

A Service of



Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre

Bandourian, Ripsy

Working Paper

Income Distributions: A Comparison across Countries and Time

LIS Working Paper Series, No. 231

Provided in Cooperation with:

Luxembourg Income Study (LIS)

Suggested Citation: Bandourian, Ripsy (2000): Income Distributions: A Comparison across Countries and Time, LIS Working Paper Series, No. 231, Luxembourg Income Study (LIS), Luxembourg

This Version is available at: https://hdl.handle.net/10419/160903

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Luxembourg Income Study Working Paper No. 231

Income Distributions: A Comparison across Countries and Time

Ripsy Bandourian

April 2000

Income Distributions: A Comparison across

Countries and Time

Ripsy Bandourian* April 12, 2000 Brigham Young University

*Great appreciation is expressed to Dr. James B. McDonald for his patience in helping me understand this subject.

Abstract

A number of distributions have been used to describe income distribution. This paper estimates eleven distributions (GB, GB1, GB2, B, B1, B2, GG, BR3, BR12, GA, and LN) using data from eight countries (Australia, Canada, Israel, Norway, Russia, Taiwan, United Kingdom, and the United States) for three different years. Relative performance of each distribution is considered as well its performance over time for each country. In addition, a measure of income inequality is also estimated, and its relationship to these distributions is discussed. Some suggestions for further study in this field are also discussed.

I. Introduction

On January 10, 2000 while addressing the UN Security Council, James D. Wolfensohn, President of the World Bank Group, stated his belief that "combating poverty" is "the true key to security and peace" (Wolfehnsohn, 2000) He then proceeded to discuss income distribution issues that plague the majority of developing countries. This issue has been pertinent throughout human history. As ancient Greek and Roman philosophers addressed it in numerous treatises (i.e., Aristotle), the great thinkers of the medieval and enlightenment periods followed them. With the development of quantitative methods, scientists from a variety of disciplines have explored this issue. This quantitative approach appealed to mathematicians, statisticians, and economists (Parker, 1999).

The purpose of this paper is to examine a number of probability density functions that previous literature considers to be successful in describing income distributions. Income data from a variety of countries representing a spectrum of development stages will be fitted to these probability density functions, and tested for best fit. In addition, a measure of income inequality (the Gini coefficient) will be estimated. The paper will proceed in the following manner: in section II, we will provide a review of the literature. In section III, we will describe the methodology, as well as the models to be used in estimation, and their functional forms. In section IV, we will provide a detailed description of the data. In section V, the estimation procedure and analysis of its results will be discussed; and finally, section VI will comprise our conclusions and suggestions for further study.

II. Literature Review

What functional form of a probability distribution function (pdf) provides the closest fit to the shape of an earnings distribution? Over the years, this question has been studied extensively in economics as well as in other social sciences. One of the first papers, authored by Pareto that dealt extensively with the quantitative estimation of income distributions appeared in 1896 and 1897. In these publications, Pareto specified a new pdf, which is currently known as the Pareto distribution, and suggested using it as a representation of income and wealth distribution (Dagum, 1988).

Pareto's research on the distribution of income provided an answer to an important question that the French and Italian scientists raised, the question of income inequality in a given country or region and also the relative degree of inequality between two countries. Corrado Gini opposed Pareto's proposition that income growth implies less income inequality and suggested that inequality indeed is an increasing function of income growth. He suggested an easily derived unit free measure that would allow one to compare inequality across countries. This coefficient, appropriately named the Gini coefficient, will be discussed in further detail in the following section (Dagum, 1988).

As research in this area continued, new distributions were discovered and suggested for use in estimating income distributions. For instance, in 1969, Brown and Aitchinson suggested using the lognormal pdf. They argued that the lognormal distribution is appropriate for the distribution of income because it satisfies a number of specified criteria. For instance, the functional form of the lognormal pdf can be easily derived based on predictions of economic theory. In addition, because the lognormal pdf is a two-parameter distribution, it is relatively

easy to estimate these parameters and provide their economic interpretation (Brown and Aitchinson, 1969).

