

Hired labour *versus* self-employment. Comparison of income distribution

Paweł Ulman¹, Agnieszka Wałęga²

Abstract

The size of labour income – the main source of earning a livelihood – determines not merely the degree of satisfaction of individual needs but also ensures households' adequate social status. The paper aims to compare the economic situation of people earning a living from hired labour with that of self-employed people. The research hypothesis assumes that those two sources of income generate different income levels and different levels of inequality of income distribution. The study identifies specific characteristics of the individual persons that differentiate distributions of income and variables characterizing their environment. In order to compare income distributions use is made of theoretical models of income distribution with variable parameters, and to identify factors differentiating income use is also made of logit models. All calculations rely on individual sets of data derived from a 2011 household budget survey.

Keywords: *labour income, income distribution model*

JEL Classification: *R2, C3, C35, D3*

AMS Classification: *62P25*

1. Introduction

Labour income is the main source of income and determines the population's wealth. Its level is crucial not only for the satisfaction of individual needs, but also for the quality of life and social status of household members. The paper aims to analyse differences in distributions of total labour income and income by source (hired labour and self-employment). The research hypothesis assumes that these two sources of income generate different levels of income and further that there are different levels of inequality in their distribution. The basic socio-demographic characteristics of people drawing these two types of income, such as their age, level of education, gender and being (or not being) a disabled person are used as the discriminating features of these income distributions. According to Mincer's theory regarding income distribution, the first two of these attributes are crucial as they have a major impact on the creation of human capital [9, 12]. Age measures (albeit imperfectly) an employee's experience, while the level of education reflects an employee's qualification to discharge occupational duties which arises from formal education. However, in the discussion of the

¹ Cracow University of Economics, Department of Statistics, Rakowicka 27, 31–510 Kraków, Poland, pawel.ulman@uek.krakow.pl

² Cracow University of Economics, Department of Statistics, Rakowicka 27, 31–510 Kraków, Poland, agnieszka.walega@uek.krakow.pl

creation of human capital, alongside its universally acknowledged constituents such as skills, experience and knowledge, a number of other elements that are specific to the economic mainstream, such as personality, appearance, and a system of values and cultural aspects have also been acknowledged [8, p. 94]. This, in turn, allowed Becker to construct demand and supply based models of the elements constituting human capital. Currently, the representatives of the so-called Chicago School have fine-tuned these models, effectively making them much more complex, by incorporating into them a number of new elements, such as the level of health, social environment or migration³.

2. Research methods

In order to describe income one can use empirical and theoretical distributions. The first approach describes income in terms of characteristics of an empirical distribution based on which conclusions are drawn as to the position of average income in the distribution, income variability, asymmetry of the distribution and concentrations of income in this distribution. The second approach involves the use of a suitable mathematical function with which the studied distribution is described. The function is called a density function. The whole difficulty in this case involves identifying a function which will best way reflect the distribution of income in the population. Information on this sought-after distribution is obtained from a study of a sample, i.e. an empirical distribution. The use of density functions, which are models of income distribution, has its advantages and disadvantages. The advantages primarily include ease of determining the particular characteristics of the distribution regardless of the manner of presentation of empirical data (individual data or aggregate data with open extreme ends), while weaknesses involve a difficulty in "fitting" a density function (model) to the empirical data. In this study of income from hired labour and self-employment use is made of the latter approach. The authors use the Singh-Maddala distribution (Burr type 12). Very much like the Dagum distribution (Burr type 3) it is a special case of the generalised Beta distribution as described by McDonald [7]. According to McDonald and Xu [6], the type two generalised Beta distribution best approximates empirical distributions. However, the above-mentioned Burr distributions prove to be slightly worse as

³ It should be noted that human capital consists of two distinct groups of components: individual capital resources of individuals and resources which arise from organization and the existence of teamwork. Explanation of the nature of individual intellectual capital consists in computing certain intangible resources that a given individual is endowed with without his active participation (membership of a particular family, hereditary occupation, social status, family's financial situation, as well as his talent, intelligence and health) and those acquired through continuous education, training and experience. In addition, these also involve an ability to use the resources for the benefit of the individual and society and include such components as resourcefulness, expertise and employees' innovation [10].

income distribution models. Similar conclusions were drawn by Bandourian, McDonald and Turley [1, p. 9] when they fitted different income distribution models for 23 countries from different periods. The researchers concluded that compared to the type 2 Beta generalised distribution they prove to be slightly inferior in terms of the applied measures of the model's fit to empirical data, whereas the density function graphs of these distributions are almost identical.

