

**Universidade de Évora**  
**Escola de Ciências Sociais**  
**Departamento de Economia**

**Resumo da Lição de Síntese**

**“Desigualdade e Discriminação: Metodologias para decomposição de diferenças salariais entre grupos”**

Apresentada para as provas públicas de Agregação em Economia

Maria Aurora Murcho Galego  
2023

## **Índice**

<b>I.</b>	<b>Enquadramento da Lição .....</b>	<b>2</b>
<b>II.</b>	<b>Objetivos e Competências a Adquirir.....</b>	<b>2</b>
<b>III.</b>	<b>Plano da Lição.....</b>	<b>3</b>
<b>IV.</b>	<b>Apresentação da lição: Síntese dos conteúdos abordados .....</b>	<b>4</b>
4.1.	Introdução: motivação e objetivos.....	4
4.2.	Decomposição na média: Decomposição de Oaxaca(1973) e Blinder(1973).....	5
4.2.1.	Forma de Cálculo e interpretação .....	5
4.2.2.	Pressupostos, principais problemas e limitações .....	6
4.3.	O Problema da invariância.....	7
4.4.	Decomposição detalhada: o problema da categoria de referência nas Variáveis binárias .....	8
4.5.	Para além da média: Decomposições distribucionais.....	9
4.5.1.	Decomposição de Juhn-Murphy-Pierce (1993).....	9
4.5.2.	Abordagem dos quantis.....	11
4.5.2.1.	Abordagem dos quantis condicionais – Machado & Mata (2005).....	11
4.5.2.2.	Abordagem dos quantis incondicionais – Fortin et al.(2011); Firpo et al. (2018).....	13
4.6.	O problema das características comuns .....	14
4.7.	Endogeneidade das profissões: Decomposição de Brown, Moon e Zoloth(1980)....	16
4.8.	Seletividade amostral .....	18
4.9.	Evidência empírica: caso das diferenças salariais de género em Portugal.....	20
4.10.	Considerações finais ... .....	23
<b>V.</b>	<b>Bibliografia.....</b>	<b>23</b>

## **I. Enquadramento da Lição**

A lição de síntese que seguidamente se apresenta, sob o tema “Desigualdade e discriminação: Metodologias para decomposição de diferenças salariais entre grupos”, tem como objetivo preencher um dos requisitos exigidos pela legislação em vigor para prestação de provas públicas de Agregação. De acordo com o estipulado no Decreto/Lei nº 239/2007 de 19 de Junho, nas referidas provas deve ser apresentado um sumário pormenorizado do seminário ou lição sobre um tema, dentro do âmbito do ramo do conhecimento ou especialidade em que são prestadas as provas.

A presente lição enquadra-se no âmbito de um dos temas a abordar na unidade curricular de Economia do Trabalho do programa de doutoramento em Economia. Na verdade, as questões da desigualdade e da discriminação são aspectos fundamentais analisados na área da Economia do Trabalho, sendo, portanto, de extrema importância o conhecimento das estratégias empíricas de análise desta problemática.

## **II. Objetivos e competências a adquirir**

Nesta lição, pretende-se apresentar as decomposições mais usadas na literatura para analisar a importância das diferenças entre grupos de indivíduos em termos de salários (ou rendimentos) e as suas causas, chamando a atenção para as vantagens e limitações destas metodologias. Em particular, é analisada em detalhe a metodologia mais usada nas últimas décadas— Decomposição de Oaxaca (1973) e Blinder (1973), suas hipóteses e problemas. Esta decomposição tem sido examinada num número relativamente elevado de artigos metodológicos e extensões desta decomposição têm sido apresentadas ao longo dos anos. O propósito desta lição é, para além da decomposição original, apresentar as principais soluções encontradas para ultrapassar as limitações da mesma, incluindo alguns dos desenvolvimentos mais recentes. Finalmente, pretende-se também ilustrar aplicações destas decomposições com uma análise dos resultados para as diferenças salariais de género em Portugal.

No final da lição os estudantes deverão:

- Perceber como calcular a decomposição de diferenças salariais (ou de rendimento) na média, nomeadamente a decomposição de Oaxaca e Blinder (OB) e como interpretar os resultados.
- Conhecer as hipóteses, limitações e problemas desta abordagem.
- Conhecer os principais desenvolvimentos posteriores propostos pela literatura para ultrapassar os problemas da decomposição de OB original.
- Compreender a intuição das decomposições distribucionais (para além da média), suas vantagens e problemas.
- Tomar consciência das limitações das decomposições em geral
- Conhecer alguma literatura empírica em que as metodologias apresentadas são aplicadas, nomeadamente no que diz respeito a Portugal.

### **III. Plano da lição**

1. Introdução: motivação e objetivos
2. Decomposição na média: Decomposição de Oaxaca(1973) e Blinder(1973)
  - 2.1. Forma de cálculo e interpretação
  - 2.2. Pressupostos, principais problemas e limitações
3. O Problema da invariância
4. Decomposição detalhada: o problema da categoria de referência nas variáveis binárias
5. Para além da média: Decomposições distribucionais
  - 5.1. Decomposição de Juhn-Murphy-Pierce (1993)
  - 5.2. Abordagem dos quantis
    - 5.2.1. Abordagem dos quantis condicionais – Machado & Mata (2005)
    - 5.2.2. Abordagem dos quantis incondicionais – Fortin et al.(2011); Firpo et al. (2018)
6. O problema das características comuns
7. Endogeneidade das profissões: Decomposição de Brown, Moon & Zoloth(1980)
8. Seletividade amostral
9. Evidência empírica: caso das diferenças salariais de género em Portugal
10. Considerações finais

## IV. Apresentação da Lição: Síntese dos conteúdos abordados

### 4.1. Introdução: Motivação e objetivos

Compreender as causas das diferenças de salários (ou de rendimento) entre os indivíduos ( ou entre países ou regiões diferentes, entre períodos de tempo diferentes, ou entre grupos diferentes do mesmo país ou região) é fundamental em vários ramos da investigação em economia, mas particularmente na área da economia do desenvolvimento ou da economia do trabalho. Um dos aspectos mais analisados no âmbito da economia do trabalho são as diferenças salariais e desigualdade entre diversos grupos de indivíduos: homens/mulheres, nacionais/imigrantes, trabalhadores no sector público/trabalhadores no setor privado, etc. Nessa análise, procura-se medir o montante dessa diferença e identificar as suas causas. Em muitos estudos, uma questão importante é se essas diferenças são ou não resultado de discriminação, colocando-se o problema de como identificar e medir a discriminação.

Entre as metodologias mais usadas para avaliar as causas das diferenças salariais estão precisamente as decomposições salariais, baseadas em regressões de equações de salários. Como tal, nesta aula apresentam-se as decomposições mais usadas na literatura, chamando a atenção para as suas vantagens e limitações. Na verdade, dada a proliferação de metodologias que têm surgido na literatura existem aspectos de ordem prática que se levantam, nomeadamente sobre as vantagens e desvantagens de cada uma delas.

