

Group Function of Income Distribution in Society

Sergey G. Fedosin

Sviazeva Str. 22-79, Perm, 614088, Perm region, Russian Federation

email intelli@list.ru

Keywords: annual income; economic modeling; distribution of income.

Abstract. Based on the similarity of properties of photons and money, and on the formula for the density of distribution of photon gas by energies, the corresponding mathematical formula for distribution of annual income per capita is obtained. Application of this formula for the data analysis reveals several independent groups of population with different average levels of their income. In particular four main groups of population contribute to the distribution of income in the economy of the USA.

Introduction

Due to variety of reasons, in every society there is inequality in income distribution, which divides people into different groups. It sets a challenge to the economics, which is trying to describe mathematically this inequality, to understand its essence and to give such recommendations, which could bring the society close to the optimal state. Among the achieved results we shall note the Lorenz curve, based on which the dependence is built of the total share of the society's income P in percentage (the vertical axis) on the share of families N in percentage (the horizontal axis). If each share of families had the same income, we would have a linear dependence:

$$P = \alpha N, \quad (1)$$

where $\alpha = 1$ is the constant coefficient of proportionality.

However, the income of families in the different groups varies. Usually, the whole society is divided into 10 shares equal by the number of people, and within each share the family incomes do not differ much. These shares are called deciles. Let's plot the Lorenz curve not for deciles but for quintiles, each of which contains 20 % of the population. We shall start with the lowest income quintile and finish with the highest income quintile. For the first quintile (and for all the rest) the share of families is $n_1 = 0.2$, and since the income share p_1 of the first quintile is small, the slope

$\alpha_1 = \frac{P_1}{N_1} = \frac{p_1}{n_1}$ of the segment from the origin to the point A in Figure 1, the coordinates of which on

the coordinate plane are (N_1, P_1) , is small too. In order to obtain the second point of the Lorenz curve, it is necessary to add the shares of families in the first two quintiles, which gives $N_2 = n_1 + n_2 = 0.4$. We also need to add the income shares of both quintiles and obtain the number $P_2 = p_1 + p_2$. After that the points $A = (N_1, P_1)$ and $B = (N_2, P_2)$ are connected on the graph with a straight line. Obviously, due to the increased incomes in the second quintile the slope of the second segment will increase: $\alpha_2 = \frac{P_2 - P_1}{N_2 - N_1} = \frac{p_2}{n_2} > \alpha_1 = \frac{p_1}{n_1}$. Indeed $n_2 = n_1 = 0.2$, and $p_2 > p_1$.

Continuing this procedure, we see that the Lorenz curve is bent upwards. If i is the number of a segment of the curve, then for each segment the slope will equal $\alpha_i = \frac{p_i}{n_i}$. The greater is the total

deviation of the Lorenz curve from the straight line of the form (1), which connects the origin and the end of the curve, the greater is the difference in income for different groups of population.

