

INCOME OF HOUSEHOLD MEMBERS IN POLAND

PAWEŁ ULMAN

Cracow University of Economics, Department of Statistics¹
e-mail: ulmanp@uek.krakow.pl

STRESZCZENIE

P. Ulman. *Dochody członków gospodarstw domowych w Polsce*. Folia Oeconomica Cracoviensia 2015, 56: 23–34.

Poziom dochodów gospodarstw domowych wpływa nie tylko na zaspokojenie indywidualnych potrzeb ich członków, ale również zapewnia odpowiedni standard życia gospodarstwa. Z tego względu wielu badaczy wykazuje zainteresowanie analizami dochodów gospodarstw domowych. W artykule podjęto tematykę kształtowania się dochodów członków gospodarstw domowych. Na podstawie indywidualnych danych o dochodach możliwą staje się analiza struktury źródeł dochodów, kształtujących ogólny dochód badanej osoby.

Celem pracy jest prezentacja rozmiaru i źródeł dochodów ludności Polski w różnych przekrojach badanej zbiorowości. Dane statystyczne zostały zaczerpnięte z *Badania Budżetów Gospodarstw Domowych* przeprowadzonego przez GUS w 2011 r. Zbiór danych zawierał 37375 obserwacji gospodarstw i ponad 100 tys. rekordów, zawierających informacje o indywidualnych osobach. W celu porównania rozkładów dochodów wykorzystane zostały modele teoretyczne tych rozkładów z uzmiennionymi parametrami, w szczególności zastosowano rozkład Burra typu XII. Ponadto wykorzystano regresję kwantylową dla identyfikacji determinant dochodów indywidualnych członków gospodarstw domowych.

ABSTRACT

The size of households income determines not merely the degree of satisfaction of individual needs but also ensures households' adequate social status. It causes an interest many researchers in analysis of households income. The subject of the study is the individual income of the each member of household. When analysing household budgets, financial income from one month is assigned to individual persons. Thus, the level and structure of individual income sources may be identified. The income of single persons acquired from numerous sources forms their individual income.

¹ Publication was financed from funds allocated to Faculty of Management (Cracow University of Economics) within grants to maintain research capacity.

The paper aims to show the size and sources of the Polish people income in the different cross section of the analysed sample. The statistical information is derived from the household budget survey of 2011 that is conducted annually by the Central Statistical Office of Poland. The entire set of individual observations of households constitutes 37,375 records and in the case of household members — over 100,000 records. In order to compare income distributions use is made of theoretical models of income distribution with variable parameters. This approach allows to capture the impact of individual characteristics on the shape of the distribution of income and its descriptive statistics. The theoretical distribution Burr XII has been used as a model of income distribution. Moreover quantile regression model are used for identification of determinants of individual income.

SŁOWA KLUCZOWE — KEY WORDS

rozkład dochodów, model regresyjny, źródła dochodów indywidualnych
income distribution, regression models, sources of personal income

1. INTRODUCTION

Household income is one of the most common economic terms. Most often it is generated through basic human labour, which provides for personal needs. For this reason, it has generated much interest in analysing income levels and its variation from person to person. D. Ricardo and T. Malthus had begun discussions regarding the allocation of income among different members of society many years ago. D. Riccardo analysed the low level of minimum wage in comparison to the large surplus of labour force available. The risk of poverty was similarly interpreted by T. Malthus. While examining the relationship between income levels and population growth, he endorsed a limitation to the birth rate, justifying it with the inevitable expansion of poverty. E. Engel conducted research on household income in a special way, analysing statistical budgets of workers' families. He observed a pattern (known as Engel's law), stating that as household income increases, the proportion of income spent on food decreases, thus increasing the proportion of expenditures on durable goods and services. Further on, studies were conducted on various aspects of income distribution, an example of which may be measuring income inequalities throughout the community. As early as in 1898, von Bortkiewicz proposed the mean deviation as a measure of income inequality, while C. Gini (1912) presented one of the most popular measures of income inequality known as the the Gini coefficient. In parallel, methods of describing income distribution using mathematical functions were developed. The first one who used such a function for models of income distribution was V. Pareto. Since then, many measures have been proposed to characterise income distribution along with many functions to describe these distributions. Various theories were also developed which considered the origin and shaping of these distributions.