As decomposições permitem identificar as causas das diferenças salariais ou de rendimentos entre dois grupos, separando essas diferenças em duas partes:

- ➔ A parte das diferenças salariais que são devidas a diferenças nas características observadas dos dois grupos em fatores relacionados com o capital humano, nomeadamente devido a diferenças ao nível da educação, formação profissional e experiência assim como devido a características da profissão e do local de trabalho, entre outras. Ou seja, permite saber qual a parte da diferença que é devida a diferenças de produtividade e que é conhecida como o **efeito de composição (ou efeito dotação)**.
- ➔ A parte da diferença salarial que é explicada por diferenças na taxa de rendibilidade (ou prémio salarial) das características observadas e não observadas dos indivíduos que é conhecida como o **efeito estrutural (ou efeito remuneração)**. Normalmente, entende-se que a existência de um efeito estrutural significativo pode ser indicativo de discriminação.

Ao permitirem identificar não só o montante das diferenças de rendimento existentes entre grupos mas, igualmente, as suas causas, nomeadamente em medir a potencial existência de discriminação no mercado de trabalho, os métodos de decomposição podem fornecer informação pertinente para a definição de políticas no sentido de, por exemplo, reduzir a desigualdade.

## 4.2. Decomposição na média: Decomposição de Oaxaca(1973) e Blinder (1973)

### 4.2.1. Forma de cálculo e interpretação

No sentido de tornar a explicação das metodologias de decomposição mais simples vamos considerar como exemplo as diferenças salariais de género, que talvez seja a situação que mais atenção tem recebido por parte da literatura.

Na decomposição de Oaxaca(1973) e Blinder(1973), doravante representada por OB, começamos por estimar separadamente as equações salariais (hedónicas) para os dois grupos que queremos comparar, neste caso homens ( $h$ ) e mulheres ( $m$ ):

$$\begin{aligned}\ln W_h &= X_h \beta_h + \varepsilon_h \\ \ln W_m &= X_m \beta_m + \varepsilon_m\end{aligned}$$

em que  $W$  será o salário horário,  $X$  o vetor das características pessoais e das características do emprego,  $\varepsilon$  é o erro aleatório e o  $\beta$  o vetor dos coeficientes. Estima-se seguidamente a média do logaritmo dos salários horários de ambos os grupos com os valores médios de  $X$ , assim como a sua diferença:

$$\overline{\ln W_h} - \overline{\ln W_m} = \bar{X}_h \hat{\beta}_h - \bar{X}_m \hat{\beta}_m$$

No sentido de identificar a parte das diferenças relacionadas com diferenças ao nível das características e ao nível dos prémios salariais das mesmas, há que definir e calcular o chamado salário contrafactual. Este é o salário médio previsto para as mulheres na ausência de discriminação, ou seja, o que estas receberiam mantivessem a sua distribuição de características se fossem remuneradas de acordo com uma estrutura salarial não discriminatória. Considerando que a estrutura salarial não discriminatória é a dos homens podemos calcular esse salário contrafactual como :  $\bar{X}_m \hat{\beta}_h$

Somando o seguinte termo ao lado direito da equação anterior:

$$-\bar{X}_m \hat{\beta}_h + \bar{X}_m \hat{\beta}_h$$

E rearranjando temos:

$$\overline{\ln W_h} - \overline{\ln W_m} = \underbrace{(\bar{X}_h - \bar{X}_m) \hat{\beta}_h}_{(A)} + \underbrace{\bar{X}_m (\hat{\beta}_h - \hat{\beta}_m)}_{(B)} \quad (1)$$

A diferença salarial entre homens e mulheres é assim dividida em duas partes:

- (A) Diferenças nas características X – efeito composição (ou de dotação)
- (B) Diferenças nos retornos das características- efeito estrutural (ou remuneração), que normalmente se considera como um indicativo de discriminação.

#### 4.2.2. Pressupostos, principais problemas e limitações

Embora a decomposição OB seja simples e forneça informação relevante, tem importantes limitações. As mais importantes têm a ver com as suas hipóteses no sentido de podermos ter uma decomposição válida e de podermos identificar a parte não explicada da decomposição como discriminação.

##### Pressupostos da decomposição Oaxaca-Blinder

1. As diferenças nos valores observados de capital humano ( $X_s$ ) não são afetados por discriminação, ou seja, as variáveis explicativas incluídas representam todas as variáveis relevantes e nada têm a ver com discriminação.
2. Os grupos potencialmente discriminados não preferem trabalhar em empregos que são mal pagos.
3. Os grupos devem ser mutuamente exclusivos;
4. Suporte comum (características comuns): cada indivíduo no grupo A tem um correspondente no grupo B.

Existem igualmente um outro conjunto de problemas e limitações:

##### Problemas e limitações adicionais

- a. O problema da invariância->relacionado com a escolha da contrafactual
- b. Decomposição detalhada: o problema da categoria de referência nas variáveis binárias
- c. A decomposição analisa as diferenças na média e apenas num só ponto do tempo
- d. Endogeneidade
- e. Seletividade

➔ Iremos ver em que consistem as maiores limitações e problemas desta metodologia e as respostas (soluções) encontradas na literatura para as resolver.

### 4.3. O Problema da invariância

A decomposição OB não é única. Na equação anterior (1) toma-se como referência a equação salarial dos homens (estrutura salarial não discriminatória) o que pode ser um considerado um pouco arbitrário. Na verdade, tanto a estrutura salarial das mulheres como a dos homens podem ser usadas e consideradas como estruturas salariais não discriminatórias.

Considerando a estrutura salarial das mulheres como não discriminatória, podíamos escrever:

$$\overline{\ln W_h} - \overline{\ln W_m} = (\bar{X}_h - \bar{X}_m) \hat{\beta}_m + \bar{X}_h (\hat{\beta}_h - \hat{\beta}_m) \quad (2)$$

O que dá normalmente uma resposta diferente da que se obtém na equação (1). Em (2) a parte não explicada reflete o que os homens auferem acima do salário não discriminatório, enquanto que na equação (1) a parte não explicada reflete o que as mulheres recebem abaixo do salário não discriminatório.

Embora a maioria dos estudos que analisa as diferenças entre géneros, assuma que a estrutura salarial não discriminatória é a dos homens, a controvérsia existe e a literatura acabou por se centrar nesta questão. A razão para a controvérsia surgiu da ideia de que a discriminação não só pode diminuir o salário das mulheres relativamente aos homens como também pode implicar aumentos de salário nos homens.

A decomposição apropriada depende do tipo de discriminação que pode ocorrer. Em particular, os empregadores podem praticar nepotismo relativamente aos homens, discriminação relativamente às mulheres ou praticar os dois. Na hipótese de nepotismo, as mulheres são pagas de acordo com o salário competitivo mas os homens são pagos um salário superior. Numa situação de discriminação os empregadores pagam aos homens o salário competitivo mas pagam as mulheres um salário inferior.

No sentido de dar uma resposta a esta questão vários autores propuseram soluções. As equações (1) e (2) podem ser escritas numa forma geral, representando por  $\beta^*$  a estrutura salarial não discriminatória (Oaxaca & Ransom ,1988, 1994):

$$\overline{\ln W_h} - \overline{\ln W_m} = (\bar{X}_h - \bar{X}_m) \hat{\beta}^* + \bar{X}_h (\hat{\beta}_h - \hat{\beta}^*) + \bar{X}_m (\hat{\beta}^* - \hat{\beta}_m) \quad (3)$$

Em que o primeiro termo mede o diferencial de produtividade, o segundo a vantagem dos homens (nepotismo) e o terceiro a desvantagem das mulheres (discriminação).