In Poland Burr distributions were used by Kot [3, 4, 5], Kośny [2], Ulman [13] to describe the empirical distributions of salaries, income and expenditures. As these distributions offer a good approximation of empirical distributions of income, in the later part of the study the authors will use the Burr type 12 distribution to describe the shape of labour income distribution.

The Burr type 12 density distribution function has the following formula: $f(y) = c \cdot b \cdot e^a \cdot y^{(b-1)} / [1 + e^a y^b]^{c+1}$, whereas the cumulative distribution function may be expressed as: $F(y) = 1 - 1/[1 + e^a y^b]^c$, where y means income (or expense) per capita, a , b , c are distribution parameters which must be determined with maximum likelihood estimation. When one has parameter estimates at one's disposal, one may determine the values of individual characteristics of the distribution [5, pp. 124–126].

The application of the above theoretical distribution can be extended when parameters a , b , c are turned into functions of characteristics of people who draw the tested income. This allows capturing the impact of these characteristics on the entire distribution and not on just one of its characteristics. The function's formula combining parameters with individuals' characteristics and the decision as to which parameters to connect with what attributes remain an open question.

In order to study the potential impact of factors on the chance (probability) of the occurrence of certain events, use is made of probability models. The most commonly used ones include the logit model, whose analytical form is as follows [11, p. 497]:

$$P(Y = 1) = \frac{\exp(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k)}{1 + \exp(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k)} \quad (1)$$

where: Y is a zero-one dependent variable, X_1, \dots, X_k are explanatory variables, while $\alpha_0, \dots, \alpha_k$ are the model's parameters. The paper uses the logit model to determine the impact of certain factors on the probability of drawing income from work.

3. Empirical results

The analysis of labour income is based on data derived from a household budget survey conducted by the Central Statistical Office in 2011. The authors take into account income from hired labour and income from the self-employment of individuals (household members) aged at least 15. In other words, in this case, the level of labour income also accounts for the involvement of each individual in the acquisition of the income (expressed, for example, in terms of the number of jobs), depending also on factors such as the individual's gender or disability. Ultimately, 38 149 people drawing income from at least one of two sources: hired labour and self-employment, are considered.

Persons drawing income from hired labour account for 38.05% of the sample, compared to a mere 4.46% of the respondents who draw income from self-employment. Of the surveyed people, income from both of these sources is drawn by a mere 0.4% of the respondents. Generally, income from self-employment is higher than that drawn from hired labour. Income from self-employment in Poland is over 35% higher than that from hired labour. For people who draw this income abroad, this income type is over 22% higher. The same situation is revealed for income drawn from casual work abroad (over 22% higher).

In order to determine the impact of certain factors on the probability of drawing income from labour, the parameters of the logit model are estimated. The dependent variable is dichotomous and assumes the value of one if the person draws income from hired labour and zero if he draws income from self-employment. The set of potential explanatory variables includes characteristics of individual persons: gender, age (in years), level of education (five zero-one variables – the benchmark: at most lower secondary education (gymnasium)) and characteristics of the household in which these individuals function: the number of persons in the household; class of household's locality (six zero-one variables - benchmark: household located in rural areas), financial situation (five zero-one variables - the benchmark: a situation described as "average")

Of all of the above variables, only the number of people in the household and most of the zero-one variables characterizing the residence of the household occur to be statistically insignificant (at the significance level of 0.05). However, an increase in the age of the person depresses the probability of the person drawing income from hired employment (*ceteris paribus*). Assuming that the other factors remain constant, a male is less likely to draw income from hired labour than a female. The probability of obtaining income from hired labour is also reduced when the level of the person's education declines. A householder better assessing his financial situation (compared to householders assessing their situation as

"average") is less likely to draw income from hired labour. The opposite situation can be ascertained in households assessing their situation as "bad". Only people living in urban areas with populations between 100 000 to 199 000 residents are more likely than residents of other localities to draw income from hired labour (*ceteris paribus*).

Description	Parameter estimates	Standard error	Wald's statistic	p-value
Constant/intercept	3.9809	0.0798	2491.675	0.0000
Gender	-0.6795	0.0371	335.488	0.0000
Age	-0.0247	0.0015	283.236	0.0000
City – 100.000 to 199.000 residents	0.1611	0.0687	5.499	0.0190
Level of education – secondary general	-0.4482	0.0650	47.619	0.0000
Level of education – secondary vocational	-0.4135	0.0436	89.962	0.0000
Level of education – higher	-0.1663	0.0488	11.605	0.0007
Material situation of the household – „very good”	-0.8570	0.0877	95.497	0.0000
Material situation of the household – „rather good”	-0.6325	0.0380	276.722	0.0000
Material situation of the household – „rather bad”	0.4635	0.0782	35.098	0.0000
χ^2	$\chi^2(9)=1185.6$ p=0.0000			

Table 1 Results of parameters estimation of the logit model for the dichotomous variable of income from labour.