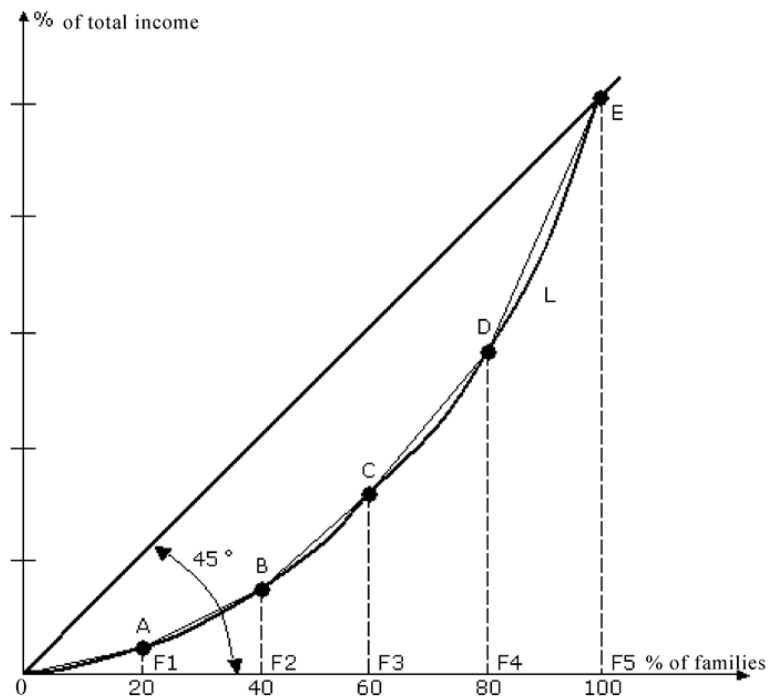


Fig. 1. The typical Lorenz curve for quintiles.

For example, we shall present the statistical data for Russia in 2006 – 2007 according to [1]. The incomes of 20 percent groups of population in 2007 were as follows: in the first group – 5.1 % (in 2006 it was 5.2 %), in the second group – 9.8 % (9.9 %), in the third – 14.8 % (15 %), in the fourth – 22.5 % (22.6 %), in the fifth group with the highest incomes, they were 47.8 % (47.3 %).

We can make the following specification for the poorest and the richest deciles: in 2007 the 10 % share of the wealthiest population was equal to 31% of the total money income (in 2006 – 30.6 %), and 10 % of the least wealthy population had only 1.9 % (1.9 %). If we divide 31 % by 1.9 % we shall obtain the decile coefficient equal to 16.3. This gap between the rich and the poor in Russia is obscenely large because in all developed countries the decile coefficient ranges from 6 to 9. For comparison in 1991 in the beginning of “perestroika” (restructuring) in Russia, this coefficient was equal to 4.5 [2].

The Lorenz curve is associated with another characteristic, which is called the Gini coefficient (the index of income concentration). This coefficient is determined as the ratio of the area of the figure lying between the segment OE and the curve $OABCDE$ to the area of the triangle OEF_5 in Figure 1. Obviously, in case of uniform distribution of income the Gini coefficient tends to zero, and in case of the extreme income inequality it reaches unity.

The Gini coefficient in Russia in 2007 equaled 0.422 as against 0.416 in 2006. The Gini coefficient is the standard tool for comparing countries with each other in the global economy. It turns out that in terms of income inequality Russia is at the level of Latin American countries [3]. If in Brazil the Gini coefficient is 0.57, in Venezuela – 0.486, then in Belarus it is 0.25, in the Czech Republic – 0.27, in Ukraine – 0.28, in Germany – 0.283, in France – 0.327, in Kazakhstan – 0.33, in the UK – 0.368, in India – about 0.37. Even lower values of the Gini coefficient are in the Scandinavian countries (Sweden and Denmark – about 0.25).

Figure 2 shows changing of the density of the distribution function of the per capita annual income in the Soviet Union and the countries of the former Soviet Union from 1970 to 2000 according to [4].

The function is shown with accuracy of the order of 1 %, as the variation range of the income was divided into 100 intervals according to centiles (each centile contains 1 % of the population, which has only slight difference in the income), and the distribution function is calibrated. This

allows us by adding the number of people in 100 intervals to find the total number of people that equals the number of population. In addition, the sum of the products of the number of people and the average income in the intervals gives the total income of the society. The vertical line represents the income level of \$ 1 a day, according to the prices in 1985, as the level of absolute poverty according to the definition of the World Bank.

From Figure 2 we see rather complicated structure of the distribution function. If we build this function on the ordinary but not on the logarithmic scale, then we could see the long falling tail in the region of high incomes. In 1897 the Italian economist Vilfredo Pareto tried to present in terms of quantity this decline with the help of the power function of the form [5]:

$$f(p) = \text{const} \cdot p^{-\kappa}, \quad (2)$$

where the index κ is of the order of and less than 2,
 p is the income of the citizens or businesses.

Function (2) is called the Pareto distribution for the number of people (companies) depending on their income, and was intended to analyze the nature of the income inequality in the society. Mathematicians and economists have also tried to describe the tail of the income distribution function by exponential functions. Thus according to the estimates in [6], in the UK and USA the income and the property are distributed mainly exponentially, and only a small part of the richest population satisfies the Pareto distribution.