Como tal, qualquer das outras equações são um caso particular desta equação.

A questão importante diz respeito à definição de  $\hat{\beta}^*$  na equação anterior. Várias alternativas têm sido propostas:

Reimers(1983)	$\hat{\beta}^* = 0,5 * \hat{\beta}_h + 0,5 * \hat{\beta}_m$	Média simples dos coeficientes de homens e mulheres
Cotton (1988)	$\hat{\beta}^* = \left( \frac{n \text{ homens}}{N} \right) * \hat{\beta}_h + \left( \frac{n \text{ mulheres}}{N} \right) * \hat{\beta}_m$	Média ponderada dos coeficientes de homens e mulheres
Oaxaca & Ransom (1988,1994) Neumark(1988)	$\hat{\beta}^* = \hat{\beta}_a$	Os coeficientes resultam de uma regressão com dados agrupados (homens+mulheres)
Fortin (2008) Jann (2008)	$\hat{\beta}^* = \hat{\beta}_x$	Os coeficientes resultam de uma regressão com dados agrupados da forma: $Y_i = \alpha^P + \beta_d^P d_i + X_i' \beta^P + \varepsilon_i^P$ $d_i$ é uma dummy, =1 se o indivíduo é homem.

A decomposição de Neumark/Oaxaca-Ransom é atrativa e tem sido a mais usada. Mas há que ter cuidado com a interpretação destas decomposições. Não é claro se os coeficientes resultantes são de facto um bom estimador para a estrutura salarial não discriminatória. A ponderação é claramente muito difícil de definir de forma precisa e uma grande área de arbitrariedade existe.

#### 4.4. Decomposição detalhada: O problema da categoria de referência nas variáveis binárias

A estrutura linear da decomposição permite decompor os efeitos estrutural e de composição por variável e assim identificar mais claramente as causas das diferenças salariais. Contudo, existem alguns problemas nessa decomposição detalhada, nomeadamente no efeito estrutural. Entre as variáveis explicativas existe normalmente um número elevado de variáveis binárias (0/1). Ao calcular a contribuição separada das variáveis binárias, há um problema de identificação. Este problema, que foi inicialmente identificado por Jones(1983) e posteriormente confirmado por Oaxaca & Ransom (1999), consiste no facto da constante e dos coeficientes das variáveis binárias serem influenciadas pela escolha do grupo de referência.

Este problema pode ser ultrapassado aplicando normalizações propostas por vários autores.

- Nielson (2000) propôs que o valor 1 e 0 das variáveis binárias fosse substituído pela proporção das observações de cada grupo
- Yun (2005) propôs uma abordagem da média, baseada na transformação dos coeficientes das variáveis binárias de forma que eles possam refletir diferenças em relação à média e não diferenças em relação ao grupo de referência
- Kim(2013) sugere usar um método de média ponderada a que ele chama de grande-média: calcula-se a média da média para os dois grupos de indivíduos em análise.

Contudo as normalizações tornam mais difíceis a interpretação dos resultados e a comparação dos resultados de diferentes estudos. Na verdade, algum grau de arbitrariedade na derivação da equação normalizada é inevitável.

#### **4.5. Decomposição para além da média: decomposições distribucionais**

A decomposição OB considera apenas a média da distribuição. Analisar as diferenças ao longo da distribuição permite melhorar o conhecimento das diferenças salariais e logo construir medidas de política mais adequadas. Para além disso, a decomposição OB não permite relacionar a discriminação com a desigualdade dos salários ou dos rendimentos (medida por exemplo pela variância, coeficiente GINI, etc)

##### **4.5.1. Decomposição de Juhn-Murphy-Pierce (1993)**

A decomposição de Juhn-Murphy-Pierce (1993), que representaremos por JMP fornece uma forma de ter em consideração as características não observadas explicitamente na decomposição através de um método de imputação de resíduos, sendo esta a sua principal contribuição.

Assim sendo, na decomposição de JMP, as diferenças salariais entre homens e mulheres são divididas em três partes: uma parte explicada por diferenças nas características observáveis, uma parte explicada por diferenças nos coeficientes (efeito preço) e uma parte atribuída a fatores não observados (efeito residual).

Tendo como base as equações salariais:

$$\ln W_h = X_h \beta_h + \varepsilon_h$$

$$\ln W_m = X_m \beta_m + \varepsilon_m$$

De acordo com Juhn et al. (1993), os resíduos  $\varepsilon_h$  e  $\varepsilon_m$  podem ser definidos como:

$$\varepsilon_g = F^{-1}(\theta_g | X_g) \equiv F_g^{-1}, \text{ com } g=h \text{ e } m.$$

Sendo  $F_g$  a função de distribuição cumulativa dos resíduos da equação dos salários, condicionada nas variáveis.

Usando esta definição dos resíduos as equações salariais podem ser escritas:

$$\ln W_h = X_h \beta_h + F_h^{-1}$$

$$\ln W_m = X_m \beta_m + F_m^{-1}$$

Como anteriormente, vamos assumir a estrutura salarial dos homens como não discriminatória e para obter a decomposição teremos que derivar duas contrafactuals.

$$\ln W^{C1} = X_m \beta_h + F_h^{-1}$$

$$\ln W^{C2} = X_m \beta_m + F_h^{-1}$$

A primeira contrafactual dá-nos o salário que as mulheres receberiam se mantivessem as suas características observadas mas fossem remuneradas de acordo com a estrutura salarial dos homens e tivessem a mesma distribuição residual dos homens. A segunda contrafactual dá-nos o salário que as mulheres receberiam se estas mantivessem as mesmas características observadas e respetivos prémios salariais mas tivessem uma distribuição dos resíduos igual à dos homens.

Usando estas duas contrafactuals podemos escrever a decomposição como:

$$\overline{\ln W_h} - \overline{\ln W_m} = \underbrace{(\bar{X}_h - \bar{X}_m)\beta_h}_{\begin{array}{c} \text{diferença nas características} \\ \text{observadas} \end{array}} + \underbrace{\bar{X}_m(\beta_h - \beta_m)}_{\begin{array}{c} \text{diferença nos preços das} \\ \text{características observadas} \end{array}} + \underbrace{(F_h^{-1} - F_m^{-1})}_{\begin{array}{c} \text{diferença em não observáveis} \\ \text{Efeito não explicado} \end{array}} \quad (4)$$

A diferença salarial entre os dois grupos é agora o resultado de diferenças em características entre homens e mulheres, como na formulação de Oaxaca, mas ao contrário dessa (que considera a totalidade da componente não explicada como discriminação), JMP consideram que parte da

componente não explicada se deve a alterações nas características não observadas e respetivos prémios salariais, ou seja corresponde a desigualdade salarial residual.

