Improving Models of Income Dynamics Using Cross-Section-Information[*]

Robert Aebi[a], Klaus Neusser[b] and Peter Steiner[b,c]

1. Introduction

Changes in the distribution of income in the US and elsewhere have kindled a lively debate on possible explanations. Special attention was given to the alleged decline of the middle income classes. Burkhauser, Cutts, Daly, and Jenkins (1999) addressed this issue by focusing on statistical inference and on measuring changes in the income distribution. They argue that sensible conclusions can only be drawn by comparing changes between years when the economy was in the same state with respect to the business cycle. Consequently, they compare the US income distributions at two peaks of the business cycle (i.e. in 1979 and 1989). Although they found significant increases in both tails of the distribution, the upper tail showed the largest gain by far. This picture is due to the fact that the authors looked at the problem from an absolute point of view in the sense that they did not control for economic growth.

The picture suggested by Burkhauser et al. (1999) remains incomplete, because their study does not shed any light on the underlying income process. An assessment of the increased inequality observed, however, must take into

account the extent of mobility. Of course, this is easier said than done. The data attrition problem becomes severe because of the long time span between the two business cycle peaks. This makes it practically impossible to obtain a representative sample from a longitudinal database, like the Panel Study of Income Dynamics (PSID). In addition, income dynamics directly derived from panel data might be biased, because on average, individuals earn more when growing older.

To close this gap, we propose an alternative method that enables researchers to get estimates on the dynamics of income from cross-section information. To make our point clear, we will look at two closely related cases in the empirical section of this paper. First, we will adjust a theory based hypothesis of income dynamics in the light of cross-section information stated as income distributions observed at two points in time. We thus get estimates on the dynamics of income from cross-section information only. This proceeding enables researchers to shed light on the dynamics of income even in the absence of panel data information. Second, we draw the initial hypothesis on the dynamics of income from panel data. Here, our method is used to enhance the quality of regularly derived transition matrices by incorporating information from cross-section databases, which usually cover a larger sample and therefore contain more information than panel data. In this case, our approach leads to a statistical test which can be used to draw inferences concerning the initial hypothesis. The test indicates whether the information inherent in the cross-section leads to statistically significant adjustments of the initial hypothesis. Because the adjustment is interpreted as a projection, the adjusted dynamics, as seen from the initial hypothesis, is always "closer" to the true but unknown dynamics. We therefore always use the adjusted income dynamics to compute descriptive measures capturing important aspects of the evolution of the income distribution.

We model the income process as a discrete-time Markov chain using transition matrices. The proposed method relies on the minimization of the entropy relative to a given initial hypothesis subject to the observed cross-section income distributions. This optimization leads to an adjusted transition matrix which remains closest (in the sense of the relative entropy) to the initial hypothesis, but is compatible with the observations. The resulting necessary adjustments of the hypothesis can be expressed in terms of the Schrödinger multipliers (exponentials of the corresponding Lagrange multipliers). Our approach can thus be interpreted as a misspecification analysis in the sense of WHITE (1994). Compared to other approaches (see among others LEE, JUDGE, and ZELLNER 1970; GOLAN, JUDGE, and MILLER 1996) which seek to extract information on transition probabilities from cross-section observations, our method is easy to implement, has

the advantage to avoid cumbersome optimization procedures, and automatically produces transition matrices.

Having presented our method in Section 2 of the paper, we apply our concepts to the same data as BURKHAUSER et al. (1999). Starting from three more or less plausible theory based hypotheses, we estimate transition matrices which map the 1979 into the 1989 income distribution. We also show how a transition matrix estimated from the PSID database must be adjusted in the light of the cross-section information contained in the very representative BURKHAUSER et al. (1999) sample. Finally, we discuss the properties of the adjusted transition matrices and their implications for the future development of the income distribution.

2. Methodological Foundations

2.1 A Probabilistic Model[1]

Suppose we observe a large number of N distinguishable individuals, whose incomes develop independently, driven by the same unknown, discrete and time invariant Markov process. This means that the transition matrix $\mathbf{P}$, representing the Markov process, doesn't change over time. Starting with some initial income distribution at time 0 with density m_0, the process is generating a time sequence of income-distributions represented by densities $\{m_t\}$, $t \in \{1, 2, \ldots\}$.

To simplify the presentation and in the face of empirical applications, these distributions are defined on a finite partition $I = \{I_i\}_{i=1,\ldots,K}$ of $\mathfrak{R}^+$, making them discrete probability distributions m_t. Thus, the income generating process conforms to a Markov chain with initial distribution m_0. Consequently, $\{m_t\} = \{(m_{t,1}, \ldots, m_{t,K})'\}$ is a sequence of $(K \times 1)$-vectors with properties:

$$m_{t,i} \geq 0 \qquad \forall\, i \in \{1, \ldots, K\}, \tag{1}$$

$$\sum_{i=1}^{K} m_{t,i} = 1, \quad t \in \{1, 2, \ldots\}.$$

In the following, we assume that this sequence is observed at two points in time only, t and $t+1$. Thus, incomes at the two points in time are distributed according to the discrete probability distributions $m_t = (m_{t,1}, \ldots, m_{t,K})'$ and $m_{t+1} = (m_{t+1,1}, \ldots, m_{t+1,K})'$ defined on I, where $m_{t,i}$ indicates the probability of

---

1   AEBI (1996, 1997) introduced the probabilistic approach into the mathematical literature.

an individual income being in income class i in period t. In addition, it is supposed that no information concerning the true transition probabilities of the Markov process is available. We then ask the following questions: observing the cross-section distributions m_t and m_{t+1}, what can be said about the individual dynamics (transition probabilities) and how should we modify any prior belief about individual dynamics?

If it were possible to observe the incomes of all individuals across time, it would be easy to count the number of individuals that are in income class i in period t and in income class j in period $t+1$. Let's denote these numbers by γ_{ij} and arrange them in a $(K \times K)$-contingency table Γ:

$$\Gamma = (\gamma_{ij})_{i,j=1,\ldots,K} \text{ with } \sum_{i,j=1}^{K} \gamma_{ij} = N \text{ and } \gamma_{ij} \in [0, N]. \tag{2}$$

We will call this matrix *income history matrix*. Γ can be represented with the corresponding 2-dimensional joint probability distribution (2-dimensional density matrix) $\mathbf{D}$:

$$\mathbf{D} = (d_{ij})_{i,j=1,\cdots,K} = \frac{\Gamma}{N}. \tag{3}$$

Neither the true income history matrix Γ_{true} nor the corresponding 2-dimensional density $\mathbf{D}_{\text{true}}$ are usually observable. It is solely known that they have to be compatible with the observed true income distributions m_t and m_{t+1}. In the following, we assume that nobody gets lost or is joining when passing from t to $t+1$.[2] Thus, each individual starting in some income class i must end up in some income class j. Likewise, each individual ending up in some income class j must have started from some income class i. Thus, the income history matrix and the corresponding 2-dimensional density matrix have to meet the following restrictions:

$$\sum_{j=1}^{K} \gamma_{ij} = N \cdot m_{t,i} \ \forall i \in \{1,\ldots,K\}; \quad \sum_{i=1}^{K} \gamma_{ij} = N \cdot m_{t+1,j} \ \forall j \in \{1,\ldots,K\}; \tag{4}$$

2   Even though this assumption seems to be unrealistic at first glance, it gets more plausible if we
    follow Champernowne (1953) and identify N as the number of incomes instead of individuals. In Champernowne's words, this means that "the incomes live on individually, although their recipients are transitory". Alternatively, we could establish some "income class" 0. At t, class 0 consists of individuals that are present in $t+1$ but not in period t (joining individuals, "births"). At $t+1$, only individuals that are present in period t but not in $t+1$ belong to class 0 (leaving persons, "deaths").

$$\sum_{j=1}^{K} d_{ij} = m_{t,i} \quad \forall i \in \{1,\ldots,K\}; \qquad \sum_{i=1}^{K} d_{ij} = m_{t+1,j} \quad \forall j \in \{1,\ldots,K\}.$$

In matrix notation and with the $(K \times 1)$-vector ι of elements 1, these restrictions simplify to:

$$\begin{aligned} \Gamma \cdot \iota &= N \cdot m_t & \Gamma' \cdot \iota &= N \cdot m_{t+1} \\ \mathbf{D} \cdot \iota &= m_t & \mathbf{D}' \cdot \iota &= m_{t+1}. \end{aligned} \qquad (4)'$$

Because $\Sigma_i m_{t,i} = 1$ and $\Sigma_j m_{t+1,j} = 1$, the above conditions impose $2 \cdot K - 2$ independent restrictions on Γ as well as on $\mathbf{D}$, referred to as *continuity restrictions* or *initial and terminal conditions*.