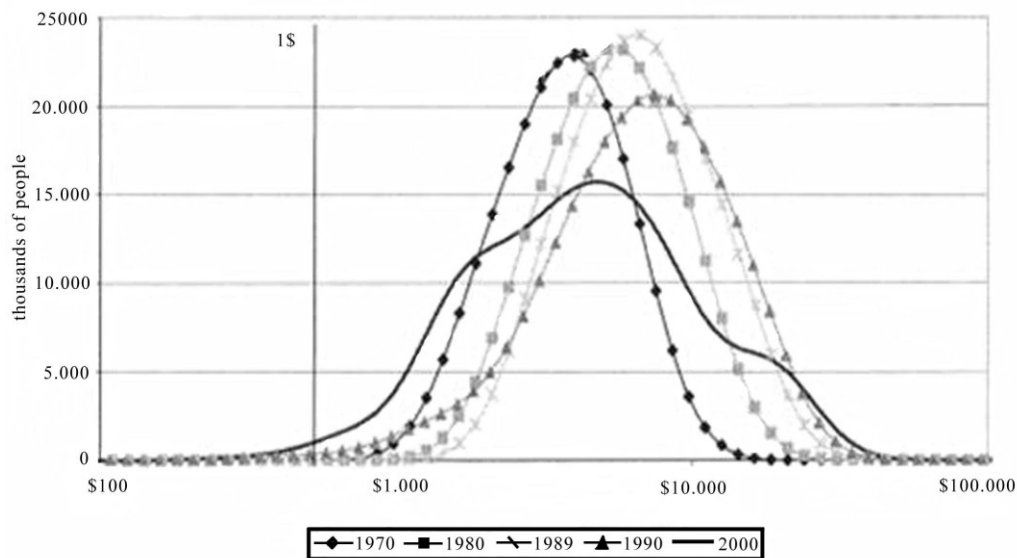


Fig. 2. The density of the distribution function of the per capita annual incomes in USSR from 1970 to 1989 and the total density of the distribution function in the countries of the former USSR on the logarithmic scale from 1990 to 2000.

However, the most important is the analytical description of the entire distribution function and not just a part of it. Some researchers modeled the general function of the income distribution density by non-parametric methods. For example, in [7] the method of nuclear estimates was used, in [8] and [9] – the diagram method, in [10] – the method of Fourier series. In the conditions when the distribution function is theoretically unknown and therefore can not be written by a simple mathematical formula with a small number of parameters, non-parametric methods give a possibility to structure large data arrays, to estimate the economic inequality and the levels of poverty and wealth. In [11] we can find the nuclear estimates of distribution densities of the logarithms of per capita incomes in nominal calculation in Russia in 1994 – 1997. In the obtained

dependences there are a number of peaks, which are not seen in the estimates, based on the available for the public average data of the State Committee on Statistics and the standard log-normal distribution. The latter stands for such representation of the per capita income distribution, which is given by:

$$f(\lg p) = \frac{1}{\sqrt{2\pi} h} \sum w_i \exp \left[-\frac{1}{2} \left(\frac{\lg p - \lg p_i}{h} \right)^2 \right], \quad (3)$$

where $h = 0.12$,

$\pi = 3.14\dots$ is the number equal to the ratio of the perimeter of the circle to the diameter of the circle,

w_i is the weight of the observed income p_i ,

p is the value of the income with respect to which the distribution is centered.

Formula (3) is written by analogy with the Gaussian normal distribution of the random variable for which the probability density is given by:

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left[-\frac{1}{2} \left(\frac{x - a}{\sigma} \right)^2 \right], \quad (4)$$

where a is the distribution center for the random variable x , which is also the point of maximum distribution and the center of symmetry.

Expression (4) for the function $f(x)$ describes the bell-shaped curve. The parameter σ is the distance from the vertical line of symmetry, specified by the equation $x = a$, to the inflection points that are on the right or left wings of the curve.

