DECOMPOSITION OF DIFFERENCES
BETWEEN PERSONAL INCOMES DISTRIBUTIONS IN POLAND

Please use MIBE\_....styles for formatting

according to the instructions in this file. How to use styles - help: <http://guides.lib.umich.edu/c.php?g=283073&p=1888265>

Full text paper should not exceed 10 pages (Portrait orientation).

* The texts of the papers are set individually by the Authors. Only minor corrections to the text may be carried out by the publisher.
* Please**, use** **style MIBE\_... to format your article.**
* Names of suitable styles are placed in blue text boxes.
* Pay attention to the shortcuts: e.g. - Ed. - Eds. -
* Use a dot as a separator of decimal part of number
* Please DO NOT use hyphenation

**DELETE ALL TEXT OBJECTS (LIKE THIS) AFTER A PREPARATION OF YOUR PAPER**

* Please save the file under the name: Nazwisko.docx (Name\_Name.docx)
in Word 2007 format 2007 (or newer version).
* **NIE** edytuj nagłówka na pierwszej stronie.

Use style **MIBE\_Title**

Use style **MIBE\_Author**

Katarzyna Author  https://orcid.org/0000-0000-0000-0000

Faculty of Applied Informatics and Mathematics
Warsaw University of Life Sciences – SGGW

Use style **MIBE\_Adress**

e-mail: katarzyna\_author@sggw.pl

Krzysztof Writer Author  https://orcid.org/0000-0000-0000-0000

Piotr Worden Author  https://orcid.org/0000-0000-0000-0000

DO NOT put your e-mail address as HYPERLINK

Faculty Name,

University Name

e-mail: krzysztof.writer@domain.pl; piotr\_worden@sggw.pl

**Abstract:** In this paper we study differences between personal incomes distributions in Poland in 2002 and 2012. The empirical data have been collected within the Household Budget Survey project. We used the Machado & Mata decomposition, which utilizes quantile regression. This method allowed us to investigate differences between income distributions in the whole range of values, going beyond simple average value decomposition. We evaluated influence of person’s attributes on the differences of incomes distributions in 2002 and 2012. By decomposing the differences into the explained and unexplained components we got information about their causes.

Use style **MIBE\_Keywords**

**Do not put a dot at the end**

**Abstract shoud not exceed 600 signs (incliding spaces).**

Use style

 **MIBE\_Abstract**

**Keywords:** quantile regression, Machado & Mata decomposition, counterfactual distribution

**JEL classification:** C51, C52

Use style **MIBE\_Section**

Use style **MIBE\_JEL**
**Do not put a dot at the end**

INTRODUcTION

**.**

**.**

Based on the works [Oaxaca 1973] and [Blinder 1973] one elaborated techniques which went far beyond simple comparison of average values, for example decomposition of the variances or the whole distributions. New techniques allowed to discover various factors influencing incomes distributions, as minimal wage [DiNardo et al. 1996]. They also have been useful in studying differences of incomes distributions for various group of people [Albrecht et al. 2003].

References in [ ] brackets

Use style **MIBE\_Text-Indent**

**.**

**.**

First Author et al. - if there are several Authors

During the decomposition of differences between the distributions one utilizes so called counterfactual distributions. They are a mixture of an conditional distribution of the dependent variable and various distributions of the explanatory

**.**

**.**

variables (see [Juhn et al. 1993, DiNardo et al. 1996]). One of them,

Use style **MIBE\_Text\_without\_Indent**

**.**

to enter PARAGRAPH without indenting the first row

... proposed in Machado and Mata [2005] decomposes differences of distributions using a quantile regression.

**.**

**.**

... property is satisfied (see [Qiang, Jiajin 2018] and [Feller 1968, 79-82]) ...

examples of references to the literature

**.**

**.**

**....** are the least probable and the most probable values for $T\_{n}^{+}$ are close to 0 or 1 (see [Feller 1966, p. 81]). It may be seen from the graph of the density $f$ in the figure

**.**

**.**

... which was identified and indicated in the late 1980s and early 1990s [Strassmann 1990; Brynjolfsson 1993; Barua, Lee 1997]. Explanation of ...