In order to describe the empirical distributions of income drawn from labour use is made of the Burr type 12 income distribution model, in which parameters b and c become linear functions of the characteristics of the surveyed persons. The results of the MLE of the model are presented in Table 2. It is assumed that parameter b will be a function of age and level of education, while c – one of gender and disability. Formally, the parameterisation of the distribution can have the following formula:

$$b = b_0 + b_1 \cdot age + b_2 \cdot age^2 + b_3 \cdot education$$

$$c = c_0 + c_1 \cdot gender + c_2 \cdot disability$$

where: age represents completed years, $education$ is a variable with three categories (1– at most secondary vocational education; 2 – secondary education; 3 – higher education), $gender$ comprises two categories (0 – females and 1 – males), while $disability$ indicates whether the person is disabled (1), or able-bodied (0). In the course of the study, all grades occurred to be

statistically significant, while the Akaike criterion shows that the model reflects the empirical data in the best way of several initially estimated models.

Parameter	Parameter estimates	Standard error	<i>t-value</i>	<i>p-value</i>
a	-23.6459	0.1117	-211.6	0.0000
b ₀	4.0967	0.0273	150.0	0.0000
c ₀	2.2846	0.0393	58.1	0.0000
b ₁	-0.0367	0.0007	-55.8	0.0000
b ₂	0.0004	0.000008	53.2	0.0000
b ₃	-0.1295	0.0015	-83.7	0.0000
c ₁	-1.1752	0.0255	-46.2	0.0000
c ₂	1.2717	0.0566	22.47	0.0000

Log *L* = -313524.9; Akaike'a = 627065.9

Table 2 Estimation results of the parameterised Burr type 12 distribution of total income drawn from labour.

Description	Total labour income						Income from	
	Group 1.	Group 2.	Group 3.	Group 4.	Group 5.	Group 6.	hired labour	self-employment
Mean	3200.15	2148.11	2107.59	1653.96	4165.41	2753.56	3081.80	3515.62
Median	2708.73	2004.45	1972.45	1437.99	3485.88	2554.87	2660.37	2887.29
Mode	2186.60	1790.14	1768.49	1202.50	2771.90	2257.46	2200.69	2005.10
V	0.7240	0.4730	0.4671	0.6389	0.7612	0.4888	0.6614	0.7818
ROP	0.2250	0.1789	0.1774	0.2068	0.2323	0.1842	0.2125	0.2665
Skewness	0.4375	0.3523	0.3445	0.4272	0.4395	0.3686	0.4322	0.5496
Gini	0.3196	0.2537	0.2515	0.2944	0.3296	0.2611	0.3023	0.3743
Sen Index	2177.49	1603.20	1577.47	1167.09	2792.56	2034.58	2150.17	2199.75

Notes: V – coefficient of variation; ROP – relative average deviation; Group 1. – able-bodied males aged 30 with higher education; Group 2. – able-bodied females aged 30 with higher education; Group 3. – disabled males aged 30 with higher education; Group 4. – able-bodied males aged 30 with secondary vocational education; Group 5. – able-bodied males aged 45 with higher education; Group 6. – able-bodied females aged 45 with higher education.

Table 3 Descriptive statistics and inequality measures of labour income distribution for selected groups of people.

Based on the estimated parameters one can determine the descriptive statistics of labour income distribution for different groups of people. The authors take into account six groups of people with different characteristics of the total labour income and one group of people for income from hired labour and one group of self-employed people. In the latter two cases, the income distribution of able-bodied males aged 30 with higher education is examined. These results correspond with estimates for total labour income for Group 1.

Based on the data in Table 3, it can be seen that the distribution of labour income is differentiated by such people's features as: age, education, gender, and disability. The higher the level of education is, the higher the level of labour income. In addition, males draw higher income than females, as do able-bodied persons in relation to people with disabilities. Typically, the level of inequality in the distribution of income is related to the size of the salary. The higher the income, the higher the level of income inequality, although in the case of group 4 one can note a low average level of income and a relatively high level of inequality in the distribution of the same. Introduction of the variable of age^2 to the model allowed (with a statistically significant parameter for this variable) capturing non-linearity in the relationship between the distribution characteristics and age. This leads, among others, to the determination of an income-age profile for a selected group of people.

