Scientific Papers (www.scientificpapers.org)
Journal of Knowledge Management, Economics and Information Technology

Vol. III, Issue 3
June 2013

A Polynomial Distribution Applied to Income and Wealth Distribution

Authors: **Elvis Oltean**, Loughborough University, Loughborough, the UK, e.oltean@lboro.ac.uk, **Fedor V. Kusmartsev,** Loughborough University, Loughborough, the UK, f.kusmartsev@lboro.ac.uk

Income and wealth distribution affect stability of a society to a large extent and high inequality affects it negatively. Moreover, in the case of developed countries, recently has been proven that inequality is closely related to all negative phenomena affecting society. So far, Econophysics papers tried to analyse income and wealth distribution by employing distributions such as Fermi-Dirac, Bose-Einstein, Maxwell-Boltzmann, lognormal (Gibrat), and exponential. Generally, distributions describe mostly income and less wealth distribution for low and middle income segment of population, which accounts about 90% of the population. Our approach is based on a totally new distribution, not used so far in the literature regarding income and wealth distribution. Using cumulative distribution method, we find that polynomial functions, regardless of their degree (first, second, or higher), can describe with very high accuracy both income and wealth distribution. Moreover, we find that polynomial functions describe income and wealth distribution for entire population including upper income segment for which traditionally Pareto distribution is used.

Keywords: *polynomial; cumulative distribution function; decile; mean income; upper limit on income*

Introduction

Wealth and income distribution are among the most important issues in a society considering that an optimal level ensures social stability, while a high degree of inequality causes multiple problems.

A recent study by Wilkinson [1] shows that for the developed countries there is a direct relation between economic inequality and all problems that impact negatively on society such as criminality, social trust, obesity, infant mortality, violence, child poverty, mental illness, imprisonment, and many others that characterise the quality of life. Thus, the countries with the lowest inequality, such as Scandinavian countries and Japan, have the best indicators regarding the social phenomena that affect the quality of life. The opposite is represented by the USA, which has the highest inequality among developed countries. Thus, the USA is characterised by the highest impact of negative phenomena affecting the society.

In the more recent years, a distinct field of Econophysics emerged. This one deals with size of firms, macroeconomic aggregates, income and wealth distribution [2], [3]. Traditionally, in Economics literature Gini coefficient is the most used measure of inequality. So far, the most used in Econophysics literature were Maxwell-Boltzmann, Bose-Einstein, Fermi-Dirac, lognormal (Gibrat), and exponential distributions.

Short History and Theoretical Background

The modern approach of income and wealth distribution first appeared in the articles of Yakovenko [4], [5], [6], and [7]. Main findings of his work are about Bose-Einstein distribution, lognormal distribution, and exponential distribution which explain income and wealth distribution for low and middle tier income population.

Kusmartsev, while exploring possible applications of Bose-Einstein distribution in the income distribution, tries to analyse possible correlations with other statistical physics variables such as chemical potential and activity coefficient [8], [9].

While most of papers claim to cover income and wealth distribution only for low and middle income part of population, there are two exceptions. Fermi-Dirac distribution [10] and Tsallis distribution [11] claim to be robust enough in order to explain income distribution for the entire range of income, including for upper income segment of population which traditionally is described by a Pareto distribution.

Methodology

The data used in our analysis is about nominal or real income, and wealth of the population, where income or wealth of each person/household is ranked in increasing order of the values. Next, a segmentation of the population is performed by dividing the entire range of income in ten segments. Each part/segment of the population represents 10 % of the total population. Thus, the lowest income segment is considered to be the first one, while the highest income one is the tenth. These segments are known in the relevant literature as deciles.

In our case, we consider data from several countries such as France, Finland, Italy, and Romania. For each country there are certain particularities.

Firstly, for each segment/part of population there are two methods for income measurement by using deciles. Thus, it can be calculated a mean income which is total income earned by all individuals/households from a certain group/segment of population (decile in our case) divided to the total number of individuals/households belonging to that segment of population (decile). Also, for each segment can be calculated upper limit on income, which represents the highest income of an individual or household from an income segment (decile). The latter term was coined by the National Institute of Statistics from Finland.