In Soviet times, the spread of the income of the population in Russia was small and was well modeled by the normal distribution law. But in the market economy with a large private sector the incomes of different groups differ significantly, depending on the level of skills in the achieving the results, on various kinds of talents, proficiency, the value of the employee for the employer, the possibility of fair pay in the corresponding sector of the economy taking into account various accepted norms of pay in different sectors, on the geographic and demographic characteristics, health, or the presence of bringing profit property, land, means of production, stocks and other tangible and intangible rights. The income level is also influenced by discrimination in its various forms, as well as by the objective phenomena (natural disasters, unemployment).

Studies show that the actual income distribution is far from the distribution of the form (4) with a single maximum. Using the sum in distribution (3) and the logarithms of the quantities instead of the quantities themselves improves the situation to some extent, allowing us to describe several peaks in the income distribution. However, we find it difficult to agree with the point of view that the incomes can be considered simple random quantities, in some way distributed around the average values. It is well known that the country's economy is not just the sum of its sub-systems such as firms, organizations and households, but is something qualitatively different. The presence of logarithms in (3) also points to the special type of distribution, different in its essence from random distribution. Apparently, the quasi-normal logarithmic distribution is only one of the possible approximations to describe the situation, without revealing its internal features.

The general function of the income distribution density

Figure 3 shows the normalized density function of the per capita annual income distribution in the USA for the year 2000, taken from [4]. The range of changing of the income is divided into 100 intervals, each interval corresponds to one centile of population. How can we describe the function of income distribution from Figure 3?

The presented function has two clear maximums and a remarkable detail on the right wing in the region of high incomes. Obviously, neither normal nor log-normal distribution are suitable, as well as the Pareto distribution or simple exponential distribution.

In search of a suitable distribution function we shall refer to the Bose-Einstein function of distribution of the particles by energies, which has the following form:

$$\bar{n} = \frac{dN}{dg} = \frac{1}{\exp\left(\frac{W - \mu}{kT}\right) - 1}, \quad (5)$$

where dN is the number of particles in the system, the energy of which is in the narrow interval from W to $W + dW$,

dg is the number of corresponding various quantum states of the particle,

μ is the chemical potential per particle,

k is the Boltzmann constant,

T is the temperature of the system of particles.

Distribution (5) is used for particles – bosons, which are, in particular, the neutral atoms and photons. The initial idea in derivation of (5) in quantum mechanics is the quantization of the possible energies of particles on the one hand, and the exponential dependence of the probability of finding particles in the state with the energy W : $P(W) = \exp\left(-\frac{W}{kT}\right)$.

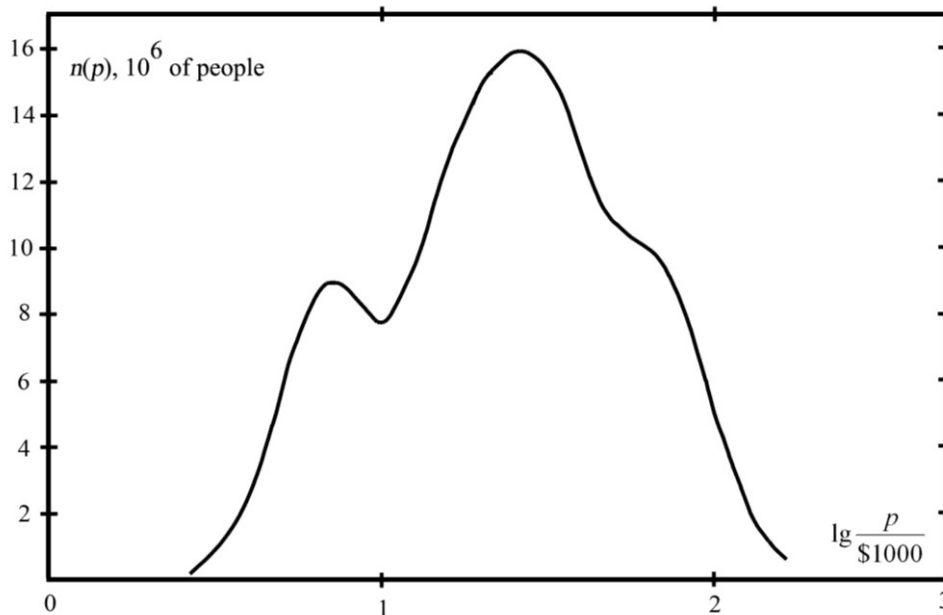


Fig. 3. The density of the function of distribution of annual per capita income in the USA in 2000.

