STATISTICS IN TRANSITION new series, June 2016
Vol. 17, No. 2, pp. 331–348

DECOMPOSITION OF DIFFERENCES IN INCOME DISTRIBUTIONS USING QUANTILE REGRESSION

Joanna Małgorzata Landmesser[1]

ABSTRACT

The paper deals with microeconometric techniques useful for the study of differences between groups of objects, methods that go beyond simple comparison of average values. Techniques for the decomposition of differences in distributions by constructing counterfactual distributions were considered. Using the Machado-Mata quantile regression approach the empirical decomposition of the inequalities in income distributions of one-person households in urban and rural areas was performed. We employed data from the Household Budget Survey for Poland in 2012. It was found that the tendency towards increased income inequalities between urban and rural residents when moving to the right of the income distribution can be observed. The rural residents are at a disadvantage. The decomposition of the inequalities revealed a growing share of the part explained by different characteristics of people and a declining share of the unexplained part, associated with the evaluation of those characteristics.

Key words: decomposition of differences, quantile regression, counterfactual distribution.

1. Introduction

Recent years have witnessed the rapid development of microeconometric techniques useful in the context of studying the differences between groups of objects. Various inequality decomposition methods are becoming more popular. Since the seminal works of Oaxaca (1973) and Blinder (1973) many procedures that go beyond simple decomposition of differences between the average values have been proposed. These are the variance decomposition techniques and the decomposition allowing the analysis of the differences with respect to the entire distribution of the outcome variable.

The main advantage of modern decomposition methods is to help to discover the factors affecting changes in the distribution of wages, for example. Studying

[1] Warsaw University of Life Sciences. E-mail: joanna_landmesser@sggw.pl.

changes in the distribution of wages has become an active area of research (see, e.g. Juhn, Murphy and Pierce, 1993; DiNardo, Fortin and Lemieux, 1996; Gosling, Machin and Meghir, 2000; Donald, Green and Paarsch, 2000; Machado and Mata, 2005; Autor, Katz and Kearney, 2005). For instance, DiNardo, Fortin and Lemieux (1996) analysed the implications of the observed changes on the labour market for specific points of the wage distribution and found that the minimum wage affects only the bottom end of this distribution. Other explanations based on de-unionization tend to affect the middle of the distribution (Card, 1992). The differences in income distributions between various groups of people, e.g. women and men, were also analysed (see Albrecht, Björklund and Vroman (2003), who look whether there is a glass ceiling in female earnings).

In particular, new techniques make it possible to carry out the decomposition of the differences in distributions, by constructing a counterfactual distribution that mixes conditional distribution for the outcome variable Y with various distributions for explanatory variables X. The most popular methods of constructing counterfactual distributions are those proposed in Juhn, Murphy and Pierce (1993), DiNardo, Fortin and Lemieux (1996), Machado and Mata (2005). Machado and Mata (2005) suggested using quantile regression in order to estimate counterfactual unconditional wage distributions.

The aim of our study was to decompose the observed inequalities in income distributions of one-person households in urban and rural areas applying the Machado-Mata technique. We employ data from the Household Budget Survey (HBS) for Poland in 2012. Applying the method to one-person households in urban and rural areas, not the typical class of the Polish households, allows interpersonal comparisons of the individual's income. An income of, say 4000 zlotys per head a month, implies a different purchasing power for a household of one and four persons. Living costs are often higher for single person households. Needs for housing space, electricity, etc. will not be four times as high for a household with four members as for a single person. Therefore, in order to properly compare the household incomes our attention has been focused only on one-person households.

The research concerning the income gap, conducted so far in Poland, was limited to decomposing mainly the average level of wage differences for men and women using the Oaxaca-Blinder method (e.g. Kot, Podolec and Uhlman, 1999; Słoczyński, 2012; Goraus, 2013; Śliwicki, Ryczkowski, 2014). Only a few studies go beyond the mean-decomposition. Grajek (2003) applied the John, Murphy and Pierce decomposition to analyse data on Polish employees from the period 1987–1996. He found that the explained component of gender pay gap is relatively small and rises slowly over the analysed period. Newell and Socha (2005), on the basis of quantile analyses using Labour Force Survey (LFS) data for 1992–2002, showed that many of the factors influencing wages, including gender, have a stronger impact in higher quantiles of wage distribution. Rokicka and Ruzik (2010) found that the inequality of earnings between women and men tends to be larger at the top of the earnings distribution (in the case of formal employees).

Nobody in Poland has made the decomposition of income inequalities for residents in urban and rural areas. In this paper we apply the Machado-Mata technique in order to move beyond estimation based on mean values. We argue that employing these techniques can provide deeper insights into the nature of income differentials.

The structure of the paper is as follows. Section 2 describes various techniques used for the decomposition of inequalities. Section 3 presents data and the results of the decomposition of differences in income densities between urban and rural inhabitants. Section 4 discusses the results and offers some concluding remarks.