Secondly, is about kind of income considered analysed. For all the countries, data is about net (or disposable) income which is the income that remains available after paying taxes and (if the case) benefits received by individual or household. In the case of France, we have an additional set of data about mean wealth.

Thirdly, in the cases of France and Finland, the data are about real income which means that nominal income is adjusted with Consumer Price Index (CPI), which correlates the evolution of nominal income with the inflation rate. For Italy and Romania, we deal with nominal income.

Fourthly, the basic element in the analysis of income and wealth is individual (in case of France and Finland), while for Italy and Romania is about households.

The methodology used in order to calculate the probability distribution for the allocation of money for each segment of population is cumulative probability distribution function (cdf). From a variety of definitions we chose the one that describes the cumulative share of population which has an income or wealth greater than a certain threshold.

$$P(x) = \int_x^\infty p(t)dt \qquad (1),$$

P represents fraction of population with nominal or real income or wealth greater than x. Thus, for income equal to 0, the cumulative density probability is 100 % as it is considered that everyone has an income higher than 0. For the first decile (the lowest income decile), P represents the population that has an income higher than mean income or upper limit on income of the first decile, hence equals 90%. Subsequently, for the highest income the cumulative distribution function is 0 % (in case of mean income). For the upper limit on income data set, we do not represent the upper decile as the value corresponding for it was not made available by any of the statistical bodies.

Thus, on the x-axis we represent cumulated nominal or real income, or wealth of the population divided in ten deciles (or nine deciles for upper limit on income data set), while on the y-axis we represent the cumulative probability density of the population (in percentages) having certain income or wealth above a certain threshold.

The distribution that we found to fit best for income the data from a variety of distributions and with minimum number of parameters is first degree polynomial function

$$y=P_1*x + P_2 \qquad (2)$$

In the case of mean wealth, we used second degree polynomial which has a significant goodness of fit compared to first degree polynomial. The cumulative distribution function is as follows:

$$y= P_1*x^2+ P_2*x + P_3 \qquad (3)$$

Data analysis

For France, Finland, and Italy we have two sets of data. Thus, one set is about mean income and the other set is about upper limit on income. For Romania, we have data only about mean income. All the data mentioned before are about net/disposable income. Lastly, in case of France we analyze data about mean wealth.

Income

France

The data available for France are about mean income [12] and upper limit on income [13]. The detailed findings are exhibited in the Appendix 1. The data from France is in euro 2009, which means that for the previous years, the nominal income was converted to real income by using CPI with reference to the prices from the year 2009.

Using cumulative distribution function, we got a very good fitting to the data. Thus, for mean income set of data the annual coefficient of determination (R2) has the lowest value 94.73% for the year 2006, while the highest value is 95.07% for the data from the year 2003. In the case of upper limit on income, the annual coefficient of determination was the highest for the year 2004 (97.27%), while the lowest is 97.13% for the year 2009.

We can notice that the values for the coefficient of determination are higher in the case of the upper limit on income data set.

In Figure 1, we present an example in order to observe graphically the goodness of the fit to the data.

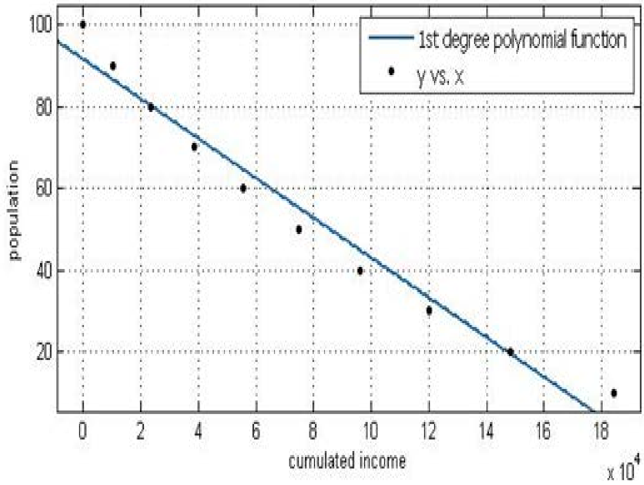

Figure 1 : Cumulative distribution probability for upper limit on income in the year 2009

The equation describing the distribution is

$$Y= (-0.0004861)*x+ 91.52, \text{ where } R^2=97.13\ \%$$

The value for the coefficient of variable x from the fitting function (P1) is very small considering that the values for income are high compared to the values from the y-axis.

