Title
AN ENTROPY BASED ANALYSIS OF EUROPEAN MICRO INCOME DISTRIBUTIONS AND INEQUALITY MEASURES

Authors
Judge, George G.
Villas-Boas, Sofia B.
Fu, Quizi

Publication Date
2017-07-13

Peer reviewed

AN ENTROPY BASED ANALYSIS OF EUROPEAN MICRO INCOME DISTRIBUTIONS AND INEQUALITY MEASURES

SOFIA B. VILLAS-BOAS, QIUZI FU, AND GEORGE JUDGE

July 12, 2017

Abstract

The focus of this paper is on recovering probability density functions from samples of time ordered household micro income data. To investigate the heterogeneity and time paths of country income distribution functions we use household micro data from European countries. For information recovery we use a family of information theoretic divergence measures that maximizes entropy under constraints. This type of quantitative income analysis is important since it provides a framework for processing information on how a country's economy is functioning, how the allocation and distribution system is performing, and in terms of dynamics, how the economic system changes over time.

JEL: D31, E21, C1, C10, C24

Keywords: Income Probability Distribution Function, Micro Income Data, Information Theoretic Methods, Cressie-Read Divergence, Entropy Maximization, Pareto's law.

Corresponding author: Sofia Villas-Boas, Professor of ARE, University of California, Berkeley, CA, 94720 (sberto@berkeley.edu), Qiuzi Fu, National School of Development, Peking University, Beijing, China, 100871 (jennyfuberkeley@gmail.com), George Judge, Professor of the Graduate School and Giannini Foundation, 207 Giannini Hall, University of California Berkeley, Berkeley, CA, 94720 (e-mail:gjudge@berkeley.edu). We thank E. Aisbett, S. Gold, M. Grendar, W. Griffiths, J. Lee, E. Saez, T. Squartini, I. Trilnick, A. Stevens, and K. Train for excellent comments. We thank the Giannini Foundation for support.

1. INTRODUCTION

In this paper we use information theoretic entropy based divergence methods to investigate the heterogeneity and time paths of income probability density-distribution functions from country based samples of micro household income data. Income probability density functions-distributions evolve from complex, uncertain, and volatile economic behavioral systems that are seldom in equilibrium. This means that behavioral system income distribution outcomes may best be viewed and analyzed in an information theory-probability context. In economic behavioral systems, markets provide a basis for processing information and determining the value of most of the components in the income portfolio and the distribution of income-wealth is the result of a very complex set of market interactions, governmental policies and interventions, and changes that are institutional in nature. Thus at the economic unit-country level, the income probability density function-distribution contains information on how the market is functioning, how the allocation and distribution system is performing, and in terms of dynamics, how the economic system has changed and is changing over time. In this paper we obtain this type of information using informational theoretic entropy based econometric methods and annual samples of European micro income data, as a basis for estimating, ordering and determining the informational content of countrywide income probability density functions-distributions. In using information theory based methods to recover a country's probability density-income distribution, we have the possibility for drawing inferences as to how the economy is functioning in an absolute and relative sense. In this entropy approach to information recovery, moment constraints provide a basis for representing our knowledge of the micro economic behavioral system in terms of a probability density function-distribution. The income distribution that can be obtained in the largest number of ways is the maximum entropy income distribution. A unique country based entropy measure of income equality-inequality results within this possible framework for defining a statistical income distribution equilibrium.

The study of income distribution recovery has a long history. In the search for a universal regularity in income and wealth, Pareto (1896) originally suggested an exponential-power law income distribution to describe the allocation of wealth among individuals, and to demonstrate that a larger portion of the wealth of any society is owned by a small percentage of the people. Pareto's description of the nature of the income distribution is sometimes expressed more simply as the Pareto fat tail principle or the "80-20 rule", which says that 20% of the population controls 80% of the wealth (see for example Gabaix, et al., 2016). Power laws are proposed as ergodic distributions for stochastic processes by Champernowne, (1953) and Levy and Solomon, (1996), to explain the Pareto exponential distribution of income. Many possible measures, such as the Gini concentration ratio, the Lorenz curve (see for example Theil, 1967 and Adamou and Peters, 2016) and the exponential distribution (Drăgulescu, and Yakovenko, 2001 and Cho, 2014), have been proposed to measure income equality-inequality and reflect the nature of the distribution of income. In addition the dynamic nature of income distribution and the evolution of income inequality has been the focus of a growing empirical literature in the context of developed countries (for a survey see Piketty and

Saez, 2014; Saez and Zucman, 2016; Piketty and Saez, 2003; and Alvaredo, et al, 2013; Roine and Waldenstroem, 2015).

Building on the productive efforts noted above, we contribute to this literature by presenting a new uniform entropy-based behavior-related method that simultaneously recovers country based income probability density-distribution functions and the corresponding income equality-inequality measure, from samples of micro income data. We extend this literature by recovering country based probability density-income distribution functions that serve as a basis for determining how the economic systems are functioning and measuring the nature of income equality-inequality. Traditionally, there is no commonly accepted definition of inequality and the basis for determining the underlying income distribution is unrelated to the equality measure. Thus, there are different interpretations of the inequality concept as well as existing questions about the underlying dynamics of income inequality and the way in which income distributions and income equality-inequality change over time. From an empirical standpoint, a Eurostat data series of about two million household-country-year observations permits us to pursue some of these questions and allows us to use a new adaptive intelligent behavior-causal entropy maximization conceptual framework and information recovery method to study income distributions and income inequality in a European country based context.

In the sections to follow, using some of the tools associated with information theory and statistical physics, we argue in section 2 that the probability distribution of income is given by the possible ways in which a collection of non interacting micro incomes may occupy a set of discrete income states. Consistent with the data generation process, we use an information theoretic entropy based method (see Judge and Mittelhammer, 2012a, 2012b) and micro sample data to recover the unknown probability density-distribution function. Using European country level micro income data that is described in Section 3, we investigate in section 4 whether economic behavior and the nature of income equality-inequality is captured by this entropy based distribution and inequality measure. In the empirical information recovery process of section 4 we recognize the connection between adaptive intelligent behavior, causal entropy maximization, and self-organized equilibrium seeking behavior in an open economic behavioral system, and use the Cressie-Read family of divergent measures to determine the nature of the distribution of European country based micro income data for the sample of years 2008 to 2013. In Section 5 we investigate the time ordered nature of income inequality for countries featuring the longest available income series. Finally, in section 6 we discuss the implications of the methods and the micro income data sample results.

2. AN ENTROPY-INFORMATION RECOVERY FRAMEWORK

As we focus on the recovery of the underlying probability density functions (PDFs) from micro income data, we recognize and emphasize the behavior related nature of the observations. This means we recognize that like prices, incomes do not behave, but that people behave. Thus a countries income distribution is one way to exhibit and summarize economic behavior and its allocative and distributive performance. In 1948, Shannon looked at a communication system and saw informational content in the context of a dynamic entropic system. In a similar way, recognizing the potential use of information theory in economics leads us to the question: does the distribution of samples of the micro income data provide a basis for recovering information regarding the unknown parameters of a micro behavioral income distribution system? In seeking an answer to this question we follow Wissner-Gross and Freer (2013) and recognize the connection between adaptive intelligent behavior, causal entropy maximization (AIB-CEM), and self-organized equilibrium seeking behavior in an open dynamic economic system. Under this optimizing criterion each microstate can be seen as a causal consequence of the macro income state to which it belongs. This connection between causal adaptive behavior and entropy maximization, which is based on a causal generalization of entropic forces, suggests that economic social systems do not evolve in a deterministic or a random way, but tend to adapt behavior in line with an optimizing principle. In the sections ahead, we exhibit new ways to think about income information recovery and the causal adaptive behavior of large complex micro economic systems, and the use entropy as the systems status equality-inequality measure.

