Diskussionsschrifte

Evaluating Theories of Income Dynamics: A Probabilistic Approach

Robert Aebi, Klaus Neusser, Peter Steiner

99-05

June 1999

Universität Bern
Volkswirtschaftliches Institut
Gesellschaftstrasse 49
3012 Bern, Switzerland
Tel: 41 (0)31 631 45 06
Web: www-vwi.unibe.ch

Version from 04.05.99

Evaluating Theories of Income Dynamics:
A Probabilistic Approach

Robert AEBI, Klaus NEUSSER, and Peter STEINER

University of Berne

Abstract

The paper proposes an approach to evaluate hypotheses about transition dynamics when only the distributions at two points in time are observed. Using principles of statistical mechanics, we show how to adjust in the "most probable" way a hypothesis such that it becomes compatible with the observed distributions. This adjustment procedure also allows to test hypotheses in a statistical sense. The test is based on the relative entropy and is equivalent to a likelihood ratio test. We apply our approach to compare the dynamics of the income distribution between men and women in the U.S. using PSID data.

JEL classification: C50, C52, D31
Keywords: income dynamics, large deviation, relative entropy, misspecification

Correspondence

Klaus Neusser
University of Berne
Department of Economics
Gesellschaftsstrasse 49
CH-3012 Berne, Switzerland

Tel.: +41-31-631 4776
Fax: +41-31-631 3992
e-mail: klaus.neusser@vwi.unibe.ch
Internet: www-vwi.unibe.ch/econometrics

Evaluating Theories of Income Dynamics:
A Probabilistic Approach [*]

1. Introduction

This paper proposes a unified statistical framework for evaluating and testing hypotheses on the evolution of a distribution, say an income distribution, over time. Our approach is based on the concept of relative entropy or Kullback-Leibler Information criterion (Kullback 1959) as a measure of "distance" from one probability distribution to another one.[1] In contrast to other similar applications, we justify this choice by an explicit probabilistic micro model as suggested by Aebi (1996; 1997) in a different context. This justification relies on the fundamental hypothesis of statistical mechanics and on a large deviation argument. As it turns out, our approach also provides a sound statistical basis for hypothesis testing.

Despite this methodological perspective, we illustrate our case by considering a concrete problem which is of some practical relevance. The problem we wish to analyze is the following. Suppose we are in a situation where the distribution of income is observed at two points in time and where no information on the incomes of individual members in the population is available. We may think of having at our disposal a repeated cross-section where individual income histories are not recorded. Suppose further that we want to evaluate some particular model (or hypothesis) of the transition dynamics. This model may have been derived from theoretical considerations or from samples drawn from another population. Although by construction no information on the income of any individual in the two periods is available, we will show that it is nevertheless possible to draw meaningful statistical inferences on the transition dynamics under these circumstances. Moreover, our approach does not only result in a statistical test, but also indicates in what respect our model is misspecified and how it can be adapted "optimally".

---

[*] We thank Boris Zürcher for helpful comments. The paper has benefited from seminar participants at the Institute for Advanced Studies in Vienna and at Stanford University.

[1] The concept of relative entropy is not widespread in econometrics. Important exceptions are White (1982), Golan, Judge, and Miller (1996), or Kitamura and Stutzer (1997).

It turns out that the above problem is equivalent to the problem of fitting the cell probabilities of a contingency table when the marginal probabilities are known and fixed. This question has been treated in the statistical literature by Deming and Stephan (1940) and Ireland and Kullback (1968) among others. These authors also propose an algorithm known as iterative proportional fitting procedure (IPFP) to solve this problem in practice. Recently, Aebi (1996; 1997) gave a probabilistic framework in terms of “large deviations” for contingency tables. He shows how to compute „most probable“ adjustments of observed contingency tables to prescribed marginals based on the fundamental hypothesis of statistical mechanics. In this paper we follow his interpretation and use a large deviation principle to operationalize the meaning of „most probable“.

We think that our approach can be fruitfully applied to such diverse issues as the convergence hypothesis in the theory of economic growth and to the theory of personal income inequality. The so-called convergence hypothesis asserts that differences across countries in per capita income are transitory, controlling for technology, preferences and population growth rates. As has been forcefully pointed out by Quah (1996), the cross-country growth equation initially advocated by Barro and Sala-i-Martin (1992) suffers from severe deficiencies which lead to unreliable conclusions. Instead, Quah (1996) suggests to “model explicitly the dynamics of the entire cross-country distribution of incomes”. The second application relates to the recently observed increase in income inequality in some countries (notably the U.S. and the U.K.). The reasons for this rise are widely debated and have brought the income distribution “in from the cold” (Atkinson 1997; Gottschalk 1997). In order to assess this rise in inequality it is important to develop a notion of mobility within the income distribution. This, however, requires again to model the dynamics of the entire distribution.

Although these two applications are related to quite different economic traditions and concerns, the analysis of the dynamics of the underlying distribution uses similar tools. In both strands of literature, the evolution of the income distribution is analyzed in terms of a transition probability matrix (or a stochastic kernel in case of a continuous state space) estimated from panel surveys. The convergence hypothesis can then be assessed by computing the stationary distribution or passage times associated with the transition matrix;[2] mobility is assessed by computing some scalar mobility measure from the transition matrix.[3]

Although these applications produce interesting insights, they are purely descriptive in nature. They lack a probabilistic foundation and do not formally test or evaluate theories of income dynamics formulated in terms of the transition matrix. We think that this is a serious drawback which hinders further progress in these fields. The purpose of our paper is therefore to provide the methodological foundations to the testing and evaluation of theories of income dynamics. Although we expose our views by investigating a concrete problem, we think that our approach can be fruitfully extended to related issues.