Finland

The data about income is about mean income [14] and for upper limit on income [15]. The detailed findings are exhibited in the Appendix 2.The data about Finland is real income expressed in euro to the value from the year 2009, which implies that the values of real income from the previous years has been adjusted by taking into account the CPI.

The annual values for coefficient of determination resulting from the cumulative distribution function in order to fit the data yielded were very good. Thus, for the mean income data set, the lowest annual value was 95.17% (corresponding to the year 2007), while the highest value was 97.64% corresponding to the year 1987. For the upper limit on income data set, the annual values for the coefficient of determination were as follows: the highest value was 98.56% for the year 1992 and the lowest was 97.52% corresponding to the year 2007.

We can notice that generally the annual values of the coefficient of determination were higher for upper limit on income than in the case of mean income data set.

In order to illustrate with an example, we chose the year that presented the lowest annual value for the coefficient of determination from the upper limit on income data set (1992).

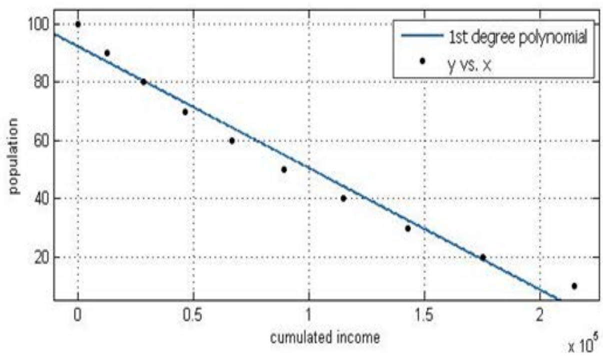

Figure 2: Cumulative distribution probability for upper limit on income in the year 1992

The equation describing the probability distribution is

$$Y= (-0.0005679)*x+ 93.95, \text{ where } R^2=98.56\%$$

We can notice that the values for the parameter P1 of the variable x is very small. This is due to the high values of income and subsequently P1 acts as a normalization constant.

Italy

The data sets describe income distribution by using mean income [17] and upper limit on income [18]. The detailed findings are exhibited in the Appendix 3. However, we are dealing with two different currencies. Thus, for the time interval 1989-1998 the currency used was Italian lira, while for the time interval 2000-2008 the data are expressed in euro. Italian Lira was a currency characterized by high inflation, though the inflation in the 90s was not as high as in the previous decade. Euro can be considered a very stable currency compared to Italian lira.

The first degree polynomial cdf fits the data very well. For mean income data set, the coefficient of determination has the highest annual value 93.19% for the year 1991 and the lowest one is 91.03% for the year 1998. Similarly, for the upper limit on income data set, the highest annual value is 95.44% for the year 1991 and the lowest one is 94.28% for the year 1993.

We can notice that generally the annual values of the coefficient of determination were higher for upper limit on income than in the case of mean income data set.

In order to illustrate these evolutions, we chose the year 1993 from the upper limit on income set which is exhibited in Figure 3.

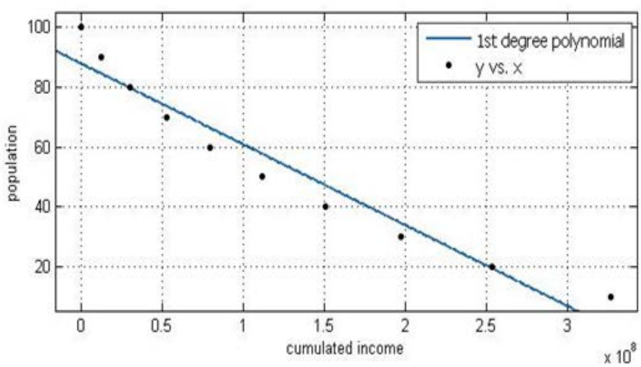

Figure 3: Cumulative distribution probability for upper limit on income in the year 1993

The equation describing the probability distribution is

$$Y= (-2.686*10^{-7})*x+87.59, \text{ where } R^2=94.28\%$$

We can observe that the values for the parameters of the x variable (P1) in the cdf are significantly lower for the time interval when Italian lira was the official currency. This statement is true when we compare it to the values of coefficients for euro values from Italian data and to the values of the coefficients of variable x (P1) from fitting the data from the countries previously considered. Moreover, this is valid for both types of data sets (mean income and upper limit on income).