Uma característica importante da abordagem do JMP é que esta equação pode ser usada para estimar diferenças em diferentes pontos da distribuição e consequentemente este método permite a análise em diferentes quantis, ou seja:

$$\overline{\ln W_h^q} - \overline{\ln W_m^q} = (\bar{X}_h^q - \bar{X}_m^q)\beta_h + \bar{X}_m^q(\beta_h - \beta_m) + (\bar{F}_h^{-1})^q - (\bar{F}_m^{-1})^q \quad (5)$$

Onde  $\bar{X}_h^q$  e  $\bar{X}_m^q$  são as médias das características de homens e mulheres em cada quantil  $q$ .

Contudo, a estimação continua a ser com base no método dos mínimos quadrados e logo é um modelo para a média condicional da distribuição salarial. Como tal, os seus resultados não se estendem naturalmente para os quantis salariais. Sendo que as estimativas ainda são feitas na média, implica que as características têm a mesma remuneração ao longo da distribuição, ou seja, os coeficientes são os mesmos ao longo da distribuição.

A abordagem de JMP tem várias outras limitações e problemas identificados na literatura (veja-se Lemieux, 2002; Suen, 1997 ou Yun, 2009).

#### **4.5.2. Abordagem dos quantis**

Trabalhos mais recentes têm generalizado a abordagem OB para ter em consideração outros momentos da distribuição na comparação entre grupos com base na regressão por quantis.

##### **4.5.2.1. Abordagem dos quantis condicionais – Machado & Mata (2005)**

A decomposição sugerida por Machado & Mata (2005) permite decompor diferenças salariais em pontos diferentes da distribuição, numa componente baseada nas diferenças nas características observadas e outra baseada nas diferenças nos retornos para essas características.

Assumindo uma relação linear entre o  $\ln(W)$  - y - e as variáveis explicativas- x-, o quantil da distribuição condicional de y dado x é:

$$Q_\tau(y_i | x_i) = x_i \beta_\tau, \tau \in (0,1)$$

A obtenção da decomposição com a regressão de quantis é mais complicada do que com a regressão linear, porque a diferença incondicional de rendimentos não é igual à diferença condicional. Dessa forma, os quantis condicionais têm que ser relacionados com as diferenças incondicionais e isso requer métodos de simulação. Machado & Mata (2005) deduzem um estimador da distribuição salarial não condicional contrafactual com base nos coeficientes estimados das regressões de quantis condicionais, recorrendo para tal ao método de simulação *bootstrap*.

Após a estimação da densidade salarial para os homens e mulheres- $Q_\tau(x^h, \beta^h(\tau))$  e  $Q_\tau(x^m, \beta^m(\tau))$  - e da densidade salarial contrafactual:  $Q_\tau(x^m, \beta^h(\tau))$ , podemos obter a decomposição para cada quantil:

$$\begin{aligned} \underbrace{Q_\tau(x^h, \beta^h(\tau)) - Q_\tau(x^m, \beta^m(\tau))}_{\substack{\text{Diferença salarial estimada} \\ \text{no quantil } \tau}} &= \underbrace{Q_\tau(x^h, \beta^h(\tau)) - Q_\tau(x^m, \beta^h(\tau))}_{\substack{\text{diferenças nas características}}} + \\ &+ \underbrace{Q_\tau(x^m, \beta^h(\tau)) - Q_\tau(x^m, \beta^m(\tau))}_{\substack{\text{diferenças nos retornos das características}}} \end{aligned} \quad (6)$$

Este estimador é muito exigente em termos de cálculo já que combina a regressão de quantis com o *bootstrapping*. Contudo, Melly(2005, 2006), propôs um estimador assintoticamente equivalente mais simples e que tem sido bastante usado na literatura.

✓ **Limitações da abordagem da regressão de quantis condicionais:**

Apesar desta abordagem fornecer uma forma de fazer uma decomposição detalhada através de simulações, o valor dos coeficientes estimados depende da sequência da simulação. A ideia é obter o valor de um coeficiente de uma determinada variável enquanto se mantém as outras variáveis constantes. Logo, nesta abordagem não é possível isolar o efeito de cada variável em qualquer das componentes da decomposição de forma fiável.

#### 4.5.2.2. Abordagem de regressão de quantis incondicionais - Fortin et al.(2011), Firpo et al. (2018)

Na metodologia clássica de Oaxaca (1973) e Blinder (1973) é possível dividir o efeito composição e estrutural na contribuição de cada regressor. Contudo, este procedimento não pode ser diretamente estendido a qualquer outra estatística distribucional diferente da média (quantis, variância, índice de Gini, etc.), já que a propriedade de linearidade não se verifica (Fortin et al., 2011).

Fortin et al. (2011) and Firpo et al. (2018) apresentaram uma solução propondo uma extensão da decomposição OB com base em regressões recentradas na função de influência, denominadas como *RIF* (do inglês recentered influence function ( propostas por Firpo et al. ,2009)). Contudo, sendo que a esperança condicional dos salários pode não ser linear (Firpo et al. ,2018; Rothe,2015), o impacto de alterações na distribuição das variáveis nos quantis pode não ser corretamente aproximado pelas regressões RIF e logo existem erros de aproximação.

Para avaliar a importância desses erros de aproximação, Fortin et al. (2011) e Firpo et al.(2018) propõem um procedimento em dois passos. Num primeiro passo, realiza-se a decomposição da diferença salarial entre os dois grupos no efeito estrutural e no efeito composição, usando uma abordagem de ponderação semelhante à de Dinardo et al. (1996). Para tal, terá que ser calculado um fator de ponderação que será aplicado aos dados do grupo dos homens para calcular a cuntrafactual da distribuição salarial ( $Q^\tau(\ln w_c)$ ).

E obtém-se a decomposição salarial para um qualquer quantil incondicional:

$$\hat{\Delta}_o^\tau = \left[ \left[ Q^\tau(\ln w_h) - Q^\tau(\ln w_c) \right] + \left[ Q^\tau(\ln w_c) - Q^\tau(\ln w_m) \right] \right] \quad (7)$$

$$\hat{\Delta}_o^\tau = \hat{\Delta}_S^\tau + \hat{\Delta}_X^\tau$$

Em que  $\hat{\Delta}_X^\tau$  representa o efeito composição e  $\hat{\Delta}_S^\tau$  o efeito estrutural.

No segundo passo, a contribuição de cada variável ou grupo de variáveis para os dois efeitos é calculada, tal como na decomposição OB normal, usando as regressões *RIF*.

O efeito estrutural pode ser decomposto de acordo com a expressão seguinte:

$$\begin{aligned}\hat{\Delta}_S^\tau &= \bar{X}_m \left( \hat{\beta}_{\tau,m} - \hat{\beta}_{\tau,h}^C \right) + \left( \bar{X}_m - \bar{X}_h^C \right) \hat{\beta}_{\tau,h}^C \\ &= \hat{\Delta}_{S,p}^\tau + \hat{\Delta}_{S,e}^\tau\end{aligned}\tag{8}$$

$\hat{\Delta}_{S,p}^\tau$  é o puro efeito estrutural no quantil  $\tau$  e  $\hat{\Delta}_{S,e}^\tau$  é o erro de reponderação.