However, in subsequent years, research on income distribution continued its evolution. In 1970, Thurow suggested using the beta of the first kind, a three-parameter distribution, as a model for income distribution (Thurow, 1970). In his estimation, Thurow examined income data for US families in 1949 and used an upper limit of \$15,000 due to the fact that he was mostly interested in estimating income distribution for the lower end. He found that the beta distribution fit the actual income distributions rather well. However, upon comparing Gini coefficient estimates obtained by Thurow with those provided by the Census Bureau, McDonald concluded that they are inconsistent with the latter. He suggested that this inconsistency could be due to incorrect modeling of "intra- and/or intergroup variation" as well as possible problems with optimization algorithm (McDonald, 1984).

Later, Salem and Mount used the gamma distribution for the same purpose (Salem and Mount, 1974). They fit a two-parameter gamma distribution, which is a limiting case of the beta distribution, to income data for the CPS Reports for the years 1960 to 1969. They estimate these two parameters and examine them in the light of their representation of scale and inequality. In this paper, the authors also compare the results of fitting the gamma and the lognormal distributions to income data. Based on likelihood ratio tests, Salem and Mount demonstrated that the gamma distribution is indeed a better probability density function to be used in income distribution (Salem and Mount, 1974).

Soon after that a number of studies showed that alternative methods using generalizations of these distributions provide significantly better "fits" for income distribution analysis. In one particular study, McDonald proposed using the generalized beta of the first and second kind

(GB1, GB2) as suitable pdf's for the distribution of income (McDonald, 1984). He suggested that since these distributions included the beta of the first kind and second kind (B1, B2), the gamma (GA), Singh-Maddala (SM), the lognormal (LN), and some other distributions as special cases, they provide at least as good a fit as the B1, B2, GA, and LN.

In a later study, McDonald and Xu (1995) showed that the generalized beta distribution that nests all of the above distributions as its special cases is much more flexible and thus is better "fit" to describe income distributions. Bordley and McDonald (1996) have also examined the relative performance of some of these distributions. In this research, they fit a variety of distributions to data obtained for five years of U.S. family income. In conclusion of this study, they asserted that of the models considered, GB2 is the model that provides the best fit for income distribution (Bordley, McDonald, and Mantrala, 1996).

This study is an extension of previously described research in that it describes relative performance of alternative models for income distribution data for eight countries, we will attempt to fit income data from eight countries to all of these distribution models, and compare the results. In fact, this study serves as an extension of research described above.

III. Methodology

Based on the information provided by the World Bank World Development Indicators database, one may investigate whether the distribution that most accurately describes the income distribution of a developing country might be very different from that of a developed country. In this paper, we will estimate income distributions using a number of pdf's and will follow their variation across countries.

A. The Generalized Beta Distribution Family

We will use the generalized beta (GB), generalized beta of type I and II (GB1 and GB2), the beta (B), beta of type I and II (B1 and B2), Burr3 and Burr12 (BR3 and BR12), the gamma (GA), and the lognormal (LN) distributions. The GB distribution is defined by the following pdf

GB(y;a,b,c,p,q) =
$$\frac{|a|y^{ap-1}(1-(1-c)(y/b)^a)^{q-1}}{b^{ap}B(p,q)(1+c(y/b)^a)^{p+q}} \text{ for } 0 < y^a < b^a$$

= 0 otherwise, where $0 \le c \le 1$, and b, q, p > 0.

The GB includes the other distributions as its special cases. For instance,

$$GB1(y; a, b, p, q) = GB(y; a, b, c=0, p, q)$$

$$GB2(y; a, b, p, q)=GB(y; a, b, c=1, p, q).$$

In addition,

$$B(y; b, c, p, q) = GB(y; a=1, b, c, p, q).$$

The remaining distributions that will be estimated represent special cases of GB1, GB2, and B distributions and can easily be derived.

B1(y; b, p, q)= GB(y;
$$a=1$$
, b , $c=0$, p , q) and

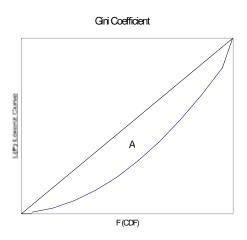
B2(y; b, p, q)= GB(y;
$$a=1$$
, b , $c=1$, p , q).