As the consequence of (5) and of zero chemical potential μ for the light, for the spectral distribution of the intensity of the electromagnetic emission inside the black box with temperature T we obtain the Planck's law:

$$I(\omega) = \frac{dE}{dt ds d\omega} = \frac{\hbar\omega^3}{\pi^2 c^2 \left[\exp\left(\frac{\hbar\omega}{kT}\right) - 1 \right]}, \quad (6)$$

where c is the speed of light,

$\hbar = \frac{h}{2\pi}$ is the Dirac constant (h is the Planck constant),

$\frac{dE}{dt}$ is the power of emission passing in one way through the surface area ds oriented perpendicular to the direction of emission in the interval of the angular frequency of light from ω to $\omega + d\omega$.

At low energies of photons $\hbar\omega$ as compared to the average energy of atoms kT , we can expand the exponent in (6) into series and confine ourselves with only the first expansion terms. This gives the following:

$$\exp\left(\frac{\hbar\omega}{kT}\right) - 1 \approx \frac{\hbar\omega}{kT}, \quad I(\omega) = \frac{\omega^2 kT}{\pi^2 c^2}. \quad (7)$$

The last relation in (7) is known as the Rayleigh–Jeans law and it approximates (6) in the low frequency region. We shall suppose now that money and working people in a certain sense have the properties of bosons and therefore can be described by distributions found for bosons. In favor of this hypothesis is the fact that money is the embodiment of the quantity, which is called in physics energy. Electromagnetic quanta or photons transfer electromagnetic energy, and money is the universal measure of the economic value. Changing the amount of energy in physics is associated either with doing the work by the system or on the system, or with accumulation or dissipation of energy. Similarly, changing of the quantity of money in the subject of economic activity is associated with performing paid work and services, and with the processes of accumulation and distribution of money. Photons can be absorbed completely in case of coincidence of their frequency with the resonant frequencies in the atom. The emission from atoms is quantized, since the photons' energies are equal to the difference between the energies of the levels of atoms. In a solid body the electromagnetic energy can be efficiently converted into other forms of energy.

Similarly, in the society the incomes of employees are also discrete and correlated with the known rates of wages and with the level of work performed. The surplus value of goods and services produced by an employee is equal to the difference between the price of their realization and the cost of paid labor, raw materials, tools, third-party services needed on their production. At the moment of realization of goods we can say that the energy of money (spent on the production of goods and their realization) is converted into the internal energy of goods in the form of its cost. The fact that the cost has not only the objective but also the subjective property follows from the cost estimation procedure by means of the market or expert evaluation. But the presence of subjectivity in evaluating the cost of goods by the seller and the buyer does not mean the absence of contained in the product corresponding energy of cost. In respect of money, just as in many things related to the human society, there is a significant subjective component in the evaluation and adoption of equivalent relations between money, cost, price, etc.

The property of atoms and photons as bosons is that a number of particles of the system can be in the same state at a time. For example, in a small spatial volume a large number of photons with the same energy can be concentrated (the example is laser). In contrast to this, the fermions due to the Pauli principle should have obligatory difference in the states of particles. It is obvious that tangible and intangible values circulating in the society, expressed in the money equivalent, are closer in the properties to bosons than to fermions.

Based on the stated above, we shall assume money similarly to photons to be the carriers of corresponding energy and referring by their properties to bosons. This property will be true for incomes, as they can be expressed in money obtained for a certain period. We shall now use the formula of the type (6) to describe the density distribution function of the per capita annual income in the society. In general terms, we can write it according to [12] as follows:

$$n(p) = \frac{dN}{dp} = \frac{A(p - \mu)^\gamma}{\exp\left(\frac{p - \mu}{M}\right) - 1}, \quad (8)$$

where dN is the number of the population referring to the range of income change dp , located in the interval from p to $p + dp$,

A is a constant coefficient,

μ is the quantity, similar in its meaning to the chemical potential in (5),

M is the quantity, specifying the characteristic "temperature" of the considered sector of the economy.