In general, these restrictions are fulfilled by an infinite number of income histories and corresponding 2-dimensional densities. All 2-dimensional density matrices compatible with the observed income distributions m_t and m_{t+1} constitute a convex subset $\overline{\mathbf{D}}$ of the set of all 2-dimensional $(K \times K)$-density matrices $\mathbf{D}$ (see Figure 1). Thus, the continuity restrictions can also be expressed in terms of set theory:

$$\mathbf{D} \in \overline{\mathbf{D}}. \qquad (4)''$$

The unknown true 2-dimensional density matrix, that is to be estimated, is an element of the set $\overline{\mathbf{D}}$. Because the true income generating process is unknown, a hypothesis or a model of this process is needed for periods t to $t+1$. This hypothesis will be formulated in terms of a 2-dimensional density matrix $\mathbf{D}_{\text{mod}} = (d_{\text{mod},ij})_{i,j=1,\ldots,K}$. This model can be gained from empirical observations or theoretical reasoning and is generated either directly or even better, with a hypothetical transition matrix $\mathbf{P}_{\text{mod}} = (p_{\text{mod},ij})_{i,j=1,\ldots,K}$. Because contrary to the 2-dimensional density, the transition matrix is time invariant by assumption, and because transition matrices are more common in economics, it seems convenient to generate the hypothetical transition matrix $\mathbf{P}_{\text{mod}}$ in order to derive the density matrix $\mathbf{D}_{\text{mod}}$. Note that a given 2-dimensional density $\mathbf{D}$ unambiguously implies the corresponding transition matrix $\mathbf{P}$. Conversely, a given transition matrix $\mathbf{P}$ implies infinitely many corresponding 2-dimensional density matrices $\mathbf{D}$. For each arbitrarily chosen initial income distribution $\mu_t = (\mu_{t,1}, \ldots, \mu_{t,K})'$, $\mu_{t,i} \cdot p_{\text{mod},ij} = d_{\text{mod},ij}$ represents the hypothesized unconditional probability for any individual, to be in income class i at time t and in class j at $t+1$. Given the hypothetical transition matrix $\mathbf{P}_{\text{mod}}$, the problem that the hypothetical 2-dimensional density $\mathbf{D}_{\text{mod}}$

is not distinct is identical with the indefiniteness of μ_t, the initial distribution of the hypothetical dynamics of income. The solution to this problem of indefiniteness is discussed in depth in STEINER (2004, pp. 33–36).[3] With a chosen distribution μ_t, the hypothetical 2-dimensional density matrix is given by

$$\mathbf{D}_{\text{mod}} = (d_{\text{mod},ij})_{i,j=1,\ldots,K} = (\mu_{t,i} \cdot p_{\text{mod},ij})_{i,j=1,\ldots,K} = \text{diag}(\mu_t) \cdot \mathbf{P}_{\text{mod}}, \qquad (5)$$

where $\text{diag}(\mu_t)$ is the diagonal matrix with the elements of μ_t on the main diagonal. Note that the hypothesis $\mathbf{D}_{\text{mod}}$ does not fulfill the continuity restrictions (4) in general. $\mathbf{D}_{\text{mod}}$ corresponds to the null hypothesis when it comes to statistical inference (see Section 2.3).

The continuity restrictions (4) are not sufficient to uniquely determine the income history Γ and the corresponding 2-dimensional density matrix, respectively. The problem to be solved could be described as follows:

Among all possibilities, we have to find the income history Γ_{opt} or the corresponding 2-dimensional density matrix $\mathbf{D}_{opt}$ compatible with the continuity restrictions (4) on one hand and having maximal probability of being realized under the chosen hypothesis $\mathbf{D}_{mod}$ on the other.

This problem is going to be solved in two steps. First, we compute the probability for some distinct income history Γ to be realized under the chosen hypothesis $\mathbf{D}_{\text{mod}}$. Second, we will solve the underlying optimization problem. However, the analysis is not straightforward because in general, the hypothesis does not meet the continuity restrictions. Viewed from the perspective of our hypothesis, the law of large numbers implies that 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 look for the income history matrix with the probability of being realized tending to zero at the slowest rate.[4]

---

3   Shortly, the solution is to take the observed true income distribution m_t as initial distribution of the hypothesis (see Section 2.2). Among other reasoning, this mixing of hypothesis and initial observation can be justified by the fact that the solution of the subsequent optimization problem has to comply with the continuity restrictions (4) anyway.

4   The field of application of the theory of large deviations is best outlined by means of a coin tossing thought experiment. Suppose a fair coin is going to be tossed N times. The probability of a rare event, e.g. more than 70% out of the N tosses yield head, decreases very fast when the number of tosses N is increased. The theory of large deviations is dealing with the speed at which the probability of such rare events tends to 0 when the number of observations (coin tossings N) is increased. In other words, this theory analyzes the behaviour in the tails of distributions. The theory of large deviations could also be labelled as theory of rare events or

2.1.1 Probabilities of Income History Matrices

Assume that individual incomes evolve independently from each others. We will now compute the probability that a given income history Γ, compatible with restrictions (4), is going to be realized by N individuals from the viewpoint of the hypothesis $\mathbf{D}_{mod}$. This given income history Γ can be realized in different ways by N individuals. At first, we will compute the number $n(\Gamma)$ of distinct possibilities to realize Γ. This number corresponds to the number of possibilities to arrange N distinct individuals in subgroups of γ_{ij} persons. Elementary combinatorics yields:

$$n(\Gamma) = \binom{N}{\gamma_{11}} \cdot \binom{N - \gamma_{11}}{\gamma_{12}} \cdot \binom{N - \gamma_{11} - \gamma_{12}}{\gamma_{13}} \cdots$$

$$\binom{N - \sum_{j=1}^{K} \gamma_{1j}}{\gamma_{21}} \cdots \binom{N - \sum_{i=1}^{K-1}\sum_{j=1}^{K-1} \gamma_{ij} - \sum_{j=1}^{K-1} \gamma_{Kj}}{\gamma_{KK}} = \frac{N!}{\prod\limits_{i,j=1}^{K} \gamma_{ij}!}. \tag{6}$$

Because individuals are distinguishable, the history Γ could be realized by N individuals in $n(\Gamma)$ ways. From the viewpoint of hypothesis $\mathbf{D}_{mod}$, we will now compute the probability $Pr(\Gamma)$ of realization of one specific member out of these $n(\Gamma)$ possibilities. As usual with multinomial distributions, this probability is given by:

$$\mathrm{Pr}(\Gamma) = \prod_{i,j=1}^{K} (\mu_{t,i} \cdot \mathrm{p}_{\mathrm{mod,\,ij}})^{\gamma_{ij}} = \prod_{i,j=1}^{K} (d_{\mathrm{mod,\,ij}})^{\gamma_{ij}}. \tag{7}$$

Thus, each of the $n(\Gamma)$ specific possibilities of realizing income history Γ has probability $\mathrm{Pr}(\Gamma)$ of being realized under hypothesis $\mathbf{D}_{mod}$. From a macroeconomic point of view, we are not interested in any specific realization of history Γ, we are just interested in the probability, that history Γ is realized in one way or another. Consequently, the probability $\mathrm{Pr}_N(\Gamma|\mathbf{D}_{mod})$ that income history Γ is going to be realized by N individuals from the viewpoint of hypothesis $\mathbf{D}_{mod}$ is given by the product of equations (6) and (7):

---

theory of analyzing the tail-distributions. The probability of realization of such rare events tends to 0 at an exponential rate. This exponential rate of convergence is of eminent interest in the theory of large deviations.

$$\Pr{}_N(\Gamma \,|\mathbf{D}_{\text{mod}}) = \Pr(\Gamma) \cdot n(\Gamma) = \frac{N!}{\displaystyle\prod_{i,j=1}^{K} \gamma_{ij}!} \cdot \prod_{i,j=1}^{K} (d_{\text{mod},\,ij})^{\gamma_{ij}}$$

$$= N! \cdot \prod_{i,j=1}^{K} \frac{(d_{\text{mod},\,ij})^{\gamma_{ij}}}{\gamma_{ij}!}. \tag{8}$$

2.1.2 Minimization and Optimally Adjusted Dynamics

Infinitely many income history matrices Γ or 2-dimensional densities $\mathbf{D}$ are compatible with the continuity restrictions (4). Out of this set and from the viewpoint of the hypothesis, we have to unambiguously identify the most probable income history or 2-dimensional density respectively. To do this, we deploy the *fundamental hypothesis of statistical mechanics* to the evolution of incomes. This general principle from particle physics means:

> *An observation on a macroscopic level (e.g. marginal distributions) is realized in the limit of infinitely many particles (e.g. individuals) by that microscopic ensemble (e.g. N-samples with N→∞), which has maximal probability given the observation.*[5]

Thus, we seek the income history which satisfies the continuity restrictions and which has, from the viewpoint of our hypothesis, the highest possibility of being realized.[6]

As already mentioned, the law of large numbers implies that every income history matrix that is compatible with the continuity restrictions has, from the perspective of our hypothesis, probability zero of being realized as N tends to infinity:

$$\Pr{}_N(\Gamma|\mathbf{D}_{\text{mod}}) \to 0 \text{ , for } N \to \infty . \tag{9}$$

This is because Γ and the corresponding 2-dimensional density $\mathbf{D}$ have to be compatible with the continuity restrictions (4) while hypothesis $\mathbf{D}_{\text{mod}}$ usually is not. However, we can find a unique solution to our optimization problem if we interpret "maximal probability" as "vanishing of the realization probability at the slowest rate". This is a so-called *large deviation* argument. The rate at which the realization probability (8) tends to zero is given by the limit of

---

5   See SCHRÖDINGER 1931 and LANFORD 1973.
6   Chapter I in ELLIS (1985) provides an insightful introduction to the concepts we will use subsequently.

