods discussed in this chapter allow us to look beyond the mean and are good at obtaining aggregate decompositions, detailed decompositions beyond the mean remain challenging, and where possible, they come at the cost of simplicity and intuitive appeal of the OB method.

3.1 Residual Imputations

The main contribution of Juhn, Murphy, and Pierce (1993), JMP henceforth, was to take into account the returns to the unobservable characteristics (also referred to as unobserved heterogeneity) explicitly in the decompositions. Starting with the regression function given by equation (1), when we decompose the mean gap between two groups we get equation (2). Residuals, υ_{ig}, are not a part of the decomposition as $\bar{\upsilon}_g = 0$, where $\bar{\upsilon}_g$ is the mean residual of group g. However, the residuals υ_{ig} are interpreted as unobservable characteristics and JMP propose a way to look at how the distributions of these unobserved characteristics differ among the two groups.

JMP define the cumulative distribution of wage residuals, conditional on covariates, as $\theta_{ig} = F(\upsilon_{ig}|X_{ig})$. This gives us, $\upsilon_{ig} = F^{-1}(\theta_{ig}|X_{ig}) \equiv F_{ig}^{-1}$, which is the inverse cumulative distribution of wage residuals. Using this definition of residuals we can rewrite equation (1) as:

$$Y_{gi} = \beta_{g0} + \Sigma_{k=1}^{K} X_{gik}\beta_{gk} + F_{ig}^{-1} \tag{13}$$

Assuming the two groups to be men and women, there are now three potential sources of the gender wage gap, differences in the observables, X; differences in returns to observables, β; and differences in the distribution of unobservables (residuals), F_g^{-1}. To obtain the decomposition JMP recommend generating two different counterfactuals:

$$Y_i^{C1} = \hat{\beta}_{M0} + \Sigma_{k=1}^{K} X_{Wik}\hat{\beta}_{Mk} + F_{iM}^{-1} \tag{14}$$

$$Y_i^{C2} = \hat{\beta}_{W0} + \Sigma_{k=1}^{K} X_{Wik}\hat{\beta}_{Wk} + F_{iM}^{-1} \tag{15}$$

The first counterfactual, Y_i^{C1}, gives us the counterfactual wages for women, if the observable characteristics are of women and the returns to the observables and the residual distribution are of men. The second counterfactual, Y_i^{C2}, gives us the counterfactual wages for women, if both the observable characteristics and returns to them are of women, but the residual distribution is of men.

Using the two counterfactuals the decomposition is given as:

$$\bar{Y}_M - \bar{Y}_W = (\bar{Y}_M - \bar{Y}^{C1}) + (\bar{Y}^{C1} - \bar{Y}^{C2}) + (\bar{Y}^{C2} - \bar{Y}_W) \tag{16}$$

The first term on the right-hand side of equation (16), $(\bar{Y}_M - \bar{Y}^{C1})$, gives us the difference in wages due to observable characteristics (the composition effect); second term, $(\bar{Y}^{C1} - \bar{Y}^{C2})$, gives us the difference in wages due to returns to observable characteristics (the wage structure effect); and the third term, $(\bar{Y}^{C2} - \bar{Y}_W)$, gives us the difference in wages due to differences in the distribution of the unobservable factors.

While proving to be very useful in looking at the role of unobserved characteristics in explaining wage differentials, the JMP method has some limitations which need to be kept in mind. First, if the number of observations between the two groups does not match it is not clear how to assign residuals from one group to another, i.e. how to generate the first counterfactual, equation (14). The solution proposed by JMP for this is to replace the i^{th} ranked residual from women's residual distribution with the i^{th} ranked residual from the residual distribution for men. Second, while theoretically, we need $\theta_{ig} = F(\upsilon_{ig}|X_{ig})$, empirically all we can get is $\theta_{ig} = F(\upsilon_{ig})$, which implies independence between observed and unobserved characteristics. This can be an unrealistic assumption. Third, the decomposition is path dependent, the order in which we generate the counterfactuals can change the size of the different effects. Lastly, in the empirical analysis the three components of the decomposition need not add up to the observed mean gap. For a full discussion of the JMP method's limitations and possible solutions, see Lemieux (2002) and Yun (2009).