With low incomes, when p tends to μ , $n(p)$ will tend to zero in the case when the exponent is $\gamma > 1$. The constant A in (8) is determined from the condition that:

$$\int n(p)dp = \int dN = A \int_{\mu}^{\infty} \frac{(p - \mu)^\gamma dp}{\exp\left(\frac{p - \mu}{M}\right) - 1} = N, \quad (9)$$

where N is the total population of the region.

With $\gamma = 3$ corresponding to the cubic degree with ω in (6), from (9) we find: $N = \frac{AM^4 \pi^4}{15}$.

The sum of the products of the average income p_i in the i -th income interval and the number of people ΔN_i in this interval should be equal to the total income of the population: $\sum p_i \Delta N_i = P$. Moving to the integral, taking into account (8), we have:

$$\int p dN = \int p n(p) dp = A \int \frac{p(p - \mu)^\gamma dp}{\exp\left(\frac{p - \mu}{M}\right) - 1} = P. \quad (10)$$

As it will be shown below, there are several groups in the society at the same time, which differ significantly from each other in their income and their work intensity. In this regard, we can apply the known in physics superposition principle, assuming each group's contribution into the total density of distribution $n(p)$ of the annual per capita income in the society relatively independent of each other. Then, instead of (8), we should use the sum of contributions of all groups:

$$n(p) = \sum_{i=1}^j \frac{A_i (p - \mu_i)^\gamma}{\exp\left(\frac{p - \mu_i}{M_i}\right) - 1}, \quad (11)$$

where the index i denotes the group number and it ranges from 1 to the number j , equal to the number of population groups; the coefficients A_i are proportional to the relative weight of the respective group; and the coefficients μ_i by their meaning specify the minimum income in each group. The coefficients M_i correspond to the temperature in (5) and (6), and, according to (7), the greater is M_i in some group, the larger are the cash flows in this group, all other conditions being equal.

In case of several population groups, instead of (9) we can write:

$$\int n(p) dp = \int dN = \sum_{i=1}^j \int_{\mu_i}^{\infty} \frac{A_i (p - \mu_i)^\gamma dp}{\exp\left(\frac{p - \mu_i}{M_i}\right) - 1} = \sum_{i=1}^j N_i = N, \quad (12)$$

and if $\gamma = 3$, we can estimate the number of people in each individual group: $N_i = \frac{A_i M_i^4 \pi^4}{15}$. For the total income of the entire population, instead of (10), we must have the following:

$$\int p dN = \int p n(p) dp = \sum_{i=1}^j \int_{\mu_i}^{\infty} \frac{A_i p (p - \mu_i)^\gamma dp}{\exp\left(\frac{p - \mu_i}{M_i}\right) - 1} = \sum_{i=1}^j P_i = P. \quad (13)$$

Income distribution in the USA in 2000

We shall now apply the function (11) to the analysis of the density of the income distribution function in Figure 3. As we can see, the curve in Figure 3 has at least two maximums. Therefore, we must take the sum of several terms in (11), and find for each of them the quantities A_i, μ_i, γ, M_i . Doing this, we arrive at the following result for the income distribution in the USA in 2000:

$$n(p) = \frac{1.53(p-2.51)^3}{\exp\left(\frac{p-2.51}{1.6}\right) - 1} + \frac{0.079(p-7.25)^3}{\exp\left(\frac{p-7.25}{5.15}\right) - 1} + \frac{0.015(p-22)^3}{\exp\left(\frac{p-22}{7.3}\right) - 1} + \frac{0.0026(p-42)^3}{\exp\left(\frac{p-42}{12.6}\right) - 1}. \quad (14)$$

We obtained four terms in (14), all of which have the same degree in the numerator $\gamma = 3$. The total number of unknown coefficients A_i, μ_i, M_i for the four terms equals 12, so in order to find them it is enough to take 12 different points on the curve in Figure (3). In this case, its own $n(p)$ corresponds to each p , and we substituted these p and $n(p)$ in (11), and then solved a system of 12 equations and determined the coefficients A_i, μ_i, M_i in (14).