$(1/N) \cdot \log[\Pr_N(\Gamma|\mathbf{D}_{mod})]$.[7] Using Stirling's formula for large factorials,[8] this limit is given by:

$$\lim_{N \to \infty} \frac{1}{N} \cdot \log\left[\Pr_N(\Gamma \mid \mathbf{D}_{mod})\right] = -H(\mathbf{D} \mid \mathbf{D}_{mod}) \quad \text{with} \quad \mathbf{D} = \frac{\Gamma}{N}. \quad (10)$$

The function $H(\mathbf{D}|\mathbf{D}_{mod})$ is known as *relative entropy* or *Kullback-Leibler divergence* of the 2-dimensional density $\mathbf{D}$ with respect to the hypothesis $\mathbf{D}_{mod}$. It is defined as follows:

$$H(\mathbf{D} \mid \mathbf{D}_{mod}) = \sum_{i,j=1}^{K} d_{ij} \cdot \log\left(\frac{d_{ij}}{d_{mod,ij}}\right) \quad \text{with} \quad 0 \cdot \log\left(\frac{0}{0}\right) = 0$$

$$\text{and} \quad d_{ij} \cdot \log\left(\frac{d_{ij}}{0}\right) = +\infty.\text{[9]} \quad (11)$$

Because $\Pr_N(\Gamma|\mathbf{D}_{mod})$ converges exponentially fast to zero at rate (11) for large N, the function $H(\mathbf{D}|\mathbf{D}_{mod})$ is also called *rate function*.

$$\Pr_N(\Gamma \mid \mathbf{D}_{mod}) \stackrel{N \to \infty}{=} e^{-N \cdot H(\mathbf{D}|\mathbf{D}_{mod})}. \quad (12)$$

The relative entropy has the following properties:

– $H(\mathbf{D}|\mathbf{D}_{mod}) \geq 0$; nonnegative function of $\mathbf{D}$.
– $H(\mathbf{D}|\mathbf{D}_{mod}) = 0$ if and only if $\mathbf{D} = \mathbf{D}_{mod}$.
– $H(\mathbf{D}|\mathbf{D}_{mod})$ is a strict convex function of $\mathbf{D}$.

---

7   We denote the natural logarithm throughout this paper with log.
8   Stirling's formula is: $n! = (n/e)^n \sqrt{2 \cdot \pi \cdot n} \cdot (1 + \varepsilon_n)$ with $e_n \to 0$ for $n \to \infty$.
9   The formula suggests generating hypothetical 2-dimensional densities $\mathbf{D}_{mod}$ without any zero elements because computing the relative entropy of some arbitrary matrices with respect to such a hypothesis always yields non-infinite results. If the hypothesis contains 0's, the optimally adjusted matrix has to inevitably exhibit 0's at the same matrix-positions. Otherwise, the relative entropy that should be minimized by the method is going to be infinite. To exclude zero elements in hypotheses generated from real data, the technique to be used in these cases is 2-dimensional kernel estimation.

– $H(\mathbf{D}|\mathbf{D}_{mod})$ has a global minimum in $\overline{\mathbf{D}}$.
– $H(\mathbf{D}|\mathbf{D}_{mod}) \neq H(\mathbf{D}_{mod}|\mathbf{D})$; no symmetry in its arguments.

The relative entropy does 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.[10] However, it is possible to give the relative entropy a geometric interpretation analogous to the common Euclidean distance. In particular and most relevant for this paper, the minimization of the relative entropy with respect to a given probability distribution $\mathbf{D}_{mod}$ over a convex subset $\overline{\mathbf{D}}$ of probability distributions can be viewed as a projection with properties similar to the projection in Euclidean or Hilbert spaces (Csiszár 1975). The relative entropy is some kind of "directed measure of distance" between two densities. This relation is shown in Figure 1.

Figure 1: Geometry of the Csiszár Projection
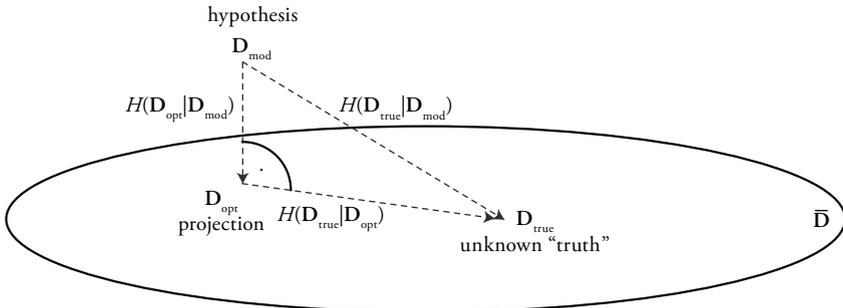
$\overline{\mathbf{D}}$ = convex set of all twodimensional (K × K)-density matrices $\mathbf{D}$ with marginal densities m_t and m_{t+1}

The analogy of the relative entropy and the Euclidian geometry becomes obvious if we look at Figure 1 and note that the relative entropy fulfills the following analogon to the theorem of Pythagoras,

$$H(\mathbf{D}_{true}|\mathbf{D}_{mod}) = H(\mathbf{D}_{opt}|\mathbf{D}_{mod}) + H(\mathbf{D}_{true}|\mathbf{D}_{opt}), \qquad (13)$$

10 Further properties of the relative entropy and a deeper discussion of its interpretation can be found among others in Kullback (1959) and Ellis (1985). By definition, a metric or distance function $M(\cdot,\cdot)$ is nonnegative, symmetric in its arguments, fulfills the triangular inequality and has the property, that $M(x,y) = 0$ if and only if $x = y$.

where $\mathbf{D}_{true}$ is the true but unknown income dynamics in the form of a 2-dimensional density matrix. $\mathbf{D}_{true}$ could be replaced by any element of $\overline{\mathbf{D}}$, equation (13) is going to hold.[11] In addition, Figure 1 shows that in terms of relative entropy, the optimally adjusted dynamics $\mathbf{D}_{opt}$ is at least as close to the true but unknown density $\mathbf{D}_{true}$ as the hypothetical dynamics $\mathbf{D}_{mod}$.

From the viewpoint of our hypothesis, the relative entropy can be interpreted as a measure of probability of observing a certain income history matrix compatible with the given marginal distributions. 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 "ill posed pure inverse problem" (GOLAN, JUDGE and MILLER 1996):

Minimize $H(\mathbf{D}|\mathbf{D}_{mod})$ over all 2-dimensional densities $\mathbf{D}$ subject to the continuity restrictions (4).

Of course, the set $\overline{\mathbf{D}}$ of all 2-dimensional densities $\mathbf{D}$ compatible with restriction (4) contains infinitely many elements. If the convexity of the set $\overline{\mathbf{D}}$ is taken in account, the 2-dimensional density $\mathbf{D}_{opt} = (\mathbf{D}_{opt,ij})_{i,j=1,\dots,K}$, that is the most probable realization fulfilling restriction (4), can be determined as the Csiszár projection of the hypothesis $\mathbf{D}_{mod}$ on $\overline{\mathbf{D}}$ (CSISZÁR 1975). This projection is computed by minimizing the relative entropy over all elements of $\overline{\mathbf{D}}$ with respect to $\mathbf{D}_{mod}$:

$$\mathbf{D}_{opt} = \arg\min_{\mathbf{D}\in\overline{\mathbf{D}}} H(\mathbf{D} \,|\, \mathbf{D}_{mod}) = \arg\min_{\mathbf{D}\in\overline{\mathbf{D}}} \sum_{i,j=1}^{K} d_{ij} \cdot \log\left(\frac{d_{ij}}{d_{mod,ij}}\right). \tag{14}$$

If the minimizing problem (14) is solvable, the solution is going to be unique because the relative entropy $H(\cdot\,|\mathbf{D}_{mod})$ is a nonnegative and strictly convex function.

In the words of the statistics literature, we have to find the minimum discrimination information under the hypothesis $\mathbf{D}_{mod}$ (KULLBACK 1959, p. 37). The solution is called the "minimum discriminant information adjustment" of $\mathbf{D}_{mod}$ (HABERMAN 1984). The Lagrangian for this optimization problem is given by

---

11 See CSISZÁR (1975) for a more detailed discussion.

$$L = \sum_{i,j=1}^{K} d_{ij} \cdot \log\left(\frac{d_{ij}}{d_{\mathrm{mod},\,ij}}\right) - \sum_{i=1}^{K} \lambda_{t,i} \cdot \left(\sum_{j=1}^{K} d_{ij} - m_{t,i}\right)$$
$$- \sum_{j=1}^{K} \lambda_{t+1,j} \cdot \left(\sum_{i=1}^{K} d_{ij} - m_{t+1,j}\right), \qquad (15)$$

where $\lambda_{t,i}$ and $\lambda_{t+1,j}$ are the $2 \cdot K$ Lagrangian multipliers associated with the constraints (4). A solution to this optimization problem exists if and only if there is at least one income history matrix that satisfies the continuity restrictions (4). In addition, this matrix has at least the same zero entries as the hypothesis $\mathbf{D}_{\mathrm{mod}}$ (CSISZÁR 1975, corollary 3.3), i.e. the solution has to be absolute continuous with respect to the hypothesis.[12] To solve the optimization problem, the set $\overline{\mathbf{D}}$ has to contain at least one element that is absolute continuous with respect to the hypothesis. The strict convexity of the relative entropy then implies the uniqueness of this solution. This solution $\mathbf{D}_{\mathrm{opt}}$ is found by differentiating (15) with respect to d_{ij} and by setting the resulting derivatives equal to zero (first order conditions):

$$d_{\mathrm{opt},ij} = \phi_{t,i} \cdot d_{\mathrm{mod},ij} \cdot \phi_{t+1,j} \ \forall i,j \in \{1,\ldots,K\}, \ \text{ with } \ \phi_{t,i} = e^{\lambda_{t,i}}$$
$$\text{and } \ \phi_{t+1,j} = e^{\lambda_{t+1,i}-1}. \qquad (16)$$

With $\phi_t = (\phi_{t,1},\ldots,\phi_{t,K})'$ and $\phi_{t+1} = (\phi_{t+1,1},\ldots,\phi_{t+1,K})'$, equation (16) can be written more compactly in matrix notation:

$$\mathbf{D}_{\mathrm{opt}} = \mathrm{diag}(\phi_t) \cdot \mathbf{D}_{\mathrm{mod}} \cdot \mathrm{diag}(\phi_{t+1}). \qquad (16)'$$

Given the observed marginal densities m_t and m_{t+1} and from the viewpoint of the hypothesis $\mathbf{D}_{\mathrm{mod}}$, $\mathbf{D}_{\mathrm{opt}}$ is the most probably realized 2-dimensional density matrix.