2.1 Problem Formulation and Solution

In the context of the above Wissner-Gross adaptive intelligent behavior causal entropy maximization framework, in this section we discuss how information theoretic methods may be used to establish a data based link to recover income probability density functions-distributions. It seems reasonable that the resulting distribution of a sequence of positive real numbers from a sample of income data should vary over countries and economic systems. In this context information theoretic methods offer a natural way to capture income distributions in the form of a probability density function.

In recovering the income probability distribution-density function from a sample of N positive real numbers, we assume the income probability to be represented by partitions-bar plots–histograms that span the income sample space. Samples from these bar plots yield histogram outcomes of the discrete random income variable d_j, for $j = 1, 2, \ldots , n$, and under repeated observation, one of n histograms-micro configurations associated with the macro state income is observed with probability p_j. Further, suppose after a large number of trials, we have first-moment sample information in the form of the mean value of a country's income:

$$\sum_{j=1}^{n} d_j \, p_j = \bar{d}. \tag{2.1}$$

Given this first-moment sample information and the inverse problem of identifying an income distribution from the sample income data, we seek the best predictions of the unknown probabilities $p_1, p_2, \ldots , p_n$. It is readily apparent that there is one data point $\bar{d}$, and n unknown p_i. From an information recovery standpoint there are an infinite number of possible discrete probability distributions with $\bar{d} \in [1, n]$. Based only on the information $\sum_{j=1}^{n} d_j \, p_j = \bar{d}$, $\sum_{j=1}^{n} p_j = 1$, and $0 \leq p_j \leq 1$, the problem cannot be solved for a unique solution. Thus, a function must be inferred from insufficient information when only a feasible set of solutions is specified. In such a situation it is useful to have an approach that allows the investigator to use sample based information recovery methods without having to choose a parametric family of probability densities on which to base the income function. If we replace the expectation value of the macro variable income by its most likely value, this is equivalent to maximizing entropy with respect to the macro state. This is the problem to which we now turn.

2.2. The Information Theoretic Family

Making use of the connection between adaptive intelligent behavior and causal entropy maximization as an optimizing criterion-status measure, in this section we discuss how information theoretic entropy based methods provide a natural basis for establishing a causal influence-econometric-inferential link to the data and solving the resulting ill posed stochastic inverse problem. In this ill posed inverse problem context, we face a system which may have more than one solution, or in which the solutions depend discontinuously upon the initial data. This type of uncertainty, regarding the economic-econometric model, the associated estimating equations and the data sampling-probability distribution function, create unsolved problems as they relate to information recovery. Although likelihood is a common loss function used in fitting econometric models, the optimality of a given likelihood method is fragile inference-wise under model uncertainty. In addition the precise functional representation of the data sampling process cannot usually be justified from economic-behavioral theory. Given this situation, a natural solution is to use estimation and inference methods that are designed to deal with systems that are fundamentally ill posed and stochastic in nature and uncertainty and random behavior are basic to information recovery. To identify estimation and inference measures that represent a way to link the model of the process to a family of possible likelihood functions associated with the income data, we suggest and use the Cressie and Read (1984) and Read and Cressie (1988) single parameter CR family of entropic function-power divergence measures given by

$$I(\mathbf{p},\mathbf{q},\gamma)=\frac{1}{\gamma(\gamma+1)}\sum_{i=1}^{n}p_i\left[\left(\frac{p_i}{q_i}\right)^{\gamma}-1\right]. \tag{2.2}$$

In (2.2), γ is a parameter that indexes members of the CR-entropy family of divergence measures-distributions, p_i's represent the subject probabilities and $q_i's$ are interpreted as reference probabilities that represent any prior pre data information. Being probabilities, the usual probability distribution characteristics of $p_i, q_i \in [0,1] \forall i$, $\sum_{i=1}^{n} p_i = 1$, and $\sum_{i=1}^{n} q_i = 1$ are assumed to hold. In (2.2), as γ varies, the resulting CR-entropy statistical family of estimators that minimize power divergence, exhibit qualitatively different sampling behavior that includes Shannon's entropy, the Kullback-Leibler measure and in general a range of independent (additive) and correlated systems (see Gorban, et al., (2010), Judge and Mittelhammer, (2012a) and (2012b)). In identifying the probability space, the CR family of power divergences is defined through a class of additive convex functions and the CR power divergence measure leads to *a broad family of likelihood functions* and test statistics. All well known divergences belong to the class of CR functions. The CR measure exhibits proper convexity in **p**, for all values of γ and **q**, and embodies the required probability system characteristics of additivity and invariance with respect to monotonic transformations of the divergence measures (see Gorban, et al., 2010).

In the context of extremum metrics, the general CR family of power divergence statistics represents a flexible family of pseudo-distance measures from which to recover the joint distribution probabilities and encompasses a wide array of empirical goodness-of-fit and information recovery criteria. As γ varies power law Pareto behavior is efficiently described and the resulting estimators that minimize power divergence exhibit qualitatively different sampling behavior. To place the CR family of power divergence statistics in an entropy perspective, we note, following Gorban, et al. (2010), that there are corresponding families of entropy functions–divergence measures. Over defined ranges of the divergence measures, the CR and entropy families are equivalent.

If in the CR family of entropy functions, in the limit as $\gamma \rightarrow 0$, the solution of the first-order condition leads to Shannon maximum entropy and the logistic expression for the conditional probabilities. Alternatively, if in the family of CR entropy functional in the limit $\gamma \rightarrow -1$, the solution of the first-order condition leads to the maximum empirical likelihood distribution for the conditional probabilities. These criterion based entropy functions permit us to use the information content of the structure of the income-probability density function-distribution as a measure of equality-inequality (Thiel, 1967 and Colwell, 2003). The income statistical system may be characterized by a macro state, for which many micro configurations exist, that are compatible with it. To obtain the probability density function income distribution we use the principle of maximizing the CR-entropy functional, subject to constraints, to identify the most likely distribution function-histograms for a given economic statistical system. In this context, recovering the income distribution from a sample of positive real numbers through the use of the CR-entropy criterion (2.2), suggests we seek a solution to the following extremum problem:

$$\hat{\boldsymbol{p}} = \arg\min_{\boldsymbol{p}}\left[I(\boldsymbol{p},\boldsymbol{q},\gamma) \mid \sum_{j=1}^{n} p_j\, d_j = \bar{d}, \sum_{j=1}^{n} p_j = 1, p_j \geq 0\right]. \quad (2.3)$$

Solving this optimization problem provides a solution to the income probability distribution function and to the entropy inequality measure. In general, the solution to this extremum problem does not have a closed-form expression and the optimal values of the unknown network parameters must be numerically determined.

3. THE EUROSTAT MICRO HOUSEHOLD INCOME DATA

As an information base for the recovery method discussed in section 2, we make use of Eurostat country based micro income data. Eurostat is a Directorate-General of the European Commission. Its main responsibilities are to provide statistical information to the institutions of the European Union (EU). Considering data availability and country characteristics, we use income household data for the following central and Northern European countries listed in Table I.

Table 3.1

COUNTRIES AND ABBREVIATIONS AND NUMBER OF OBSERVATIONS

Source: EUROSTAT.

Abbreviation	Country	Number Observations
AT	Austria	58,669
BE	Belgium	59,053
CH	Switzerland	44,035
DE	Germany	119,883
DK	Denmark	57,735
EL	Greece	45,245
ES	Spain	102,211
FI	Finland	106,547
FR	France	77,053
IE	Ireland	51,895
IT	Italy	138,943
NL	Netherlands	89,436
NO	Norway	56,708
PT	Portugal	37,395
SE	Sweden	67,536
UK	United Kingdom	83,797
Total in 16 Country Sample		1,196,141
Total Negative		804

Total Missing	59
Percent Data Used From Raw Data	99.94%

Among all income related variables in the Eurostat's micro survey data base, we use the variable titled "HY010: Total household gross income" to measure the income level. It is measured in Euros without an inflation factor. If we take a closer look at this variable, it measures the sum for all household members of gross personal income components (gross employee cash or near cash income; gross non-cash employee income; employers' social insurance contributions; gross cash benefits or losses from self-employment including royalties). So it is a comprehensive and well-defined variable for the study, consistent with income measures used in previous studies such as Piketty and Saez (2003, 2014), who focus on the long-run evolution of the inequality of gross income, that is income before taxes and government transfers.