De forma semelhante, o efeito de composição pode ser expresso da seguinte forma:

$$\begin{aligned}\hat{\Delta}_X^\tau &= \left( \bar{X}_h^C - \bar{X}_h \right) \hat{\beta}_{\tau,h} + \bar{X}_h^C \left( \hat{\beta}_{\tau,h}^C - \hat{\beta}_{\tau,h} \right) \\ &= \hat{\Delta}_{X,p}^\tau + \hat{\Delta}_{X,e}^\tau\end{aligned}\tag{9}$$

$\hat{\Delta}_{X,p}^\tau$  é o puro efeito composição no quantil  $\tau$  e  $\hat{\Delta}_{X,e}^\tau$  é o erro de especificação, que deverá ser zero no caso do modelo ser linear.

#### 4.6. O problema das características comuns

A decomposição OB assume que cada indivíduo no grupo A tem um correspondente no grupo B. Este aspeto é relevante pois existem combinações de características individuais para as quais é possível encontrar indivíduos num dos grupos e não no outro.

Logo, a decomposição OB (e suas variações) não tem em consideração a relação entre a “abundância” ou “escassez” de determinadas características e a forma como elas são remuneradas em função disso. Em particular, se algumas características são relativamente mais abundantes para um grupo do que para outro (Ex: mais para homens do que para mulheres) a sua remuneração possivelmente refletirá a sua abundância, para além da avaliação de mercado.

De salientar que a inclusão de variáveis binárias para as profissões e para os setores industriais não ajuda a ultrapassar o problema, numa situação em que se verifica a indisponibilidade de salários para certas profissões em que um dos grupos não está representado. Este problema, normalmente conhecido como “problema do suporte comum” – é negligenciado por todas as decomposições clássicas. Se homens e mulheres não são estritamente comparáveis ou se as comparações não fazem sentido, então os resultados podem ser enganadores.

No sentido de ultrapassar este problema Black et al. (2008) e Nopo (2008) enfatizam a comparabilidade entre homens e mulheres (entre os dois grupos) antes do cálculo da diferença

de salário ajustada. Ambos os autores usam uma abordagem não paramétrica através de um estimador *matching*.

Estas técnicas assumem que a função de distribuição salarial pode ser dividida em duas partes: uma onde as características dos membros dos dois grupos coincidem (características comuns – “common support”) e outra que representa a variação (ou diferença) de cada grupo relativamente às características comuns.

Black et al (2008) usa o *matching* exato (com base nas características observadas) e as observações para as quais não há *matching* para ambos homens e mulheres não são usadas. Contudo, eliminar os indivíduos para os quais não existe *matching* retira informação importante da amostra.

No sentido de evitar isto Nopo(2008) desenvolveu uma decomposição que permite também identificar a parte da diferença salarial que é atribuída às características dos homens e às características das mulheres para os quais não existe *matching*.

Nopo(2008) obtém a distribuição contrafactual para as mulheres escolhendo uma mulher de cada vez e combinando-a com todos os homens que são estatisticamente iguais. Desta forma é construída uma observação salarial contrafactual que iguala o salário médio dos homens combinados.

Nopo (2008) constrói assim a seguinte decomposição da diferença salarial entre homens e mulheres:

$$\Delta = E[Y | M] - E[Y | F] = (\Delta H + \Delta X + \Delta M) + \Delta O \quad (10)$$

Onde  $\Delta O$  é a parte da diferença salarial devida a diferenças nas características não observadas para indivíduos com suporte comum. A parte devida a diferenças nas características tem três componentes.  $\Delta X$  é a parte devida a diferenças nas características observadas para indivíduos com suporte comum.  $\Delta H$  é a parte da diferença salarial explicada pelas diferenças entre os homens que estão fora do suporte comum e os que estão no suporte comum (calculada pela diferença entre os salários esperados dos dois grupos, ponderada pela medida de probabilidade de não existência de suporte comum).

Finalmente  $\Delta M$  é a parte da diferença salarial entre géneros que pode ser explicada pelas diferenças das características entre as mulheres que têm com suporte comum e as que estão fora do suporte comum. Mais uma vez, é calculada pela diferença nos salários esperados ponderada pela medida da probabilidade de não existência de suporte comum.

#### ✓ Limitações desta metodologia

Uma limitação refere-se à metodologia usada, dado que emprega a técnica do *matching* e a mesma implica um *trade-off* entre o número de características para os quais se deve controlar o *matching* e o rácio entre as observações combinadas e não combinadas para homens e mulheres. Na verdade, quanto mais variáveis são usadas na correspondência entre homens e mulheres menor probabilidade existe em encontrar uma combinação exata. Sendo assim, um maior número de características usadas implica uma estimativa mais confiável do diferencial salarial ajustado, mas menor será a fração da amostra, o que limita a validade do resultado. Este problema é referido como a “*dimensionality curse*”.

### 4.7. Endogeneidade das profissões: A decomposição de Brown, Moon & Zoloth(1980)

Todas as decomposições apresentadas até agora, sejam as clássicas sejam as distribucionais, assumem que a distribuição das profissões por género é exógena. Contudo, a distribuição das profissões por género pode ser endógena já que pode ser um reflexo da própria discriminação existente. Este argumento foi apresentado por Brown et al. (1980) sugerindo uma abordagem que possa examinar a sensibilidade da decomposição da desigualdade salarial à estrutura da segregação profissional. O aspeto inovador destes autores é o facto de decompor a diferença salarial ao longo de toda a distribuição das profissões o que permite a endogeneidade da distribuição das mulheres pelas diferentes profissões (eles assumem que a distribuição relativa aos homens é exógena).

Brown et al. (1980) formularam uma abordagem empírica em que diferenças no capital humano e de produtividade explicam simultaneamente a variação nos salários e a ocupação profissional de homens e mulheres. Sendo assim, nesta decomposição a diferença de salário entre homens e mulheres é calculada através da diferença entre as médias dos logaritmos dos salários horários ponderados pelas respetivas  $K$  profissões.

$$\overline{\ln W^h} - \overline{\ln W^m} = \sum_{j=1}^k P_j^h \overline{\ln W_j^h} - \sum_{j=1}^k P_j^m \overline{\ln W_j^m}$$

Onde  $P_j^h$  e  $P_j^m$  representam as proporções de homens e mulheres nas profissões  $j$ , respectivamente e  $\overline{\ln W_j^h}$  e  $\overline{\ln W_j^m}$  são as médias no logaritmo dos salários horários para homens e mulheres na profissão  $j$ .

Rearranjando os termos, obtemos:

$$\overline{\ln W^h} - \overline{\ln W^m} = \underbrace{\sum_{j=1}^k P_j^m (\overline{\ln W_j^h} - \overline{\ln W_j^m})}_{\text{efeito intra-profissões}} + \underbrace{\sum_{j=1}^k (P_j^h - P_j^m) \overline{\ln W_j^h}}_{\text{efeito inter-profissões}} \quad (11)$$

A primeira parte do lado direito da equação mede o que é causado pelas diferenças no salário médio entre homens e mulheres dentro de cada profissão. A segunda parte da equação indica a parte que é devida a diferenças na estrutura profissional entre homens e mulheres. No caso de não segregação a proporção de homens e mulheres em cada profissão seria igual e este segundo termo desapareceria.