---

12 I.e. $\mathbf{D}_{\mathrm{mod},ij} = 0$ implies $\mathbf{D}_{\mathrm{opt},ij} = 0$. From (11) follows that the relative entropy of all matrices $\mathbf{D}$, not absolute continuous with respect to the hypothesis, is going to be infinite. To avoid this, $\mathbf{D}_{\mathrm{opt}}$ must have zeros at the same positions as the hypothesis $\mathbf{D}_{\mathrm{mod}}$. Hence, it is advantageous to generate the hypothesis without zero-elements to get the most realistic solution when adjusting the hypothesis to the observed marginal densities. This also proves beneficial in view of the iterative proportional fitting procedure (IPFP, see Section 2.1.3 and DEMING and STEPHAN (1940)), which is going to be used as the solving algorithm. This algorithm is easier implemented and always yields a unique solution when using strictly positive hypotheses (SINKHORN, 1967).

From the viewpoint of the hypothesis, $\mathbf{D}_{\text{opt}}$ characterizes the most probable estimation of the true but unknown dynamics of income generation.

In the theory of quantum mechanics, the ϕ's are known as *Schrödinger multipliers* (AEBI and NAGASAWA 1992). They indicate how to adjust "in the most probable way" the 2-dimensional density $\mathbf{D}_{\text{mod}}$, representing our hypothesis about income dynamics, to satisfy the continuity restrictions (4). The Schrödinger multipliers adjust the probabilities $\mathbf{D}_{\text{mod},ij}$ of our hypothesis downward, if $\phi_{t,i} \cdot \phi_{t+1,j} < 1$, and upward, if $\phi_{t,i} \cdot \phi_{t+1,j} > 1$. The matrix $\phi_t \cdot \phi_{t+1}{}'$ may therefore reveal patterns of adjustment and indicates the entries, where our hypothesis is misspecified. In addition, the Schrödinger multipliers contain some kind of "time separability property" because the $\phi_{t,i}$ depend only on the initial distribution m_t, whereas the $\phi_{t+1,j}$ depend only on the final distribution m_{t+1} (see equations (15) and (16)). The relative size of ϕ_t and ϕ_{t+1} thus indicates whether the misspecification is primarily due to the initial or to the terminal restriction.

The Schrödinger multipliers are found after differentiating L (equation (15)) with respect to the Lagrangian multipliers ($\lambda_{t,i}$ and $\lambda_{t+1,j}$ respectively) and setting the derivatives equal to zero. The resulting equation system is the so-called *Schrödinger system*:

$$
\begin{aligned}
m_{t,i} \quad &= \phi_{t,i} \cdot \sum_{j=1}^{K} d_{\text{mod},ij} \cdot \phi_{t+1,j} \\
&= \phi_{t,i} \cdot \mu_{t,i} \sum_{j=1}^{K} \mathrm{p}_{\text{mod, ij}} \cdot \phi_{t+1,j} \qquad \forall i = 1,\ldots,K \\
m_{t+1,j} \quad &= \left(\sum_{i=1}^{K} \phi_{t,i} \cdot d_{\text{mod},ij} \right) \cdot \phi_{t+1,j} \\
&= \left(\sum_{i=1}^{K} \phi_{t,i} \cdot \mu_{t,i} \cdot \mathrm{p}_{\text{mod, ij}} \right) \cdot \phi_{t+1,j} \qquad \forall j = 1,\ldots,K.
\end{aligned} \qquad (17)
$$

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 $\phi_{t,1} = \phi_{t+1,1}$.

As mentioned above, transition matrices are more common in economics than 2-dimensional densities. From the viewpoint of the hypothesis, we can reformulate the adjustment equation (16) in terms of the "most probable" transition matrix $\mathbf{P}_{\text{opt}} = (p_{\text{opt},ij})_{i,j=1,\ldots,K}$. Given the initial distribution m_t, the elements of

the 2-dimensional density and the corresponding transition matrix are related by $\mathbf{D}_{\text{opt},ij} = m_{t,i} \cdot p_{\text{opt},ij}$. The elements of $\mathbf{P}_{\text{opt}}$ are therefore derived from the hypothesis $\mathbf{P}_{\text{mod}}$ as follows:

$$
\begin{aligned}
p_{\text{opt},ij} &= \frac{d_{\text{opt},ij}}{m_{t,i}} = \frac{\phi_{t,i} \cdot d_{\text{mod},ij} \cdot \phi_{t+1,j}}{\phi_{t,i} \cdot \sum_{j=1}^{K} d_{\text{mod},ij} \cdot \phi_{t+1,j}} \\[2mm]
&= \frac{\phi_{t,i} \cdot \mu_{t,i} \cdot p_{\text{mod},ij} \cdot \phi_{t+1,j}}{\phi_{t,i} \cdot \mu_{t,i} \cdot \sum_{j=1}^{K} p_{\text{mod},ij} \cdot \phi_{t+1,j}} = \frac{p_{\text{mod},ij} \cdot \phi_{t+1,j}}{\sum_{j=1}^{K} p_{\text{mod},ij} \cdot \phi_{t+1,j}}.
\end{aligned}
\tag{18}
$$

In matrix notation, equation (18) summarizes to:

$$
\mathbf{P}_{\text{opt}} = \left(diag(\widetilde{\phi}_{t+1}) \right)^{-1} \cdot \mathbf{P}_{\text{mod}} \cdot diag(\phi_{t+1})
$$

$$
\text{with } \widetilde{\phi}_{t+1} = \left(\sum_{j=1}^{K} p_{\text{mod},1j} \cdot \phi_{t+1,j}, \cdots, \sum_{j=1}^{K} p_{\text{mod},Kj} \cdot \phi_{t+1,j} \right)'.
\tag{18'}
$$

Note that $\mathbf{P}_{\text{opt}}$ satisfies the definition of a transition matrix, i.e.

$$
p_{\text{opt},ij} \geq 0 \text{ and } \sum_{j=1}^{K} p_{\text{opt},ij} = 1 \text{ for all } i.
$$

Moreover, $\mathbf{P}_{\text{opt}}$ is obtained from $\mathbf{P}_{\text{mod}}$ only through the Schrödinger multipliers $\phi_{t+1,j}$ related to the terminal restrictions. The adjustment factors are now given by $(\widetilde{\phi}_{t+1,i}^{-1} \cdot \phi_{t+1,j})_{i,j}$. As with the 2-dimensional density, the optimally adjusted transition matrix $\mathbf{P}_{\text{opt}}$ results from multiplying each $p_{\text{mod},ij}$ with its corresponding adjustment factor.

2.1.3 Iterative Proportional Fitting Procedure (IPFP)

The optimally adjusted dynamics $\mathbf{D}_{\text{opt}}$ can be computed directly as the solution of the optimization problem (14) subject to the restrictions (4) by means of numerical optimization techniques. However, it is more convenient to implement a method known as *iterative proportional fitting procedure* (IPFP). In this context, the computing of $\mathbf{D}_{\text{opt}}$ is interpreted as the estimation of the cell-probabilities

of a $(K \times K)$-contingency table with known and given marginal distributions m_t and m_{t+1}. This problem was first treated by DEMING and STEPHAN (1940). They suggested the IPFP method to solve this optimization problem. The proceeding is an iterative one:

1. By element-wise division of the observed initial distribution m_t with μ_t, the vector of row-sums of $\mathbf{D}_{\text{mod}}$, we compute the vector θ_t.
2. The first estimation of $\mathbf{D}_{\text{opt}}$ is computed according to $\mathbf{D}_1 = \text{diag}(\theta_t) \cdot \mathbf{D}_{\text{mod}}$. The row-sums of the resulting density $\mathbf{D}_1$ are identical with m_t.
3. Element-wise division of the observed final distribution m_{t+1} with μ_{t+1}, the vector of column-sums of $\mathbf{D}_1$, generates the vector θ_{t+1}.
4. The second estimation of $\mathbf{D}_{\text{opt}}$ is computed according to $\mathbf{D}_2 = \mathbf{D}_1 \cdot \text{diag}(\theta_{t+1})$. The column-sums of $\mathbf{D}_2$ match m_{t+1}.
5. Like in step 1, μ_t and θ_t are computed where μ_t is now the vector of row sums of $\mathbf{D}_2$ instead of $\mathbf{D}_{\text{mod}}$. Concretely, steps 1 to 4 are repeated with the 2-dimensional density of the respective prior step until this process converges.