We do not only develop the theoretical concepts, but we also illustrate our approach by a practical example. In particular, we compare the income dynamics of men and women in the U.S. using the PSID data set. These data encompass more information than we actually need because the PSID data trace individual incomes over time. This additional information will, however, allow us to assess and document the validity of our approach.

2. Concepts and theoretical background

2.1 A Probabilistic Model

Suppose that for a population consisting of a large number of N independent individuals we observe the true distribution of income at two points in time t and s with $t < s$. As our exposition relies on a finite state space, we take a finite partition $I = \{I_i\}_{i=1,\ldots,k}$ of $\mathbf{R}^+$ and assume that income is distributed in the two time periods according to the discrete probability distributions $q_t = (q_{1t}, \ldots, q_{kt})'$ and $q_s = (q_{1s}, \ldots, q_{ks})'$ defined on I, i.e. q_{it} is the probability that income in period t falls in the i-th interval.

If we were actually in a position to trace the income of each individual in the population, we could count how many persons starting in income class i in period t arrive in income class j in period s. Denote these numbers by Γ_{ij} and arrange them in a $k \times k$ matrix

$$\Gamma = (\Gamma_{ij})_{i,j=1,\ldots,k}$$

---

[2] Durlauf and Quah (1998) provide extensive references and a critical assessment of the literature.

We call this matrix the *income history matrix*. Note that the income history matrix is unobserved. We only know that it must be compatible with the observed income distributions at time t and s, q_t and q_s. Thus if nobody gets lost or is joining in going from period t to s, each person starting in income class i must end up in some income class j, likewise each person ending up in income class j must have started in some income class i. When the number of persons N is large, these restrictions on the income history matrix can be stated as follows:

$$\text{(1)} \qquad \begin{aligned} \sum_{j=1}^{k} \Gamma_{ij} &= Nq_{it} \qquad 1 \le i \le k \\ \sum_{i=1}^{k} \Gamma_{ij} &= Nq_{js} \qquad 1 \le j \le k \end{aligned}$$

If we denote by ι the k-vector of ones, these restrictions can be written more compactly as

$$\text{(1')} \qquad \begin{aligned} \Gamma\iota &= Nq_t \\ \Gamma'\iota &= Nq_s \end{aligned}$$

Because $\Sigma_i\, q_{it} = 1$ and $\Sigma_j\, q_{js} = 1$, the above conditions impose 2k – 2 independent restrictions on Γ, referred to as *continuity restrictions* or *initial and terminal conditions*.

The theory or model of the dynamics of the income distribution between the two periods t and s is formulated in terms of a two-dimensional joint probability distribution. This can be done either directly or, more conveniently, indirectly via a transition probability matrix.[4] If we denote by $P = (p_{ij})_{i,j=1,\dots,k}$ the transition matrix representing our model, the elements p_{ij} are just the probabilities of moving to income class j given that the individual was in income class i. For any given income distribution $\pi = (\pi_1, \dots, \pi_k)'$ in period t, $\pi_i\, p_{ij}$ is the probability that an individual is in income class i in period t and in class j in period s. The two-dimensional joint probability distribution is then given by the matrix $(\pi_i\, p_{ij})_{i,j=1,\dots,k} = \text{diag}(\pi)\, P$. Note that our model will in general not satisfy the continuity restrictions. When it comes to statistical inference, this probability distribution will be our null hypothesis.

---

[3] For a theoretical discussion see Shorrocks (1978). Schluter (1998) and Trede (1998) provide examples of empirical applications.

[4] Champernowne (1953) was the first one to view the income distribution as the equilibrium outcome of a Markov process specified by a transition matrix. He presented conditions on the transition matrix such that the ergodic distribution satisfies Pareto´s law. Later Wagner (1978), and more recently Conlisk (1990) and Dardanoni (1994), discussed alternative hypotheses about the form of the transition matrix.

The continuity restrictions (1) are not sufficient to determine the income history matrix Γ uniquely. The problem we address in this paper can therefore be stated in the following way: find the income history matrix Γ which would have the *maximum likelihood* of being observed under our maintained hypothesis, diag(π) P, subject to the continuity restrictions (1). We solve this problem in two steps. First, we compute the probability of observing a particular income history matrix and then solve the underlying maximization problem. The analysis is, however, not straightforward because our model does not in general satisfy the continuity restrictions. The law of large numbers then implies that, viewed from the perspective of our model, the probability of every income history matrix goes to zero as N tends to infinity. We resolve this indeterminacy by relying on a *large deviation principle*, i.e. we seek the income history matrix whose probability goes to zero at the slowest rate.

2.2 Probability of Income History Matrices

Assuming that the evolution of individual incomes is independent from each other, the probability that a particular history of N persons belongs to a given income history matrix Γ is

$$\prod_{i,j=1}^{k}\left(\pi_i p_{ij}\right)^{\Gamma_{ij}}$$

This given income history matrix Γ can be realized in several ways from the N individual income histories. The number of such possibilities corresponds to the number of arrangements of N distinguishable individuals as subsets of Γ_{ij} persons. It is obtained by an elementary combinatorial argument:

$$\binom{N}{\Gamma_{11}}\binom{N-\Gamma_{11}}{\Gamma_{12}}\binom{N-\Gamma_{11}-\Gamma_{12}}{\Gamma_{13}}\cdots\binom{N-\sum_{j=1}^{k}\Gamma_{1j}}{\Gamma_{21}}\cdots\binom{N-\sum_{i=1}^{k-1}\sum_{j=1}^{k-1}\Gamma_{ij}-\sum_{j=1}^{k-1}\Gamma_{kj}}{\Gamma_{kk}}=\frac{N!}{\prod_{i,j=1}^{k}\Gamma_{ij}!}$$