Not all countries report data for all the years. In Table 3.2 we specify which countries have data available for each of the years from 2004 to 2013.

Table 3.2

AVAILABLE YEARLY INCOME DATA SAMPLES BY COUNTRY

Years	Countries
2004	AT,BE,DK,IE,NO,FI,SE
2005	AT,BE,DE,DK,FI,IE,NL,NO,SE,UK
2006	AT,BE,DE,DK,ES,FI,HU,IE,NL,NO,SE,UK

2007	AT,BE,DE,DK,ES,FI,FR,GR,IE,IT,NL,NO,
2008	AT,BE,CH,DE,DK,EL,ES,FI,FR,IE,IT,NL,NO,PT,SE,UK
2009	AT,BE,CH,DE,DK,EL,ES,FI,FR,IE,IT,NL,NO,PT,SE,UK
2010	AT,BE,CH,DE,DK,EL,ES,FI,FR,IE,IT,NL,NO,PT,SE,UK
2011	AT,BE,CH,DE,DK,EL,ES,FI,FR,IE,IT,NL,NO,PT,SE,UK
2012	AT,BE,CH,DE,DK,EL,ES,FI,FR,IE,IT,NL,NO,PT,SE,UK
2013	AT,BE,CH,DE,DK,EL,ES,FI,FR,IE,IT,NL,NO,PT,SE,UK

Source: EUROSTAT.

The household level income data are in Euros for all of the 16 countries. Summary statistics of the household income data by country and by year are reported in the appendix in Table A.1, along with the percent negative income and percent missing income data by country and year. There are a total number of 1,196,141 observations in the data. As shown in the bottom of Table I, 59 of the observations for income are missing, 804 of which are negative. Table I also contains the total number of non-missing and non-negative household income observations by country used in the analysis. In the end we use a total of about 1.6 million observations of yearly household income data without sample weights. As can be seen in the appendix, once broken up by country year, less than 0.82 percent of country year income observations are negative. Furthermore, 0.46 percent of the observations are missing for Spain 2006 and 0.05 percent for Norway in 2006. By country-year almost all data are quite complete and clean. After removing the negative and missing entries we keep 99.94 percent of the original household sample data and have a total 1,445,520 household level income observations by country year.

4. AN INFORMATION THEORETIC DATA INCOME ANALYSIS

In this section building on the method and data of sections 2 and 3, we focus on the analysis of samples of micro household income data from sixteen European countries that range over the years 2004 to 2013. In the analysis of the income data, we make use of the information theoretic methods of Section 2 as a basis for summarizing the income data in the form of a probability density function, and use entropy as an income distribution measure of equality or inequality. In the analysis ahead, using an integrated information theoretic entropy measure, we use twelve income levels-histograms to span the micro sample spaces to investigate the patterns of country probability density-income distributions and to ,make cross country and time comparisons in terms of income equality-inequality

4.1. Maximum Empirical Exponential Likelihood (MEEL) Formulation

Two information-theoretic variants of the CR-entropy $I(\boldsymbol{p}, \boldsymbol{q}, \gamma)$ discrepancy-distance measure are prominent in the literature. The choice of γ is concerned with a measure of uncertainty about the realization of the micro sample data. Out of all the distributions consistent with the constraint set, we choose in the limit $\gamma \rightarrow 0$, which leads to the distribution that can be achieved in the greatest number of distinctive ways (Jaynes, 1978, Judge and Mittelhammer, 2012a). Thus in the limit $\gamma \rightarrow 0$, the Maximum Exponential

Empirical Likelihood (MEEL), is the most likely distribution to be observed from a statistical or combinatorial point of view. It is also the most appropriate measure of effective support size (Grendar, 2006).

Consistent with the discussion of identifying the micro configurations compatible with the income macro state in section 2.2, in the comparative analysis of the Eurostat data, twelve income histogram levels are used to span the range of the micro income data space. Consequently, in analyzing the Eurostat samples of income data we use the CR-MEEL in the limit criterion, γ→0, a uniform reference distribution $\boldsymbol{q}$ $(q_j = \frac{1}{n}, \forall j)$, twelve income-histogram levels, and first-moment information $\sum_{j=1}^{12} d_j\, p_j = \bar{d},$ as a basis for recovering discrete income probability density function-distributions. Under this specification, when in the limit γ→0, the CR $I(\boldsymbol{p}, \boldsymbol{q}, \gamma)$ converges to an estimation criterion equivalent to the maximum exponential empirical likelihood (MEEL) metric $H(\boldsymbol{p}) = -\sum_{j=1}^{12} p_j \ln{(p_j)}$. Our extremum problem likelihood-entropy function may then be formulated as

$$\max_{\boldsymbol{p}} \left[-\sum_{j=1}^{12} p_j \ln\, p_j \;\middle|\; \sum_{j=1}^{12} p_j\, d_j = \bar{d}, \sum_{j=1}^{12} p_j = 1, \mathrm{p} > 0 \right]. \qquad (4.1)$$

The corresponding Lagrange function-extremum problem is

$$L(\boldsymbol{p}, \eta, \lambda) \equiv -\sum_{j=1}^{12} p_j \ln(p_j) + \lambda\left(\bar{d} - \sum_{j=1}^{12} p_j\, d_j\right)$$
$$+ \eta\left(1 - \sum_{j=1}^{12} p_j\right). \tag{4.2}$$

Solving the first-order conditions yields the exponential result

$$\hat{p}_j = \frac{exp\,(-d_j\hat{\lambda})}{\sum_{j=1}^{12} exp\,(-d_j\hat{\lambda})} \tag{4.3}$$

for the jth income outcome and the mean-related income distribution. As the mean of income varies over a range of micro data sets, an exponential family of distributions results. In equation (4.3), $\hat{p}_j$ is a function of $\hat{\lambda}$, the Lagrange multiplier for constraint (4.2). This information may be used as a basis for modifying the distribution of income probabilities. The CR-MEEL-entropy criterion provides an empirical representation of the joint income probability distribution function, where the p_j are chosen to assign the maximum joint probability among all of the possible probability assignments. Using the CR ($\gamma \rightarrow 0$) entropy functional and the mean of a country's income data, we can then recover the resulting probability density income distribution.

4.2. Entropy Measure of Income Inequality

In order to provide the information that is needed to group and compare the income distributions of the 16 European countries, we make use of the entropy measure $E = \gamma \rightarrow 0 = \sum_{j=1}^{12} p_j \ln\, p_j$. In the entropy criterion-measure we seek a probability density function solution for each country over the combined years 2008-2013, that is as close to a uniform PDF-distribution of income (as equal distribution with the least inequality), as the sample data will permit.

To eliminate yearly fluctuations, in Figure 4.1 we use the E entropy measure for the combined 2008-2013 years, to provide an inequality basis for ranking the 16 European countries. The Central and Northern Europe countries displayed on the right half of Figure 4.1, suggest a significant difference in the equality of income when compared to the Southern European countries on the left half of Figure 4.1. For instance, Norway's high relative entropy measure means that Norway has the lowest level of inequality of the European countries studied. On the other hand, the Portugal and Greece low relative entropy measures indicates that these country have the highest level of inequality of the European countries studied. It is interesting to note that these two countries are and have been having major economic problems.