Esta equação poderá ainda ser decomposta numa componente explicada e noutra não explicada, tal como no caso da decomposição de Oaxaca-Blinder. Cada um dos termos do efeito composição e do efeito estrutural inclui um efeito inter-profissões e um efeito intra-profissões.

Os autores propõem usar um logit multinomial para produzir a distribuição contrafactual das profissões para as mulheres, ou seja a proporção de mulheres em cada profissão que aconteceria se a distribuição das mulheres pelas profissões fosse igual á dos homens. Estima-se a probabilidade de um homem trabalhar na profissão  $j$  em função das suas características e os coeficientes deste modelo estimado são combinados com as características das mulheres para simular a distribuição por profissões não discriminatória para as mulheres.

Seguidamente, podemos construir uma decomposição do diferencial salarial entre géneros não só em diferenças nos coeficientes e em diferenças nas características médias de homens e mulheres, mas também em diferenças das partes explicada e não explicada derivadas da distribuição das profissões para homens e mulheres:

$$\begin{aligned}
\ln \overline{W}^h - \ln \overline{W}^m &= \underbrace{\sum_{j=1}^k P_j^m (\overline{X}_j^h - \overline{X}_j^m) \hat{\beta}_j^h}_{\text{efeito composição}} + \underbrace{\sum_{j=1}^k (P_j^h - \hat{P}_j^m) \overline{X}_j^h \hat{\beta}_j^h}_{\text{efeito inter-profissões}} \\
&\quad + \underbrace{\sum_{j=1}^k P_j^m (\hat{\beta}_j^h - \hat{\beta}_j^m) \overline{X}_j^m}_{\text{efeito estrutural}} + \underbrace{\sum_{j=1}^k (\hat{P}_j^m - P_j^m) \overline{X}_j^h \hat{\beta}_j^h}_{\text{efeito intra-profissões}}
\end{aligned} \tag{12}$$

Onde  $\hat{P}_j^m$  representa a estrutura profissional não discriminatória para as mulheres,  $\overline{X}_j$  é a média das características para a profissão  $j$ , e  $\hat{\beta}_j$  são os coeficientes estimados da função salarial para a profissão  $j$ .

#### 4.8. Seletividade amostral

As estimativas das equações salariais sofrem muitas vezes de seletividade. Este enviesamento surge quando fatores não observados que influenciam a probabilidade de participação no mercado de trabalho estão correlacionados com fatores não observados que afetam o salário. Consequentemente, este problema poderá também enviesar a medida da discriminação na decomposição salarial.

##### ✓ Decomposição na média

A modelização mais usada para ter em consideração este problema é o método em dois passos sugerido por Heckman(1979). No primeiro passo é estimada a probabilidade de participação no mercado de trabalho (através de um modelo probit) e estimado o termo de correção de seletividade que é introduzido como regressor adicional das equações salariais. A estimação da equação salarial com esse termo adicional consiste no segundo passo desta metodologia.

Em consequência do facto das regressões salariais conterem o termo para a correção da seletividade as decomposições incluem também esse termo de correção. Mas, o problema é como interpretar os efeitos desse termo, ou seja a questão é se a seletividade deverá ser atribuída a diferenças nas características ou se deve ser incluída no efeito estrutural (discriminação), ou se deve ser uma categoria à parte.

Várias soluções podem ser encontradas na literatura. Muitos autores consideram o termo de seletividade como um termo à parte da discriminação e do efeito composição. Considerando a decomposição de Oaxaca e Blinder (1973):

$$\overline{\ln W_h} - \overline{\ln W_m} = (\bar{X}_h - \bar{X}_m)\hat{\beta}_h + \bar{X}_m(\hat{\beta}_h - \hat{\beta}_m) + (\hat{\sigma}_h \hat{\lambda}_h - \hat{\sigma}_m \hat{\lambda}_m) \quad (13)$$

Em que o termo:  $(\hat{\sigma}_h \hat{\lambda}_h - \hat{\sigma}_m \hat{\lambda}_m)$ , é devido à seletividade.

Uma forma alternativa, também muito usada na literatura, é subtrair o termo da seletividade à diferença dos salários observados:

$$(\overline{\ln W_h} - \overline{\ln W_m}) - (\hat{\sigma}_h \hat{\lambda}_h - \hat{\sigma}_m \hat{\lambda}_m) = (\bar{X}_h - \bar{X}_m)\hat{\beta}_h + \bar{X}_m(\hat{\beta}_h - \hat{\beta}_m) \quad (14)$$

Mas com esta abordagem o lado esquerdo da equação não é mais a diferença do salário observado:  $(\overline{\ln W_h} - \overline{\ln W_m})$ . Logo esta abordagem não nos dá uma decomposição do diferencial do salário observado, para os que participam no mercado de trabalho mas sim para uma medida do diferencial dos salários potenciais ou salários oferecidos.

Existem, contudo, mais alternativas que levam a decomposições mais complexas.<sup>1</sup> Não há forma de dizer qual das formas de ter em consideração o termo de seletividade é a mais correta.

#### ✓ Decomposições distribucionais

A correção da seletividade também tem sido considerada no âmbito das decomposições distribucionais, por exemplo com a decomposição de Machado & Mata(2005) e também com a distribuição proposta por Fortin et al.(2011) e Firpo et. al.(2009,2018)

Uma dessas extensões foi proposta por Albrecht, Vuum & Vroman(2009) que estende a decomposição de Machado & Mata(2005) e usam o método de correção semi-paramétrico de Buchinsky(1998) para a regressão de quantis.

---

<sup>1</sup> Ver Neuman, S. & Oaxaca, R. (2004) para uma apresentação de outras alternativas

Um método recente proposto no âmbito da regressão de quantis condicionais é o de Arellano & Bonhomme (2017) que propôs um método em três passos baseado em cópulas para ter em consideração seletividade.

Também para a decomposição do Fortin et al., (2011), Firpo et al.(2018) foi estudada a possibilidade de existência de seletividade, considerando três métodos de correção de seletividade que usam um modelo de escolha binária: probit (semelhante ao de Heckman) e métodos semi-paramétricos como Ichimura (1993) e Klein & Spady (1993). Um dos autores que propôs esta solução foi Topfer(2017).

Contudo, deverá ser dito que a correção da seletividade também está sujeita a críticas. Uma das mais comuns diz respeito à forma como é modelada a participação ou não no mercado de trabalho – mais concretamente às variáveis que são consideradas. Na verdade, se o procedimento de correção da seletividade é incorreto pode introduzir mais problemas na equação estimada do que aqueles que resolve.

#### **4.9. Evidência empírica: caso das diferenças salariais de género em Portugal**

Do ponto de vista empírico a questão que se coloca é se os resultados são muito diferentes consoante a metodologia usada e como podemos comparar os resultados das diferentes metodologias (se é que podem ser comparados). Apresentam-se, como exemplo de aplicação de algumas das decomposições, resultados de estudos que analisam a discriminação salarial de género em Portugal. A apresentação não pretende ser exaustiva, mas sim exemplificativa das diversas metodologias.