Romania

The data set about Romania contains information about nominal income using mean income data set [16]. The detailed findings are exhibited in the Appendix 4.

Similarly to the Italian data, we are dealing here with two different currencies. Thus, up to the year 2005, the national currency of Romania was leu. After middle of the year 2005, the new currency introduced was called heavy leu and 1 heavy leu=10000 leu. The time interval considered before 2005 was inflationary (inflation rate was up to 30%).

Cumulative distribution function used to fit the data has yielded the following annual values for coefficient of determination: the lowest was for the year 2007 (95.24%), while the highest value was 96.81% for the year 2010. In order to have a graphical image about the goodness of the fit, we present the function fitting the upper limit on income data set from the year 2007 in Figure 4.

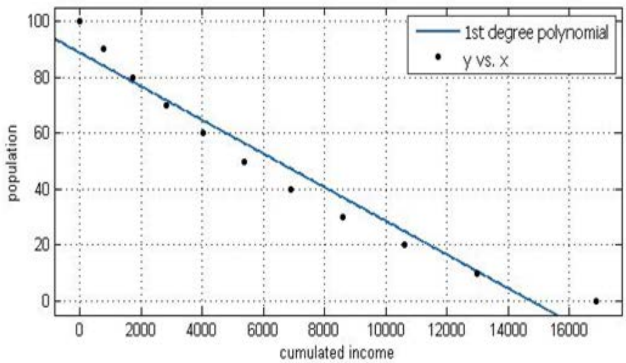

Figure 4 :. Cumulative distribution probability for mean income in the year 2007

The equation describing the distribution function is

$$Y = (-0.006014)*x + 88.62, \text{ where } R^2 = 95.24\%$$

Regarding the values of the parameter P1 for the variable x, we can notice that its values are significantly lower for the time interval before the year 2005 compared with other countries considered in this paper. This case, similar to the Italian data in the case of Italian Lira, is explainable by the higher prices and higher inflation in the time interval before the year 2005.

Wealth

Unfortunately, the only data regarding wealth we have available is from France. In the following, we present our findings regarding mean wealth from the years 1998, 2004, and 2010.

For the data set about wealth [19], we used mean wealth in order to represent the wealth distribution. The detailed findings are exhibited in the Appendix 5. Unfortunately, the first degree polynomial does not fit very well the data, the annual values for the coefficient of determination being around 70 %. Therefore, we decided to use second degree polynomial to fit the data.

Subsequently, the annual values for the coefficient of determination were 88%-89%, so without large variations.

In order to illustrate that, we present the case for the year 2010 in Figure 5.

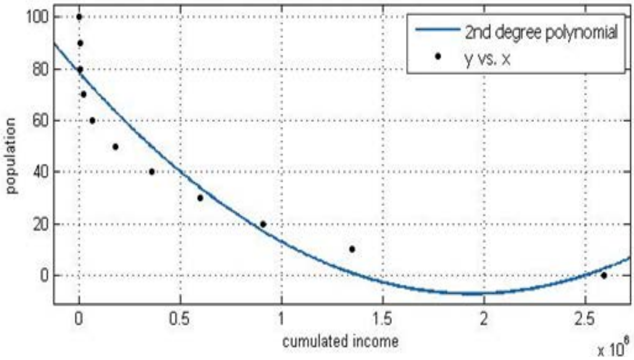

Figure 5: Second degree cdf probability for mean wealth in the year 2010

The equation describing the probability distribution is

$$Y= 2.276*10^{-11} + (-8.817*10^{-5})*x+ 78.16, \text{ where } R^2=88.94\%$$

In spite of the limited quantity of data (unlike the case of income), we can notice that the second degree polynomial distribution yields high annual values for the coefficient of determination.