2. Analysis method

This section outlines the methodology to be employed. First, we present the Oaxaca-Blinder decomposition of differences in mean wages. Then, we explain the idea of the decomposition of differences along the entire distribution. Finally, we present the conditional quantile decomposition techniques developed by Machado and Mata (2005).

2.1. Oaxaca-Blinder decomposition of differences in mean wages

There are two groups given, *A* and *B*, an outcome variable *y*, and a set of predictors *X*. The variable *y* may present log wages and predictors *X* may concern such individual socio-demographic characteristics of people as age, education level or work experience. 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). The authors of the methods assume that the expected value of *y* conditionally on *X* is a linear function of *X*:

$$y_g = X_g \beta_g + v_g, \qquad g = A, B, \tag{1}$$

where X_g are the characteristics of people in group *g* and β_g are the returns to these characteristics. The idea of Oaxaca-Blinder decomposition of the difference $\Delta^{\mu} = E(y_A) - E(y_B)$ is as follows:

$$\hat{\Delta}^{\mu} = \overline{X}_A \hat{\beta}_A - \overline{X}_B \hat{\beta}_B = \underbrace{(\overline{X}_A - \overline{X}_B)\hat{\beta}_A}_{\hat{\Delta}^{\mu}_{\text{explained}}} + \underbrace{\overline{X}_B(\hat{\beta}_A - \hat{\beta}_B)}_{\hat{\Delta}^{\mu}_{\text{unexplained}}} \tag{2}$$

The above equation is based on characteristics of one group and the estimated coefficients of the equation of another group. The first term on the right-hand side of the equation gives the effect of characteristics and expresses the difference of the potentials of both groups (the so-called explained, endowments or composition effect). The second term represents the effect of coefficients, typically interpreted as discrimination in numerous studies (the so-called unexplained, wage structure effect). This is the result of differences in the estimated parameters, and consequently in the "prices" of individual characteristics of representatives of a group. Blinder argued that "the latter sum […] exists only because the market evaluates differently the identical bundle of traits if possessed by different demographic groups" (Blinder, 1973, pp. 438-439).

One important drawback of this technique is that it focuses only on average effects, and this may lead to a misleading assessment if the effects of covariates vary across the wage distribution (Salardi, 2012).

2.2. Beyond the mean - decomposition of differences in distributions

The preceding scalar decomposition analysis may be extended to the case of differences along the entire distribution. Let $f^A(y)$ and $f^B(y)$ be the density functions for the outcome variable *y* in group *A* and *B*, respectively. The distribution $f^i(y)$, *i* = *A*, *B*, is the marginal distribution of the joint distribution $\varphi^i(y,X)$:

$$f^i(y) = \int_{C(X)} \dots \int \varphi^i(y,X)dX, \tag{3}$$

where *X* is a vector of individual characteristics observed and *C*(*X*) is the domain on which *X* is defined (cf. Bourguignon and Ferreira, 2005, p.28). Denoting $g^i(y|X)$, the conditional distribution of *y*, an equivalent expression for (3) is:

$$f^i(y) = \int_{C(X)} \dots \int g^i(y|X)h^i(X)dX, \tag{4}$$

with $h^i(X)$ as the joint distribution of all elements of *X* in group *i*.

The observed difference between the two distributions may be decomposed into

$$f^A(y) - f^B(y) = [f^A(y) - f^C(y)] + [f^C(y) - f^B(y)], \tag{5}$$

where $f^C(y)$ represents the counterfactual distribution, which can be constructed for example as

$$f^C(y) = \int_{C(X)} \dots \int g^A(y|X)h^B(X)dX. \tag{6}$$

The first term on the right-hand side of equation (5) gives the effect of different endowment's distributions in group *A* and group *B*. The second term describes the inequalities between two distributions of *y* conditional on characteristics *X*. The main difference with respect to the Oaxaca-Blinder decomposition is that this decomposition refers to full distributions, rather than just to their means. The formula (5) may be applied to any statistic defined on the distribution of outcome variable y: mean, quantiles, summary measures of inequality such as the variance or the Gini coefficient.

Several approaches have been suggested in the literature for estimating the counterfactual distribution $f^C(y)$ (cf. Fortin, Lemieux and Firpo, 2010). An approach proposed by Juhn, Murphy and Pierce (1993) is based on the residual imputation procedure. DiNardo, Fortin and Lemieux (1996) suggested to use a reweighting factor. Donald, Green and Paarsch (2000) used a hazard model approach, Fortin and Lemieux (1998) applied an ordered probit. Machado and Mata (2005) proposed using quantile regression to transform a wage observation y into a counterfactual observation y^C.

2.3. Decomposition of differences in distributions using quantile regression

The standard linear regression assumes the relationship between the regressors and the outcome variable based on the conditional mean function. This, however, gives only a partial insight into the connection. The quantile regression (Koenker and Bassett, 1978) allows the description of the relationship at different points in the conditional distribution of y.