**.**

... works of H. Markowitz [Markowitz 1952; Markowitz 1991]. The main ...

**.**

**.**

... customers [Jałowiecki, Gostkowski 2013]. Most likely, this means that the use

**.**

**.**

To determine the advancement of used IT solutions level proposed in the paper Jałowiecki and Jałowiecka [2013] coefficient was used. It was constructed on

... (see [Porter, Fuller 1989; Vizjak 1990; Rupprecht-Däullary 1994]). A characteristic feature ...

In this paper we compared incomes of the employees running the one-person households in 2012 with those in 2002. The data have been collected in the Household Budget Survey project in Poland. The aim of the work is to study differences between income distributions in year 2002 and 2012. By use of the [Machado, Mata 2005] decomposition method we investigated differences in the whole range of income values. The past studies in Poland were mostly focused on the decomposition of the average values by using the Oaxaca and Blinder method [e.g. Śliwicki, Ryczkowski 2014].

DECOMPOSITION METHODS

Use style **MIBE\_Subsection**

Oaxaca & Blinder decomposition of average incomes differences

We consider two groups of one-person households[[1]](#footnote-1). The first one contains data for 2002, the second one – for 2012, denoted by *T*1 and *T*2 respectively. We also deal with the outcome variable *y*, and a set of predictors *X*. The variable *y* is individual income and predictors *X* are individual sociodemographic characteristics of households (people) such as sex, age, education and others. The idea of Oaxaca & Blinder decomposition can be applied whenever we need to explain the differences between the expected values of dependent variable *y* in two comparison groups [Oaxaca 1973, Blinder 1973]. We assume that the expected value of *y* conditionally on *X* is a linear function of *X*:

Use style **MIBE\_Equation**

**Use Tab button
to adjust an equation and a number**

 , (1)

where *Xi* are characteristics of objects in the year *i* and *βi* is the vector of parameters. The equation (1) can be estimated for both years:

 . (2)

insert punctuation between formulas

The difference between the expected values of *y* in both years is as follows:

 . (3)

**.**

**.**

Decomposition of differences between distributions

The mean decomposition analysis may be extended to the case of differences along the whole distribution. Let  and  be the density functions for the variable *y* in 2002 and 2012, respectively. The distribution , *i* = *T*1, *T*2, is the marginal distribution of the joint distribution :

 , (5)

Use style **MIBE\_Footnote**

where *X* is a vector of individual characteristics observed [Bourguignon, Ferreira 2005]. Let  be the conditional distribution of *y*. Then one can (5) express as:

 , (6)

where  is the joint distribution of all elements of *X* in year *i*. The difference between the two distributions may be decomposed onto

 , (7)

where  is the counterfactual distribution, which can be constructed as

 . (8)

The first component in (7) gives the effect of the unequal different personal characteristic’s distributions in 2012 and 2002. The second component describes the inequalities between two distributions of *y* conditional on *X*. The difference with respect to the Oaxaca & Blinder decomposition is that this decomposition refer to full distributions, rather than just to their means.

**.**

**.**

assume that all quantiles of *y* conditional on *X* are linear in *X*. The quantile regression estimator for quantile *θ* minimizes the sum:

 , (10)

where .

The sum (10) gives asymmetric penalties for over and under prediction. For each quantile other parameters are estimated. We interpret these coefficients as the returns to different characteristics *X* at given quantiles of the distribution of *y*. The standard errors of parameters are calculated using bootstrap method [Gould 1992].