Results

The annual values for coefficient of determination are above 88% (in case of wealth analysis), while all the annual values are above 94% in the case of income analysis.

In the case of the upper limit on income data sets, the annual values regarding coefficient of determination are generally higher than in the case of mean income data sets. A possible explanation for this is due to the fact that upper income population (which is comprised in the tenth decile) is not included in the upper limit on income data sets. Moreover, upper income population depends not only on income but on assets prices as well, so the underlying mechanism for this segment of population works differently than in the case of the low and middle income part.

Generally, the values of the parameters of the fitting functions look similar with no relative great discrepancies. There are two exceptions in the values of the parameter P1 of the x variable. This occurs in the case of Romania for time interval before the year 2005 and in the case of Italy for time interval before the year 2000. Thus, both cases are characterized by the introduction of a new and more stable currency. For these currencies and time intervals, the values of the parameter P1 of the x variable obtained from fitting the function were significantly lower. This could be explained by the fact salaries and prices are high in nominal terms, and inflation was relatively high.

We performed the same analysis using a log-log scale for the values on both x and y axes. Surprisingly, the results were very similar regarding both coefficients of the distribution functions and the values obtained for coefficient of determination.

In the case of Italy for the time interval before the year 2000 (introduction of euro) and Romania for the time interval before the year 2005 (introduction of heavy leu), the annual values for coefficient of determination were very high. Thus, we can say that goodness of the fit did not change significantly for currencies and time interval marked by inflation. Moreover, if we compare the values for coefficient of determination for Italy from Italian Lira on one hand and Euro on the other hand, no significant differences can be observed.

First degree polynomial can describe successfully the income distribution. In the case of wealth distribution, the first degree polynomial cannot describe it with very high accuracy. The second degree polynomial can be used successfully to fit the wealth data. However, the third degree polynomial and higher degree polynomials can fit even more successfully the data. But since all the annual values for the coefficient of determination are higher than 88%, we chose for distribution a function with minimum number of parameters.

Conclusions

Polynomial cumulative distribution function fits well the data. Subsequently, this function is robust and can describe with fairly high accuracy the wealth and income distribution for population. Moreover, the different economic characteristics of the countries considered (such as different levels of development, different macroeconomic characteristics, data reliability) and similar values obtained when log-log scale was used can be considered as an additional proof regarding the robustness of the distribution function.

It is very important that polynomial distribution function can describe successfully both wealth and income distribution. Thus, most of the distributions can solely explain income distribution. This is remarkable since the underlying mechanism for both economic variables is different and may suggest a partial resemblance. However, wealth distribution can be described with high accuracy by second degree polynomials unlike income distribution which is explainable using first degree polynomial. A possible explanation for this phenomenon is that the underlying mechanisms for wealth and income are different.

Polynomial distribution function confirms the main findings of the analysis performed on the same set of data by using Fermi-Dirac distribution regarding income distribution [10].

Given the high values obtained for coefficient of determination, we can conclude that first and second degree polynomial distribution functions (as well as higher degree ones) can be used with a high degree of success to explain income and wealth distribution for the entire range of income and wealth among population. Thus, both low and middle income part of population and upper income part that traditionally were explained by different distribution functions can be explained by a single polynomial function.

Polynomial distribution function is not affected by inflation, so slightly lower values for coefficient of determination from the annual analyses may indicate a lower degree of credibility regarding data.

Upper limit on income data can be used more successfully to explain the income and wealth distribution than mean income data for low and middle income segment of population. However, if the analysis aims to look at the entire income and wealth range, mean income data is more suitable.