This paper presents the issue of income, the source and variation among members of Polish households. The analyses of income are often based on data related to household income, which introduces certain limitations. The characteristics of individual household members and their contribution to the general income of the household are thus omitted. The objective of this paper is to present the level of income and its variation in different social groups of Polish society. In particular, this differentiation has been analysed in terms of the factors which shape it, assuming that their effect is not identical in different income groups of the analysed individuals. The Polish Household Budgets Survey data were used for a statistical analysis. They are described in short in the latter part of the paper. The statistical methods used in this paper are also presented there. The third chapter presents the results obtained from the conducted analyses, followed by conclusions in the final part.

2. MATERIAL AND METHODS

Data was retrieved from household budgets survey (HBS) conducted by the Polish CSO in 2011 and was used for analysing individual income of household members. It is one of the basic surveys for public statistics in Poland. It has been conducted on a continuous basis since 1957, although its method and scope of execution has been altered several times since its beginning. In recent years, it has provided data sets for ca. 37,500 household and over 100,000 individuals. From among the total number of Polish household members, this paper takes into account those at least 18 years of age, bringing the number of total observations to 84,316. Data regarding the income and characteristics of these individuals were acquired in this way. The Burr type XII distribution (or simply the Burr distribution) will be used for analysing the income distribution of household members. It is one of the twelve distributions generated by the Burr distribution system, and is also known as the Singh–Maddala distribution; see Kleiber, Kotz (2003). The Burr type III distribution is another distribution in this family, also known under the name of the Dagum distribution. The paper by McDonald and Xu (1995) p. 139 — includes classification of most of the theoretical distributions used for modelling the empirical distribution of income (wages). Utilizing a tree diagram, they presented the dependencies between particular distributions, from the most general, 5-parametric beta distributions to one-parameter distributions, e.g. the χ^2 distribution. Basing on data related to family income, they concluded that the generalised beta type II distribution best approximates empirical distributions of income. The Dagum distribution proved to be slightly inferior according to the used statistical adjustments measures. Bandourian, McDonald and Turley (2003), p. 9 — reached similar conclusions adjusting various income distribution models for data from

23 countries obtained in several one-year periods. They found that the Dagum distribution is more favourable for 3-parametric distributions rather than when used for 2-parametric distributions. As compared with the generalised beta type II distribution, it is slightly worse according to the adjustments needed to the empirical data, whereas the density function charts of both distributions are almost identical. The Burr type III distribution (Dagum) and Burr type XII distribution (Singh–Maddala) are very similar (also due to the approximation possibilities of empirical distributions) models of income distribution and both were used numerous times to describe the shaping of income and wages as indicated by Kleiber Kotz (2003), p. 197–222. For this reason, it was decided to utilize one of these methods in this paper as a model of income distribution for individual household members in Poland.

The cumulative Burr type XII distribution function may be defined as follows:

$$F(y) = 1 - 1/[1 + e^{ay^b}]^c,$$

where y is income, and a , b , and c refer to distribution parameters which may be estimated by ML. Some of the theoretical distribution parameters may be made a function of the characteristics of the studied units (household members). By following this approach, it allows for a more thorough analysis to the changes in income distribution, depending on the characteristics of the individuals obtaining this income. As a result, we have a conditional income distribution in terms of individual characteristics. The selection of these “conditional” parameters is made on the basis of statistical criteria on adjusting the model to empirical data. Most often, reliable results can be obtained by making parameters a and b as a function of explanatory variables, which in turn allows for examine their impact on shaping all distribution descriptives.

To study the effect of various factors on shaping individual income of household members, regression modelling has been used based on so-called quantile regression. Although it was proposed by Koenker and Bessett in 1978 (1978), its popularity came into light in recent years; see Koenker (2000), Koenker, Hallock (2001), Koenker (2005). A quantile regression produces a more robust parameter estimation when compared to a traditional regression based on the expected value of the dependent variable. As a method of robust statistics, it takes into account the existence of non-typical observations (outliers); moreover, the problem of biased estimation disappears, which in traditional regression modelling could arise and exist when sub-samples were extracted on the basis of income criterion. In reference to a classic regression, the quantile approach allows for the expansion of the analysis of changes in the value of the dependent variable into many points (the quantiles of the dependent variable) and not only to the expected value of this variable. Estimation of

parameters of a quantile regression has a semi-parametric nature, which allows rejecting premises about the type of distribution for the residual vector in the model. In a general case, quantile regression parameters for any quantile are estimated by the minimization of the weighted sum of the value of absolute residues according to the formula:

$$\min_{\beta \in R^k} \sum_{i=1}^n \rho_{\tau}(|y_i - f(x_i\beta)|),$$

where:

$$\rho_{\tau}(z) = \begin{cases} \tau z & \text{for } z \geq 0 \\ (1 - \tau)z & \text{for } z < 0 \end{cases}.$$

The estimation of several quantiles each time for the whole sample results in several vectors of parameter estimates (β). In this way, non-typical (outliers) observations are assigned lower weights, which solve the problem of accounting for them in the model. Most often, standard error of estimates in quantile regressions are obtained using the bootstrap method according to the Gould approach; see Gould (1992), (1997).