Further relationships between GG, BR3, BR12, GA, LN are described in greater detail by McDonald and Xu (1994).

B. A Measure of Inequality: the Gini Coefficient

In order to assess the degree of inequality in any given country, we will estimate the Gini coefficient. Prior to the discussion about the Gini coefficient, it is necessary to define the Lorenz

curve, one of the most important tools in measuring income inequality. The Lorenz curve relates the cumulative proportion of population to the cumulative proportion of income, after the population is ordered according to increasing levels of income. The Gini coefficient is shown to be twice the area between the equidistribution line and the Lorenz curve (this area is designated A in the following diagram).



Camilo Dagum (1988) suggests that this measure of inequality has become widely accepted due to the fact that it fulfills desirable theoretical properties: (i) transfer, (ii) proportional addition to incomes, (iii) equal addition incomes, (iv) proportional addition to persons, (v) symmetry, (vi) normalization, and (vii) operationality.

IV. Description of the Data

The analysis is based on Gross Income variable obtained from the Luxembourg Income Study database. The database maintains income information from countries who are members of the Organization for Economic Co-Operation and Development (OECD), hence the choice of countries for this analysis. The variable that was chosen for this analysis is a summary income variable that includes the following components:

- All forms of cash wage and salary income, including bonuses, health insurance benefits, accident and disability pay, net of taxes and mandatory contributions;
- Self-employment income net of social insurance contributions;
- Cash-property income, including interest, rent, dividents, and annuities, excluding any form of lump sum payments (i.e., lottery winnings, insurance settlements, etc.);
- Cash social security benefits for senior citizens, retirees, survivors, as well as alimony and child support, as well as private and public pensions;
- Cash payments for child or family allowances;
- Unemployment, maternity and military benefits provided by the government and by the employer on a non-mandatory basis.

Income data is provided in nominal values and in local currency units. Only grouped data was available; hence, the sample has been divided into twenty equal probability intervals. In fact,

$$I_i = (Y_{i-1}, Y_i)$$
, where $Y_0 = 0$ and $Y_{20} = \infty$.

Descriptive statistics for the data are provided in the following table.

Variable	Obs	Mean	Std. Dev.	Min*	Max	
Australia 85	8369	25859.38	21653.82	10	458340	
Australia 89	16244	35316.98	30266.71	3	527370	
Australia 94	7318	37796.44	31100.35	1	391539	
Canada 87	11927	33558.44	23793.57	150	272800	
Canada 91	21568	41264.89	31491.72	100	1107936	
Canada 94	38940	44732.3	34224.82	199	900000	
UK 86	7156	11179.45	8605.493	2	148027	
UK 91	7043	17505.15	17145.67	5.356	561693.4	
UK 95	6778	19897.1	17945.33	2.756	304968.5	
USA 86	12482	29624.22	24426.11	10	393380	
USA 91	15915	35526.94	29170.85	1	411765	
USA 94	65286	39310.44	36949.15	1	653494	
Israel 79	2271	33303.99	25343.52	910	367310	
Israel 86	4998	23258.15	22193.04	460	808530	
Israel 92	5212	5602.243	4603.519	158.53	68192.29	
Taiwan 86	16434	365325.7	256539.6	2120	9342800	

Taiwan 91	16434	638757.7	484209.1	11142 3.27E+07
Taiwan 95	14706	877236.1	584723.2	984 1.32E+07
Russia 92	6292	83329	269612.7	44.85981 1.84E+07
Russia 95	3372	63746.68	70720.54	21.72 821551.3
Norway 86	4969	249556.5	149451.4	892 1565626
Norway 91	8059	372922.6	245537.4	436 4920045
Norway 95	10106	365696.9	328056.4	1 1.46E+07

Table 1. Descriptive Statistics.