Let us consider the relationship between the regressors and outcome using the conditional quantile function:

$$Q_\theta(y|X) = \Phi^{-1}_{y|X}(\theta, X) = X\beta(\theta)\,, \tag{7}$$

where $Q_\theta(y|X)$ - the θ^{th} quantile of a variable y conditional on covariates X, $\theta \in (0,1)$; $\Phi_{y|X}$ - the joint distribution function for the variable y.

A different quantile θ may be specified (most frequently from three to nine). For each quantile other parameters $\beta(\theta)$ are estimated. These coefficients can be interpreted as the returns to different characteristics X at given quantiles of the distribution of y. Bootstrap standard errors are often used (Gould, 1992; 1997). Quantile regression is more robust than least squares regression to non-normal errors and outliers. This method also provides a richer characterization of data, considering the impact of covariates on the entire distribution of y, not only its conditional mean.

Machado and Mata (2005) used quantile regression in order to estimate counterfactual unconditional wage distributions. Since the unconditional quantile is not the same as the integral of the conditional quantiles, authors provide a simulation-based estimator where the counterfactual distribution is constructed from the generation of a random sample. This estimator is widely used in various applications (cf. Albrecht, Björklund and Vroman, 2003; Melly, 2005). The idea underlying this technique is the probability integral transformation theorem. If U is a uniform random variable on [0,1], then $F^{-1}(U)$ has distribution F. Thus, if θ_1, θ_2, …, θ_m are drawn from a uniform (0,1) distribution, the corresponding m

estimates of the conditional quantiles of wages at X, $\left\{ X\hat{\beta}(\theta_i) \right\}, i = 1,...,m$, constitute a random sample from the (estimated) conditional distribution of wages given X (Machado and Mata, 2005, p.448).

The Machado-Mata approach to generate a random sample from the wage density that would prevail in group A if model (7) was true and covariates were distributed as $h^A(X)$ is as follows:

1. Generate a random sample of size m from a $U[0,1]$: $u_1, \ldots, u_m$.
2. Using the dataset for group A, estimate m different quantile regression $Q_{u_i}(y|X_A)$, obtaining coefficients $\hat{\beta}_A(u_i), i = 1,...,m$.
3. Generate a random sample of size m with replacement from the rows of X_A, denoted by $\{X^*_{Ai}\}, i = 1,..., m$.
4. $\left\{ y^*_{Ai} \equiv X^*_{Ai}\hat{\beta}_A(u_i) \right\}, i = 1,...,m$ is a random sample of size m from the unconditional distribution $f^A(y)$.

Counterfactual distributions could be estimated by drawing X from another distribution and using different coefficient vectors. For example, to generate a random sample from the wage density that would prevail in group A and covariates that were distributed as $h^B(X)$ (e.g. the men log wage density that would arise if men were given women's labour market characteristics but continued to be paid like men), we generate a random sample from the rows of X_B, denoted by $\{X^*_{Bi}\}, i = 1,..., m$, and $\left\{ y^{C*}_{Ai} \equiv X^*_{Bi}\hat{\beta}_A(u_i) \right\}$ is a random sample from the counterfactual distribution $f^C(y)$.

Consequently, the idea of Machado-Mata decomposition of the difference between the wage densities in two groups, $\Delta^\theta = Q_\theta(y_A) - Q_\theta(y_B), \forall\theta$, is as follows:

$$\begin{aligned} \hat{\Delta}^\theta &= Q_\theta(y^*_A|X^*_A) - Q_\theta(y^*_B|X^*_B) = \\ &= Q_\theta(y^*_A|X^*_A) - Q_\theta(y^{C*}_A|X^*_B) + Q_\theta(y^{C*}_A|X^*_B) - Q_\theta(y^*_B|X^*_B) = \\ &= \underbrace{(X^*_A - X^*_B)\hat{\beta}_A(\theta)}_{\hat{\Delta}^\theta_{\text{explained}}} + \underbrace{(\hat{\beta}_A(\theta) - \hat{\beta}_B(\theta))X^*_B}_{\hat{\Delta}^\theta_{\text{unexplained}}} \end{aligned} \qquad (8)$$

We can also use the Machado-Mata approach to estimate standard errors for the estimated densities by repeating the procedure many times and generating a set of estimated densities. The standard error of the estimator diminishes as we increase the number of replications, but estimating a large number of replications is time-consuming especially when the number of observations is high (Melly, 2006).

3. Results of empirical analysis

After outlining the methodology, we now provide the results of our empirical analysis. First, we present the data. Then, we discuss the results of Oaxaca-Blinder decomposition for mean incomes in urban and rural area. In the next step, we present the quantile regressions estimates for models with various predictors. Finally, we implement the Machado and Mata quantile decomposition technique for differences in income densities between inhabitants in urban and rural area.