A maior parte dos estudos sobre a discriminação salarial em Portugal usa a metodologia de Oaxaca e Blinder (1973) ou a extensão OB(1994) (ou a proposta de Cotton,1988 ou Neumark,1988). A esmagadora maioria dos estudos exclusivamente dedicados a Portugal usa os dados dos *Quadros de Pessoal* – concentrando-se, pois, no setor privado. Os estudos de comparação entre países europeus, incluindo Portugal, usam as bases de dados da Eurostat: ECPH ou EU-SILC.

## A. Decomposição de Oaxaca e Blinder

OB 1973

	Dados	Diferença Salarial	Composição	Discriminação
Vieira et al. (2005)	QP -1999 16-65 anos	0.2585	0.0944	0.164
Aláez-Aller et al. (2011)	ECPH-2000 (Privado) >16 anos Part-time+fulltime	0.21	-0.036	0.246
Galego & Pereira(2010)	ECHP -2001 (publico+privado) 16-65 anos	0.16	-0.03	0.19
Boll et al..(2016)	EU-SILC 2013 (privado) 14-74 anos	0.141	0.0026	0.138

Cotton (1988)

	Dados	Diferença Salarial	Composição	Discriminação		
				Vantagem homens	Desvantagem mulheres	soma
Vieira et al. (2005)	QP 1999 16-65 anos	0.2585	0.0990	0.0637	0.0958	0.16

OB(1994)

	Dados	Diferença Salarial	Composição	Discriminação		
				Vantagem homens	Desvantagem mulheres	soma
Christofides et al. (2013)	EU-SILC 2007 (privado+publico) 25-54 anos	0.092	-0.074	0.081	0.085	0.166

## B. Decomposição Brown et al. (1980)

	Dados	Diferença Salarial	Componente intra- profissões			Componente entre-profissões (distribuição ocupacional)		
			Total	Composição	Discriminação	Total	Composição	Discriminação
Mendes, R. (2009)	QP 2004 Privado sem agricultura 16-64 anos	0.204	0.191	0.037	0.154	0.013	-0.037	0.05

### C. Decomposição por quantis

	Dados	Decomposição	Quantis	Diferença Salarial	Composição	Discriminação
Christofides et al. (2013)	EU-SILC 2007 (público+privado) 25-54 anos	Melly (2005)	10	0.140	-0.132	0.272
			25	0.169	-0.148	0.317
			50	0.164	-0.184	0.348
			75	0.001	-0.31	0.311
			90	-0.080	-0.291	0.203
Cardoso et al. (2016)	QP 2013 18-65 anos Sem part-time	Machado e Mata (2005)	10	0.099	0.021	0.078
			20	0.130	0.014	0.117
			50	0.205	-0.004	0.209
			70	0.244	-0.015	0.259
			90	0.288	-0.013	0.301
Galego (forthcoming)	QP 2019 18-65 anos Part-time+full-time	Firpo et al.(2018)	10	0.008	0.002	0.007
			30	0.041	-0.001	0.042
			50	0.112	-0.004	0.115
			70	0.167	0.019	0.148
			90	0.180	0.054	0.127

### D. Selectividade

#### Decomposição de Oaxaca e Blinder com controlo de seletividade

	Dados	Decomposição	Método	Diferença Salarial <sup>(a)</sup>	Composição	Discriminação
Galego & Pereira(2010)	ECHP 2001 (Público+privado) 16-65 anos	OB 1973	Semi-paramétrico: Li e Wooldridge(2002) (Seletividade significativa)	0.08	-0.03	0.11
Ponthieux & Meurs (2005)	ECHP-2000 (Público+privado) 25-55 anos	OB 1994	Heckman(1979) (Seletividade não significativa)	0.172	-0.025	0.202(b)
Christofides et al. (2013)	EU-SILC 2007 (Público+privado) 25-54 anos	OB 1994	Heckman(1979) (Seletividade sem influência significativa nos resultados)	0.104	-0.112	0.216(b)

(a) Corrigida para a seletividade; (b) soma da vantagem dos homens com a desvantagem das mulheres

**Principais conclusões relativamente aos estudos apresentados:** Independentemente do método de decomposição, em geral os estudos concluem que existe evidência de discriminação de género no mercado de trabalho português. Igualmente, a maioria dos estudos concluem que as mulheres apresentam características observadas semelhantes ou melhores que os homens. As maiores diferenças nos resultados devem-se à utilização de dados diferentes – uns concentram-se no setor privado enquanto outros consideram os setores privado e público.

#### **4.10. Considerações Finais**

- As metodologias de decomposição são instrumentos poderosos para identificar as causas das diferenças entre grupos, nomeadamente na identificação de potencial discriminação. Ao permitir identificar as causas concretas das diferenças registadas entre grupos será possível desenhar as políticas mais adequadas para combater as desigualdades.
- Contudo, as decomposições assentam em hipóteses que nalgumas circunstâncias podem ser consideradas fortes e irrealistas.
- Em geral, em todas as decomposições subsiste a questão de saber se a separação entre características observadas e não observadas pode ou não capturar exclusivamente o fenómeno da discriminação. A parte não observada, reflete também, entre outras coisas, erros de medida nas variáveis, problemas de especificação, assim como a existência de variáveis omitidas.
- A boa notícia no trabalho aplicado em anos mais recentes é o facto de estarem disponíveis bases de dados cada vez mais ricas e com mais informação sobre os indivíduos e sobre as empresas, que reduz os erros de medida e o problema das variáveis omitidas.
- ➔ Contudo também devemos ter consciência que em todo o trabalho empírico existem sempre problemas e questões metodológicas a considerar.

#### **V. Bibliografia**

Relativamente a este tópico apresentam-se as leituras principais recomendadas aos alunos.. Para além disso, é fornecida a lista dos artigos referenciados na aula e que os alunos poderão consultar caso desejem aprofundar ou esclarecer algum dos tópicos apresentados.

## **Leituras recomendadas**

- Bazen, S. (2011), *Econometric methods for Labor Economics*, Oxford University Press. (capítulo 2)
- Brown, R. S., Moon, M. & Zoloth, B. (1980) Incorporating Occupational Attainment in Studies of Male-Female Earnings Differentials, *The Journal of Human Resources*, 15(1), 3-28
- Cahuc, Pierre & Zylberberg, André (2014) , *Labor Economics* , The MIT Press (parte 2, capítulo 8)
- Fortin N., Lemieux T. & Firpo S. (2011) “Decomposition Methods in Economics” in David Card, Orley Ashenfelter(eds) , *Handbook of Labor Economics*, North-Holland , Vol.4, Part A, pp. 1–102
- Nopo, H. (2008) Matching as a tool to decompose wage gaps. *The Review of Economics and Statistics*, 90(2), 290-299.
- Oaxaca, R (2007) The challenge of measuring labor market discrimination against women. *Swedish Economic Policy Review* 14: 199-231
- Popli, Gurleen (2023) "Decompositions: Accounting for Discrimination," in Ashwini Deshpande(ed.), *Handbook on Economics of Discrimination and Affirmative Action*, Springer, chapter 8, pp 133-150.