3.2 Reweighting methods

DiNardo, Fortin, and Lemieux (1996), DFL henceforth, generalised the OB method for the entire distribution of wages. We start by estimating the distribution of observed wages for the two groups,

$$f(Y_g) \equiv f(Y|g) = \int f(Y|g,x)h(x|g)dx \tag{17}$$

where $f(Y|g)$ is the distribution of the wages for group g; $f(Y|g,x)$ is the wage distribution for group g given individual characteristics, $X = x$; and $h(x|g)$ is the distribution of individual characteristics for group g. The empirical counterpart to equation (17) can be given as:

$$\hat{f}(Y_g) = \sum_{j=1}^{n_g} K\left(\frac{Y_{g,i}-Y_{g,j}}{b_g}\right), \text{ for all } i = 1, \ldots, n_g \tag{18}$$

where $K(.)$ is the kernel function and n_g is the number of observations in group g. The actual distributions are estimated for both groups, let's say men and women.

DFL then propose an estimation of a counterfactual distribution for one of the groups, we show it for women, defined as:

$$f^C(Y_W) \equiv f^C(Y|M) = \int f(Y|M,x)h(x|W)dx = \int \omega(x)f(Y|M,x)h(x|M)dx \tag{19}$$

where $f^C(Y_W)$ is the counterfactual distribution of women, such that the distribution of individual characteristics is as that of women, but they are paid as men would be. The counterfactual distribution suggested by DFL is the distributional equivalent of counterfactual wage regression defined in equation (3). The counterfactual distribution is simply a reweighted distribution of men, where $\omega(x)$ is the reweighting function defined as $\omega(x) \equiv \frac{h(x|W)}{h(x|M)}$. To estimate the counterfactual distribution, we need an estimate of the reweighting function, which using the Bayes rule can be written as:

$$\omega(x) = \frac{\Pr(W|x)\,/\Pr(W)}{\Pr(M|x)\,/\Pr(M)} \tag{20}$$

To estimate equation (20), we pool the data and estimate a probit model, where the dependent variable is the binary gender variable, and the covariates are the individual characteristics, X. $\Pr(W|x)$ and $\Pr(M|x)$ are simply the predicted probabilities from the probit model, and $\Pr(W)$ and $\Pr(M)$ are the unconditional probabilities. Once we have the estimate of the reweighting function empirically the counterfactual distribution can be estimated as:

$$\hat{f}^C(Y_M) = \sum_{j=1}^{n_M} \hat{\omega}(x)K\left(\frac{Y_{M,i}-Y_{M,j}}{b_M}\right), \text{ for all } i = 1, \ldots n_M \tag{21}$$

Once we have estimates of the actual and the counterfactual distributions we can estimate the distributional composition effects ($\hat{\Delta}_X^f$) and the wage structure effects ($\hat{\Delta}_S^f$) as:

$$\hat{\Delta}_X^f = f(Y_M) - f^C(Y_W) \quad \text{and} \quad \hat{\Delta}_S^f = f^C(Y_W) - f(Y_W) \tag{22}$$

The DFL method is simple to implement, intuitive, and can decompose statistics other than the mean. The reweighting function proposed by them has been used in the context of many other decomposition methods, including the Recentered Influence Function approach discussed below. While aggregate decompositions are easy to do with the DFL method, detailed decompositions are still an issue. For binary variables detailed decomposition is feasible, and the authors discuss how to do them in their 1996 paper, however detailed decompositions for continuous covariates remain a challenge. See Butcher and DiNardo (2002) and Altonji et al. (2012) for some solutions to detailed decomposition within the DFL framework. Other methods, other than DFL, have been proposed in the literature to look at the decomposition of distribution, for one such method see Jenkins (1994) who takes the Generalised Lorenz Curve approach to estimate both the actual distributions and the counterfactual distributions.