3.1 Empirical data

We employ data from the Household Budget Survey (HBS) for Poland in 2012. This representative survey is conducted by the Central Statistical Office, Social Surveys and Living Conditions Statistics Department. It is one of the most comprehensive sources of socio-economic information on Polish households and it plays an important role in the analysis of the living standards of the population. It is the source of information on the revenues and outgoings of such socio-economic groups like employees' households, farmers' households, households of the self-employed, households of retirees and pensioners and households living on unearned sources (GUS, 2013).

Our data consist of a sample of 7056 one-person households (5146 town residents and 1910 village residents) containing information on household's monthly available income as well as on reference persons' attributes, such as gender, age, education level, place of residence. Household's available income is defined as the sum of household's current incomes from various sources reduced by taxes, and it comprises: income from hired work, income from a private farm in agriculture, income from self-employment, income from property and social insurance benefits.

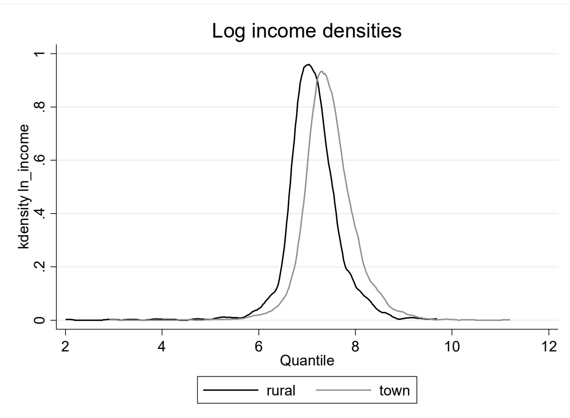


Figure 1. (Log) income densities for the residents of urban and rural areas.

Source: own elaboration based on GUS (2012).

In our empirical decomposition analysis the logarithm of the average monthly available income (*ln_income*) constitutes the outcome variable. Figure 1 illustrates the kernel estimates of the log income densities for the residents of urban and rural areas. The data indicate that the income inequality can be observed. When we look at both distributions, we find that urban residents have the level of the log income higher than the rural residents. The average monthly available income of a person in urban area was PLN 2,028.11, whereas for a person in rural area it was only PLN 1,444.47 (see Table 1).

Table 1. Descriptive statistics for the sample

	Whole sample	Urban area	Rural area
Number of observation	7056	5146	1910
Household's available	1870.13 (1602.47)	2028.11	1444.47
ln_income	7.39 (0.55)	7.46 (0.53)	7.12 (0.55)
sex (% men)	29	28	33
age	60.58 (17.54)	58.88 (18.44)	65.13 (13.86)

Sample averages; standard errors in parentheses.

Source: own elaboration based on GUS (2012).

We establish three explanatory variables in our models: *sex* – the dichotomous variable encoding gender of a person (number 1 coded the male sex), *age* – age of a person in years, *education* – an ordinal variable describing the educational level.

It is useful to look at summary statistics for some covariates (in Table 1). Among the urban residents there were less men (28%) than it was in the case of the rural residents (33% of men). The average age of a person in urban area was only 58.88 years, whereas for a person in rural area it was 65.13 years. The average educational level in the countryside was lower than in cities.

3.2. Results of Oaxaca-Blinder decomposition for differences in mean log incomes

Many authors examined the determinants of income and the income gap in the urban and rural areas. For example, Sicular et al. (2007) and Su and Heshmati (2013) analysed the urban-rural income gap in China using the Oaxaca-Blinder decomposition method. Ali et al. (2013) used this method to analyse the income gap between urban and rural Pakistan. Haisken-DeNew and Michaelsen (2011) investigated the differences in wages between rural and urban workers in the informal and formal sectors of Mexico's labour market. The set of regressors in their papers included conventional human capital characteristics (e.g. education, occupation or experience), personal characteristics (e.g. age, gender, marital status) or regional labour market conditions (Adamchik and Bedi, 2003). The demographic characteristics such as the household size, the proportion of dependents versus working-age household members may also be important for the household incomes (Knight, Song, 1999; Miles, 1997).

The first step of our analysis also included the decomposition of the income inequalities observed between residents of urban and rural areas using the Oaxaca-Blinder technique. The results on an aggregated and detailed basis are presented in Table 2.

Table 2. Oaxaca-Blinder decomposition of mean differences in log incomes for residents in urban and rural areas

Average *ln_income* in urban area			7.460
Average *ln_income* in rural area			7.124
Raw gap (differential observed)			0.336
Aggregate decomposition			
Explained effect			0.185
Unexplained effect			0.151
% explained			55
% unexplained			45
Detailed decomposition			
due to characteristics		due to returns	
sex	-0.002	*sex*	0.035
age	0.008	*age*	-0.268
education	0.179	*education*	0.039
cons	0.000	*cons*	0.344
Total	0.185	Total	0.151

Source: own elaboration using the Stata command 'decompose'.