Viewed from the perspective of our model, the income history matrix Γ is realized with probability $P_N(\Gamma \mid \text{diag}(\pi)P)$ given by

$$P_N\left(\Gamma\middle|\,\text{diag}(\pi)P\right)=\frac{N!}{\prod_{i,j=1}^{k}\Gamma_{ij}!}\prod_{i,j=1}^{k}\left(\pi_i p_{ij}\right)^{\Gamma_{ij}}=N!\prod_{i,j=1}^{k}\frac{\left(\pi_i p_{ij}\right)^{\Gamma_{ij}}}{\Gamma_{ij}!} \tag{2}$$

2.3 Maximization and adjusted Dynamics

There are many income history matrices which are compatible with the continuity restrictions (1). To determine the income history matrix Γ uniquely, we adopt the fundamental hypothesis of *statistical mechanics* to the evolution of incomes: an observation at the macroscopic level is realized in the limit of infinitely many individuals by that microscopic ensemble which has maximal probability (i.e. is „most probable") given the observation. This principle means that we want to choose the income history matrix which has the highest probability of being realized, viewed from the perspective of our model, and which satisfies the continuity conditions. Chapter I in Ellis (1985) provides an insightful introduction to the concepts we will use subsequently.

As explained previously, the law of large numbers implies that every income history matrix has probability zero of being realized as N tends to infinity, $P_N(\Gamma | \mathrm{diag}(\pi)P) \to 0$ as $N \to \infty$, because Γ satisfies the continuity restrictions whereas our conjecture diag(π)P does not. We can nevertheless obtain a unique solution to our maximization problem if we interpret "most probable" as "vanishing at the slowest rate". This is a so-called *large deviation* argument. The rate at which the probability (2) goes to zero is given by the limit of (1/N) log $P_N(\Gamma | \mathrm{diag}(\pi)P)$. Using Stirling's formula for large factorials[5], this limit is

$$\text{(3)} \qquad \lim_{N\to\infty} \frac{1}{N} \log P_N\left(\Gamma \middle| \mathrm{diag}(\pi)P\right) = -H\left(\gamma \middle| \mathrm{diag}(\pi)P\right)$$

where $\gamma = (\gamma_{ij})$ denotes the matrix $\Gamma/N = (\Gamma_{ij}/N)$. The function $H(\gamma | \mathrm{diag}(\pi)P)$ is known as the *relative entropy* or Kullback-Leibler divergence of the two-dimensional distribution γ with respect to diag(π)P and is defined as

$$\text{(4)} \qquad H\left(\gamma \middle| \mathrm{diag}(\pi)P\right) = \sum_{i,j=1}^{k} \gamma_{ij} \log\left(\frac{\gamma_{ij}}{\pi_i p_{ij}}\right)$$

where it is understood that 0 log(0) equals 0 and that $\gamma_{ij} \log(\gamma_{ij}/(\pi_i p_{ij}))$ equals infinity if $\pi_i p_{ij}$ equals 0 and $\gamma_{ij} \neq 0$. The function $H(. | \mathrm{diag}(\pi)P)$ is also called the rate function because $P_N(\Gamma | \mathrm{diag}(\pi)P)$ decays to zero exponentially fast at a rate given by (4). It can be shown that

---

[5] Stirling´s formula is $x! = \left(\frac{x}{e}\right)^x \sqrt{2\pi x}\,(1+\varepsilon_x)$ with $\varepsilon_x \to 0$ as $x \to \infty$.

$H(.\,|\,\mathrm{diag}(\pi)P)$ is a nonnegative and strictly convex function. Moreover, $H(.\,|\,\mathrm{diag}(\pi)P)$ equals zero if and only if $\gamma = \mathrm{diag}(\pi)P$. Thus $H(.\,|\,\mathrm{diag}(\pi)P)$ attains its infinum at the unique measure $\gamma = \mathrm{diag}(\pi)P$. These properties suggest to interpret the relative entropy $H(\gamma|\,\mathrm{diag}(\pi)P)$ as a distance or a measure of discrepancy from the distribution $\mathrm{diag}(\pi)P$ to the distribution γ. The relative entropy does, however, not define a metric in the space of probability distributions because it is not symmetric in its arguments and because it violates the triangular inequality.[6] The relative entropy can, nevertheless, be given a geometric interpretation analogous to the usual Euclidean distance. In particular and most relevant for this paper, the minimization of the relative entropy with respect to a given probability distribution over a convex subset of probability distributions can be viewed as a projection with properties similar the projection in Euclidean or Hilbert spaces (Csiszár 1975).

The relative entropy can be interpreted as a measure of the probability of observing a given income history matrix viewed from the standpoint of our model. The principle of statistical mechanics then advises us to take the "most probable" income history matrix subject to the continuity restrictions. We are thus led to consider the following pure inverse problem (Golan, Judge and Miller 1996): minimize $H(\gamma|\,\mathrm{diag}(\pi)P)$ over all two-dimensional distributions γ subject to the continuity restrictions (1). In the words of the statistics literature, we have to find the minimum discrimination information under the hypothesis $\mathrm{diag}(\pi)P$ (Kullback 1959, 37). The solution is called the minimum discriminant information adjustment of $\mathrm{diag}(\pi)P$ (Haberman 1984). The Lagrangian L for this optimization problem is

$$(5)\qquad L = \sum_{i,j=1}^{k} \gamma_{ij} \log\left(\frac{\gamma_{ij}}{\pi_i p_{ij}}\right) - \sum_{i=1}^{k} \lambda_{it}\left(\sum_{j=1}^{k} \gamma_{ij} - q_{it}\right) - \sum_{j=1}^{k} \lambda_{js}\left(\sum_{i=1}^{k} \gamma_{ij} - q_{js}\right)$$

where λ_{it} and λ_{js} are the 2k Lagrangian multipliers associated with the constraints (1). A solution to this optimization problem exists if and only if there is at least one income history matrix which satisfies the continuity restrictions (1) and which has at least the same zero entries as $\mathrm{diag}(\pi)$ P (Csiszar 1975, corrolary 3.3). The strict convexity of the relative entropy then implies that this solution is unique. The optimum, denoted by $G = (g_{ij})_{i,j=1,\ldots,k}$, is found by differentiating (5) with respect to γ_{ij} and setting the derivative equal to zero:

$$(6)\qquad g_{ij} = \phi_{it}\, \pi_i p_{ij}\, \phi_{js}$$

---

[6] Further properties of the relative entropy and a deeper discussion of its interpretation can be found among

where ϕ_{it} and ϕ_{js} equal $\exp(\lambda_{it})$ and $\exp(\lambda_{js}-1)$. In matrix notation the above relation becomes

$$G = \Phi_t \operatorname{diag}(\pi)\, P\, \Phi_s \tag{6´}$$

where Φ_t and Φ_s denote $\operatorname{diag}((\phi_{1t},...,\phi_{kt}))$ and $\operatorname{diag}((\phi_{1s},...,\phi_{ks}))$. According to our terminology we call G the „most probable“ income history probability density matrix.

In the theory of quantum mechanics the ϕ´s are known as *Schrödinger multipliers*. They indicate how to adjust “in the most probable” way the two-dimensional density $\operatorname{diag}(\pi)\, P$, representing our model about income dynamics, to satisfy the continuity restrictions (1). The Schrödinger multipliers adjust the probabilities of our model ($\pi_i p_{ij}$) downward if $\phi_{it} \times \phi_{js}$ is smaller than one and upward if $\phi_{it} \times \phi_{js}$ is larger than one. The matrix $(\phi_{it}\phi_{js})_{i,j=1,...,k}$ may therefore reveal patterns of adjustment and indicates the entries where our model is misspecified. In addition the Schrödinger multipliers have a kind of “separability property” because the ϕ_{it}´s depend only on the distribution at time t whereas the ϕ_{is}´s depend only on the distribution at time s. The relative size of ϕ_t and ϕ_s thus indicates whether the misspecification is primarily due to the initial or to the terminal restriction.

The Schrödinger multipliers are found after differentiating L with respect to the Lagrangian multipliers (λ_{it}) and setting the derivatives equal to zero. The resulting equation system is the so-called *Schrödinger system*:

$$\begin{aligned} \phi_{it}\pi_i \sum_{j=1}^{k} p_{ij}\phi_{js} &= q_{it} \\ \left(\sum_{i=1}^{k} \phi_{it}\pi_i p_{ij}\right)\phi_{js} &= q_{js} \end{aligned} \tag{7}$$

This equation system shows that the Schrödinger multipliers are unique only up to a multiplicative constant. In the following we normalize the ϕ´s such that ϕ_{1t} equals ϕ_{1s}.

In empirical applications it is often more convenient to deal with transition probabilities instead of two-dimensional densities. We can reformulate the adjustment equation (6´) in terms of the ”most probable” transition matrix $R = (r_{ij})$. Given the initial distribution q_t, the elements of the two-dimensional density and of the transition matrix are related by $g_{ij} = r_{ij}\, q_{it}$. The elements of R are therefore obtained from P as follows

---

others in Kullback (1959), Ellis (1985), and Hillman (1996).

$$(8) \qquad r_{ij} = \frac{g_{ij}}{q_{it}} = \frac{\phi_{it}\pi_i p_{ij}\phi_{js}}{\phi_{it}\pi_i \sum_{j=1}^{k} p_{ij}\phi_{js}} = \frac{p_{ij}\phi_{js}}{\sum_{j=1}^{k} p_{ij}\phi_{js}}$$

$$R = \tilde{\Phi}_s^{-1} P \Phi_s$$

where $\tilde{\Phi}_s = \mathrm{diag}\left(\sum_{j=1}^{k} p_{ij}\phi_{js}\right)$. Note that R satisfies the definition of a transition matrix, i.e. $r_{ij} \geq 0$ and $\sum_{j=1}^{k} r_{ij} = 1$ for all i. Moreover, R is obtained from P only through the Schrödinger multipliers ϕ_{js} related to the terminal restrictions.

2.4 Statistical inference

From a statistical point of view, we do not only want to know how to best adjust our model, but also if these adjustments are significant in a statistical sense. For this purpose, it is convenient to interpret the computation of G as estimating the cell probabilities of a k×k contingency table for which the marginal probabilities, in our case q_t and q_s, are given. This problem was first treated by Deming and Stephan (1940) who also suggest an iterative procedure, known as iterative proportional fitting procedure (IPFP), to solve the Schrödinger system (7). Taking the ϕ_{js} equal to one as starting values, the ϕ_{it} can be computed from the first part of (7). Inserting these values in the second part of (7), new values for ϕ_{is} are obtained. These can then be used to update the ϕ_{it}. This procedure is then repeated until convergence is achieved.[7] Having found the Schrödinger multipliers, it is straightforward to compute G and R using equations (6) and (8). It can be shown that this procedure converges geometrically fast, generates best asymptotically normal (BAN) estimates and is equivalent to maximum likelihood estimates (Smith 1947; Ireland and Kullback 1968). With a reference sample of size n, these latter authors show that the statistic 2n times the relative entropy function is asymptotically distributed as chi-squared and can thus be used to test the null hypothesis given by our model:

$$(9) \qquad 2n\, H\left(G \middle| \mathrm{diag}(\pi)P\right) \sim \chi^2_{2k-2}$$

According to Ireland and Kullback (1968), the degrees of freedom, 2k–2, are given by the difference between the degrees of freedom in the unrestricted model, k^2–1, and in the

---

[7] The procedure assumes that $\pi_i p_{ij} > 0$. Clearly, if $\pi_i p_{ij} = 0$, $g_{ij} = 0$.

restricted model, $k^2–2k+1$. Therefore the degree of freedom corresponds to the number of restrictions imposed by the continuity restrictions (1).[8]