Machado & Mata decomposition of differences in distributions

Machado & Mata [2005] have used quantile regression in order to estimate counterfactual unconditional income distributions. The unconditional quantile is not the same as the integral of the conditional quantiles. Therefore, authors provide a simulation based estimator where the counterfactual distribution is constructed from the generation of a random sample. The approach is the as follows:

1. generate a random sample of size *m* from a *U*[0,1]: *θ*1, *θ*2, …, *θm*;
2. using the dataset for *T*1 estimate *m* different quantile regression , obtaining coefficients ;

Use style

 **MIBE\_Text\_Numbered**

1. generate a random sample of size *m* with replacement from , denoted by ;
2.  is a random sample from the unconditional distribution .

.

.

.

* Appropriate definitions and terminologies, in particular of such categories as: edible/inedible parts of food; avoidable/unavoidable food waste;
* Borders of the particular stages and links of the supply chain of raw materials and agri-food products;
* Measurement units, so as the existing data between the regions, countries and types of the products could be comparable.

.

Use style

 **MIBE\_Text\_Bullet**

.

.

EMPIRICAL DATA

The data were collected in the Household Budget Survey project for 2002 and 2012, group *T*1 and *T*2 respectively. The analyzed data regards households running by one person whose main source of earning comes from a work as an employee. The annual disposable incomes (variable *INC*, in thousands of PLN) in 2012 were compared with those obtained in 2002. The incomes in 2002 during the analysis were expressed in prices in 2012 (variable *INCREAL*). The sample consisted of 834 and 1594 persons in 2002 and 2012 respectively. For each person: sex, age, education, place of residence, type of labor position have been obtained. Based on the obtained attributes one defined the following describing variables:

*SEX* (0 − woman, 1 − man),

*AGE* (years),

*EDU* (education, 1−9, 1 − primary, . . ., 9 − tetriary),

*RES* (residence, 1−6, 1 − village, . . ., 6 − town ≥ 500k of inhabitants),

**Table in vertical orientation**

**Center
a Table**

*POS* (0−1, 0 − manual labor position, 1 − non-manual labor position).

Features of the variables have been collected in the Table 1.

Use style **MIBE\_Table\_Title**

**Do not put a dot at the end**

Table 1. The mean values and the standard deviations for the selected variables

First row centered

All rows aligned vertically

NUMBERS aligned by dot

|  |  |  |  |
| --- | --- | --- | --- |
|  | Whole sample | 2002 | 2012 |
| Number of observation | 2428 | 834 | 1594Use font TNRoman 10 ptUse single bordersUse a dot as a separator of decimal part of number |
| INC | 27.92 (21.32) | 19.76 (17.92) | 32.19 (21.71) |
| INCREAL | 30.09 (22.58) | 26.07 (23.64) | 32.19 (21.71) |
| SEX (% men) | 40.28 | 38.49 | 41.22 |
| AGE | 41.65 (12.12) | 40.65 (11.29) | 42.16 (12.51) |
| POS (% non-manuals) | 67.26Use style **MIBE\_Table\_Source****Do not put a dot at the end** | 64.51 | 68.70 |

Source: own calculations

RESULTS

We compared the personal incomes distributions for years 2002 and 2012.
In the first step of the analysis the Oaxaca & Blinder decomposition has been applied for the average values. The results are listed in the Table 2.

Table 2. The Oaxaca & Blinder decomposition of the average incomes differences

|  |  |
| --- | --- |
| Average *INCREAL* in 2002  | 26.072References to the Table and the Figure:Table number, Figure numberor (Table number, Figure number) |
| Average *INCREAL* in 2012  | 32.189 |
| Raw gap | 6.117 |
| Aggregate decomposition |
| Explained effect | 1.302 |
| Unexplained effect | 4.816 |
| % explained | 21.3 |
| % unexplained | 78.7 |
| Detailed decomposition |
| explained component | unexplained component |
| *SEX* | 0.217 | *SEX* | 6.251 |
| *AGE*  | 0.199 | *AGE*  | 0.445 |
| *EDU* | 1.038 | *EDU* | 10.852 |
| *RES* | −0.369 | *RES* | 1.462 |
| *POS* | 0.216 | *POS* | −6.804 |
| *const*  | 0.000 | *const*  | −7.390 |
| Total  | 1.302 | Total | 4.816 |