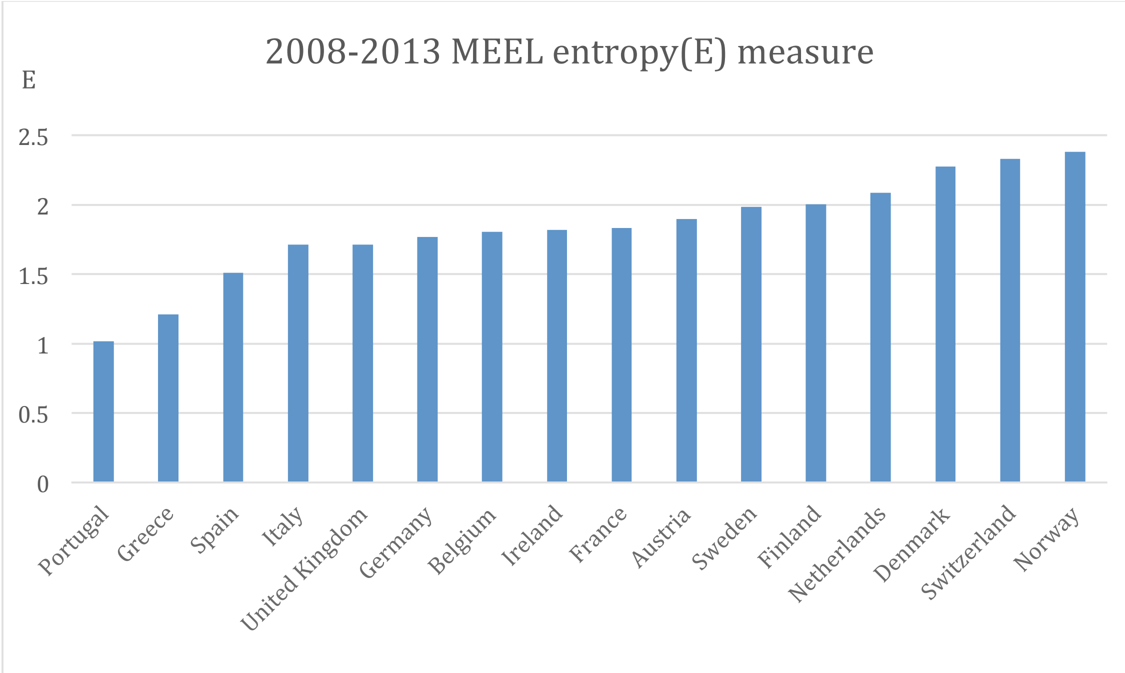

FIGURE 4.1 ENTROPY MEASURE OF INCOME INEQUALITY BY COUNTRY

Given the country based entropy measures in Figure 4.1, the nature of European income distributions is dramatically demonstrated and compared by the income probability density functions of the next sections.

4.3. Income Probability Density Functions

To develop income probability density functions that attempt to capture a statistical measure for each of the 16 European countries, we use a 2008-2013 data sample, and in line with the information recovery method of Section 4.1, divide the income data for each country into 12 bins, bar plots-histogram categories. To make sure the income probability density function for each 16 countries is sufficiently differentiated for the years 2008-2013, and in line with the entropy measures in Figure 4.1, we find that they fall broadly into three main types of income probability density distributions: the Non Euro Zone, Central Europe, and Southern Europe (Portugal, Spain and Greece).

4,3.1 Three Non Euro Zone countries—Entropy measure greater than 2

In Figure 4.2 we display the information theoretic (IT) probability density-income distributions for NO-Norway, CH-Switzerland and DK-Denmark. This group represents two non-European Union (EU) members (NO and CH), and Denmark that have not adopted the Euro as their common currency and sole legal tender. Using the information recovery methods discussed in Sections 2 and 4.1, these three countries conform to the exponential income probability density distribution form, and have an entropy measure value greater than 2, but vary in terms of their probabilities over the income levels. Interestingly, the income density function-distribution for CH-Switzerland lies between the other two income distributions as did its entropy based inequality measure as shown in Figure 4.2. As the entropy measure indicates, NO-Norway has the most uniform probability income distribution with an entropy measure approaching 2.5 and thus displays lower probabilities of being in the lower income levels and higher probability of being in the higher income levels. It would be interesting to comment in a comparative way on the economic and social systems of European countries not using the Euro, but that is another topic.

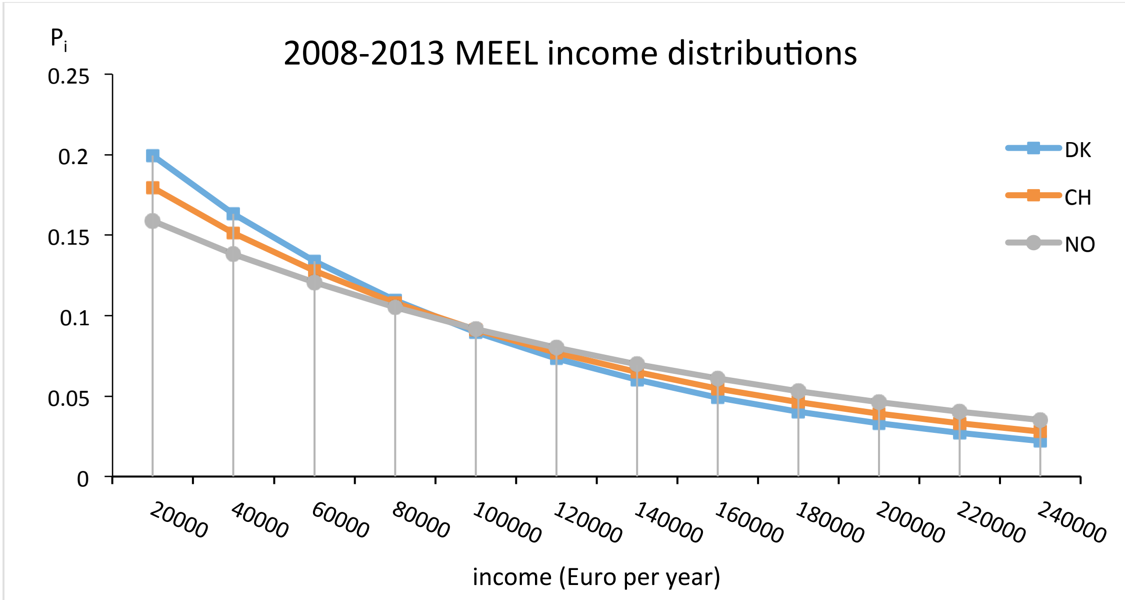

Figure 4.2 INCOME DISTRIBUTIONS FOR LOWEST INEQUALITY NON EURO EUROPEAN COUNTRIES

4.2 Central Europe with Entropy Measures between 1.5 and 2

In Figures 4.3 4.4 and 4.5 we display the income probability density functions for the next block-group of European countries with inequality measures between 1.5 and 2. This group consists of the following 10 European Union and one non-EU country: AT-Austria, BE-Belgium, FI-Finland, DE- Germany, FR-France, IE-Ireland, IT-Italy, NL-Netherlands, SE-Sweden (not in EU), and the UK-United Kingdom (not in the Eurozone). As seen in Figure 4.3 all these countries conform to the exponential distribution shape and share a similar feature relative to their probability density-income distribution functions and entropy measures.