V. Estimation and Analysis

The estimators of the distribution parameters were obtained using maximum likelihood estimation by maximizing the following multinomial likelihood function

$$L(\mathbf{q}) = N! \prod_{i=1}^{g} \frac{(p_i(\mathbf{q}))^{n_i}}{n_i!}, \text{ where}$$

$$p_i(\mathbf{q}) = F(Y_i; \mathbf{q}) - F(Y_{i-1}; \mathbf{q}),$$

over the unknown distributional paramaters q, where F() represents the cumulative distribution function, q denotes the distribution parameters, and Y_i and Y_{i-1} represent a particular group of data (McDonald, 1996). Closed form solutions for q do not exist; hence, iterative techniques are used. The Newton-Raphson, David-Fletcher-Powell, Simplex, and Greadient Search algorithms were using the program GQOPT with a convergence criterion of 10^-8.

These results are presented in tables 2-11.

^{*} Observations of zero and below zero were dropped before the estimation.

The distributions are compared based on the values of the sum of squared (SSE) and absolute errors (SAE), the chi², and the log-likelihood function, where

$$SSE = \sum_{i=1}^{21} \left(\frac{n_i}{N} - p_i(\hat{\Theta}) \right)^2,$$

and

$$SAE = \sum_{i=1}^{21} \left| \frac{n_i}{N} - P_i(\hat{\Theta}) \right|,$$

Whereas, the χ^2 is evaluated using the following formula:

$$\mathbf{c}^2 = N \sum_{i=1}^{21} \left(\frac{n_i}{N} - P_i(\hat{\Theta}) \right)^2 / P_i(\hat{\Theta}).$$

Based on these criteria, in the majority of cases generalized beta of the first kind outperforms all other distributions. At first this result might seem puzzling, since it has been shown before that generalized beta of the second kind provided a better fit for income distribution. Indeed, the results of this paper do not contradict these previous findings: in fact, for the United States in both 1986 and 1991 GB1 and GB2 provide the same test statistics, and if 1994 income data is used, then GB2 is clearly a better fit. These results confirm those of McDonald, who in his 1984 paper showed that GB2 gave a better fit for income data in the US. According to our findings, since a number of economic factors affect income distribution, this variation in best fitting distributions is not surprising when considered across a wide spectrum of countries.

Upon careful examination of estimation results of the three-parameter distributions, (BR3, BR12, B1, B2, GG) one can easily see that GG is generally the optimal distributions (in

14 out of 17 cases regardless of the criteria investigated). However, it is interesting to note that income data from Taiwan and Norway suggests that BR3 would give the best fit. Since both of these countries are characterized by a much lower Gini coefficient, and hence a smaller income dispersion, these findings yet again suggest that the choice of an optimal functional form for estimation of income distribution is indeed dependent on a number of economic factors.

In most cases, the gamma distribution is much better than the lognormal. This could be due to the fact that the lognormal weighs income in the upper tail more than the gamma, and thus overestimates both the mean, the variance, and the Gini coefficient.

We can compare nested models using the log-likelihood ratio test. This test statistic is calculated as follows

$$2[\ln L(\Theta_{\rm ML}) - \ln L(\Theta_{\rm R})],$$

where Θ_{ML} and Θ_R represent parameter estimates of the general and the restricted models, respectively. This statistic is asymptotically distributed as a χ^2 with degrees of freedom equal to the number of restrictions imposed on GB or another more general distribution. Based on this test, we may conclude that the differences between GB2 and GG are statistically insignificant in the majority of cases. Some exceptions to this case include distributions of income in Taiwan and Norway. Recall that both these countries are characterized by a relatively low Gini coefficient, which we believe causes these differences between the results for these two and the rest of the countries. It is also notable that the results obtained in this study are in accordance with that of McDonald conducted in 1984 where he used US family income data. In fact, our findings are consistent with his findings that that statistical differences between GB1 and GG are insignificant, and that GB1, B1, GG and GA share many similarities.

The χ^2 statistic, which is used to provide a goodness of fit measure, is distributed as a χ^2 with degrees of freedom equal to the difference between the number of income groups in the data and the number of parameters estimated. The fact that the χ^2 statistic obtained from the estimations above varies anywhere from 100 to 3000 allows us to reject the null hypothesis.