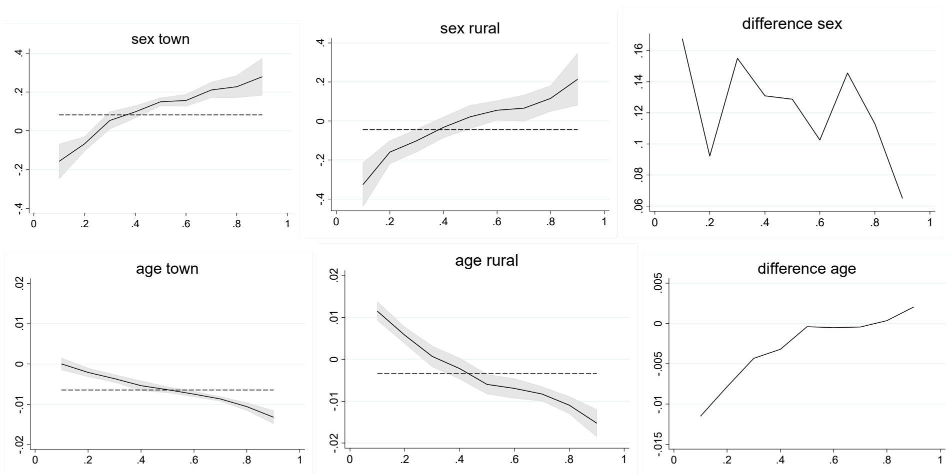
There is a positive difference between the mean values of the log income for urban and rural residents, meaning that the inhabitants in rural area have lower average incomes than inhabitants in urban area. The inequalities examined should be assigned to a similar extent to the characteristics (55%) as well as to the coefficients (45%) of the estimated regression models. Decomposition, which was carried out, made it possible to isolate the factors explaining the inequality observed to a different extent. The strong effect of different education levels of people living in rural and urban areas can be noticed (see the value of 0.179). A different "evaluation" of personal characteristics allow the conclusion that the residents in rural area are discriminated against residents in urban area, but not because of the age of people (due to the negative value -0.268). A large part of the unexplained component lies in the intercept differences (that is, the inter-group differences in other factors were not captured in the model).

Our findings are mainly consistent with that reported elsewhere. Sicular et al. (2007) found for China that education was the only characteristic whose contribution to the income gap was significant. The contribution of education was largely due to differences in the endowments and not in the returns. According to the results of Su and Heshmati (2013), the urban-rural income gap can be explained by attributes of individuals, especially by the level of education and the type of occupation. The educational returns were higher among urban residents. The gender income gap was evident, showing males had higher income than females. For the formal sector in Mexico, Haisken-DeNew and Michaelsen (2011) revealed that only differences in education contribute to the explanation of the wage gap and no differences in coefficients can be identified.

3.3. Estimation of quantile regressions

Some recent studies have decomposed the urban-rural income gap by focusing on the entire distribution of income and not just on the means. Nguyen et al. (2007) and Huong and Booth (2010) adopted the quantile regression method to analyse urban-rural consumption expenditure inequality in Vietnam. Shilpi (2008) and Chamarbagwala (2010) used this method to analyse income gap between urban and rural Bangladesh and India. Matita and Chirwa (2009) analysed the extent of urban-rural welfare inequalities in Malawi using the Machado and Mata decomposition technique.

Therefore, we analyse the quantile regression results for the outcome variable *ln_income* in the second step of our research. The plots in Figure 2 show the coefficient estimates $\hat{\beta}_i(\theta)$ with the associated 95% confidence intervals, obtained by the bootstrap method with 100 replications. For the variables *sex*, *age* and *education* the plots provide information on the coefficients in models estimated using the data from urban area (left column), rural area (central column) and the difference between the parameters $\hat{\beta}_i^{town}(\theta) - \hat{\beta}_i^{rural}(\theta)$ (right column). Additionally, in the first two columns the coefficients estimated by mean regression (OLS) are reported (dotted horizontal lines). The graphs illustrate what is the impact of each covariate on income inequality.



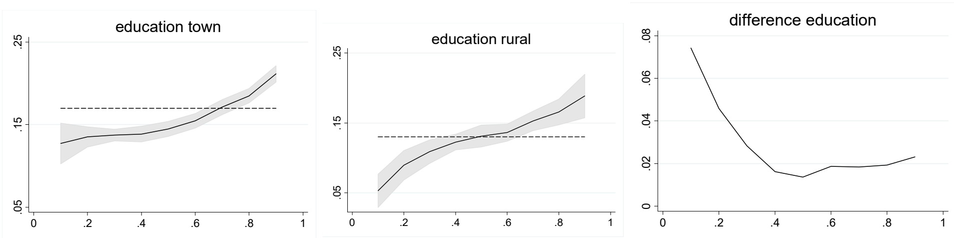


Figure 2. Quantile regression coefficients with 95% confidence intervals for the deciles; the dotted horizontal line represents the least squares (conditional mean) estimate.

Source: own elaboration using the Stata command 'sqreg'.