References

[1] R. Wilkinson, K.Pickett: *The Spirit Level: Why More Equal Societies Almost Always Do Better*, Allen Lane, 2009.
[2] Z. Burda, J. Jurkiewicz, M.A. Nowak: *Is Econophysics a Solid Science?* www. arXiv:cond-mat/ 0301096 v1, 08.01.2003.
[3] N. de Liso, G. Filatrella, *Econophysics: The emergence of a new field?* www.unifr.ch
[4] A. Dragulescu, V. M. Yakovenko, *Statistical Mechanics of Money, Income, and Wealth: A Short Survey,* http://www2.physics.umd.edu/~yakovenk/econophysics.html
[5] A. Dragulescu, V.M. Yakovenko, *Statistical mechanics of money*, Eur. Phys. J. B 17, 723-729 (2000).
[6] A.Dragulescu, V.M. Yakovenko, *Evidence for the exponential distribution of income in the USA*, Eur. Phys. J. B 20, 585-589 (2001)
[7] A. C. Silva, V. M. Yakovenko, *Temporal evolution of the "thermal" and "superthermal" income classes in the USA during 1983–2001,* Europhys. Lett., 69 (2), pp. 304–310 (2005)
[8] K. E. Kurten and F. V. Kusmartsev: *Bose-Einstein distribution of money in a free-market economy*. Physics Letter A Journal, EPL, 93 (2011) 28003
[9] F. V. Kusmartsev: *Statistical mechanics of economics,* Physics Letters A 375 (2011) 966–973
[10] E. Oltean, F. V. Kusmartsev: *A study of methods from statistical mechanics to income distribution*, ENEC, 2012
[11] J. C. Ferrero, *The equilibrium and nonequilibrium distribution of money, International Workshop on "Econophysics of Wealth Distributions"*, Saha Institute of Nuclear Physics, Kolkata, India, 15-19 March, 2005.

[12] *Niveau de vie moyens par décile en France*, http://www.insee.fr/fr/themes/tableau.asp?reg_id=0&ref_id=NATnon04249
[13] *Distribution des niveaux de vie en France,* http://www.insee.fr/fr/themes/tableau.asp?reg_id=0&ref_id=NATnon04247
[14] *Mean income by decile group in 1987–2009 in Finland,* http://www.stat.fi/til/tjt/2009/tjt_2009_2011-0520_tau_003_en.html
[15] *Upper limit on income by decile group in 1987–2009 in Finland*, http://www.stat.fi/til/tjt/2009/tjt_2009_2011-0520_tau_005_en.html
[16] *Buletin statistic*, Institutul National de Statistica, Romania, 1999-2011
[17] *Supplementi al Bollettino Statistico - Note metodologiche e informazioni statistiche,* Numero 14 – 8 Ottobre 1991, Numero 44 - 14 Luglio 1993, Numero 9 - 10 Febbraio 1995, Numero 14 - 20 Marzo 1997, Banca d'Italia, Centro Stampa, Roma
[18] *Supplements to the Statistical Bulletin-Methodological notes and statistical information*, Number 22 – 18 April 2000, Banca d'Italia, Centro Stampa, Roma
[19] *Patrimoine moyen par décile en 2010*, http://www.insee.fr/fr/themes/tableau.asp?reg_id=0&ref_id=patrmoyendecile

Vol. III, Issue 3
June 2013

Appendix 1. Coefficients from fitting a first degree polynomial distribution to data regarding mean income and upper limit on income from France

	France mean income			France upper limit on income		
	P1	P2	R^2 (%)	P1	P2	R^2 (%)
2003	-0.0004894	87.84	**95.07**	-0.0005206	91.67	97.21
2004	-0.0004905	87.88	95.07	-0.0005226	91.76	**97.27**
2005	-0.0004844	87.61	94.93	-0.0005169	91.68	97.24
2006	-0.0004733	87.46	**94.73**	-0.0005073	91.59	97.16
2007	-0.0004653	87.48	94.78	-0.0004977	91.59	97.17
2008	-0.0004569	87.56	94.81	-0.0004897	91.74	97.27
2009	-0.0004549	87.41	94.77	-0.0004861	91.52	**97.13**

Appendix 2. Coefficients from fitting a first degree polynomial distribution to data regarding mean income and upper limit on income from Finland

	Finland mean income			Finland upper limit on income		
	P1	P2	R^2 (%)	P1	P2	R^2 (%)
1987	-0.0006026	90.99	**97.64**	-0.0006152	93.71	98.49