In the process of estimation, some difficulties were encountered. For instance, unusually low estimates of the Gini coefficient were obtained for GB1 using data from Canada 1991 and for GB2 for Norway 1986 (.25 and 2.20E-7, respectively). The Gini coefficient for the GB1, GB2, and GG involves evaluating an infinite series. It appears in the problem cases that computation problems are encountered in evaluating the corresponding infinite series. In fact, in a number of cases, for instance while estimating GB2 for Israel 1979, Norway 1995, Canada 1987, and GB1 for Taiwan 1986 and 1995, Australia 1985, United Kingdom and Russia 1995, the Gini coefficient was not obtained because the series never converged.

VI. Conclusions and Suggestions for Future Study

In summary, we have been able to show that the best functional forms for probability density functions describing income distributions are GB1 and GB2. As restrictions are placed on the parameters of these distributions, some three-parameter distributions can be shown to be better than other three-parameter for describing income distribution of particular countries, although the most general distributions exceed these distributions in relative performance. This is due to the fact that economic indicators, specific to each country, can influence the income distribution in each particular country. Hence, one may easily conclude, that unless the most general functional forms are used, it is best to study what factors affect the parameters of income distribution, and then estimate a specific distribution based on this information. Measures of

poverty and income dispersion can then be easily derived and used for policy analysis, thus leading to correct results.

To further this study, we plan to try to estimate the relationship between distribution parameters and economic factors that seem to affect income distribution. Some of these indicators that have been previously considered include measures of total employment, gross national product, transfer payments, government purchases of goods and services, etc (Thurow, 1970). In addition, demographic variables such as age of the head of the household, his/her education level, work experience, number of children in the household, and investments, could be used to predict the parameters of income distributions.

Bibliography

- 1. Aitchinson, J., and Brown, J.A.C., 1969, The lognormal distribution with special reference to its uses in economics, Cambridge University Press.
- 2. Bordley, R.F., McDonald, J.B. and Mantrala, A., 1996, Something new, something old: parametric models for the size distribution of income, Journal of Income Distribution 6, 91-103.
- 3. Dagum, Camilo, 1988, Gini ratio, The New Palgrave: A dictionary of Economics, Eds: Eatwell, John, Murray Milgate and Peter Newman, volume I.
- 4. Grilliches, Zvi and Intrilligator, Michael, 1983, Handbook of Econometrics, Volume I, Elsevier Science Publishers, Amsterdam, the Netherlands.
- 5. McDonald, J.B. Lecture Notes: Econometrics 588, Winter 2000.
- 6. McDonald, J.B. and Xu, Y.J., 1995, A generalization of the beta distribution with applications, Journal of Econometrics 66, 133-152.
- 7. McDonald, J.B., 1984, Some generalized functions for the size distribution of income, Econometrica 52, 647-663.
- 8. Parker, S.C., 1999, The generalized beta as a model for the distribution of earnings, Economics Letters 62, 197-200.
- 9. Salem, A.B. and Mount, T.D., 1974, A convenient descriptive model of income distribution: the gamma density, Econometrica 42, 1115-1127.
- 10. Thurow, L.C., 1970, Analyzing the American income distribution, Papers and Proceedings, American Economics Association 60, 261-269.
- 11. Wolfehnson, James D., 2000, Combating poverty. < http://www.worldbank.org/> 2000, March15.

Appendix A
The following diagram has been obtained from James B. McDonald and represents the relationships between distribution estimated in this paper.