3.3 Conditional Quantiles

While the DFL method allows us to do aggregate decomposition at the distributional level, detailed distribution beyond the mean remains a challenge. An alternative to looking at the entire distributions and doing detailed decomposition is proposed by Machado and Mata (2005), MM henceforth, who based their decomposition on conditional quantile regressions (Koenker and Bassett 1978). In this, instead of estimating a regression for the mean (OLS) we start by estimating a quantile regression for each group, men and women, given as:

$$Q_\tau\left(Y_g|X_g\right) = X_g\beta_g^\tau, \quad \tau \in (0,1) \tag{23}$$

where β_g^τ gives us the returns to the characteristics, X, on the τ^{th} quantile of the wage (Y) distribution. As in the OB method, we next estimate a counterfactual distribution for women if they had their own characteristics but are paid as men would be:

$$Q_\tau(Y^C|X_W) = X_W\beta_g^\tau \tag{24}$$

Once we have estimates of the actual and the counterfactual quantile regressions, we can estimate the composition effects ($\hat{\Delta}_X^\tau$) and the wage structure effects ($\hat{\Delta}_S^\tau$) at different quantiles as:

$$\hat{\Delta}_X^\tau = Q_\tau(Y_M|X_M) - Q_\tau(Y^C|X_W) \quad \text{and} \quad \hat{\Delta}_S^\tau = Q_\tau(Y^C|X_W) - Q_\tau(Y_W|X_W) \tag{25}$$

To do the quantile decomposition, MM suggest the following simulation:

1. Sample τ from a standard uniform distribution.
2. Estimate the quantile regression for the τ^{th} quantile, and obtain $\hat{\beta}_g^\tau$.
3. Compute $\hat{Q}_\tau(Y^C|X_W) = X_W\hat{\beta}_M^\tau$ and $\hat{Q}_\tau(Y_W|X_W) = X_W\hat{\beta}_M^\tau$. The difference between the two gives us the wage structure effect ($\hat{\Delta}_S^\tau$), composition effect is then computed as residual, $\hat{\Delta}_X^\tau = \left(\hat{Q}_\tau(Y_M|X_M) - \hat{Q}_\tau(Y_W|X_W)\right) - \hat{\Delta}_S^\tau$
4. Repeat steps 1 to 3 M times.

MM method is computationally very intensive, further, while it allows for the detailed decomposition of the wage structure effect, we cannot obtain the detailed decomposition of the composition effect. Melly (2005) provides a way to reduce the computation time and do a detailed decomposition of the composition effect at the median. Chernozhukov et al. (2013) provide a further extension of the MM method, giving detailed decomposition for both the wage structure and the composition effect. A key limitation of this method is that the detailed decompositions based on conditional quantile regressions are path dependent, the order in which the various covariates are considered in decomposition can alter their contribution to the explained and the unexplained components.

3.4 Recentered Influence functions

Firpo et al. (2007, 2009) proposed a way to look at the unconditional distributions and do both the aggregate and detailed decomposition, by using Recentered Influence Function (RIF) regressions. The RIF regression for the τ^{th} quantile, q_τ, of the wages, Y_g, for group g, is defined as:

$$RIF\left(Y_g, q_\tau\right) = q_\tau + \left[\tau - d_{g,\tau}\right]/f_{Y_g}(q_\tau), \ldots \tau \in (0,1) \tag{26}$$

where $f_{Y_g}(q_\tau)$ is the density function of Y_g computed at quantile q_τ, and $d_{g,\tau}$ is the dummy variable taking value one if $Y_g \leq q_\tau$ and zero otherwise. The $RIF\left(Y_g, q_\tau\right)$ has two properties that make it particularly useful; first, its expectation is the actual τ^{th}-quantile, $E_Y\left[RIF\left(Y_g, q_\tau\right)\right] = q_\tau$; and second, the expectation of the conditional RIF, when conditioning on the vector X_g, is also the actual τ^{th}-quantile, $E_X\left[E_y\left[RIF\left(Y_g, q_\tau\right)|X_g\right]\right] = q_\tau$. Assuming RIF to be a linear function of covariates we have,

$$RIF\left(Y_g, q_\tau\right) = \beta_{g0}^\tau + \Sigma_{k=1}^K X_{gik}\beta_{gk}^\tau + \upsilon_g^\tau \tag{27}$$

where β_g^τ is the vector of coefficients for the τ^{th}-quantile, and υ_g^τ is the error term. Given the two properties of the RIF function equation (27) is the unconditional quantile regression, which is estimated separately for the two groups, men and women.