SINKHORN (1967, 1964) proved that this process of alternated adjustments of the row- and column-sums of a positive matrix $\mathbf{X}$ to prescribed row- and column-totals respectively, always converges towards a positive matrix $\mathbf{X}^* = \Delta_1 \cdot \mathbf{X} \cdot \Delta_2$, where Δ_1 and Δ_2 are diagonal matrices. In addition, he proved that this convergence is unique and that both diagonal matrices are unique up to a multiplicative constant.

In our terminology, this means that applying the IPFP method to $\mathbf{D}_{\text{mod}}$ leads to convergence against the matrix $\text{diag}(\phi_t) \cdot \mathbf{D}_{\text{mod}} \cdot \text{diag}(\phi_{t+1})$, that is, according to (16)', identical with $\mathbf{D}_{\text{opt}}$, the sought-after solution of the optimization problem (14).[13] Thus, the IPFP method always converges for positive 2-dimensional densities $\mathbf{D}_{\text{mod}}$. According to $\mathbf{D}_{\text{mod}} = \text{diag}(\mu_t) \cdot \mathbf{P}_{\text{mod}}$, this implies that both, the hypothetical transition matrix $\mathbf{P}_{\text{mod}}$ as well as the chosen initial distribution μ_t have to be positive to guarantee the convergence of the IPFP method, where $\mu_t \gg 0$ is essential for convergence.[14] The positiveness of the hypothetical transition matrix $\mathbf{P}_{\text{mod}}$ is not essential, but as we have seen earlier, there are reasons why positive hypotheses are advantageous.

It can be shown that the IPFP method converges geometrically fast, generates best asymptotically normal (BAN) estimates that are equivalent to maximum

13 Compared with SINKHORN (1967, 1964), the analogies $\Delta_1 = \text{diag}(\phi_t)$ and $\Delta_2 = \text{diag}(\phi_{t+1})$ hold.
14 A zero element in μ_t produces a row of zeros in the hypothesis $\mathbf{D}_{\text{mod}}$, which stops the IPFP method in step 1.

likelihood estimates and that these estimates minimize the relative entropy (11) (Smith 1947; Ireland and Kullback 1968).

2.2 Choosing Best Possible Hypothetical Density Matrices

In the following, it is assumed that the hypothetical Markov chain is set by choosing the corresponding transition matrix $\mathbf{P}_{mod}$. As mentioned in Section 2.1, there exists an infinite number of 2-dimensional density matrices that are compatible with the given transition matrix $\mathbf{P}_{mod}$ and thus could be used as a starting point to determine optimally adjusted densities $\mathbf{D}_{opt}$ and corresponding transition matrices $\mathbf{P}_{opt}$ respectively. All these hypothetical density matrices compatible with the chosen transition matrix $\mathbf{P}_{mod}$ constitute a convex set labeled $\bar{\mathbf{D}}_{mod}(\mathbf{P}_{mod})$:

$$\bar{\mathbf{D}}_{mod}(\mathbf{P}_{mod}) = \{\mathbf{D}_{mod} \mid \mathbf{D}_{mod} = diag(\mu) \cdot \mathbf{P}_{mod} \text{ and } \mu >> 0\}. \tag{19}$$

When it comes to pinpoint the hypothetical density $\mathbf{D}_{mod}$, a problem of indeterminacy arises. At this point, there are two questions of interest:

– Given the hypothetical transition matrix $\mathbf{P}_{mod}$, how is the choice of the arbitrary initial distribution μ_t and thus the hypothetical 2-dimensional density $\mathbf{D}_{mod}$ affecting the resulting adjusted dynamics $\mathbf{D}_{opt}$?
– Is it generally possible to solve this indeterminacy problem in a satisfactory way and if yes, how do we have to proceed?

Steiner (2004, pp. 33–36) proved that the optimally adjusted dynamics $\mathbf{D}_{opt}$ and $\mathbf{P}_{opt}$, respectively, are not affected by the choice of the initial distribution $\mu_t >> 0$. Given the hypothetical transition matrix $\mathbf{P}_{mod}$, every $\mu_t >> 0$ and thus every corresponding hypothetical density $\mathbf{D}_{mod}$ yield the same optimally adjusted dynamics. In addition, he showed that of all members of the set $\bar{\mathbf{D}}_{mod}(\mathbf{P}_{mod})$, the hypothesis generated from the observed initial distribution m_t lies, in terms of relative entropy, closest to the set $\bar{\mathbf{D}}$ of all 2-dimensional densities $\mathbf{D}$ compatible with restriction (4) and thus closest to the true but unknown dynamics $\mathbf{D}_{true}$. Because it is the only hypothesis that lies this close to $\bar{\mathbf{D}}$, it is specially labeled $\mathbf{D}_{mod}^{*} = diag(m_t) \cdot \mathbf{P}_{mod}$ to distinguish it from the other possible hypotheses.

Hence, Steiner suggests to determine the hypothetical dynamics in a two step procedure:

1. Creating an easy to justify hypothetical transition matrix $\mathbf{P}_{mod}$. Empirically derived matrices are always a good choice because they can not be questioned

a priori. If no income panel is at hand, there is always the possibility to generate the hypothetical transition matrix from theoretical reasoning. Besides the fact that theory based hypotheses are always open to criticism, they have the advantage to be at hand in any situation.

2. Pre-multiplication of the hypothetical transition matrix with the diagonal matrix generated from the observed initial distribution according to $\mathbf{D}^*_{mod} = \text{diag}(m_t) \cdot \mathbf{P}_{mod}$.

2.3 Statistical Inference

From a statistical point of view, we do not only want to know how to best adjust our hypothesis to given data, but also whether these adjustments are significant in a statistical sense. IRELAND and KULLBACK (1968) show how to test the statistical significance of these necessary adjustments of the hypothesis to the observed marginal distributions. If the hypothesis $\mathbf{D}_{mod}$ is generated directly from a sample of n individuals, the statistics

$$2 \cdot n \cdot H(\mathbf{D}_{opt} \mid \mathbf{D}_{mod}) \sim \chi^2_{2 \cdot K - 2} \qquad (20)$$

is asymptotically χ^2-distributed with $2 \cdot K - 2$ degrees of freedom. According to Ireland and Kullback, the number of degrees of freedom is given as the difference in the numbers of degrees of freedom of the unrestricted model $\mathbf{D}_{mod}$ $(K^2 - 1)$ and the restricted model $\mathbf{D} \in \bar{\mathbf{D}}$ $((K^2 - 1) - (2 \cdot K - 2))$. Thus, the number of degrees of freedom is equivalent to the number of restrictions in (4).

If the hypothetical transition matrix $\mathbf{P}_{mod}$ is generated from a sample of n individuals and the hypothetical 2-dimensional density, incorporating the observed true initial distribution, is built according to $\mathbf{D}^*_{mod} = diag(m_t) \cdot \mathbf{P}_{mod}$, the test statistics

$$2 \cdot n \cdot H(\mathbf{D}_{opt} \mid \mathbf{D}^*_{mod}) \sim \chi^2_{K-1} \qquad (21)$$

is asymptotically χ^2-distributed with just $K - 1$ degrees of freedom. Again, the number of degrees of freedom is calculated from the difference of the respective numbers in the unrestricted and the restricted models. The number of degrees of freedom of the unrestricted model $\mathbf{D}^*_{mod}$ is $(K^2 - 1) - (K - 1) = K^2 - K$. The respective number of the restricted model $\mathbf{D} \in \bar{\mathbf{D}}$ is $(K^2 - 1) - (2 \cdot K - 2) = K^2 - 2 \cdot K + 1$. Hence, we obtain the resulting $K - 1$ degrees of freedom in test statistics (21).

These statistics can be used to test the null-hypothesis given by the chosen hypothetical 2-dimensional density $\mathbf{D}_{mod}$ and $\mathbf{D}^*_{mod}$ respectively. A large value

of the test statistic indicates that significant adjustments to the initial hypothesis are necessary. Because the adjusted matrix is always closer to the unknown true dynamics than the hypothesized one, we always use the adjusted transition matrix for the evaluation of income dynamics (e.g. mobility indices, projections in the future, etc.).

Of course, the literature proposes several alternative methods to extract information on transition probabilities from cross-section observations (for example Adelman, Morley, Schenzler, and Warning 1994; Golan, Judge, and Miller 1996; Kalbfleisch and Lawless 1984; Lee, Judge, and Zellner 1970). Our approach, however, has several important advantages over these alternatives. First, because we observe the distribution only at two points in time, there are more unknown elements in the transition matrix $\mathbf{P}$ (respectively in the density $\mathbf{D}$) than observations in the restrictions. We are thus facing an ill-posed inverse problem which precludes the application of least-squares estimation. In this situation the maximum entropy principle arises as a natural criterion (Golan, Judge, and Miller 1996). Second, our approach together with the IPFP guarantees that the resulting adjusted matrices $\mathbf{D}_{opt}$ and $\mathbf{P}_{opt}$ are 2-dimensional densities and transition matrices respectively. Thus we can avoid cumbersome constrained minimization problems. Third, although our approach requires the specification of a hypothesized density $\mathbf{D}_{mod}$, this is not restrictive because setting all elements of $\mathbf{D}_{mod}$ equal to $1/K^2$ results in a non-informative prior. Finally, the minimization by the iterative proportional fitting procedure is robust and easy to implement.