	GB	GB1	GB2	В	B1	B2	$\mathbf{G}\mathbf{G}$	BR3	BR12	GA	LN
а	1.00E+00	7.52E-03	1.00E+00		1.00E+00	1.000	0.9468	3.35	1.526	1.000	9.871
b	6.70E+10	1.74E+05	6.70E+10		5.69E+05	2.98E+05	1.17E+04	3.53E+04	1.14E+05	1.36E+04	0.812
p	1.87E+00	4.30E+02	1.87E+00		1.80E+00	2.09E+00	2.05E+00	0.385	1.000	1.87	NA
q	4.94E+06	7.23E+00	4.94E+06		3.87E+01	2.58E+01		1.000	9.21E+00		NA
c	1.00E+00										NA
Mean	25,380.9	25,229.40 2	25,380.9		25,319.80	25,088.60	25,409.30	26,334.00	25,393.90	25,382.70	26,914.10
SSE	2.78E-03	2.61E-03	2.78E-03		2.76E-03	2.93E-03	2.78E-03	3.59E-03	2.97E-03	2.78E-03	4.94E-03
SAE	0.16	0.1515	0.16		0.155	0.16	0.163	0.197	0.1585	0.1604	0.241
Chi^2	9.17E+02	8.55E+02	9.17E+02		9.18E+02	9.49E+02	9.12E+02	1.17E+03	9.75E+02	9.17E+02	1.11E+03
#NAME?	4.01E+02	3.81E+02	4.01E+02		3.99E+02	4.13E+02	4.00E+02	4.85E+02	4.19E+02	4.00E+02	5.37E+02
Australia											_
89											
a	0.7977	0.7738	0.7977		1.00E+00	1.000	0.7889	2.987	1.5217	1.000	10.158
b	4.94E+07	2.28E+07	4.94E+07		2.44E+07	8.06E+05	9.39E+03	4.48E+04	1.05E+05	1.95E+04	0.836
p	2.61E+00	2.75E+00	2.61E+00		1.76E+00	1.82E+00	2.67E+00	.441	1.000	1.77E+00	
q	8.95E+02	4.36E+02	8.95E+02		1.25E+03	4.36E+01		1.00E+00	5.44E+00		
c	3.65E-09		3.65E-09								
Mean	34,546.50	34,563.40	34,546.50		34,389.00	34,478.10	34,561.50	36,128.90	34,609.70	34391.8	36,569.80
SSE	2.12E-03	2.12E-03	2.12E-03		2.16E-03	2.15E-03	2.12E-03	3.01E-03	2.39E-03	2.16E-03	3.84E-03
SAE	0.1421	0.143	0.1421		0.134	0.1383	0.1428	0.1853	0.1428	0.1342	0.215
Chi^2	1.02E+03	1.02E+03	1.02E+03		1.06E+03	1.05E+03	1.02E+03	1.46E+03	1.15E+03	1.06E+03	1.36E+03
#NAME?	5.01E+02	5.01E+02	5.01E+02		5.14E+02	5.12E+02	5.02E+02	6.74E+02	5.53E+02	5.14E+02	7.13E+02
Australia											
94	1 0002	2.025.02	1.0002		1.00E : 00	1.000	0.0757	2.015.00	1 467	1.000	1.000.01
a	1.0003	2.02E-02	1.0003		1.00E+00	1.000	0.8757	3.01E+00	1.467	1.000	1.02E+01
b	3.12E+10	2.86E+05	3.12E+10		3.43E+06	3.12E+10	1.49E+04	4.97E+04	1.58E+05	2.16E+04	0.84789
p	1.73E+00	1.53E+02	1.73E+00		1.72E+00	1.73E+00	2.18E+00	4.26E-01	1.000	1.73E+00	
q	1.45E+06	7.43E+00	1.45E+06		1.56E+02	1.45E+06		1.000	8.02E+00		
c	1.00E+00	25 4 40 00	1.00E+00		27 210 10	27 20 7 00	25, 120, 2	20.220.50	27.42 < 40	27 22 4 60	20.041.00
Mean	37,326.70	37,149.80	37,326.70		37,310.40	37,295.80	37,438.3	39,238.60	37,426.40	37,334.90	39,861.90
SSE	1.89E-03	1.71E-03	1.89E-03		1.89E-03	1.89E-03	1.91E-03	2.85E-03	2.11E-03	1.90E-03	4.09E-03
SAE	0.1512	0.1426	0.1512		0.1493	0.1514	0.1595	0.1888	0.15	0.1512	0.2362
Chi^2	3.66E+02	3.22E+02	3.66E+02		3.66E+02	3.65E+02	3.57E+02	5.56E+02	5.68E+02	3.66E+02	5.52E+02
#NAME?	2.35E+02	2.17E+02	2.35E+02		2.35E+02	2.35E+02	2.33E+02	3.16E+02	3.24E+02	2.35E+02	3.49E+02
Gini	0.3722	0.3999	0.3722		0.3986	0.3987	0.4029	0.4222	0.4001	0.3993	0.4512