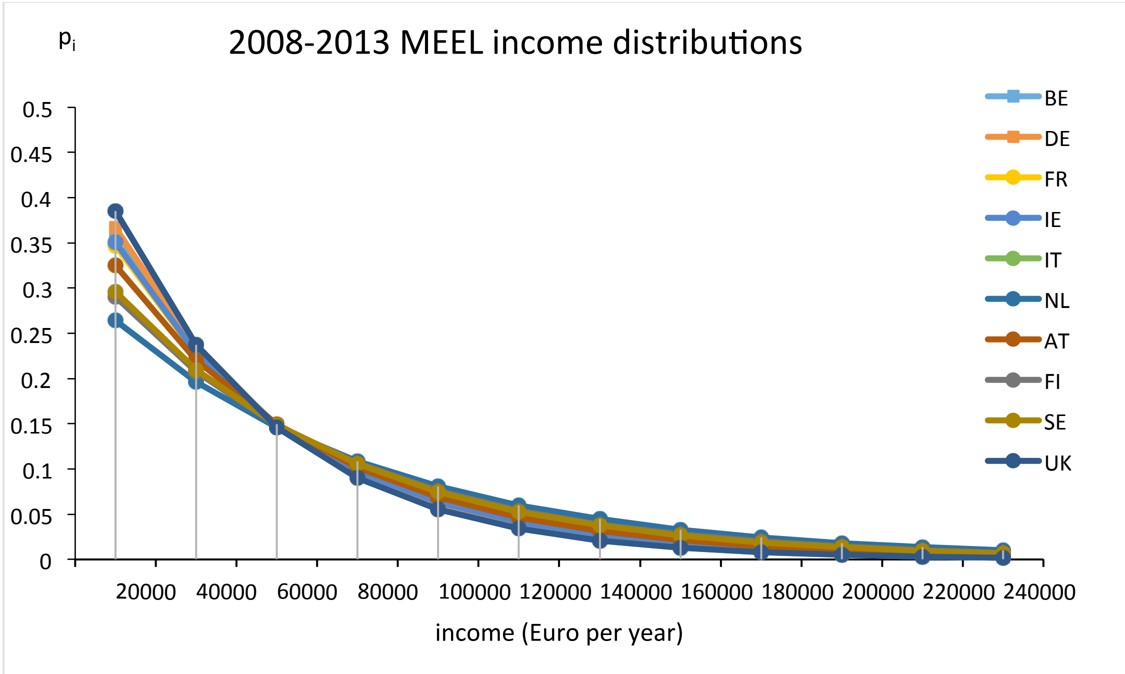

Figure 4.3 INCOME DISTRIBUTIONS FOR CENTRAL EUROPEAN COUNTRIES

There is a good deal of variability in the income density functions in Figure 4.3. Therefore, for clarification and comparative purposes and in line with the entropy measures of Figure 4.1, we break the ten countries into two groups in Figures 4.4 and 4.5. Distribution wise in Figure 4.4 Austria is a bit of an outlier and in Figure 4.5 Italy is a bit of an outlier. In general the agreement of the country income probability density functions in Figure 4.5 is especially noteworthy.

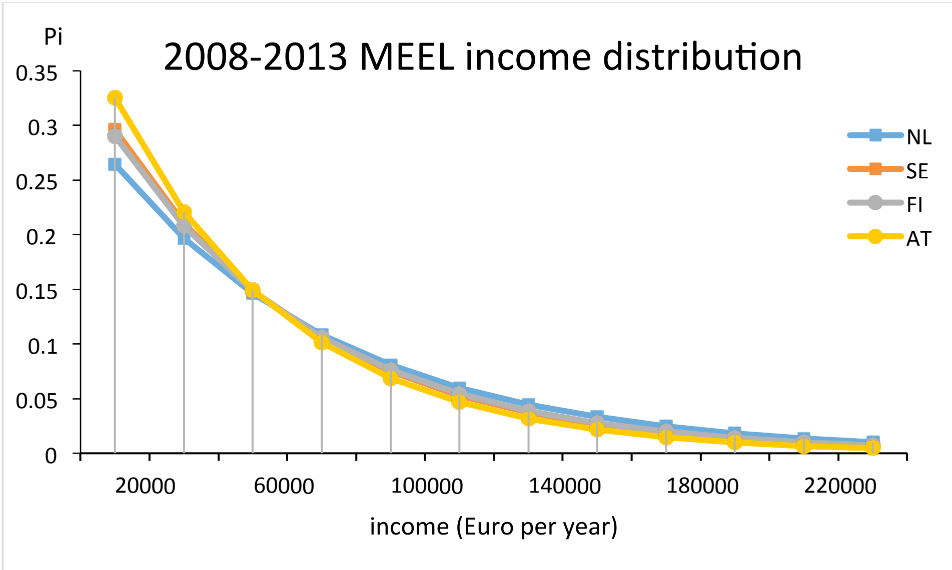

Figure 4.4 NORTH CENTRAL EUROPEAN INCOME DISTRIBUTIONS.

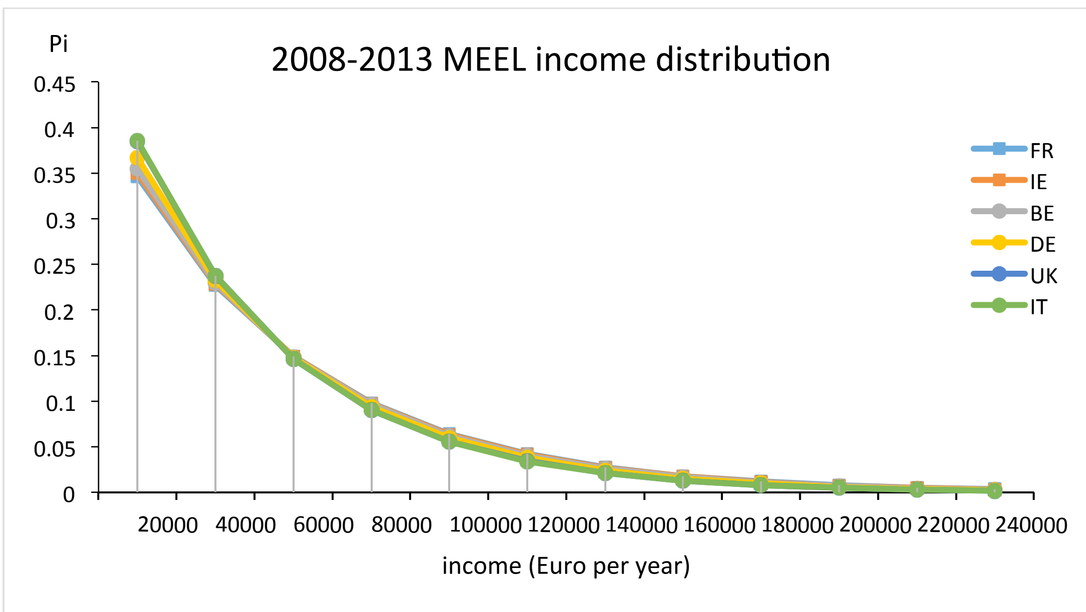

Figure 4.5 CENTRAL EUROPEAN ICOME DISTRIBUTIONS.

4.3.3 Greece, Spain and Portugal with entropy measures less than 2

Finally in Figure 4.6 we display the income distributions for ES-Spain, EL-Greece and PT- Portugal. The income probability density functions for these countries differ sharply from those of the previous two Central European groups, in terms of having high probabilities for the low-income segments and almost uniform probabilities for the higher income segments. The income probability density function for Spain reflects a shift toward a flatter probability density function and thus a higher level of entropy than Greece and Portugal. The entropy equality measures in Figure 4.6 are less than 1.5 for these three countries and reflect the income inequality nature of these income distributions.