3. Comparing the income dynamics of women and men

3.1 The data

We illustrate our approach by asking whether the observed distributions of women's income are compatible with the income dynamics estimated for men over the same period. To answer this question we use data from the panel study of income dynamics (PSID).[9] The "1968-1993 individual file" records, among other information, the annual income of 53'013 individuals from 1967 through 1992. We divided the sample period into 5-year intervals and extracted the variables "total annual work hours", "type of income", "total annual income" and "age of individual". Due to a change in data collection, we retrieved in 1992 the variable "total annual labor income" instead of "total annual income". In order to save space, this paper focuses on the last 5-year interval (1987 to 1992).[10]

To obtain sensible and meaningful results, we used only a subset of the whole sample. In particular, we applied the following restrictions:

- We focus on labor income only.
- Individuals have to be at least of age 20 in the starting year and at most of age 60 in final year of the 5-year intervals.
- We only look at fully employed individuals. People with less than 1800 hours worked per year are eliminated from the sample.
- Despite these restrictions some extreme outliers remained in the sample[11]. To eliminate them, we require a minimum annual income of 1´000 USD in 1967. This minimum is inflated in subsequent years by the growth rate of the mean income.

---

[8] The same result can be obtained by observing that (5) is just the Neyman-Pearson statistic subject to the restrictions (1) (see Billingsley 1961, chapter 5).

[9] URL: http://www.isr.umich.edu/src/psid/maindata.html; file 68_93ind.zip.

[10] The other 5-year intervals give similar conclusions and are available upon request.

[11] The following examples illustrate two cases of extreme outliers. Individual number 2'059 worked 2'728 hours in 1992 but earned an annual income of only 15$. Individual number 32'416 worked 2'080 hours in 1987 but earned an annual income of only 14$. While such cases should definitely not occur in the sample years prior to 1992, this could happen in 1992 due to the change in data collection. It is for instance possible that somebody invested a lot of time to manage his financial assets without being employed. Such a person could earn a lot of asset income and only little labor income.

After processing the restrictions mentioned above, the male data set, which is our reference sample, contains 1’180 individuals. The female data set 935 consists of individuals. To construct transition matrices and two-dimensional discrete distributions, we had to choose partitions for the starting and the final year. Setting k arbitrarily equal to 10, we chose the income interval bounds in both years such that the number of men is equally distributed among the 10 cells. Thus the i-th interval is the interval with bounds given by the (i–1)-th and i-th decile of men’s income distribution.

The female incomes are distributed according to the partitions defined for men. This procedure resulted in the marginal densities of the beginning and the final year for women. In case several female incomes happen to be exactly equal to some bound of the partition, the incomes are equally split between the two adjacent cells of the marginal density.

The income distributions of women in the two years 1987 and 1992 are plotted in figure 1. These two distributions correspond to q_t and q_s, respectively. For comparison purposes we have also plotted the distribution of men’s income which is uniform and equal to 0.1 by construction. This figure reveals that the mode of the density shifted from the first to the second income class. In addition, more women are now in the upper income classes. These two simple observations suggest that women’s income distribution has obviously changed over these five years. The question we want to address is whether these changes can be explained by the income dynamics estimated for men.

3.2 Empirical results

The income dynamics for men is represented by the two-dimensional density matrix in table 1 and the corresponding transition matrix in table 2. The cell probabilities are estimated by the method of maximum likelihood which just equals the corresponding sampling frequency. These estimates are asymptotically normally distributed so that asymptotic standard errors are easily computed. As for the marginal distributions of women q_t and q_s, we ignore, for the sake of exposition, the sampling error associated with the estimation. Accordingly, we treat men’s density matrix and men's transition matrix as given. In the following they play the role of our model and serve as our null hypothesis. They therefore correspond to the matrices diag(π) P and P, respectively. For comparison purposes we also computed Shorrocks' mobility index for transition matrices (Shorrocks 1978).[12]

We can now formulate the objective of our empirical investigation in terms of the language from the previous section. Estimate the "most probable" adjustment of the men's two-dimensional density or transition matrix taking the income distribution of women in the years 1987 and 1992 as given. The PSID data would, of course, allow us to estimate the two-dimensional density matrix and the transition for women directly and to conduct a traditional statistical analysis. We chose, however, to ignore this information at this stage, but use it to check if our approach delivers sensible and meaningful results.

Given these preliminaries, we solve the Schrödinger system (7) by the method of iterative proportional fitting. This gives the "most probable" adjusted two-dimensional density matrix G reported in table 3 with the corresponding Schrödinger multipliers plotted in figure 2. Table 4 reports all cross-products of the Schrödinger multipliers, i.e. the matrix of adjustment coefficients $(\phi_{i,1987}\times\phi_{j,1992})_{i,j = 1,...,10}$. These numbers show by how much one must multiply a cell of men's density matrix to get the "most probable" adjusted density. A closer examination of this matrix reveals that large values (values greater than 2) are concentrated in the north-west corner of the matrix whereas small values (values lower than 0.5) are concentrated in the south-east corner of the matrix.[13] This means, for example, that the probability of being in the lowest income class in 1987 and in the second income class in 1992 is nearly three times as large for women compared to men, according to the "most probable" adjustment. Similarly, the probability of being in both years in the highest income class is five times lower for women compared to men. Generally speaking, one must increase the probabilities to be in the low income classes and reduce those for being in the high income classes.

The plots of the Schrödinger multipliers in figure 2 show that the downward adjustments are due to the distribution in 1987 ($\phi_{i,1987} < 1$ for $i \geq 4$) whereas the upward adjustments are primarily due to the distribution in 1992 ($\phi_{i,1992} > 1$ for $i \leq 4$ and ($\phi_{i,1992} \approx 1$ for $i \geq 5$). This makes sense given the observed shift in the distribution documented in figure 1.