3. RESULTS AND DISCUSSION

As it has already been reported above, results of empirical research presented below were obtained on the basis of household budget survey of 2011. These mostly refer to income of individual household members of at least 18 years of age. There were 84,316 observations. These individuals obtained income from the following sources: hired work (HW), self-employment other than a private farm in agriculture and free profession (SEMP), private farm in agriculture (AGR), property (PRO), pension of a property or land (REN), social insurance benefits (SIB), other social benefits (OSB) and other sources (OSC). Figure 1 presents the structure of sources of income of the studied individuals categorized by their age. The highest share in total income of the population is shaped by two sources: hired work and social insurance benefits. The former predominates for individuals below 60 years of age, whereas the latter refers to more elderly individuals. Income from self-employment other than a private farm in agriculture and free profession (SEMP) and private farm in agriculture (AGR) form a relatively unsubstantial share of total income. The highest value of this share in both cases is recorded for persons aged 25 to about 50. Thus, two sources of income are of basic significance for the income distribution of the Polish population: hired work and social insurance benefits, which provide almost 80% of the total income observed.

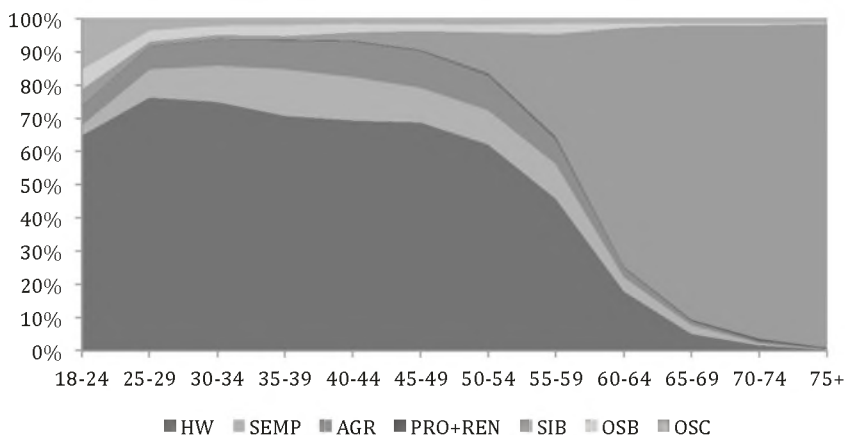


Figure 1. The structure of the individual income from different sources categorized by their age (in %) — Poland 2011

In discussion of the average level of income from the above sources, it should be mentioned that the highest average income is generally obtained by individuals aged 35–39. The main source of income is obtained through hired work and self-employment. A basic decrease in the average income from work and its share in total income is reported for individuals over the age of 50. These individuals often enjoyed the privilege of early retirement or were provided grounds to be awarded pension benefits, especially when a job is lost. Some research examples show that for elderly individuals, the loss of work results in a more difficult task of once again being employed. Thus, the alternative solution is the possibility of obtaining social benefits. Said individuals often decide to terminate professional activity entirely.

The level and structure of the obtained income also varies by gender. The average general income of men is 45% higher than that observed for women. In particular, income of women is clearly lower in both hired work and self-employment, while in the case, generally speaking, unearned sources women have got higher income on average. The share of income from work comprising of total income for men is almost 68%, while for women this share is about 55%. It was observed that women clearly have a higher share of income obtained from social insurance, which is mostly a result of a relatively larger number of women at or above pension age.

The size of the location where the studied individuals reside is directly connected with the variation of the average income. This income is mostly obtained through hired work and self-employment. Its share of total income amounted to 72% in cities and almost 51% in rural areas. The ratio of the average income for residents of large cities to those in the rural areas in Poland in 2011

was 1.46. The higher portions of total income for residents in rural areas were reported from agriculture activities (20.7%) and other social benefits (3.2%).