Source: own calculations

One can observe the positive difference between average values of the real incomes in 2012 and 2002. In the next step one tried to explain the observed difference. Using the decomposition method, one evaluated strength of the influence of the analyzed factors onto the average incomes. Generally, the differences are explained by the factors being studied in 21.3%. The *SEX*, *AGE*,

**.**

Use a dot as a separator
of decimal part of number

**.**

**.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Place in the table objects,**  **which are to be placed in two columns next to each other** |

|  |  |  |  |
| --- | --- | --- | --- |
| Lp. | Województwo | $$Q\_{i}$$ | Grupy |
| 1234567 | MazowieckieŚląskieMałopolskiePomorskieKujawsko-pomorskieDolnośląskieZachodniopomorskie | 115107102,5979291,587,5 | I (7 woj.) |
| 8.58.5101112 | OpolskieWielkopolskieLubuskieŁódzkiePodkarpackie | 777772,571,568 | II (5 woj.) |
| 13141516 | LubelskieWarmińsko-mazurskieŚwiętokrzyskiePodlaskie | 49,54838,529,5 | III (4 woj.) |
| I ($Q\_{i}$) | 3,898 |

 |

**.** **.**

Use a dot as a separator
of decimal part of number

**.**

**.**

Studying the unexplained components of the quantile differences, related to the different parameters values in both years, one can conclude that for *EDU* as well as for *POS* they are quite similar. They increase with the values of the quantiles. However the differences for the *POS* parameters are relatively small for the left part of the distribution rising strongly for the higher quantiles.

On the other hand the unexplained part of the *SEX* parameters is negligible. The big rise of the differences for the last quantile is within the statistical error.

Figure 1. The influence (vertical axis) of the selected attributes on the income distribution as a function of the quantile range (horizontal axis)

Use style

**MIBE\_Figure\_Title**

**Do not put a dot at the end**





Center a figure



\* The shaded areas represent 95% confidence intervals. The dashed lines represent results of the classical linear regression model.

Use style

**MIBE\_Explanations**

Source: own preparation

Use style **MIBE\_Figure\_Source**

**Do not put a dot at the end**

In the last step of the analysis one performed the decomposition
of differences between income distributions. The differences are expressed as the sum of the explained and unexplained components along the whole income distributions. The Machado & Mata method has been used to estimate quantile regressions for 19 percentiles. The results for deciles are presented in the Table 3. The errors have been evaluated using the bootstrap method.

Table 3. The results of the Machado & Mata decomposition of the differences of the incomes distributions for 2002 and 2012

| Quantile | Raw gap | Diference M-M | Explained part | Unexplained part | Explained % | Unexplained % |
| --- | --- | --- | --- | --- | --- | --- |
| 0.10 | 3.9567 | 4.3506 (0.2590) | 0.5032 (0.3057) | 3.8473 (0.2559) | 12% | 88% |
| 0.20 | 4.5382 | 4.6094 (0.2784) | 0.5934 (0.2674) | 4.0160 (0.2707) | 13% | 87% |
| 0.30 | 4.8726 | 4.8689 (0.2608) | 0.6819 (0.2748) | 4.1870 (0.2473) | 14% | 86% |
| 0.40 | 4.8204 | 5.0937 (0.2955) | 0.8079 (0.3076) | 4.2858 (0.2825) | 16% | 84% |
| 0.50 | 4.7140 | 5.4314 (0.3305) | 0.9372 (0.3436) | 4.4942 (0.3103) | 17% | 83% |
| 0.60 | 5.3322 | 5.8808 (0.4041) | 1.0257 (0.4118) | 4.8551 (0.3791) | 17% | 83% |
| 0.70 | 6.4007 | 6.7484 (0.4282) | 1.0990 (0.4992) | 5.6494 (0.4517) | 16% | 84% |
| 0.80 | 8.3482 | 8.1739 (0.6231) | 1.2908 (0.6968) | 6.8831 (0.5822) | 16% | 84% |
| 0.90 | 11.2275 | 11.228 (1.1327) | 1.6515 (1.1115) | 9.5765 (1.0202) | 15% | 85% |