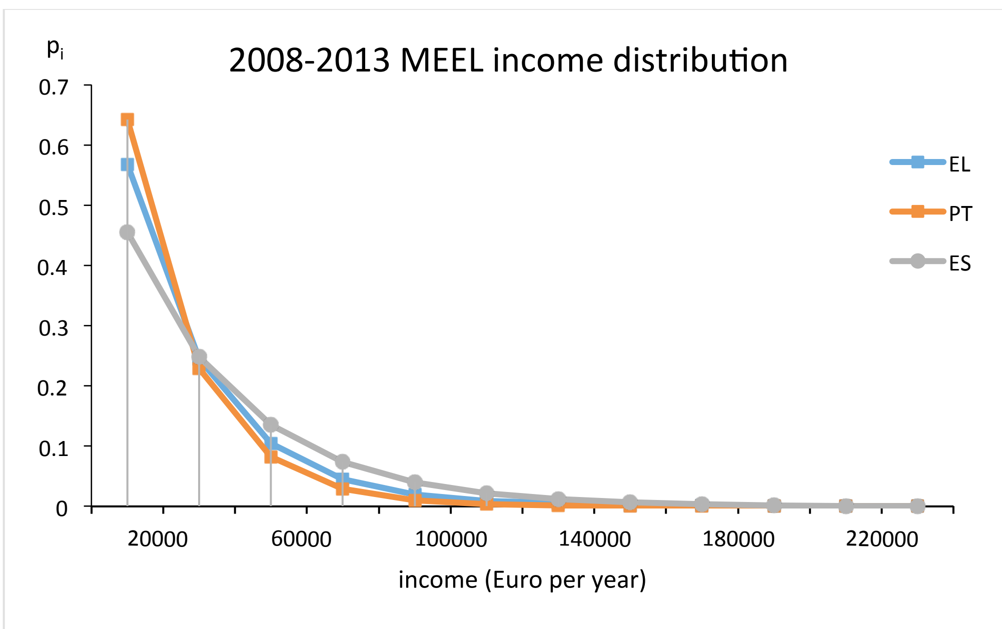

Figure 4.6 INCOME DISTRIBUTIONS: PORTUGAL, SPAIN, AND GREECE

5. TIME DATED MEASURES OF INCOME INEQUALITY

Using the fact that we observe micro income data as a yearly ordered time series, we can estimate the distribution functions for each year for each country and the corresponding entropy measure to investigate the time evolution of connected markets and the changes in country specific income inequality. One way to get the impact of time is to compare the probability density function of the first and the last available yearly income micro data. By making this time ordered density income functions comparison we are not only able to observe the time ordered country impact of connected markets and networks of credit, investment and information, but also able to make generalizations concerning the impact of new technology and networks of trade, finance and information.

5.1. Illustration of Time dated Income Distributions for Sweden

In this subsection we focus on the income time path for SE-Sweden for the decade 2004-2013. As denoted in Figure 5.1, in Sweden from 2004 to 2013, there has been an important shift toward income equality. Many economic studies in measuring income inequality use a ratio of the highest and lowest quintiles (for example see Pickering and Saez, 2014 and Dorling, 2016). These studies are strangely silent in terms of what happens over the remainder of the distribution. The time density comparison in Figure VI, illustrates the *importance* of looking at individual countries and considering the *entire* *probability density-income distribution function,* when drawing static and time related inferences concerning income inequality.

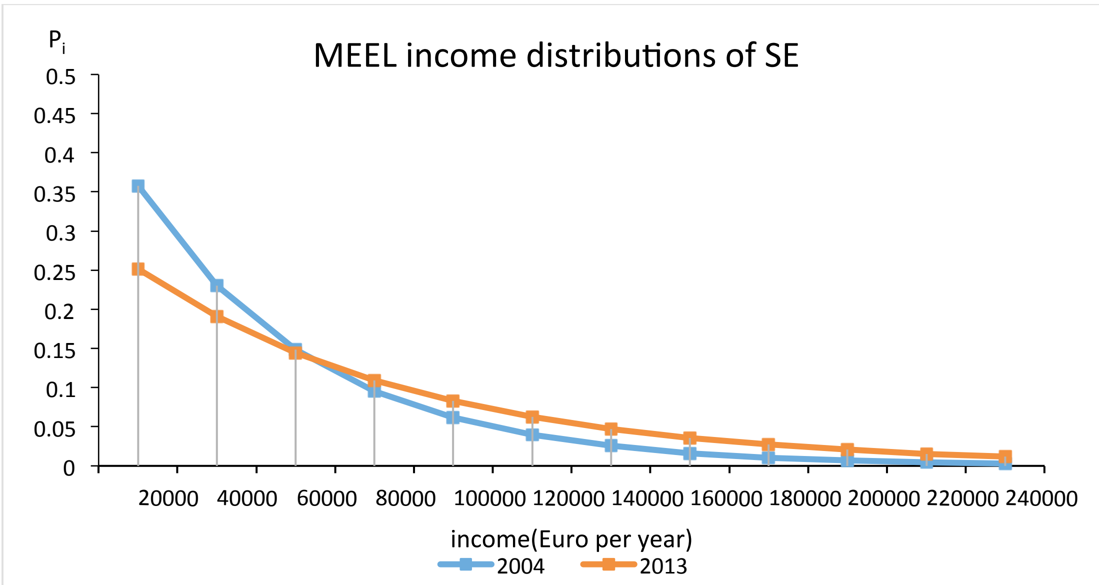

FIGURE 5.1. 2004-2013 IT INCOME PROBABILITY DENSITY FOR SWEDEN

The time implication of this change is displayed further in the yearly entropy measures of Figure 5.2 and in Figure 5.3 where the complete set of income probability density functions for the years 2004-2013 are displayed. In terms of entropy measures, in Figure 5.2, except for the financial crash years of 2009 and 2010, the higher entropy values indicate there has been a steady change over time toward a more equal income distribution in Sweden.

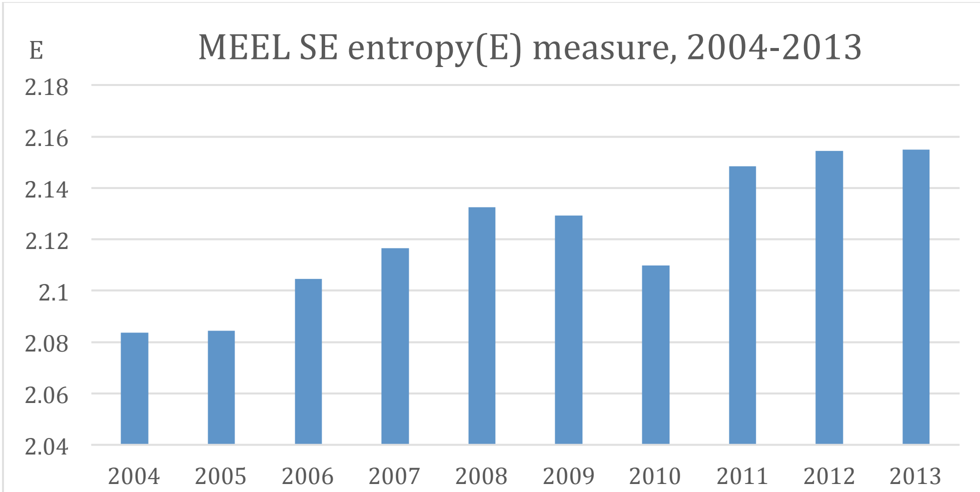

FIGURE 5.2 - ENTROPY MEASURES BY YEAR FOR SWEDEN

Finally, in Figure 5.3 we display the same pattern over time, with the more recent income density functions flattening out relative to the early 2000 years. This time pattern is consistent with figure 5.2 and indicates a shift over time towards a more uniform income distribution in Sweden.

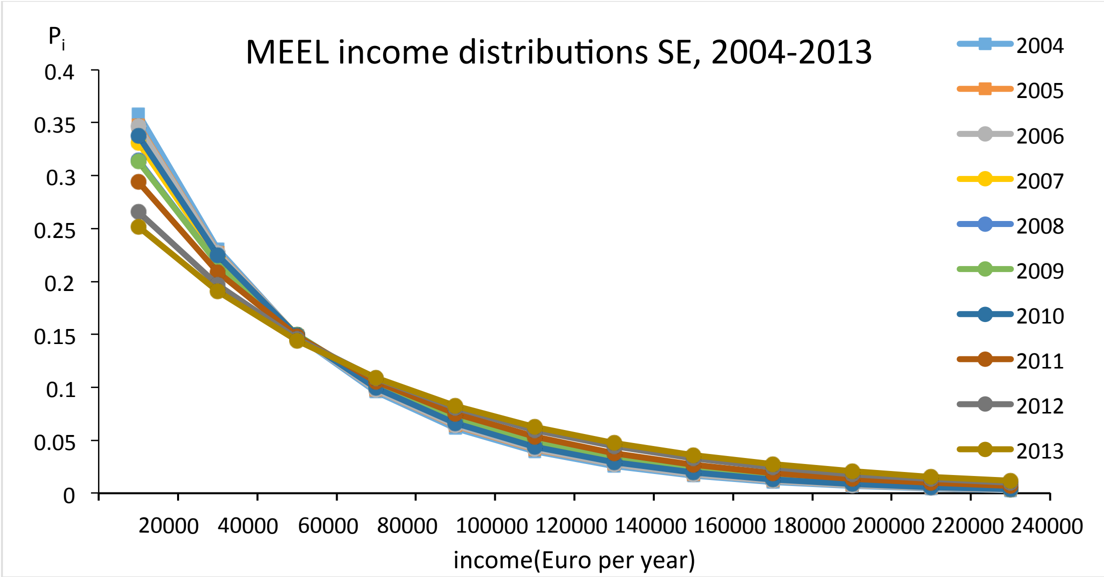

FIGURE 5.3 IT INCOME PDF'S FOR SWEDEN OVER THE YEARS

Taken together, these entropy based income distribution information recovery tools present a new basis for measuring and presenting the dynamic nature of income equality-inequality.