On average, married women report lower income than single women, whereas the situation is reversed for men. The comparison of income for wives and husbands shows that the average income of men is significantly higher than that of women (by about 66%), while the structure of income sources remains quite similar for both genders. Men are slightly more prevalent regarding income obtained through self-employment, whereas women prevail from other social benefits. Studying the differentiation of income for men and women is an interesting research issue related to the changes in social roles of both men and women in modern societies. The analysis of this problem in the context of costs and benefits in households would allow for an assessment of economic effectiveness of the given family model. However, assessment of some elements of said costs and benefits is certainly difficult to say the least.

Figure 2 presents the Burr type XII density function for three cases: total income, income from hired work and income from social insurance benefits. The parameters of these distributions have been estimated with the ML using GRETL. All the estimates proved to be statistically significantly different from zero. The figure clearly shows the difference between the distribution of income from hired work and the distribution of income from social insurance. The distribution of income from hired work is skewed towards higher values and demonstrates more variation of income in reference to the distribution of income from social insurance. The distribution of income is largely determined by that obtained through hired work. That being said, its shape is similar to the distribution of income from this source rather than to that from social insurance. The proportion of income obtained from hired work out of total income is ca. 54%, whereas social insurance income is about 25%.

The above observations are confirmed in descriptives of the studied distributions (see Table 1): the average income from hired work is significantly higher than income from social insurance. It is interesting to note that the median and most frequent income shows a lesser variation in both absolute and relative views. This analysis indicates less inequality of income from social insurance than from hired work. All the utilized measures of income inequality confirm this observation. It is interesting to note that distribution of total income is most strongly differentiated due to the fact that the total income is obtained from various sources which contain their own individual internal level of inequality. Moreover, the summary of this income additionally generates a range of inequalities resulting from the differences in shaping income from various sources. Asymmetry of distributions is right-skewed, whereas an abbreviated measure of welfare (the Sen index) shows that income from hired work generates higher levels of welfare than income from social insurance.

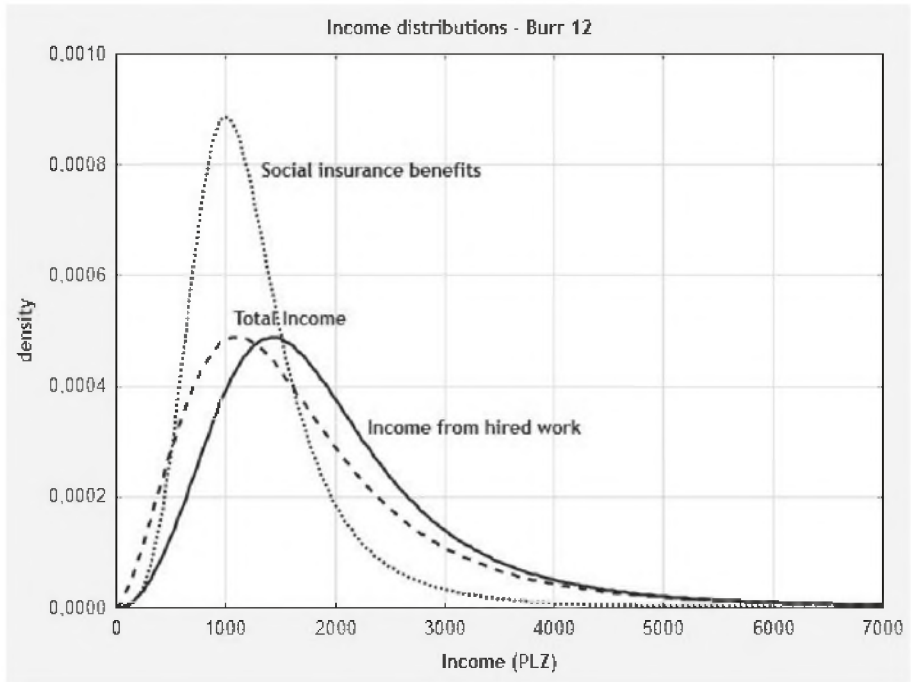


Figure 2. Income distributions (Burr XII) — selected sources of income

Table 1

Descriptives of the selected income distributions on the base Burr XII distribution (in Polish zł)

	Total income	Hired work income	Social insurance benefits
Average	1849.9	2046.4	1309.3
Median	1494.2	1748.3	1157.8
Mode	1100.5	1424.2	1001.0
Coefficient of variation	0.8612	0.6940	0.5772
Relative average deviation	0.2603	0.2202	0.1860
Skewness	0.4704	0.4382	0.4080
Gini	0.3675	0.3130	0.2656
Sen index	1170.0	1406.0	961.5