## **Leituras adicionais**

- Aláez-Aller, R., Longás-Garcia, J.C. & Ullibarri-Arce, M. (2011), Visualising Gender Wage Differences in the European Union, *Gender, Work and Organization*, 18(s1), e49-e85
- Albrecht, J., van Vuuren, A. & Vroman, S. (2009), Counterfactual Distributions with sample selection adjustments: econometric theory and an application to the Netherlands, *Labour economics*, 16, 383-396
- Arellano, M., & Bonhomme, S. (2017). Quantile selection models with an application to understanding changes in wage inequality. *Econometrica*, 85(1), 1-28.
- Black, D., A., Haviland, S. Sanders & Taylor, L. (2008), Gender wage disparities among the highly educated. *Journal of Human Resources*, 43(3), 630–659.
- Blinder, A. (1973), Wage Discrimination: Reduced Forms and Structural Estimates, *The Journal of Human Resources*, 8, 4, 436-55.
- Boll, C., Leppin, J., Rossen, A. & Wolf, A. (2016) Magnitude and Impact factors of the gender pay gap in EU countries, European commission, Directorate-General for Justice, Brussels, accessed at: [http://hwwi.org/fileadmin/hwwi/Publikationen/Publikationen\\_PDFs\\_2016/Magnitude\\_And\\_Impact\\_Factors\\_All.pdf](http://hwwi.org/fileadmin/hwwi/Publikationen/Publikationen_PDFs_2016/Magnitude_And_Impact_Factors_All.pdf)
- Buchinsky, M. (1998) The dynamics of changes in the female wage distribution in the USA: a quantile regression approach, *Journal of Applied Econometrics*, 13, 1-30.
- Cardoso, A. R. , Guimarães, P., Portugal, P. & Raposo, P. (2016) , The sources of the gender wage gap, *Revista de Estudos Económicos* , 2(2), 47-68, Banco de Portugal
- Christofides, L.N. et al.(2013) Gender wage gaps, ‘sticky floors’ and ‘glass ceilings’ in Europe, *Labour Economics*, 21, 86–102

- Cotton, J. (1988). On the Decomposition of Wage Differentials. *The Review of Economics and Statistics* 70: 236–243
- 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.
- Firpo, S. N. Fortin & Lemieux, T. (2007), *Decomposing wage distributions using recentered influence functions regressions*, working paper.
- Firpo S, Fortin N & Lemieux, T. (2009), Unconditional quantile regressions, *Econometrica* , 77, 953-973
- Firpo S, Fortin N & Lemieux, T. (2011), Occupational Tasks and Changes in Wage Structure, IZA Discussion Paper No. 5542
- Firpo, S., Fortin, N. & Limieux, T. (2018), Decomposing Wage Distributions using Recentered influence function regressions, *Econometrics*, 6(28), 1-40.
- Fortin, N. (2008) The Gender Wage Gap among Young Adults in the United States: The Importance of Money versus People, *The Journal of Human Resources*, 43(4), 884-918
- Galego, A. & Pereira, J. (2010) Evidence on gender wage discrimination in Portugal: parametric and semi-parametric approaches, *Review of Income and Wealth*, Series 56, Number 4, 651-666.
- Galego, A. (forthcoming) “Recent Trends in the Gender Wage Gap in Portugal: A Distributional Analysis”, *Applied Economics Letters*.
- Heckman, J. (1976), The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models, *Annals of Economic and Social Measurement*, 5, 4, 475-492.
- Heckman, J. (1979), Sample Selection Bias as a Specification Error, *Econometrica*, 47, 153-161.
- Ichimura, H. (1993), Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics*, 58(1–2), 71-120
- Jann, B. (2008) The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4), 453-479.
- Jones, F. (1983) On decomposing the wage gap: A critical comment on Blinder’s method. *Journal of Human Resources*, 18, 126-130
- Juhn, C., Murphy, K. & Pierce, B. (1993) Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101(3), 410-442
- Kim, ChangHwan (2013). Detailed Wage Decompositions. Revisiting the Identification Problem. *Sociological Methodology*, 43, 346–363.
- Klein, W. & Spady, R. (1993) An Efficient Semiparametric Estimator for Binary Response Models, *Econometrica*, 61(2), 387-421
- Li Q. & Wooldridge, J. (2002), Semiparametric Estimation of Partially Linear Models for Dependent Data with Generated Regressors, *Econometric Theory*, 18, 625-645.

Lemieux, T. (2002), Decomposing changes in wage distributions: a unified approach, *Canadian Journal of Economics*, 35(4), 646-688

Machado, J. & Mata, J. (2005), Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression, *Journal of Applied Econometrics*, 20(4), 445-465.

Melly, B. (2005) Decomposition of differences in distribution using quantile regression, *Labour Economics*, 12, 577-590.

Melly, B. (2006) *Estimation of counterfactual distributions using quantile regression*, mimeo, University of St. Gallen, <http://www.alexandria.unisg.ch/Publikationen/22644>.

Mendes, R. V. (2009), Gender wage differentials and occupational distribution, *Notas Económicas*, junho, 26-41.

Neumark, D. (1988) Employers' Discriminatory Behavior and the Estimation of Wage discrimination. *The Journal of Human Resources*, 23, 279–295

Neuman, S. & Oaxaca, R. (2004), Wage Decompositions with Selectivity-Corrected Wage Equations: a Methodological Note, *Journal of Economic Inequality*, 2, 3-10.

Nielson, H. (2000) Wage discrimination in Zambia: an extension of the Oaxaca-Blinder decomposition, *Applied Economics Letters*, 7(6), 405-408.

Oaxaca, R. (1973), Male-Female Wage Differentials in Urban Labour Markets, *International Economic Review*, 14, 693-709.

Oaxaca, R. & Ransom, M. (1988), Searching for the effect of unionism on the wages of union and nonunion workers, *Journal of Labor Research*, 9, 139-148.

Oaxaca, R. & Ransom, M. (1994), On discrimination and the decomposition of wage differentials, *Journal of Econometrics*, 61, 5-21.

Oaxaca, R. & Ransom, M. (1998), Calculation of Approximate Variances for Wage Decomposition Differentials, *Journal of Economic and Social Measurement*, 24, 55-61.

Oaxaca, R. & Ransom, M. (1999) Identification in detailed wage decompositions. *Review of Economics and Statistics*, 81, 154-157.

Ponthieux, S. & Meurs, D. (2005), The Gender Wage Gap in Europe: Women, Men and the Public Sector, Institut National de la Statistique et des Études Économiques, Série des Documents de Travail F0502.

Reimers, C. W. (1983). Labor Market Discrimination Against Hispanic and Black Men. *The Review of Economics and Statistics*, 65, 570–579.

Rothe, C. (2015). Decomposing the composition effect: The role of covariates in determining between-group differences in economic outcomes. *Journal of Business & Economic Statistics* 33, 323–337.

Suen, W. (1997) Decomposing wage residuals: Unmeasured skill or statistical artefact? *Journal of Labor Economics*, 15(3), 555-566

Topfer, M. (2017), Detailed Rif decomposition with selection- the gender pay gap in Italy.Discussion paper 26-2017, University of Hohenheim

Vieira, J., Cardoso, A. & Portela, M. (2005), Gender Segregation and the Wage Gap in Portugal: an Analysis at the Establishment Level, *Journal of Economic Inequality*, 3, 2, 145-168.

Yun, M. (2005), A simple solution to the identification problem in detailed wage decompositions, *Economic Inquiry*, 43, 766-772

Yun, M. (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