6., USE OF A TRANSITIONAL REFRENCE DISTRIBUTION, BMEEL

The income probability density functions-distributions reported in sections five and six were obtained using a uniform reference distributions **q.** In regard to country based income distributions, in many situations found in practice we may have non-sample or pre-sample information about the unknown income probabilities $\mathbf{p} = (p_1, p_2, \cdots, p_k)'$ in the form of a prior distribution of probabilities $\mathbf{q} = (q_1, q_2, \cdots, q_k)'$. That is, before making use of the sample of micro data that enters in the consistency relations, in (4.1) and (4.2), there may be some non uniform reference distribution $\boldsymbol{q}$ that seems reasonable as a probability density-distribution. When such prior non uniform reference information-knowledge exists, we may wish to follow Kullback and Liebler (1951), Kullback (1959) and Good (1963) and incorporate this information into the MEEL formalism of sections four and five, in the form of the principle of minimum cross-relative entropy or Kullback-Liebler directed divergence. This minimal discriminability principle implies one would choose, given the constraints, the estimate of $\mathbf{p}$ that can be discriminated from the non-uniform reference distribution $\mathbf{q}$, with a minimum of difference. Recovering this transition probability density, $when\ \gamma \rightarrow 0$, leads to the minimum cross-directed divergence entropy $\mathbf{p}$ and $\mathbf{q}$ that is defined as:

$$\mathrm{I}(\mathbf{p},\mathbf{q}) = \sum_{\mathrm{k}=1}^{\mathrm{K}} \mathrm{p_k} \ln(\mathrm{p_k}/\mathrm{q_k}) = \sum_{k} \mathrm{p_k} \ln(\mathrm{p_k}) - \sum_{k} \mathrm{p_k} \ln(\mathrm{q_k})$$

$$= \mathbf{p}' \ln(\mathbf{p}) - \mathbf{p}' \ln(\mathbf{q}). \qquad \textbf{(6.1)}$$

This criterion leads to a natural measure of the deviation of the distribution of probabilities $\mathbf{p}$ and $\mathbf{q}$. Under the principle of minimum discriminability the difference $\mathrm{I}(\mathbf{p},\mathbf{q})$ is minimized. To take account of both the prior non-uniform reference distribution and the micro data sample information, the minimum cross-entropy solution may be obtained from the minimization problem

$$\min_p \mathrm{I}(\mathbf{p},\mathbf{q}) = \textstyle\sum_k p_k \ln(\mathrm{p_k}/\mathrm{q_k}) = \mathbf{p}' \ln(\mathbf{p}) - \mathbf{p}' \ln(\mathbf{q}) \qquad \textbf{(6.2),}$$

subject to the moment consistency constraints and the adding up-normalization constraint in (4.1). The applicability of this directed divergence measure is indicated in sections 6.1 and 6.2.

6.1. A Transitional Income Probability Density Function for Germany

As noted in the entropy measures presented in Figure 4.1, Sweden is one of the more egalitarian countries. Sweden has managed to remain efficient in a production context, while keeping its inequality low. The principle of minimum cross-relative entropy or Kullback-Liebler directed divergence (6.2), makes it possible to use the entropy measure and income distribution information of one country, to gauge the impact on the income probability density function-distribution of another country. In this regard in Figure 6.1 we demonstrate the impact of using Sweden's income probability density function-distribution, which has an entropy measure of 1.987, as a reference distribution for the income probability density function-distribution for Austria, which has more income inequality and an entropy measure of 1.809. As a result, the entropy measure of the BMEEL in figure 6.1 to 1.890 and thus reflects more income equality. This suggests that the BMEEL income transitional probability density function may be used as an indicator of a change in a country's entropy measure and provides a basis for processing information about possible transitional income distribution dynamics.

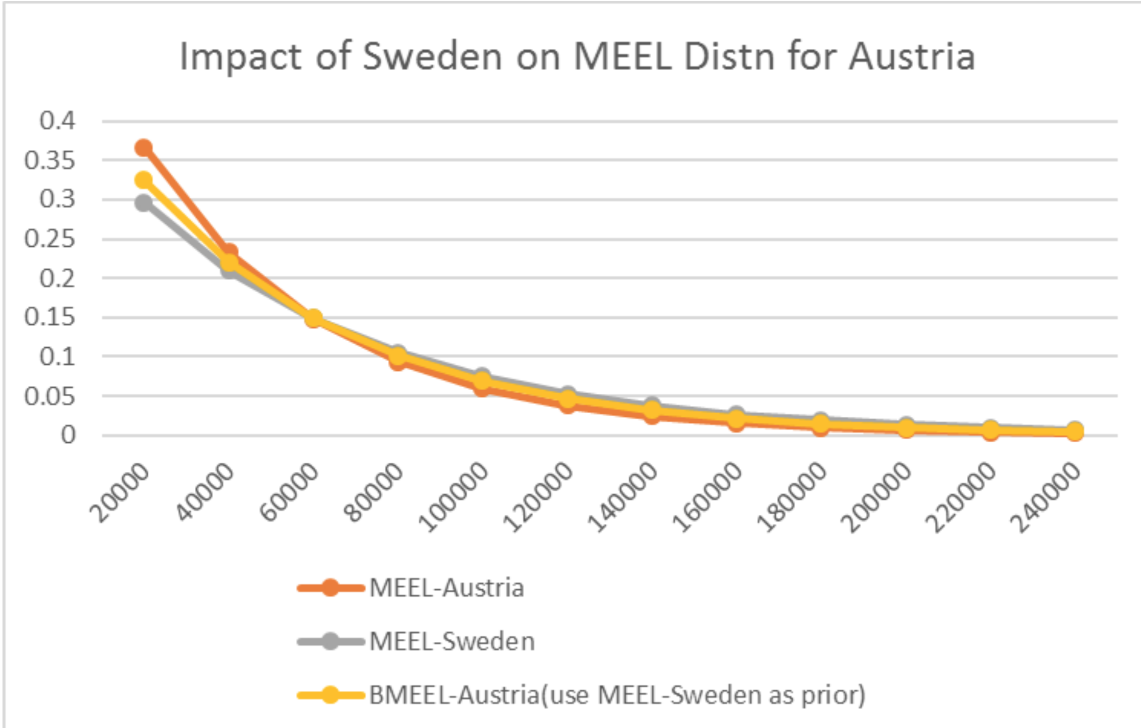

FIGURE 6.1 The BMEEL PREDICTED INCOME DISTRIBUTION FOR AUSTRIA

6.2 A Transitional Income Probability Density Function For Greece

In another context, recently Greece has been facing economic problems and these difficulties are reflected in terms of its relatively low entropy measure and income probability density function(see Figures 4.1 and 4.6). Given the economic problems facing Greece, the European Union imposed a list of economic restrictions on Greece in the hopes of improving its economy. As the possible economic impact on Greece of these economic conditions, one might use in the context of the transition-relative entropy principle (6.2), the income probability density function-distribution of Germany as a prior non-uniform reference distribution for Greece. As a result of using the income distribution for Germany with an entropy measure of 1.768 as a prior reference distribution, the entropy measure for Greece increases from 1.204, indicating a possible decrease in income inequality. This permits us to note again that the transitional-relative entropy measure provides one way of indicating possible changes in country based income probability density function-distributions and a framework for processing information about the underlying entropic income system dynamics and statistical equilibrium.