3. Empirical Results

We are now in a position to apply the method described in the previous section to real data. In order to gain information on the unobserved process of income dynamics from cross-section data, our approach needs the following two ingredients:

Income distributions observed at two points in time. These distributions are labeled m_{1979} and m_{1989} respectively. They are extracted from the Current Population Survey (CPS).

An initial hypothesis of the process of income dynamics. This hypothesis is stated in the form of a transition matrix $\mathbf{P}_{mod}$ or equivalently, by incorporating the information inherent in the initial distribution m_{1979}, as a 2-dimensional density matrix $\mathbf{D}_{mod}$.

Our approach is then used to adjust the initial hypothesis to the observed income distributions in order to get an adjusted transition matrix $\mathbf{P}_{opt}$ and an

adjusted 2-dimensional density $\mathbf{D}_{opt}$. The adjusted transition matrix is then used to gain information on the process of income dynamics such as mobility measures, projections into the future and measures of inequality.

For our empirical investigation we used the same data as BURKHAUSER et al. (1999). These data come from the Current Population Survey (CPS) and are based on pre-tax post-transfer household incomes measured in 1989 US dollars. In contrast to Burkhauser et al., we controlled for overall economic growth by adjusting the 1979 data with the mean of the 1989 data. The household income is converted to individual income using an equivalence scale with 0.5 elasticity (square-root scale). In all our calculations we also take into account the weight associated with each sample observation. For the sake of exposition, we have re-estimated the densities of the income distributions in 1979 and 1989 using the adaptive kernel density estimator described in SILVERMAN (1986) and adopted by BURKHAUSER et al. (1999). The estimates are plotted in Figure 2. They clearly document the relatively large decline of the middle to upper middle income classes with a corresponding increase of the highest and lowest income classes.

Figure 2: Income Distribution in the US (Mean Adjusted),
Individual Equivalent Pre-Tax Post-Transfer Household Income
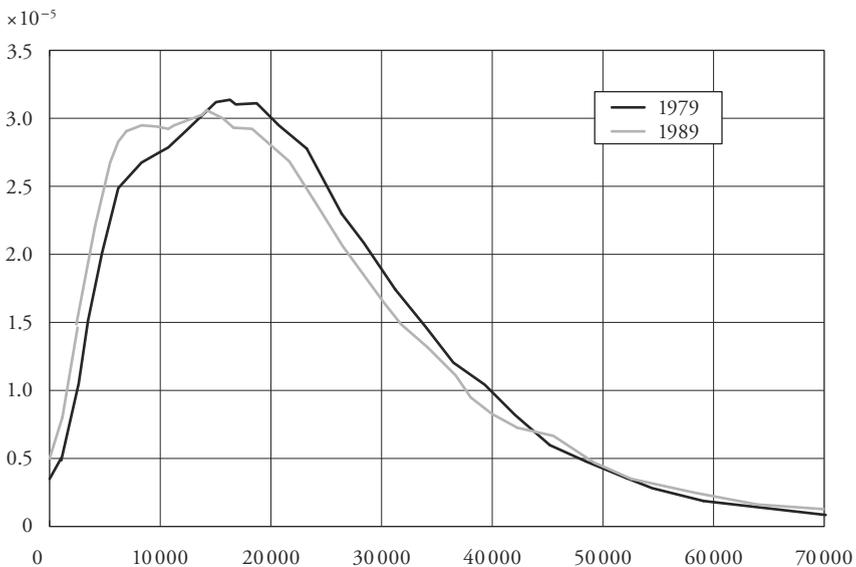
As our approach is based on a finite state space, we partitioned the real line into K intervals. We have chosen K arbitrarily equal to 10 presuming that this is enough to capture the main characteristics of the income process. Furthermore we have fixed the boundaries of the intervals in such a way that they are of equal length on a logarithmic scale, except the first and the tenth interval, of course. The resulting partition is documented in Table 1 which also shows how the incomes are distributed into the income classes in the years 1979 and 1989.

Table 1: Income Classes and Corresponding Probability Density Functions

income classes	grid points for AR(1) process	probability density function (%)	
		1979	1989
(-∞–5,790]	5,000	8.23	10.14
(5,790–7,762]	6,704	4.81	5.64
(7,762–10,407]	8,988	7.13	7.82
(10,407–13,954]	12,051	10.27	10.50
(13,954–18,708]	16,157	14.73	14.19
(18,708–25,083]	21,662	18.23	16.63
(25,083–33,630]	29,044	16.89	14.97
(33,630–45,090]	38,941	11.71	10.67
(45,090–60,454]	52,210	5.46	5.82
(60,454–∞)	70,000	2.54	3.62
		summary statistics	
sample size		65,238	59,941
mean		23,061	23,061
median		20,254	19,314

Of course, Table 1 documents the same phenomenon as Figure 2: the relatively important reduction of the middle income classes. There are approximately 5 percent less individuals in the income classes 5 to 8 in 1989 compared to 1979. As a result, not only the top two income classes increased, but also the low and lower middle income classes, i.e. classes 1 to 4. The purpose of our analysis is to explain this development by estimating appropriate transition matrices which will serve as the basis for further inference.

The critical issue is to specify a reasonable hypothesis $\mathbf{D}_{mod}$. In order to encompass alternative views on the dynamics of the income process, we have investigated three "theory" and one empirically based specifications $\mathbf{P}_{mod}$. We chose the observed initial distribution m_{1979} to transform the hypothetical transition matrices $\mathbf{P}_{mod}$ into the corresponding hypothetical 2-dimensional densities $\mathbf{D}^*_{mod}$. Thus we are left with the specification of the hypothetical transition matrices $\mathbf{P}_{mod}$.

Each specification is presented in a separate table together with the corresponding adjusted transition matrix $\mathbf{P}_{opt}$. Each table is linked to a Figure where the associated Schrödinger multipliers ϕ_{1979} and ϕ_{1989} are plotted. Although the Schrödinger multipliers corresponding to the initial conditions are not necessary for the calculations of the adjusted transition matrices, they are nevertheless plotted to give a more complete picture and because we always have to compute them in an intermediate step of the iterative proportional fitting procedure (IPFP).

The first specification starts from a transition matrix $\mathbf{P}$ with maximum mobility. These are transition matrices where the probability of attaining a certain income class is independent from the initial income class (Prais 1955). These transition matrices thus have equal rows so that their Prais' mobility index equals 1.[15] It can be shown that our entropy approach adjusts any transition matrix with maximum mobility in such a way that it becomes equal to a matrix with rows equal to $\mathbf{D}^*_{mod}$. Moreover, using the normalization $\phi_{t,1} = \phi_{t+1,1}$ and taking μ equal to m_t, the Schrödinger multipliers related to the initial restriction are constant and equal to

$$\phi_{t,1} = \sqrt{m_{t+1,1}/m_{t,1}}.$$

The Schrödinger multipliers related to the terminal restrictions then become

$$\phi_{t+1,j} = \sqrt{m_{t,1}/m_{t+1,1}} \cdot (m_{t+1,j}/m_{t,j}).$$

Given that the adjusted transition matrix is the same for any maximum mobility transition matrix, we may well take our hypothesis $\mathbf{P}_{maxmob}$ as the matrix with rows equal to m'_{1979}. Thus $\mathbf{P}_{maxmob}$ has m_{1979} as its steady state distribution. The adjusted transition matrix $\mathbf{P}_{opt}$ then equals $(m_{1989},\ldots,m_{1989})'$ where each row corresponds to the last column of Table 1. The Schrödinger multipliers plotted in Figure 3 show that the income distribution has indeed become more dispersed. The probabilities of moving into the middle income classes (classes 5 to 8) have

15 Prais' mobility index is defined as $(K - tr(\mathbf{P}))/(K - 1)$ (Prais 1955).

to be reduced whereas the probabilities of moving into the highest and lowest income classes have to be increased.

Figure 3: Schrödinger Multipliers for the Maximal Mobility Hypothesis
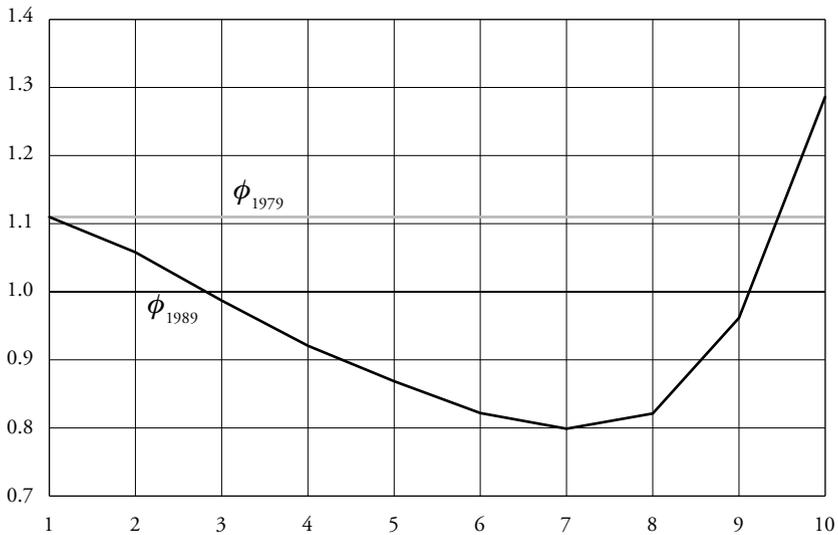
The second "theory" based hypothesis represents a more interesting specification. Suppose that individuals cannot move more than one class up or down from one year to the next. This implies that the one period hypothetical transition matrix is a 3-band matrix denoted by $\mathbf{P}_{3band}$. Suppose further that the observed distribution in 1979, m_{1979}, is invariant to $\mathbf{P}_{3band}$.[16] Taking $\mu = m_{1979}$ as before, the joint probability density function $\mathbf{D}^*_{3band}$ over the decade 1979 to 1989 is then given by

$$\mathbf{D}^*_{3band} = \text{diag}(m_{1979}) \cdot \mathbf{P}^{10}_{3band} \tag{22}$$

This second hypothesis is labeled 3-band hypothesis. The corresponding transition matrix $\mathbf{P}^{10}_{3band}$ is displayed in Table 2 (numbers in parentheses) and offers

---

16 Of course, there are infinitely many 3-band transition matrices preserving a given density. Our construction follows the recommendation of BOYARSKY and GÓRA (1997, 258).

several interesting features. First the probability of staying in the lowest income class is with 0.51 relatively high. Even for the next two income classes, there is a substantial probability of falling back into the lowest class. From class three to five, the probability of moving up clearly outweighs the probability of falling down. This changes from class seven on where the chances of falling down are now higher than the chances of moving up further.