The per capita annual income p in (14) is measured in thousands of dollars, and $n(p)$ – in the millions of people per income interval, corresponding to one centile. In Figure 4, we present four curves based on four functions in (14). The algebraic addition of these curves gives exactly the same envelope as in Figure 3.

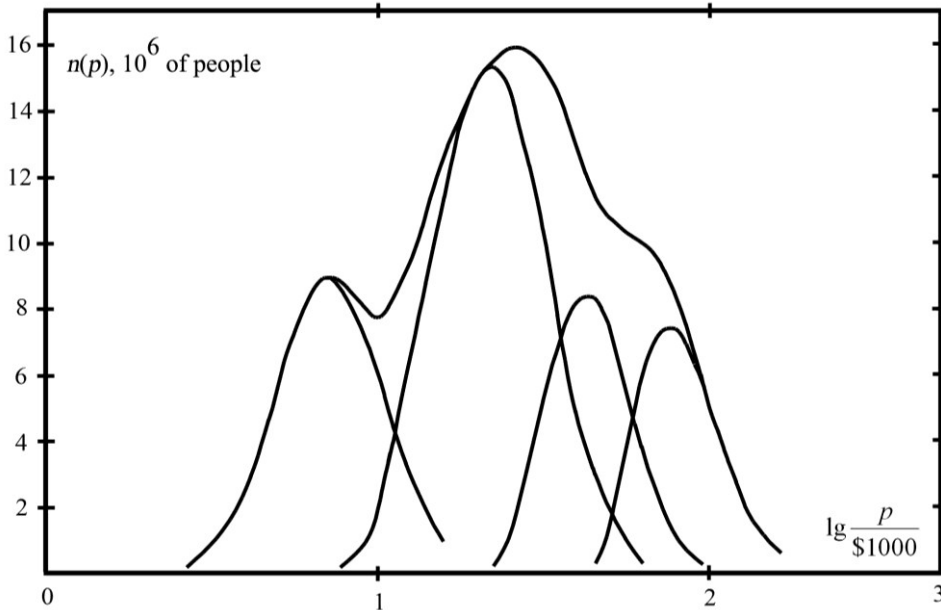


Fig. 4. Expansion of the density of the general function of distribution of per capita annual incomes in the USA in 2000 for four groups of population.

From (14) it follows that the US population consists of four major groups, which mainly form the density of the total income distribution function. The first group apparently consists of low-skilled people, or part-time employees. The incomes of this group are small, on the average about \$ 7,000 per year (we shall remind that we analyze available for us data for 2000). The second group is more qualified employees who have obtained the minimum general education and have a specialty. The income of this group at the peak of the corresponding curve is about \$ 22,000 a year. The third group mainly consists of people with higher education. The income in this group varies around the average of \$ 43,000. Finally, the fourth group includes all highly skilled and highly paid jobs, with an average income of \$ 76,000 a year.

In all groups, to this or that degree, there are also businessmen and owners of capitals, who get additional income as interest on their business, dividends or rent. A more precise analysis of these income groups is possible in those areas of income, in which the curves in Figure 4 do not cross.

From (14) we can find the coefficients of some distribution functions: $A_1 = 1.53$, $A_2 = 0.079$, $A_3 = 0.015$, $A_4 = 0.0026$. These coefficients should have the dimension of a million people divided by the fourth power of the income, according to the income scale, divided into centiles. Taking the relations $\frac{A_1}{A_4}$, $\frac{A_2}{A_4}$ etc., we find that on the average per one rich American there are 588 low-paid,

30 middle-paid and 6 well-paid Americans. However, these figures are not absolutely accurate, since they would be valid only in case if the coefficients M_i in all the groups would coincide. If the

coefficients M_i are different, the more informative are the ratios $\frac{N_1}{N_4} = \frac{A_1 M_1^4}{A_4 M_4^4}$, $\frac{N_2}{N_4} = \frac{A_2 M_2^4}{A_4 M_4^4}$, etc., where N_i is from (12). The constants $\mu_1 = 2.51$, $\mu_2 = 7.25$, $\mu_3 = 22$, $\mu_4 = 42$ in (14) define the lower limit of per capita annual income in thousands of dollars of the corresponding group.