We can see that among the poorest, the income of men is lower than that of women (see negative coefficients for the urban and rural area), but among the richest being the man gives higher income (see positive coefficients values). The differences between parameter estimates are positive, but from the 70th quantile to the right, they are smaller. This means that the market evaluation of the gender of people is responsible for the existing but decreasing (moving to the right of the distribution) income inequalities between urban and rural inhabitants. It turns out that gender is an important unexplained contributor to the observed income gaps. It is worth mentioning that also for China the gender has greater effects on people with lower level of income in rural area (Su and Heshmati, 2013).

Age is more rewarded among the poorest, but among the richest the growing age leads to the decline in income (both in urban and rural area). We find such an influence of the role of age on the income gap at the bottom of the income distribution. The unexplained part of the gap can decrease due to the negative differences between parameter estimates (compare this with the conclusion of subsection 3.2).

Finally, we find that incomes increase with education across the whole distribution. The education level is the significant contributor to the income differences in urban and rural areas not only in endowments, which favour urban inhabitants (see Table 1), but also in returns of that individual characteristic (note that the differences between parameter estimates are positive although increasingly smaller). These results are contrary to the results obtained by Su and Heshmati (2013) for China, where the education exerts heterogeneous effects on different percentiles of the income distribution. In urban areas, education is more valued for high income earners, whereas for rural areas, specialized or tertiary education is more beneficial for poorer households.

3.4. Empirical decomposition of differences in income distributions for one-person households in urban and rural area

In the third step, we decompose the inequalities in the income distributions into differences in the covariates (individual attributes) and differences attributable to the coefficients (remuneration of individual characteristics). We follow the Machado-Mata procedure as carefully as possible, except that rather than drawing m numbers at random from $U[0,1]$ and then estimating m quantile regression coefficient vectors, we simply estimate the quantile regressions every one percentile (in Table 3 we provide decomposition results only for nine deciles). Then, we make 100 draws at random from the X matrix for each estimated coefficient vector $\hat{\beta}_i(\theta)$.

The results are summarized in Figure 3. The graph on the left plots the differences between pairs of distributions of interest (these are the raw gap, the Machado-Mata differences with the associated 95% confidence interval and the Oaxaca-Blinder mean difference for comparison purposes). The graph on the right presents the decomposition results.

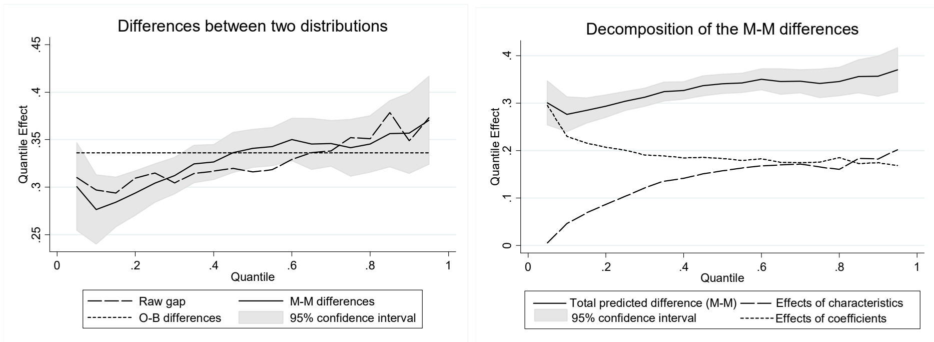


Figure 3. Decomposition of differences in income distributions for residents in urban and rural areas.

Source: own elaboration using the Stata command 'mmsel'.

These results are presented in greater detail in Table 3. The first column of this table refers to the decile number, the second column presents the raw gap between log income distributions for inhabitants in urban and rural areas. The third column contains estimates of the Machado-Mata differences with standard errors in parentheses. The next two columns decompose total inequalities into differences due to the covariates and differences due to changes in the coefficients (standard errors in parentheses). The last two columns give the respective proportions of the total inequalities explained by both kinds of differences.

The findings confirm that in urban area monthly available income of one-person household is on average greater than in rural area, and its inequality at the highest quantiles of the income distribution is also larger among the former than among the latter.

Table 3. Decomposition of differences in income distributions

Decile	Raw gap	M-M differences	Explained effect	Unexplained effect	% Explained	% Unexplained
0.10	0.2971	0.2764 (0.0185)	0.0463 (0.0175)	0.2301 (0.0193)	17%	83%
0.20	0.3096	0.2937 (0.0119)	0.0867 (0.0116)	0.2071 (0.0123)	30%	70%
0.30	0.3046	0.3124 (0.0097)	0.1216 (0.0098)	0.1908 (0.0109)	39%	61%
0.40	0.3168	0.3265 (0.0093)	0.1417 (0.0091)	0.1848 (0.0103)	43%	57%
0.50	0.3161	0.3409 (0.0102)	0.1575 (0.0111)	0.1834 (0.0112)	46%	54%
0.60	0.3292	0.3503 (0.0113)	0.1676 (0.0117)	0.1827 (0.0124)	48%	52%
0.70	0.3382	0.3461 (0.0123)	0.1716 (0.0114)	0.1745 (0.0127)	50%	50%
0.80	0.3511	0.3455 (0.0152)	0.1602 (0.0161)	0.1853 (0.0164)	46%	54%
0.90	0.3488	0.3569 (0.0215)	0.1825 (0.0197)	0.1744 (0.0198)	51%	49%

Standard errors in parentheses.