The difference in the τ^{th}-quantile wage for men and women, $q_{\tau,M} - q_{\tau,W}$, can then be decomposed as follows:

$$\underbrace{q_{\tau,M} - q_{\tau,W}}_{\hat{\Delta}^\tau} = \underbrace{\left(\hat{\beta}_{M0}^\tau - \hat{\beta}_{W0}^\tau\right) + \sum\nolimits_k \bar{X}_{Wk}\left(\hat{\beta}_{Mk}^\tau - \hat{\beta}_{Wk}^\tau\right)}_{\hat{\Delta}_S^\tau} + \underbrace{\sum\nolimits_k (\bar{X}_{Mk} - \bar{X}_{Wk})\hat{\beta}_{Mk}^\tau}_{\hat{\Delta}_X^\tau} \tag{28}$$

On the left-hand side of equation (6), $\hat{\Delta}^\tau$, is the gap in the wages of men and women at the τ^{th}-quantile. The first term on the right-hand side, $\hat{\Delta}_S^\tau$, is the wage structure effect at the τ^{th}-quantile, and the second term $\hat{\Delta}_X^\tau$, is the composition effect at the τ^{th}-quantile. Like the OB decomposition, we can obtain both the aggregate decomposition and the detailed decomposition, which are path independent. When the RIF is evaluated at the mean of Y_g, we get OB decomposition as a special case.

The RIF regressions can be biased as the assumption of linearity holds true only locally. To correct for the specification error, the RIF regression is combined with the DFL reweighting function. This requires estimating the RIF regression for women with the reweighting function such that the covariates of women have the same distribution as of men. This yields the specification error and the reweighting error, separately from the composition and the wage structure effect, respectively. Empirically these errors tend to be very small.

4. Conclusion

Oaxaca and Blinder decomposition was first popularised in the 1970s, the use and popularity of this method have not waned over the last five decades. This chapter summarises the OB decomposition's main facets, its main limitations, cautions that should be exercised when using it, and the formal identification assumptions underlying it. We also link OB decomposition to the treatment effect literature. This chapter is not an exhaustive discussion of all methods of decomposition and the issues underlying them. For example, we have only focused on continuous outcomes, and not discussed the various methods that have been proposed for limited dependent variables (Fairlie 2005). Instead, this chapter's focus has been to introduce the regression based decompositions and highlight, up to date, key innovations in this method.

This main limitation of regression based methods, from the perspective of discrimination studies, is the interpretation of the unexplained gaps. On the one hand, not all of the unexplained gap can be discrimination, there often are unobservable factors at play making the unexplained gap an overestimation of discrimination. On the other hand, the estimated coefficients (return to endowments) from regressions already taken into account the feedback from the market, so any estimated discrimination is likely to be understated. The decomposition methodology has come a long way from the original work of Oaxaca and Blinder, increasingly we have methods that help us address the issues of unobserved productivity differences, self-selection, and omitted variables.

In recent years, in an attempt to separate discrimination from unobserved productivity differences, self-selection, and omitted variables, there has been substantial growth in the experimental literature on discrimination (for a review, see Neumark 2018). While the experimental methods help us identify discrimination more robustly and explore the underlying mechanisms, they have their limitations. Most of the experimental literature in labour economics focuses on hiring, which may not impact earnings, whereas the decomposition methods can look at earnings directly.