---

[12] Shorrocks' mobility index for a transition matrix T is defined as (k – tr(T))/(k – 1) where k denotes the number of states. Schluter (1998) and Trede (1998) provide a statistical approach to the analysis of mobility indices.

[13] This pattern is typical. If we repeat this exercise for other time periods, we obtain similar results.

As mentioned in the theoretical part, we can use the relative entropy of the "most probable" adjusted density matrix (matrix in table 3) with respect to men's density matrix (matrix in table 1) to test whether the adjustments are statistically significant. The value of relative entropy is $H = 0.2069$ and the value of the corresponding test statistic (9) is $2n\, H = 2 \times 1180 \times 0.2069 = 488.28$. Given that the critical value is 28.87 for the 5 percent significance level, we must clearly reject our hypothesis.

Because the PSID contains more information, we can check our approach by computing the relative entropy of the "true" female density matrix estimated from the data. This gives a relative entropy of 0.2829 which is larger than the relative entropy of the "most probable" adjusted density matrix. This shows that our adjustment points in the "right" direction.

Often it is more convenient to interpret the transition matrices instead of the two-dimensional densities. We have therefore computed the "most probable" adjusted transition as indicated in equation (8). The result is reported in table 5. It shows only two significant changes at the 5 percent level: cells (2,2) and (3,2). In both cases the probabilities are adjusted upwards meaning that women have a significantly higher propensity to stay in the second income class and to fall back from the third income class to the second. Given the great similarity between the transition matrices which is also reflected in similar mobility indices, we conclude that the differences between the two-dimensional density matrices are largely due to the differences in the initial income distribution inherited from the past than to the income dynamics per se.

4. Conclusion

This paper has proposed a new approach to evaluate theories on the dynamics of income distributions. We hope to have demonstrated the validity and the usefulness of our method. Of course, further applications are necessary to arrive at a final judgement. The example of this paper was just a first test. The PSID data provided more information than we actually needed. We could have, in principle, estimated the transition matrix for women from the data and compared it to the transition matrix of men using conventional statistical methods. The advantage of using the PSID data was that it allowed us to check whether our adjustments went into the "right" direction, as they actually did.

In the future we hope to apply our method to issues where such additional information is not available. We could for example investigate the differences in the dynamics of income distributions across economies or across time. Or we could use our approach to evaluate specific theories of income dynamics as proposed by Conlisk (1990), Dardanoni (1994) or Wagner (1978).

The approach should also provide new insights in the "empirics of economic growth" which studies the evolution of the cross-country income distribution (Quah 1996; Durlauf and Quah 1998). This literature has not yet gone beyond the simple estimation of transition matrices.

On the methodological side it would perhaps be desirable to extend our analysis to continuous state space. This would circumvent the problem of choosing a somewhat arbitrary partition of the state space. Although the combinatoric argument presented in this paper cannot be carried over, the extension seems feasible but conceptually difficult (Föllmer 1988; Aebi and Nagasawa 1992) and goes far beyond this paper.

References

Aebi, R. 1996. Schrödinger’s Time-Reversal of Natural Laws. *The Mathematical Intelligencer* 18: 62-67.

Aebi, R., and M. Nagasawa. 1992. Large Deviations and the Propagation of Chaos for Schrödinger Processes. *Probability Theory and Related Fields* 94:53-68.

Aebi, R. 1997. Contingency Tables with Prescribed Marginals. *Statistical Papers* 38:219-29.

Atkinson, A.B. 1997. Bringing Income Distribution in from the Cold. *Economic Journal* 107:297-321.

Barro, R.J., and X. Sala-i-Martin. 1992. Convergence. *Journal of Political Economy* 100:223-51.

Billingsley, P. 1961. *Statistical Inference for Markov Processes*. Chicago: University of Chicago Press.

Champernowne, D.G. 1953. A Model of Income Distribution. *Economic Journal* 63:318-51.

Csiszár, I. 1975. I-Divergence Geometry of Probability Distributions and Minimization Problems. *Annals of Probability* 3:146-58.

Conlisk, J. 1990. Monotone Mobility Matrices. *Journal of Mathematical Sociology* 15:173-91.

Dardanoni, V. 1994. Income Distribution Dynamics: Monotone Markov Chains Make Light Work. Discussion Paper 94-16. University of California, San Diego.

Deming, W.E., and F.F. Stephan. 1940. On a Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are Known. *Annals of Mathematical Statistics* 11:427-44.

Durlauf, S.N., and D.T. Quah. 1998. The New Empirics of Economic Growth. In *Handbook of Macroeconomics*, eds. J.B. Taylor and M. Woodford. Forthcoming.

Ellis, R.S. 1985. *Entropy, Large Deviations, and Statistical Mechanics*. New York: Springer-Verlag.

Föllmer, H. 1988. *Random Fields and Diffusion Processes*. École d´Été de Saint Flour XV-XVII. Lecture Notes in Mathematics 1362. New York: Springer-Verlag.

Golan, A., G. Judge, and D. Miller. 1996. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. New York: John Wiley.

Gottschalk, P. 1997. Inequality, Income Growth, and Mobility: The Basic Facts. *Journal of Economic Perspectives* 11:21-40.

Haberman, S.J. 1984. Adjustment by Minimum Discriminant Information. *Annals of Statistics* 12:971-88.

Hillman, C. 1996. An Entropy Primer. Mimeo retrieved from URL: <http://www.math.washington.edu/~hillman>

Ireland, C.T., and S. Kullback. 1968. Contingency Tables with given Marginals. *Biometrika* 55:179-88.

Kitamura, Y., and M. Stutzer. 1997. An Information-Theoretic Alternative to Generalized Method of Moments Estimation. Econometrica 65:861-74.

Kullback, S. 1959. *Information Theory and Statistics*. New York: John Wiley.