The discussion of the methods includes the comment that analysis of income distribution may be extended with the assumption that some parameters will become functions of characteristics of the analysed individual. It was decided that both parameters a and b will depend on: the level of education measured on four levels (basic, basic vocational, secondary and higher, with basic education used as the reference level); the place of residence of the analysed individual measured in three categories (large city, small and medium-size cities and villages, with individuals residing in rural areas used as the reference category); gender (a binary variable with female as the reference category); and the age of the analysed individuals in years. The procedure for estimating the parameters of the model would provide statistically significant values. Table 2 presents conditional characteristics of distribution of income from the sources researched earlier and income from self-employment depending on gender. It has been assumed that the analysed individual lives in a large city, has higher education and is 35 years old in the case of income obtained from work, and 70 years old in the case of income obtained from social insurance.

The obtained results confirm the differentiation of income for individual persons by gender, with the smallest range found in income from social insurance. On average the highest level of income applies to self-employed individuals. It is interesting to note that women in this case achieve an average income at the level similar to the average income of men obtained from hired work. The individuals obtaining income from self-employment are certainly characterized by the largest inequalities of income levels, which results in relative equalisation of the values of the Sen index between the incomes from both types of work. This index takes into consideration the average level of income and the level of income inequalities measured by the Gini coefficient. A growth in the value of the average increases the value of the Sen index, whereas a growth in the level of inequalities reduces the value of the said index.

Table 3 presents the results of estimating the quantile regression model for the natural logarithm of income from three sources: hired work, self-employment and social insurance. Three basic factors have been taken into account that mostly determine income from work, i.e. age, gender and level of education. For age, an additional variable has also been used — the square of age, with the objective of capturing non-linearity of income to age relationships according to the theory of wages and age profiles. It assumes that the average income from work (wages) increases with age up to a certain level and decreases in the case of older individuals. The gender variable is by default a binary variable with a female reference category. In order to describe the relationship of income and education, a binary variable set has been used (Educ2–Educ4) which refer to basic vocational, secondary and higher education, respectively. Basic education is the reference category.

Table 2

Descriptives of the selected conditional income distributions on the base Burr XII distribution

	Hired work income		Self-employment income		Social insurance benefits	
	Man	Woman	Man	Woman	Man	Woman
Average	2853.0	1968.9	3706.2	2822.9	2121.8	1901.7
Median	2534.4	1769.4	3001.0	2217.2	1954.9	1762.9
Mode	2142.6	1522.4	1934.4	1443.5	1776.2	1611.6
Coefficient of variation	0.5828	0.5499	0.8058	0.9000	0.4352	0.4175
Relative average deviation	0.2016	0.1925	0.2792	0.2862	0.1484	0.1441
Skewness	0.4272	0.4124	0.5933	0.5429	0.3743	0.3653
Gini	0.2867	0.2741	0.3907	0.4014	0.2127	0.2067
Sen index	2035.1	1429.2	2258.2	1689.8	1670.5	1508.6

The complete estimates of the parameters proved to be statistically significant even for the squared age variable, in the case of which the estimate for income from social insurance proved to be at a very low level. This shows that this income has almost a linear course in reference to age. It is also interesting to observe the levels of the constant value for various quantiles, especially in the case of income from self-employment. These values are clearly varied here, especially in the first and third quartiles. This situation indicates a high level of variation in income from this source, which is also found in Table 2. However, on the other hand, this variation is compensated with a higher positive effect of individual characteristics of the analysed individuals on income from self-employment in the regression model determined for the 25th quantile in reference to the 75th quantile model, which is expressed in estimates of parameters of these models.

In all cases, women obtain lower income. At the same time, the difference grows with an increase in income class from both hired work and social insurance. In the case of income from self-employment, the largest variation in income in terms of gender is found in the lowest income class. It is also interesting to note that a variation in income in this approach is not uniform in terms of the source of income. Women are often observed behind men in the case of income from self-employment and ahead for income from social insurance. The problem of income (wages) variation for men and women is one of more important research issues in which the possibility of discrimination of women is often analysed.

The education level of individual persons is an important factor that affects income. The results presented in Table 3 clearly show that income grows with the increase in the level of education. The most significant increases are observed

in the case of income from self-employment and, similar to that of the gender variable, decreases as the income class progresses. In the case of income from hired work and social insurance, the inter-quartile variation is relatively small. The above results show that variation in the total income of the society has different sources which generate inequalities in the distribution of income with varied strength. For this reason, changes in the structure of the sources of income may result in changes in the level of income inequalities for the total population, even when the level of inequalities generated in the distributions of income from individual sources does not increase.