Source: own elaboration using the Stata command 'mmsel'.

Figure 3 also shows that the income gaps are wider at the top of distribution. Both covariates and coefficients, contribute to the explanation of the total inequalities sum and their effects are significantly different from zero in all of the estimated deciles (the confidence intervals are not provided because of lack of space, but they do not include zero). We can clearly see that the effect of coefficients is more important than that of covariates at the bottom of the income distribution. However, the unexplained differential shrinks as we move toward the top of the income distribution. By contrast, the percentage of the explained (due to the characteristics) income differential is considerably greater as we move to the right-side of the distribution.

Our findings can be compared with the results of Huong and Booth (2010) who found evidence of significant urban-rural expenditure inequality in Vietnam. In this case, the urban-rural gap monotonically increased across the expenditure distribution. Also, Nguyen et al. (2007) analysed the urban-rural consumption expenditure inequality in this country and showed that the returns due to the covariates were larger at the top of the distribution of household consumption expenditure per capita. Regarding the studies of urban-rural income inequalities, Shilpi (2008) and Chamarbagwala (2010) found that both the covariates and the returns were relevant in explaining the observed income gap, although their behaviours were different across the distribution of welfare.

4. Conclusions

The objective of the study was to perform the decomposition of income inequalities between one-person households in urban and rural areas. In order to extend the Oaxaca-Blinder decomposition procedure to different quantile points along the income distribution, we applied the Machado-Mata decomposition technique and constructed the counterfactual income distribution.

It is worth mentioning that the decomposition method applied was computationally intensive. The calculation could be simplified by estimating a specific number of quantile regressions (i.e. 99) instead of generating a random sample of size m from $U[0,1]$. Another limitation was the assumption of the linearity of the quantile regression model. Besides this, the Machado-Mata approach does not provide a way of performing the detailed decomposition for the endowment effect (Fortin, Lemieux and Firpo, 2010, p.61).

Turning to the main findings from the aggregate decomposition of the income disparities in urban and rural areas, we found that the income differentials tend to increase. There are higher differentials at the top of the income distribution, which are driven by the endowment effects as well as by the structure effects. However, the widening income inequalities at higher income quantiles are mainly possible due to growing differences in characteristics of people (especially in educational level) in favour of urban residents. The discrimination affecting rural residents by the higher incomes becomes less important (due to descending structure effects associated with the coefficients of the model). Moving to the bottom of the distribution the total income gap declines, owing primarily to a decline in the explained components.

REFERENCES

ADAMCHIK, V. A., BEDI, A. S., (2003). Gender pay differentials during the transition in Poland. Economics of Transition, 11(4), pp. 697–726.

ALBRECHT, J., BJÖRKLUND, A., VROMAN, S., (2003). Is There a Glass Ceiling in Sweden? Journal of Labor Economics, 21, pp. 145–177.

ALI, L. R., RAMAY, M. I., NAS, Z., (2013). Analysis of the determinants of income and income gap between urban and rural Pakistan. Interdisciplinary Journal of Contemporary Research in Business, 5(1), pp. 858–885.

AUTOR, D. H., KATZ, L. B., KEARNEY, M. S., (2005). Rising Wage Inequality: The Role of Composition and Prices. Cambridge: NBER Working Paper, No. 11628.

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

BOURGUIGNON, F., FERREIRA, F. H. G., (2005). Decomposing Changes in the Distribution of Household Incomes: Methodological Aspects. In: F. Bourguignon, F. H. G. Ferreira and N. Lustig, eds. 2005. The Microeconomics of Income Distribution Dynamics in East Asia and Latin America. Washington: World Bank and Oxford University Press. pp. 17–46.

CARD, D., (1992). The Effects of Unions on the Distribution of Wages: Redistribution or Relabelling? Cambridge: NBER Working Paper, No. 4195.

CHAMARBAGWALA, R., (2010). Economic Liberalization and Urban-Rural Inequality in India: a Quantile Regression Analysis. Empirical Economics, 39, pp. 371–394.

DINARDO, J., FORTIN, N. M., LEMIEUX, T., (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. Econometrica, 64, pp. 1001–1044.

DONALD, S. G., GREEN, D. A., PAARSCH, H. J., (2000). Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates. Review of Economic Studies, 67, pp. 609–633.

FORTIN, N. M., LEMIEUX, T., (1998). Rank Regressions, Wage Distributions, and the Gender Gap. Journal of Human Resources, 33, pp. 610–643.

FORTIN, N., LEMIEUX, T., FIRPO S., (2010). Decomposition methods in economics. Cambridge: NBER Working Paper, No. 16045.