Quah, D.T. 1996. Convergence Empirics Across Economies with (Some) Capital Mobility. *Journal of Economic Growth* 1: 95-124.

Schluter, C. 1998. Statistical Inference with Mobility Indices. *Economics Letters* 59:157-62.

Shorrocks, A.F. 1978. The Measurement of Mobility. *Econometrica* 46:1013-24.

Smith, J.H. 1947. Estimation of Linear Functions of Cell Proportions. *Annals of Mathematical Statistics* 13:166-78.

Trede, M. 1998. Statistical Inference in Mobility Measurement: Sex Differences in Earnings Mobility. *Jahrbücher für Nationalökonomie und Statistik* forthcoming.

Wagner, M. 1978. On Comparisons of Distribution Processes. In *Personal Income Distribution*, eds. W. Krelle and A.F. Shorrocks, 141-58. Amsterdam: North-Holland.

White, H. 1982. Maximum Likelihood Estimation of Misspecified Models. *Econometrica* 50:1-25.

Table 1: two-dimensional density of men’s income in 1987 and 1992

		Income class in 1992									
		1	2	3	4	5	6	7	8	9	10
Income class in 1987	1	0,044	0,016	0,012	0,011	0,008	0,004	0,001	0,001	0,001	0,002
	2	0,021	0,025	0,015	0,008	0,011	0,007	0,008	0,002	0,002	0,001
	3	0,010	0,026	0,019	0,009	0,014	0,009	0,007	0,003	0,002	0,001
	4	0,011	0,017	0,020	0,013	0,010	0,008	0,009	0,007	0,002	0,003
	5	0,004	0,005	0,017	0,018	0,016	0, 014	0,012	0,004	0,007	0,003
	6	0,002	0,006	0,007	0,019	0,018	0,018	0,018	0,007	0,003	0,004
	7	0,003	0,003	0,003	0,009	0,016	0,017	0,019	0,014	0,011	0,004
	8	0,002	0,001	0,004	0,008	0,003	0,011	0,020	0,028	0,014	0,008
	9	0,003	0	0,001	0,003	0,002	0,008	0,006	0,023	0,039	0,016
	10	0	0	0,002	0,002	0,001	0,004	0,001	0,012	0,021	0,058

Table 2: men's income transition matrix between 1987 and 1992

Shorrocks mobility index and its standard deviation: 0.80226 (0.01378)

Income class in 1992

Income class in 1987	1	2	3	4	5	6	7	8	9	10
1	0,44	0,16	0,12	0,11	0,08	0,04	0,01	0,01	0,01	0,02
2	0,21	0,25	0,15	0,08	0,11	0,07	0,08	0,02	0,02	0,01
3	0,10	0,26	0,19	0,09	0,14	0,09	0,07	0,03	0,02	0,01
4	0,11	0,17	0,20	0,13	0,10	0,08	0,09	0,07	0,02	0,03
5	0,04	0,05	0,17	0,18	0,16	0,14	0,12	0,04	0,07	0,03
6	0,02	0,06	0,07	0,19	0,18	0,18	0,18	0,07	0,03	0,04
7	0,03	0,03	0,03	0,09	0,16	0,17	0,19	0,14	0,11	0,04
8	0,02	0,01	0,04	0,08	0,03	0,11	0,20	0,28	0,14	0,08
9	0,03	0	0,01	0,03	0,02	0,08	0,06	0,23	0,39	0,16
10	0	0	0,02	0,02	0,01	0,04	0,01	0,12	0,21	0,58

Table 3: "most probable" adjusted two-dimensional density matrix

		Income class in 1992									
		1	2	3	4	5	6	7	8	9	10
Income class in 1987	1	0,103	0,047	0,027	0,023	0,012	0,008	0,001	0,002	0,001	0,002
	2	0,035	0,052	0,025	0,012	0,011	0,009	0,007	0,002	0,002	0,001
	3	0,017	0,056	0,031	0,014	0,015	0,013	0,007	0,004	0,002	0,001
	4	0,015	0,029	0,027	0,015	0,009	0,009	0,008	0,008	0,001	0,002
	5	0,004	0,006	0,015	0,014	0,009	0,010	0,006	0,003	0,004	0,002
	6	0,002	0,007	0,006	0,015	0,010	0,013	0,010	0,006	0,001	0,002
	7	0,002	0,004	0,003	0,008	0,009	0,013	0,010	0,012	0,006	0,002
	8	0,001	0,001	0,002	0,004	0,001	0,005	0,007	0,014	0,005	0,003
	9	0,003	0	0,001	0,002	0,001	0,005	0,003	0,015	0,019	0,007
	10	0	0	0,001	0,001	0,000	0,001	0,000	0,004	0,005	0,012

Shading indicates a value significantly different at the 5 percent level from those of men's density matrix in table 1

░ values above the 95%-confidence-interval for the two-dimensional density of men

▓ values below the 95%-confidence-interval for the two-dimensional density of men

Table 4: "most probable" adjustments by cell (crossing of ϕ_{1987} and ϕ_{1992})

	1	2	3	4	5	6	7	8	9	10
1	2,35	2,94	2,31	2,07	1,47	1,89	1,38	2,07	1,49	1,29
2	1,65	2,06	1,62	1,45	1,03	1,33	0,97	1,45	1,05	0,90
3	1,69	2,12	1,66	1,49	1,06	1,36	1,00	1,49	1,07	0,93
4	1,38	1,72	1,35	1,21	0,86	1,11	0,81	1,21	0,87	0,75
5	0,88	1,10	0,86	0,78	0,55	0,71	0,52	0,77	0,56	0,48
6	0,92	1,16	0,91	0,82	0,58	0,75	0,54	0,81	0,59	0,51
7	0,93	1,16	0,91	0,82	0,58	0,75	0,54	0,82	0,59	0,51
8	0,55	0,69	0,54	0,49	0,35	0,45	0,33	0,49	0,35	0,30
9	0,75	0,94	0,74	0,67	0,47	0,61	0,44	0,66	0,48	0,41
10	0,37	0,47	0,37	0,33	0,23	0,30	0,22	0,33	0,24	0,20