Table 2: Hypothesized and Adjusted Transition Matrix for the 3-band Hypothesis

0.5700	0.2133	0.1103	0.0541	0.0287	0.0151	0.0064	0.0019	0.0002	0.0000
(0.5102)	(0.2126)	(0.1291)	(0.0711)	(0.0408)	(0.0229)	(0.0101)	(0.0029)	(0.0003)	(0.0000)
0.4383	0.1863	0.1337	0.0908	0.0658	0.0463	0.0260	0.0105	0.0022	0.0002
(0.3640)	(0.1723)	(0.1452)	(0.1107)	(0.0869)	(0.0652)	(0.0379)	(0.0149)	(0.0026)	(0.0002)
0.2055	0.1213	0.1476	0.1423	0.1353	0.1171	0.0792	0.0392	0.0109	0.0015
(0.1490)	(0.0979)	(0.1399)	(0.1514)	(0.1559)	(0.1440)	(0.1007)	(0.0486)	(0.0116)	(0.0012)
0.0853	0.0696	0.1204	0.1468	0.1686	0.1704	0.1308	0.0758	0.0268	0.0054
(0.0570)	(0.0518)	(0.1052)	(0.1440)	(0.1790)	(0.1931)	(0.1532)	(0.0866)	(0.0261)	(0.0039)
0.0355	0.0397	0.0899	0.1325	0.1772	0.1962	0.1670	0.1058	0.0442	0.0120
(0.0228)	(0.0284)	(0.0755)	(0.1249)	(0.1809)	(0.2136)	(0.1880)	(0.1162)	(0.0414)	(0.0082)
0.0164	0.0245	0.0682	0.1174	0.1720	0.2089	0.1847	0.1276	0.0596	0.0207
(0.0103)	(0.0172)	(0.0563)	(0.1088)	(0.1726)	(0.2236)	(0.2044)	(0.1378)	(0.0549)	(0.0140)
0.0078	0.0154	0.0519	0.1012	0.1645	0.2075	0.1981	0.1437	0.0759	0.0338
(0.0049)	(0.0108)	(0.0425)	(0.0932)	(0.1639)	(0.2207)	(0.2178)	(0.1541)	(0.0694)	(0.0227)
0.0032	0.0088	0.0361	0.0825	0.1465	0.2016	0.2020	0.1607	0.0982	0.0606
(0.0020)	(0.0061)	(0.0296)	(0.0760)	(0.1462)	(0.2146)	(0.2223)	(0.1725)	(0.0899)	(0.0408)
0.0008	0.0032	0.0180	0.0521	0.1094	0.1682	0.1905	0.1754	0.1408	0.1417
(0.0005)	(0.0023)	(0.0151)	(0.0492)	(0.1118)	(0.1835)	(0.2149)	(0.1930)	(0.1320)	(0.0977)
0.0001	0.0004	0.0035	0.0151	0.0426	0.0840	0.1223	0.1558	0.2039	0.3725
(0.0000)	(0.0003)	(0.0032)	(0.0156)	(0.0478)	(0.1007)	(0.1515)	(0.1884)	(0.2102)	(0.2822)

☐ Adjustment factors above 1.25   ▨ Adjustment factors below 0.8

hypothesized transition matrix in parenthesis

The Schrödinger multipliers plotted in Figure 4 as well as in Table 2 indicate how the hypothesis P^{10}_{3band} has to be adjusted. In particular the probabilities of remaining in or moving into the highest income class must be substantially revised upwards. The same is true for the probabilities of falling down into one of the lowest two classes. At the same time, the probabilities of moving up from the

lowest three classes have to be strongly reduced. In addition, as the adjusted transition matrix in Table 2 makes clear, even the probability of staying in the lowest class has to be increased. Compared with the initial hypothesis, this shows a strong movement towards segregation. This feature is also reflected in the slightly lower Prais' mobility index which decreased from 0.8694 to 0.8546 (see Table 5). However, the Prais' mobility-index considers the elements in the main-diagonal of a transition matrix only. If we take a look at the diagonal in Table 2, we can easily see that the decrease in mobility in the optimally adjusted transition matrix with respect to the hypothesis $\mathbf{P}_{\text{3band}}^{10}$ happened because of the upward-adjustments in the upper two and in the lowest four income classes. The downward-adjustments in income classes 5 to 8 suggest that within these classes, mobility is actually increasing. So, the overall decreasing mobility happens because the decrease in mobility in the highest and lowest income classes outweighs the increase in the middle income classes. Besides Prais' mobility index, we report two alternative mobility measures in Table 5. The first index is the well known measure by Bartholomew.[17] While Bartholomew's index measures mobility in equilibrium, the period mobility index measures mobility associated with convergence towards the ergodic state. Bartholomew's index and the period mobility index are closely related and can be viewed as a pair of indices where each single index measures a different aspect of mobility.[18] While Bartholomew's index decreases during the process of adjustment of the hypothesis, the period mobility increases (see Table 5). This means that while there is less mobility in the ergodic state of the adjusted dynamics compared to the hypothesis, there is more mobility while the income distribution converges towards its equilibrium.

Our third "theory" based hypothesis is a Gaussian AR(1) model for the logged incomes:

$$\log(y_t) = (1 - \rho) \cdot \overline{y} + \rho \cdot \log(y_{t-1}) + \varepsilon_t, \ \ \varepsilon_t \sim \text{IIN}(0, \sigma^2) \tag{23}$$

where $\overline{y}$ denotes the mean of $\log(y_t)$ and where the autoregressive parameter ρ is strictly smaller than one in absolute value. Contrary to the previous hypotheses, this approach is not based on the assumption that the system is in its steady state in 1979. For any given ρ, there is a one-to-one correspondence between the

17 Bartholomew's mobility index is defined as $\displaystyle\sum_i \pi_i \sum_j p_{ij} |i - j|$ where π is the invariant distribution of $\mathbf{P}$ (BARTHOLOMEW 1982).

18 In AEBI, NEUSSER, and STEINER (2006) we show how both aspects of mobility, namely equilibrium and convergence mobility, can be captured simultaneously by one pair of interdependent indices.

Figure 4: Schrödinger Multipliers for the 3-Band-Hypothesis
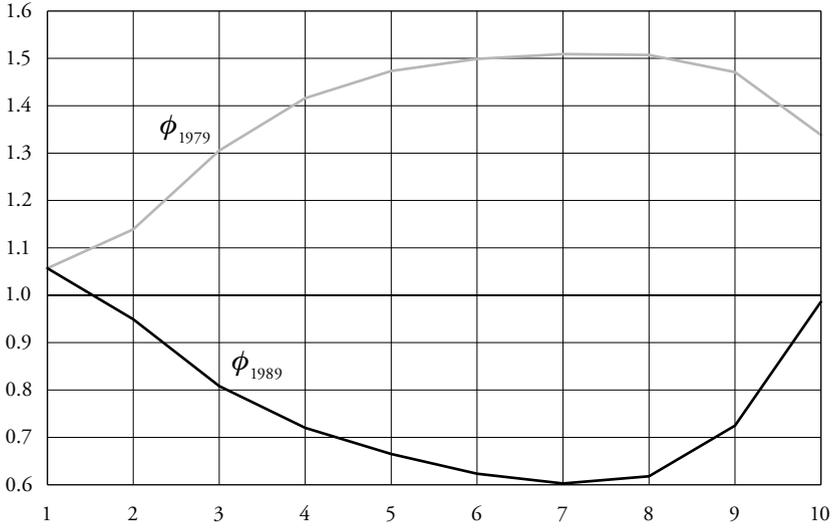
unconditional variance of $\log(y_t)$, $V_{\log(y)}$, and σ^2: $\sigma^2 = (1-\rho^2) \cdot V_{\log(y)}$. Given our initial question, we take $V_{\log(y)}$ as the estimated cross-section variance in 1979. Of course, we can not work with the AR(1) model directly, but we have to approximate it by a Markov chain, denoted by $\mathbf{P}_{AR}$. For this step we use the method proposed by TAUCHEN (1986) where the grid points are just the mid-points on a logarithmic scale of our income classes (see Table 1). This procedure leads to a family of transition matrices indexed by ρ, $\mathbf{P}_{AR}(\rho)$. As for our previous hypotheses, we set μ equal to m_{1979} to obtain the hypothesized 2-dimensional density $\mathbf{D}^*_{AR}(\rho)$:

$$\mathbf{D}^*_{AR}(\rho) = diag(m_{1979}) \cdot \mathbf{P}_{AR}(\rho) \tag{24}$$

We estimated ρ by minimizing the relative entropy also over this parameter (see equation (14)). This resulted in an estimate of 0.764 for ρ. The corresponding transition matrix is displayed in Table 3 (values in parentheses). It is similar to the 3-band transition matrix but it shows less mobility as can be seen from Table 5. The chances to move out of low income classes as well as the chances to remain in the highest income class are higher.