GORAUS, K., (2013). Luka płacowa miedzy kobietami a mężczyznami w Polsce. Available at: <http://grape.uw.edu.pl/gender/wp-content/uploads/sites/35/2013/09/

Łódź_prezentacja.pdf> [Accessed 20 September 2015].

GOSLING, A., MACHIN, S., MEGHIR, C., (2000). The Changing Distribution of Male Wages in the UK. Review of Economic Studies, 67, pp. 635–686.

GOULD, W. W., (1992). Quantile Regression with Bootstrapped Standard Errors. Stata Technical Bulletin, 9, pp. 19–21.

GOULD, W. W., (1997). Interquantile and Simultaneous-Quantile Regression. Stata Technical Bulletin, 38, pp. 14–22.

GRAJEK, M., (2003). Gender Pay Gap in Poland. Economic Change and Restructuring, 36(1), pp. 23–44.

GUS, (2013). Budżety Gospodarstw Domowych w 2012 r. [Household Budget Survey 2012] Warsaw: Central Statistical Office, Social Surveys and Living Conditions Statistics Department.

HAISKEN-DENEW, J. P., MICHAELSEN, M.M., (2011). Migration Magnet: The Role of Work Experience in Rural-Urban Wage Differentials in Mexico. Bochum: Ruhr Economic Papers No. 263.

HUONG, T. L., BOOTH, A. L., (2010). Inequality in Vietnamese Urban-Rural Living Standards, 1993-2006. Bonn: IZA Discussion Paper, No. 4987.

JUHN, CH., MURPHY, K. M., PIERCE, B., (1993). Wage Inequality and the Rise in Returns to Skill. Journal of Political Economy, 101, pp. 410–442.

KNIGHT, J., SONG, L., (1999). The Urban-Rural Divide: Economic Disparities and Interactions in China. New York: Oxford University Press.

KOENKER, R., BASSETT, G., (1978). Regression Quantiles. Econometrica, 46, pp. 33–50.

KOT, S. M., PODOLEC, B., ULMAN, P., (1999). Problem dyskryminacji płacowej ze względu na płeć [Problem of earning discrimination by gender]. In: Kot, S.M., ed. 1999. Analiza ekonometryczna kształtowania się płac w Polsce w okresie transformacji. Warszawa, Kraków: Wydawnictwo Naukowe PWN.

MACHADO, J. F., MATA, J., (2005). Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression. Journal of Applied Econometrics, 20, pp. 445–465.

MATITA, M. M., CHIRWA, E. W., (2009). Rural-Urban Welfare Inequalities in Malawi: Evidence from a Decomposition Analysis. University of Malawi Chancellor College, Department of Economics, Working Paper No. 2009/05.

MELLY, B., (2005). Public-Private Sector Wage Differentials in Germany: Evidence from Quantile Regression. Empirical Economics, 30, pp. 505–520.

MELLY, B., (2006). Estimation of counterfactual distributions using quantile regression. Review of Labor Economics, 68 (4), pp. 543–572.

MILES, D., (1997). A household level study of the determinants of incomes and consumption. The Economic Journal, 107(440), pp. 1–25.

NEWELL, A., SOCHA, M., (2005). The Distribution of Wages in Poland. Bonn: IZA Discussion Paper, No. 1485.

NGUYEN, B. T., ALBRECHT, J. W., VROMAN, S. B., WESTBROOK, M. D., (2007). A Quantile Regression Decomposition of Urban-Rural Inequality in Vietnam. Journal of Development Economics, 83, pp. 466–490.

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

ROKICKA, M., RUZIK, A., (2010). The Gender Pay Gap in Informal Employment in Poland. Warszawa: CASE Network Studies and Analyses, No. 406.

SALARDI, P., (2012). Wage Disparities and Occupational Intensity by Gender and Race in Brazil: An Empirical Analysis using Quantile Decomposition Techniques. Brighton: University of Sussex, Job Market Paper. Available at: <http://www.iza.org/conference_files/worldb2012/salardi_p7646.pdf> [Accessed 20 September 2015].

SICULAR, T., YUE, X., GUSTAFSSON, B., LI, S., (2007). The Urban-Rural Income Gap and Inequality in China. Review of Income and Wealth, 53 (1), pp. 93–126.

SHIPI, F., (2008). Migration, Sorting and Regional Inequality: Evidence from Bangladesh. World Bank Policy Research Working Paper No. 4616.

SŁOCZYŃSKI, T., (2012). Wokół międzynarodowego zróżnicowania międzypłciowej luki płacowej [The issue of international differentiation of inter-gender earning gap]. International Journal of Management and Economics, 34, pp. 169–185.

SU, B., HESHMATI, A., (2013). Analysis of the Determinants of Income and Income Gap between Urban and Rural China. Bonn: IZA Discussion Paper, No. 7162.

ŚLIWICKI, D., RYCZKOWSKI, M., (2014). Gender Pay Gap in the micro level – case of Poland. Quantitative Methods in Economics, Vol. XV, No. 1, pp. 159–173.