1988	-0.0005874	90.88	97.52	-0.0006021	93.65	98.45
1989	-0.0005584	90.8	97.45	-0.0005731	93.55	98.41
1990	-0.0005328	90.92	97.5	-0.0005464	93.64	98.42
1991	-0.0005265	90.96	97.54	-0.0005396	93.77	98.49
1992	-0.0005534	91.19	97.61	-0.0005679	93.95	**98.56**
1993	-0.0005698	90.83	97.29	-0.000589	93.83	98.44
1994	-0.0005688	90.84	97.28	-0.0005876	93.73	98.4
1995	-0.0005547	90.59	97.12	-0.000575	93.58	98.33
1996	-0.0005407	90.7	97.05	-0.0005619	93.32	98.22
1997	-0.0005186	89.76	96.53	-0.0005431	93.04	98.04
1998	-0.0004994	89.33	96.24	-0.0005246	92.76	97.93
1999	-0.0004791	89.04	95.9	-0.0005092	92.74	97.92
2000	-0.000471	88.72	95.62	-0.0005024	92.49	97.78
2001	-0.0004641	88.91	95.95	-0.0004903	92.53	97.83
2002	-0.0004511	88.93	96.00	-0.0004753	92.51	97.83
2003	-0.0004382	88.84	95.87	-0.0004638	92.46	97.78
2004	-0.0004172	88.57	95.66	-0.0004429	92.29	97.71
2005	-0.0004043	88.55	95.65	-0.000429	92.25	97.7

2006	-0.0003996	88.32	95.44	-0.0004259	92.13	97.6
2007	-0.0003892	88.03	**95.17**	-0.000416	91.99	**97.52**
2008	-0.0003918	88.43	95.64	-0.0004159	92.21	97.69
2009	-0.0003808	88.7	95.92	-0.0004012	92.27	97.73

Appendix 3. Coefficients from fitting a first degree polynomial distribution to data regarding mean income and upper limit on income from Italy

	Italy nominal mean income			Italy upper limit on income		
	P1	P2	R^2 (%)	P1	P2	R^2 (%)
1989 lire	$-2.837*10^{-7}$	84.5	92.57	$-3.046*10^{-7}$	88.82	95.21
1991 lire	$-2.647*10^{-7}$	84.96	**93.19**	$-2.808*10^{-7}$	88.95	**95.44**
1993 lire	$-2.48*10^{-7}$	83.06	91.34	$-2.686*10^{-7}$	87.59	**94.28**
1995 lire	$-2.3*10^{-7}$	83.33	91.52	$-2.504*10^{-7}$	87.98	94.64
1998 lire	$-2.039*10^{-7}$	82.81	**91.03**	$-2.233*10^{-7}$	87.77	94.49
2000 euro	-0.0003776	83.43	91.66	-0.0004098	88.21	94.79
2002 euro	-0.0003541	83.57	91.68	-0.0003842	88.22	94.75
2004 euro	-0.0003371	84.12	91.95	-0.0003672	88.68	95.01

2006 euro	-0.0003124	84.41	92.24	-0.0003404	88.99	95.33
2008 euro	-0.0003082	84.01	91.96	-0.000334	88.52	94.92

Appendix 4. Coefficients from fitting a first degree polynomial distribution to data regarding mean income from Romania

Romania mean income	P1	P2	R^2 (%)
2000 leu	$-2.994*10^{-6}$	89.98	96.21
2001 leu	$-1.967*10^{-6}$	90.78	96.43
2002 leu	$-1.55*10^{-6}$	90.31	96.39
2003 leu	$-1.29*10^{-6}$	90.72	96.45
2004 leu	$-9.412*10^{-7}$	90.18	96.16
2005 heavy leu	-0.008409	89.08	95.44
2006 heavy leu	-0.00733	88.28	94.88
2007 heavy leu	-0.006014	88.62	**95.24**
2008 heavy leu	-0.004752	89.2	95.83
2009 heavy leu	-0.004379	90.03	96.41
2010 heavy leu	-0.004408	90.7	**96.81**

Appendix 5. Coefficients from fitting a second degree polynomial distribution to data regarding mean wealth from Italy

France mean wealth	P1	P2	P3	R^2 (%)
1998	$1.044*10^{-10}$	-0.0001886	78.8	89.77
2004	$5.423*10^{-11}$	-0.0001352	78.22	89.11
2010	$2.276*10^{-11}$	$-8.817*10^{-5}$	78.16	88.94