References

1. Althauser RP, Wigler W (1972) Standardisation and Component Analysis. *Sociological Methods and Research*, 1: 97-135.
2. Altonji JG, Bharadwaj P, Lange F (2012) Changes in the characteristics of American youth: Implications for adult outcomes. *Journal of Labor Economics*, 30(4): 783-828.
3. An W, Glynn AN (2019) Treatment effect deviation as an alternative to Blinder–Oaxaca decomposition for studying social inequality. *Sociological Methods & Research*, doi:10.1177/0049124119852387.
4. Arrow KJ (1998) What has economics to say about racial discrimination?. *Journal of Economic Perspectives* 12(2): 91-100.
5. Barsky R, Bound J, Charles KK, Lupton, JP (2002) Accounting for the black–white wealth gap: a nonparametric approach. *Journal of the American Statistical Association*, 97(459): 663-673.
6. Becker GS (1957) *The Economics of Discrimination*. Chicago and London: University of Chicago Press.
7. Blinder AS (1973) Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 8(4): 436-455.
8. Butcher KF, DiNardo J (2002) The immigrant and native-born wage distributions: evidence from United States censuses. *Industrial and Labour Relations Review*, 56(1): 97-121.
9. Chernozhukov V, Fernández-Val I, Melly B (2013) Inference on counterfactual distributions. *Econometrica*, 81(6): 2205-2268.
10. Cotton J (1988) On the decomposition of wage differentials. *The Review of Economics and Statistics*, 70: 236-243.
11. DiNardo J, Fortin N, Lemieux T (1996) Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5): 1001-1044.
12. Fairlie RW (2005) An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of economic and social measurement*, 30(4): 305-316.
13. Firpo S, Fortin N, Lemieux T (2007) Decomposing wage distributions using recentered influence function regressions. Working paper, University of British Columbia.
14. Firpo S, Fortin N, Lemieux T (2009) Unconditional quantile regressions. *Econometrica*, 77(3): 953-973.
15. Fortin N, Lemieux T, Firpo S (2011) Decomposition methods in economics. In Handbook of Labor Economics, vol. 4, pp. 1-102. Elsevier.
16. Gardeazabal J, Ugidos A (2004) More on identification in detailed wage decompositions. *Review of Economics and Statistics*, 86: 1034–1036.
17. Heckman J (1979) Sample selection bias as a specification error. *Econometrica*, 47: 153-162.
18. Jenkins SP (1994) Earnings Discrimination Measurement—A Distributional Approach. *Journal of Econometrics*, 61: 81–102.

19. Jones FL (1983) On decomposing the wage gap: A critical comment on Blinder's method. *Journal of Human Resources*, 18: 126–130.
20. Jones FL, Kelley J. (1984) Decomposing differences between groups: A cautionary note on measuring discrimination. *Sociological Methods & Research*, 12(3): 323-343.
21. Juhn C, Murphy KM, Pierce B (1993) Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101(3): 410-442.
22. Kitagawa E.M (1955) Components of a Difference between Two Rates. *Journal of American Statistical Association*, 50: 1168-1194.
23. Koenker R, Bassett G (1978) Regression Quantiles. *Econometrica*, 46: 33–50.
24. Lang K, Kahn-Lang Spitzer A (2020) Race discrimination: An economic perspective. *Journal of Economic Perspectives*, 34(2): 68-89.
25. Lemieux T. (2002) Decomposing changes in wage distributions: a unified approach. *Canadian Journal of Economics/Revue canadienne d'économique*, 35(4): 646-688.
26. Machado JA, Mata J (2005) Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20(4): 445-465.
27. Melly B (2005) Decomposition of differences in distribution using quantile regression. *Labour Economics*, 12(4): 577-590.
28. Neal DA, Johnson WR (1996) The role of premarket factors in black-white wage differences. *Journal of Political Economy*, 104(5): 869-895.
29. Neuman S, Oaxaca RL (2004) Wage decompositions with selectivity corrected wage equations: A methodological note. *Journal of Economic Inequality*, 2: 3-10.
30. Neumark D (1988) Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources*, 23: 279–295.
31. Neumark D (2018) Experimental research on labor market discrimination. *Journal of Economic Literature*, 56(3): 799-866.
32. Ñopo H (2008) Matching as a tool to decompose wage gaps. *The Review of Economics and Statistics*, 90(2): 290-299.
33. Oaxaca R (1973) Male-female wage differentials in urban labor markets. *International Economic Review*, 14: 693-709.
34. Oaxaca RL, Ransom MR (1994) On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61: 5–21.
35. Oaxaca RL, Ransom MR (1999) Identification in detailed wage decompositions. *Review of Economics and Statistics*, 81(1): 154-157.
36. Oaxaca RL (2007) The challenge of measuring labor market discrimination against women. *Swedish Economic Policy Review*, 14(1): 199-231.

37. Small ML, Pager D (2020) Sociological perspectives on racial discrimination. *Journal of Economic Perspectives*, 34(2): 49-67.
38. Yun MS (2005) A simple solution to the identification problem in detailed wage decompositions. *Economic Inquiry*, 43: 766–772.
39. Yun MS (2009) Wage differentials, discrimination and inequality: A cautionary note on the Juhn, Murphy and Pierce Decomposition Method. *Scottish Journal of Political Economy*, 56(1): 114-122.