The fact that the found groups exist relatively independent on each other is confirmed by different constants $M_1 = 1.6$, $M_2 = 5.15$, $M_3 = 7.3$, $M_4 = 12.6$ in (14). These constants are measured as a per capita annual income in thousands of dollars and similar in their meaning to the temperature in (6).

Apparently, the job market offers such jobs and niches, which in different groups are fundamentally different in their ability to provide income to the employee.

Conclusion

Our goal was to determine the mathematical formulas to describe the dependence of the distribution of annual income per capita. This was achieved in equations (8) – (14). These formulas are in good agreement with the statistical data for the distribution of per capita annual income in the United States in 2000, and four large economically independent groups were discovered in the data.

To some regret, the dependence of the density of the income distribution function presented in Figure 3 is not sufficient in general for the purposes of economic research. The fact is that it was built based on the use of centiles. Each centile contains 1 % of population, and the average income in a given centile depends on the position of the centile on the income axis. Therefore, the width of the income interval Δp corresponding to some centile is different in different parts of the income axis. Now if we want to use the formulas (12) or (13), substituting in them $n(p)$ from (14), the income intervals Δp should be transformed into differentials dp for further integration. However, the information for such transformation is available only in the statistical data, used by the authors of [4], so that the direct integration by the formulas (12) or (13), without taking into account the difference between Δp and dp in Figure 3, is inaccurate, giving too much deviation. In order to fully use the expansion of the form of (14) into the groups for the density of income distribution function, we must initially have other dependencies of the income distribution. They should be built not on the basis of centiles, but by determining the number of people with the same range of income change Δp , while moving along the income axis p . This will lead to some change in the parameters in (14), without changing the quality of the general picture. The result is the ability to use quickly and easily the distributions of the form of (14) in any economic research concerning the obtaining, accumulating and spending money in the society, as well as during the monitoring of the economy.

References

- [1] The socio-economic situation in Russia, the State Statistics Committee of Russia, Moscow, 2007.
- [2] M.I. Zelensky, The origins of right, St. Petersburg, 2003.
- [3] The World Factbook, Field Listing, Distribution of family income, Gini index, 2015. <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2172rank.html>.
- [4] Xavier Sala-I-Martin, The World Distribution of Income: Falling Poverty and ... Convergence, Period, Quarterly Journal of Economics, CXXI (2006) 351-397.
- [5] V. Pareto, Cours d'Economie Politique, Vol. 2. F. Rouge, Editeur, Lausanne and Paris, 1897.
- [6] A. Dragulescu, V.M. Yakovenko, Exponential and power-law distributions of wealth and income in the United Kingdom and the United States, Physica A, 299 (2001) 213-221.
- [7] E. Parzen. On the Estimation of a Probability Density Function and Mode, Annals of Math. Statistics, 33 (1962) 1065-1076.
- [8] J. Van Ryzin, Classification and clustering. Academic Press, N.Y., C.F., London, 1977.
- [9] J. Van Ryzin, A Histogram Method of Density Estimation, Comm. Statist. 2 (1973) 493-506.
- [10] E. Wegman, Non-parametric Probability Density Estimation: I. A Summary of Available Methods, Technometrics, 14 (1972) 533-546.
- [11] A.Y. Sheviakov, A.J. Kiruta, Economic inequality, poverty and the standard of living of the Russian population: methods of measurement and analysis of causal relationships, EERC, Moscow, 2001.
- [12] Sergey Fedosin. The physical theories and infinite hierarchical nesting of matter, Volume 2. – LAP LAMBERT Academic Publishing, pages: 420. (2015).