░ values higher than 2.0

▓ values lower than 0.5

Table 5: “most probable” adjusted transition matrix

Shorrocks mobility index and its standard deviation: 0.80684 (0.01988)

next period’s income class

initial income class	1	2	3	4	5	6	7	8	9	10
1	0,45	0,21	0,12	0,10	0,05	0,04	0,01	0,01	0,01	0,01
2	0,22	0,33	0,16	0,08	0,07	0,06	0,05	0,02	0,01	0.00
3	0,11	0,35	0,20	0,09	0,10	0,08	0,04	0,02	0,01	0.00
4	0,12	0,23	0,22	0,12	0,07	0,08	0,06	0,07	0,01	0,02
5	0,05	0,08	0,21	0,19	0,12	0,14	0,09	0,05	0,05	0,02
6	0,02	0,10	0,09	0,21	0,14	0,18	0,13	0,08	0,02	0,03
7	0,03	0,06	0,04	0,11	0,13	0,18	0,15	0,17	0,09	0,03
8	0,02	0,01	0,06	0,1	0,03	0,12	0,16	0,33	0,11	0,06
9	0,05	0	0,01	0,03	0,01	0,09	0,05	0,28	0,35	0,12
10	0	0	0,03	0,02	0,01	0,05	0,01	0,17	0,21	0,50

Shading indicates a value significantly different at the 5 percent level from those of men’s transition matrix in table 2. Both values lie above the 95%-confidence-interval for the transition matrix of men.

Figure 1

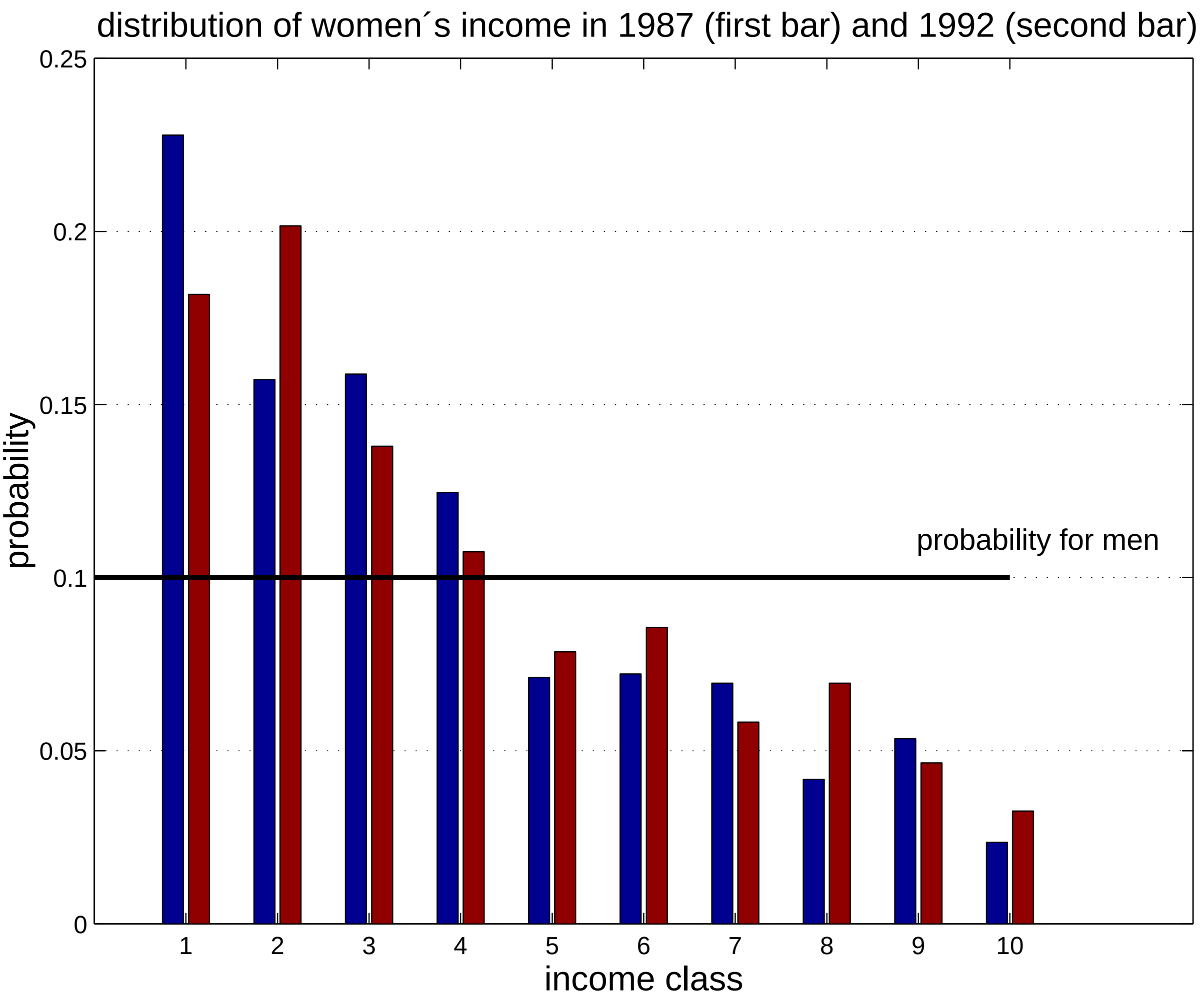
distribution of women´s income in 1987 (first bar) and 1992 (second bar)
probability
0.25
0.2
0.15
0.1
0.05
0
probability for men
1
2
3
4
5
6
7
8
9
10
income class

Figure 2: “most probable” adjustments