7. SUMMARY AND REMARKS

In this paper we have presented an information theoretic behavior related methodological basis for recovering country based income probability density-distribution functions from samples of time ordered European household micro income data. In terms of information recovery, the Cressie-Read family of entropy based information divergence measures have been used to provide a flexible family of functions to recover the unknown country based probability density functions-income distributions and yield an integrated entropy measure of income inequality. A combined 2008-2013 data sample was used to develop probability density income functions for 16 European countries and to provide an entropy measure of income inequality and a measure of statistical equilibrium. Annual data from 2004 to 2013 for three countries was used as a basis for discussing the time path of income inequality. The importance of using the entire income distribution when measuring income inequality is emphasized. Given that most countries have household micro income survey data sets collected over time, the information theoretic method provides a general approach for this type of time dated information recovery.

We find that when it comes to the entropy measure, there is heterogeneity in terms of income inequality across countries. More specifically, we have found that Portugal and Greece have the highest levels of inequality in their income distributions, while Northern European countries outside the European Union and the Euro zone have the lowest levels of inequality in their income distributions. When time ordered samples of income data are used in terms of information recovery, the problems of income probability density estimation and inference and the measure of inequality, should be analyzed as a one joint behavior related problem.

The income probability density functions in Figures 4.2 to 4.6 reflect the variation in the allocation and distributive nature of this set of European countries. The graphs reflect the country differences in terms of market performance, governmental policies and interventions and differences that are institutional in nature. Information theory-entropy based methods of the type discussed in Section 2 seem well designed for this type of information recovery. and the variability in the resulting country income probability density functions that are recovered provide a statistical systems basis for economic analysis, explanation and discussion. Many discussions of the implications of income inequality only focus on the first and last income boxplot-histogram and are strangely silent in terms of the remainder of the distribution. In income questions that are of an inequality-equality nature, this paper makes clear the importance of considering the complete income distribution.

Using Sweden as an example, we have investigated a possible time shift in the income distribution and noted the heterogeneous path in terms of the evolution of the entropy measure and income density functions-distributions. The dynamics of income equality-inequality is a very important economic and econometric problem (see Gabaix, et al., 2016). To our knowledge, a satisfactory way of handling income equality- dynamics does not currently exist. The information theoretic relative entropy-directed divergence framework used in this paper provide one basis for moving from a static analysis and incorporating time into the inequality discussion. In an information theoretic context one possible way to introduce dynamics is to use the income distribution in time T as the reference-prior distribution in time T+1. This means we can make use of the divergence measure $\sum_{i=1}^{n} p_i \ln\left(p_i/q_i\right)$, where the distribution q = {qi} is the priori in maximum entropy based inferences, or a stationary distribution in agent self-organized dynamics. In this context one possibility is to make use of data consistent empirical sample moments-constraints such as

$$\mathbf{h}(\mathbf{Y},\mathbf{X},\mathbf{Z};\boldsymbol{\beta}) = n^{-1}\left[\mathbf{Z}'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\right] \xrightarrow{p} \mathbf{0} \tag{7.1}$$

and reformulate the income distribution recovery extremum problems (2.3) and (4.2) in a more general moment form

$$\hat{\boldsymbol{\beta}}(\gamma) = \underset{\boldsymbol{\beta}\in\mathbf{B}}{\arg\min}\left[\min_{\mathbf{p}}\left\{I(\mathbf{p},\mathbf{q},\gamma) \mid \sum_{i=1}^{n} p_i \mathbf{Z}'_{i.}\left(Y_i - \mathbf{X}_{i.}\boldsymbol{\beta}\right) = \mathbf{0}, \sum_{i=1}^{n} p_i = 1, p_i \geq 0\,\forall i\right\}\right], \tag{7.2}$$

where Y is income related, **X** and **Z** are economic time dated explanatory variables and instruments and $\boldsymbol{\beta}$ is an unknown parameter vector**.** Evaluating these and other entropy based income information recovery alternatives are major topics for future research.

8. REFERENCES

Adamou, A. and O. Peters, 2016, Dynamics of Inequality, Significance, Royal Statistical Society, June, 32-35.

Alvaredo, F., Atkinson, A. B., T. Piketty, and E. Saez, 2013. "The Top 1 Percent in International and Historical Perspective," *The Journal of Economic Perspectives*, Volume 27 (3), Summer: 3-20(18).

Champernowne, D., 1953,A Model of Income Distribution. *Economic Journal*, 63:318-361.

Cho, A., 2014, Physicists Say Its Simple, Science, The Science of Inequality, May 2014,Vol. 344,Issue 6186:828.

Colwell, F., 2003, Thiel and the Structure of Income Distribution, Memorial Conference, London School of Economics.

Cressie, N. and T. Read, 1984. Multinomial Goodness of Fit Tests. *Journal of Royal Statistical Society of Series* B 46, 440-464.

Dragulesco, A. and V. Yakovenko, 2001, Evidence for the Exponential Distribution of Income in the U. S. A., *Eur. Phys. Journal* B 20, 585-589.

Gabaix, X., J. Lasry, P. Lions and B. Nolli, 2016, The Dynamics of Inequality, Econometrica, 84: 2071-2111

Gorban, A., P., Gorban, and G. Judge, 2010, Entropy: The Markov Ordering Approach. *Entropy,* 12, 1145-1193.

Good, I., 1963, Maximum Entropy for Hypothesis Formulation ,Ann. Math. Stat., 34:

911-934

Grendar, M., 2006, Entropy and Effective Support Size, Entropy, 8:169-174.

Jaynes, E., 1985, Where do we Stand on Maximum Entropy?, In C. Smith and W. Grandy, Eds., Maximum Entropy and Bayesian Methods in Inverse Problems, D. Riedel, Boston.

Judge, G.G. and R.C Mittelhammer, 2012a, An Information Theoretic Approach To Econometrics, Cambridge University Press.

Judge, G.G. and R.C. Mittelhammer, 2012b, Implications of the Cressie-Read Family of Additive Divergences for Information Recovery. *Entropy* 14, 2427-2438.

Kullback, S. and R. Liebler, 1951, On Information and Sufficiency, Ann. Math. Stat., 22:79-86.

Kullback, J., 1959, Information Theory and Statistics, John Wiley, New York.

Levy, M. and S. Solomon, 1996, Power Laws are Logarithmic Boltzmann Laws, International Journal of Modern Physics, C, 7, 595-601.

Pareto, V., 1896, Cours d'Economique Politique, Library Droz, Reprinted 1964

Piketty, T. and E. Saez, 2014, Inequality in the Long Run, *Science* 344 (6186), 838-843.

Piketty, T. and E. Saez, 2003, Income Inequality in the United States, 1913-1998, *Quarterly Journal of Economics*, 118:1-39.

Ravallion, M. 2014. "Income Inequality in the Developing World," *Science*, 344 (6186), 851-855.

Read, T.R. and N.A. Cressie, 1988, Goodness of Fit Statistics for Discrete Multivariate Data, New York: Springer Verlag.

Roine, J., and D. Waldenstroem, ‘‘Long-run Trends in the Distribution of Income and Wealth,’’ in Handbook of Income Distribution, Vol. 2, Anthony B. Atkinson and Francois Bourguignon, ed. (Amsterdam and New York: Elsevier/North Holland, 2015).

Saez, E. and G. Zucman, 2016, Income Inequality in the United States Since 1913, *Quarterly Journal of Economics*, 131:519-578.

Shannon, C., 1948, The mathematical Theory of Communications, Bell Syst. Techn. Jour, 27:379-423.

Thiel, H., 1967, Economics and Information Theory, pages 121-128, Amsterdam: North Holland.

Wissner-Gross, A. and C.E. Freer, 2013, Causal Entropic Forces. *Physical Review Letters* PRL 110, 168702-1-5.