Source: own calculations

The whole model approximates the data well. The differences between income distributions are rising with incomes. Their decomposition onto the explained and unexplained parts indicate low share of its explained part (12% to 17%). The share of the model’s unexplained part is relatively high (83% to 88%) and increasing what indicates on the increase of the “labor market value” of the households’ attributes. However, the explained part of the model also increases with income level.

SUMMARY

In this paper one studied differences between personal’s income distributions in Poland in 2002 and 2012. The households of the single employers have been taken into account. The Oaxaca & Blinder and Machado & Mata decompositions of the average values and the whole incomes distributions respectively have been used. The Oaxaca & Blinder decomposition showed the positive influence of the most analyzed variables (*SEX*, *AGE*, *EDU*, *POS*) on the average income ...

References

Use style **MIBE\_References**

Abrahamowicz M., Ząjąc K. (1986) Metoda ważenia zmiennych w taksonomii numerycznej i procedurach porządkowania liniowego. [in:] Prace Naukowe AE we Wrocławiu, 328, 5-17 (in Polish).

(in Polish) only as addition to polish language publication

**.**

**.**

**.**

Atamowicz A. (2007) Modele i algorytmy. [in:] Janek T. (Ed.) Wybrane zagadnienia. PWE, Warszawa (in Polish).

**short dashes in page numbering**

Blinder A. (1973) Wage Discrimination: Reduced Form and Structural Estimates. Journal of Human Resources, 8, 436-455.

Borkowski B. i Szczesny W. (2005) Metody wizualizacji danych wielowymiarowych jako narzędzie syntezy informacji. Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu, 7(5), 11-16 (in Polish).

Authors (year) Title of article. Title of Journal, volume number(issue number), pages.

**Put a dot at the end. DO NOT put a space between number and (issue number)**

**Title - used as a proper noun**

Gould W. W. (1992) Quantile Regression with Bootstrapped Standard Errors. Stata Technical Bulletin, 9, 19-21.

**.**

**Title - used as a proper noun**

**.**

Juhn Ch., Murphy K. M., Pierce B. (1993) Wage Inequality and the Rise in Returns to Skill. Journal of Political Economy, 101, 410-442.

Kerlinger F. N. (1986) Foundations of Behavioral Research (3rd Edition). Holt, Rinehart & Winston.

Koenker R., Bassett G. (1978) Regression Quantiles. Econometrica, 46, 33-50.

Loveman G. (1994) An Assessment of the Productivity Impact on Information Technologies. [in:] Allen T. J., Morton, M. S. S. (Ed.) Information Technology and the Corporation of the 1990’s: Research Studies. MIT Press, Cambridge, MS, USA.

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

Newell A., Socha M. (2005) The Distribution of Wages in Poland. IZA Discussion Paper, 1485.

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

Pflug G. Ch. (2000) Some Remarks on the Value-at-Risk and Conditional Value-at-Risk. [in:] Uryasev S. (Ed.) Probabilistic Constrained Optimization: Methodology and Applications, Kluwer, 272-281.

Rokicka M., Ruzik A. (2010) The Gender Pay Gap in Informal Employment in Poland. CASE Network Studies and Analyses, p. 406.

Śliwicki D., Ryczkowski M. (2014) Gender Pay Gap in the Micro Level – Case of Poland. Quantitative Methods in Economics, XV(1), 159-173.

Authors (year) Title of Article. Title of Journal, volume number(issue number), pages.

**Put a dot at the end. DO NOT put a space between number and (issue number)**

**Title - used as a proper noun**

1. Ibid., p.2. [↑](#footnote-ref-1)