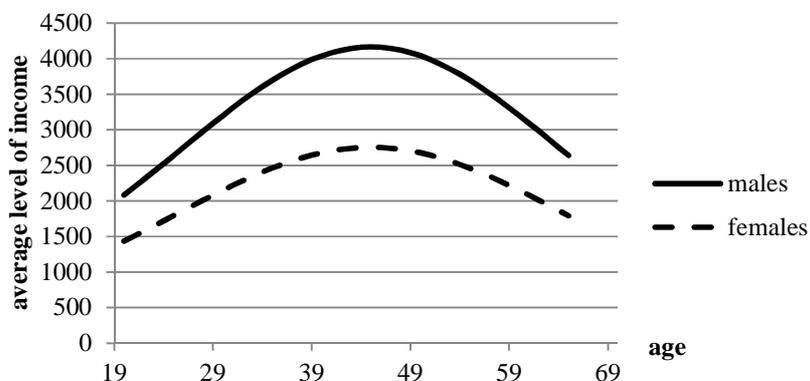


Fig. 1. Income-age profile for males and females determined with the theoretical Burr type 12 distribution.

Figure 1 shows the shape of such a profile (based on the average value) for a group of able-bodied males and females with higher education. It should be noted that the result is consistent with the shape of a typical income-age profile in which the income initially increases and then, once the maximum value has been reached, falls.

4. Conclusions

Based on the study it can be concluded that income from self-employment is higher than that obtained from hired labour. Distributions of income from hired labour and self-employment differ not only on account of location, but also in terms of the level of inequality. A high average income and high level of inequality in the case of income from hired labour causes the social welfare resulting from labour income (measured by means of the Sen index) for these groups to be similar. The logit model reveals that the older the working person, the greater the chance that he will draw income from self-employment. Males and people with lower education are more likely to draw income from this type of work. The use of the parameterized theoretical Burr type 12 distribution also allows capturing the impact of such factors as age, education, gender, and disability on all of the characteristics of the distribution. This gives room for an overall examination of changes in the distribution between different social groups.

References

- [1] Bandourian, R., McDonald, J.B., Turley, R.S., 2003. A Comparison of Parametric Models of Income Distribution Across Countries and Over Time. *Estadistica* 55, Luxemburg Income Study Working Paper No. 305, www.lisproject.org/publications/liswps/305.pdf.
- [2] Kośny, M., 2001. Probabilistyczne modele rozkładu dochodów. Weryfikacja empiryczna. *Wiadomości Statystyczne* 7, 20-35.
- [3] Kot, S.M., 1995. Modelowanie poziomu dobrobytu. Teoria i zastosowanie. Wrocław-Warszawa-Kraków: Ossolineum.
- [4] Kot, S.M., 1999. Empiryczna weryfikacja postaci rozkładu płac w Polsce. In: Kot, S.M. (ed.), *Analiza ekonometryczna kształtowania się płac w Polsce w okresie transformacji*, Warszawa-Kraków: Wydawnictwo Naukowe PWN, 64-98.
- [5] Kot, S.M., 2000. *Ekonometryczne modele dobrobytu*. Warszawa-Kraków: PWN.
- [6] McDonald, J.B., Xu, Y.J., 1995. A Generalization of the Beta Distribution with Application. *Journal of Econometrics* 66, 133-152.
- [7] McDonald, J.B., 1984. Some Generalized Functions for the Size Distribution of Income. *Econometrica* 52 (3), 647-663.
- [8] McMahan, G.C., Wright, P.M., 2011. Exploring human capital: putting human back into strategic human resource management. *Human Resource Management Journal* 21 (2), 93-104.
- [9] Mincer, J., 1958. Investment in human capital and personal income distribution. *Journal of Political Economy* 66, 281-302.
- [10] Ployhart, R.E., Moliterno, T.P., 2011. Emergence of the human capital resource: a multilevel model. *Academy of Management Review* 36, 127-150.
- [11] Scott, A.J., Wild, C.J., 1991. Fitting Logistic Regression Models in Stratified Case-Control Studies. *Biometrics* 47, 497-510.
- [12] Toda, A.A., 2012. The double power law in income distribution: Explanations and evidence. *Journal of Economic Behavior & Organization* 84, 364-381.
- [13] Ulman, P., 2008. Nierówności ekonomiczne w Polsce w 2005 roku. Ujęcie obiektywne i subiektywne. In: Woźniak, M.G. (ed.), *Nierówności społeczne a wzrost gospodarczy w kontekście spójności społeczno-ekonomicznej*, Uniwersytet Rzeszowski, 13, 167-178.