Table 3

The outcomes of quantile regression for selected income

Income	Quantile	Const.	Age	sq Age	Sex	Educ2	Educ3	Educ4
Hired work	25th	4.6874	0.0971	-0.0011	0.2948	0.2376	0.4048	0.7519
	50th	5.3627	0.0761	-0.0009	0.3231	0.1470	0.3418	0.7099
	75th	5.5926	0.0749	-0.0008	0.3749	0.1409	0.3504	0.7843
Self em-employment	25th	3.2093	0.1333	-0.0015	0.4709	0.6677	0.8452	1.1531
	50th	4.5826	0.1034	-0.0012	0.3585	0.4654	0.6570	0.9406
	75th	5.6154	0.0833	-0.0010	0.3525	0.2575	0.4807	0.7998
Social insurance	25th	4.9727	0.0339	-0.0002	0.1418	0.1419	0.3554	0.6827
	50th	5.4350	0.0273	-0.0001	0.2257	0.1926	0.4071	0.6885
	75th	6.0891	0.0191	-0.0000	0.3010	0.1661	0.3593	0.6346

4. CONCLUSION

Research of social phenomena requires continuous actualization. This is due to dynamic changes in both the analysed phenomenon and the impact of the surroundings within which the phenomenon is analysed. Household income is such a phenomenon and has many research studies devoted to it. Despite this, its nature has not been thoroughly explained.

The paper presents the results of the research on the shaping of income of household members in Poland, whose level has significantly increased, especially after Poland joined the European Union. Income of the Polish population comes mostly from two sources: hired work (ca. 54%) and social insurance (ca. 25%). The largest average level of income from hired work was observed for persons

aged 35–39, whereas persons of over 60 years of age obtained their main source of income from social insurance. While analysing various social groups of individuals included in the study, income may be shown as differing in both the level and the structure of sources as a result of gender, place of residence and marital status. Studying the variation of income in different household types due to their demographic composition may bring about interesting observations in the context of evolving social roles of both men and women. However, such discussions would require special statistical research.

It is also interesting to note that the distribution of income from different sources differs not only with the level of average values, but also with the level of income inequality in these distributions. This then poses the question of the origins of income inequalities in the given society, which may in part be due to changes in the structure of income sources. Thus, assuming an increase in the number of elderly people in the society who receive income from social benefits, one may believe that the general level of income inequality will be lower due to a smaller variation in income from this source. On the other hand, these trends cannot be foreseen in research of social phenomena not only with certainty, but even with high probability.

The application of a quantile regression allowed finding the relationships between income from various income classes and the select characteristics of the studied individuals. It turns out that the effect of these characteristics is not uniform for individuals of different income levels.

The issues related to income in the population, although studied for years, continuously bring about many challenges and observations.

REFERENCES

- Gini C. (1912), *Variabilità e Mutabilità*, Tipografia di Paolo Cuppi, Bologna.
- Kleiber Ch., Kotz S. (2003), *Statistical Size Distribution in Economics and Actuarial Sciences*, John Wiley & Sons, Inc, Hoboken, New Jersey.
- McDonald J.B., Xu Y.J. (1995), *The Generalized Beta Distribution with Application*, *Journal of Econometrics*, 66, 133–152.
- Bandourian R., McDonald J.B., Turley R.S. (2003), *A Comparison of Parametric Models of Income Distribution Across Countries and Over Time*, *Estadística*, 55, 135–152.
- Koenker R., Bassett G. (1978), *Regression Quantiles*, *Econometrica*, 46 (1), 33–50.
- Koenker R. (2000), *Galton, Edgeworth, Frisch, and prospects for quantile regression in econometrics*, *Journal of Econometrics*, 95 (2), 347–374.
- Koenker R., Hallock K.F. (2001), *Quantile Regression*, *Journal of Economic Perspectives*, 15 (4), 143–156.
- Koenker R. (2005), *Quantile Regression*, Cambridge University Press.
- Gould W.W. (1992), *Quantile Regression with Bootstrapped Standard Errors*, *Stata Technical Bulletin*, 9, 19–21.
- Gould W.W. (1997), *Interquantile and Simultaneous-Quantile Regression*, *Stata Technical Bulletin*, 38, 14–22.