Table 3: Hypothesized and Adjusted Transition Matrix for the AR(1) Model ($\rho = 0.764$)

|									
0.5660	0.1714	0.1286	0.0771	0.0386	0.0142	0.0034	0.0006	0.0001	0.0000
(0.3791)	(0.2251)	(0.1943)	(0.1220)	(0.0557)	(0.0185)	(0.0045)	(0.0008)	(0.0001)	(0.0000)
0.3888	0.1767	0.1690	0.1292	0.0826	0.0387	0.0120	0.0026	0.0004	0.0000
(0.2281)	(0.2032)	(0.2237)	(0.1792)	(0.1044)	(0.0442)	(0.0136)	(0.0030)	(0.0005)	(0.0001)
0.2248	0.1473	0.1797	0.1752	0.1428	0.0853	0.0338	0.0093	0.0017	0.0002
(0.1186)	(0.1523)	(0.2139)	(0.2185)	(0.1623)	(0.0877)	(0.0345)	(0.0098)	(0.0020)	(0.0003)
0.1064	0.0976	0.1519	0.1889	0.1963	0.1496	0.0756	0.0264	0.0062	0.0011
(0.0527)	(0.0948)	(0.1698)	(0.2212)	(0.2096)	(0.1444)	(0.0724)	(0.0264)	(0.0070)	(0.0016)
0.0411	0.0516	0.1025	0.1625	0.2154	0.2094	0.1349	0.0602	0.0182	0.0043
(0.0199)	(0.0490)	(0.1120)	(0.1860)	(0.2248)	(0.1976)	(0.1263)	(0.0587)	(0.0198)	(0.0059)
0.0131	0.0222	0.0562	0.1138	0.1924	0.2384	0.1959	0.1114	0.0429	0.0136
(0.0063)	(0.0210)	(0.0613)	(0.1299)	(0.2002)	(0.2244)	(0.1829)	(0.1085)	(0.0468)	(0.0186)
0.0035	0.0079	0.0256	0.0662	0.1427	0.2255	0.2364	0.1715	0.0842	0.0366
(0.0017)	(0.0075)	(0.0279)	(0.0753)	(0.1481)	(0.2117)	(0.2201)	(0.1664)	(0.0915)	(0.0499)
0.0008	0.0024	0.0098	0.0324	0.0891	0.1796	0.2401	0.2221	0.1391	0.0845
(0.0004)	(0.0022)	(0.0105)	(0.0363)	(0.0909)	(0.1658)	(0.2198)	(0.2120)	(0.1487)	(0.1133)
0.0002	0.0006	0.0032	0.0134	0.0471	0.1211	0.2064	0.2436	0.1945	0.1700
(0.0001)	(0.0005)	(0.0033)	(0.0145)	(0.0464)	(0.1078)	(0.1824)	(0.2243)	(0.2007)	(0.2201)
0.0000	0.0001	0.0009	0.0047	0.0210	0.0689	0.1497	0.2253	0.2295	0.3000
(0.0000)	(0.0001)	(0.0009)	(0.0048)	(0.0196)	(0.0582)	(0.1256)	(0.1971)	(0.2249)	(0.3688)

☐ Adjustment factors above 1.25   ☐ Adjustment factors below 0.8

hypothesized transition matrix in parenthesis

The necessary adjustments are again best summarized by looking at the plots of the Schrödinger multipliers in Figure 5 as well as at Table 3. It is easy to see that the probabilities to move upwards from the lowest income classes must be revised downwards. The same holds for the probabilities to move into the highest income class. At the same time the probabilities to stay in the lowest income class as well as those to fall back to the lowest class must be substantially increased. For the middle income classes, the AR(1) model needs only small adjustments. More mobility seems necessary in the highest income classes. Again the adjustment leads to a reduction in Prais' mobility index. A brief look at the diagonal in Table 3 shows that this decrease in Prais' mobility index is due to the increased probability of staying in the lowest as well as in the upper middle income classes (classes 6 to 8) whereas the probability of staying in the lower middle income

classes and in the highest classes has to be reduced. A look at Bartholomew's index and the period mobility index shows that like in the 3-band hypothesis, equilibrium mobility is reduced and convergence mobility is increased as a result of the adjustment process.

Figure 5: Schrödinger Multipliers for the AR(1)-Hypothesis ($\rho = 0.764$)
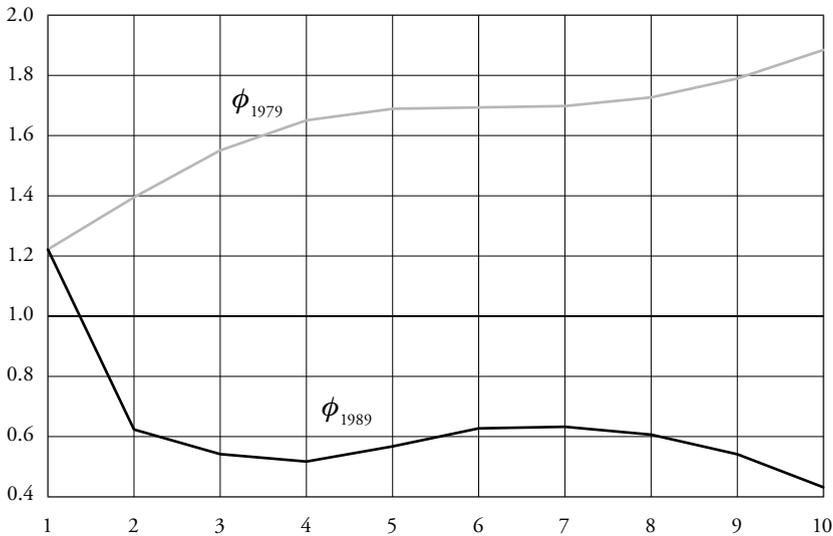
The three previously examined cases aimed at answering the question how to get information on the underlying process of income dynamics from cross-section information alone. The fourth example addresses a slightly different question: how to improve or update panel estimates of the income dynamics in the light of much more complete cross-section information. Somehow this seems to be the most natural application of our method. We take the estimated transition matrix from the PSID-panel as our starting point and seek for the most probable adjustment in the light of the larger and thus more complete information from the cross-section data used by BURKHAUSER et al. (1999).

For this purpose, we selected those persons from the PSID data which reported an income in both 1979 and in 1989. As in the Burkhauser et al. data, we computed pre-tax post-transfer incomes. This resulted in a sample of 692 individuals

out of 53,013. This type of data attrition is typical for panels over a long time span and demonstrates the usefulness of combining panel and cross-section information. We adjusted the PSID-data with the mean of the 1989 Burkhauser et al. data. Based on the reported incomes, we estimated a transition matrix defined on the same income classes as before.[19] This matrix is labeled PSID model and reported in Table 4 (numbers in parentheses). The transition matrix estimated from the PSID data delivers a reasonable and perfectly valid specification. Unlike for the previous models, it is not possible to question a priori the PSID-model because it relies on actual data.

Table 4: Hypothesized and Adjusted Transition Matrix for the PSID Model

0.3753	0.1513	0.1160	0.0953	0.0868	0.0630	0.0503	0.0443	0.0160	0.0018
(0.2817)	(0.1655)	(0.1407)	(0.1221)	(0.1018)	(0.0690)	(0.0534)	(0.0469)	(0.0170)	(0.0018)
0.2769	0.1427	0.1539	0.1316	0.1051	0.0765	0.0564	0.0381	0.0149	0.0039
(0.1987)	(0.1492)	(0.1785)	(0.1610)	(0.1178)	(0.0801)	(0.0573)	(0.0385)	(0.0151)	(0.0038)
0.1812	0.1183	0.1629	0.1626	0.1534	0.1106	0.0576	0.0286	0.0145	0.0103
(0.1248)	(0.1188)	(0.1814)	(0.1910)	(0.1651)	(0.1112)	(0.0562)	(0.0278)	(0.0142)	(0.0096)
0.1265	0.0897	0.1318	0.1597	0.1841	0.1571	0.0891	0.0371	0.0155	0.0095
(0.0859)	(0.0888)	(0.1446)	(0.1849)	(0.1953)	(0.1558)	(0.0856)	(0.0355)	(0.0150)	(0.0087)
0.0833	0.0580	0.0980	0.1379	0.1872	0.2083	0.1418	0.0595	0.0206	0.0055
(0.0563)	(0.0572)	(0.1071)	(0.1590)	(0.1977)	(0.2056)	(0.1358)	(0.0567)	(0.0197)	(0.0050)
0.0477	0.0355	0.0676	0.1132	0.1825	0.2338	0.1808	0.0942	0.0346	0.0102
(0.0322)	(0.0350)	(0.0738)	(0.1304)	(0.1927)	(0.2307)	(0.1730)	(0.0898)	(0.0331)	(0.0093)
0.0349	0.0190	0.0382	0.0795	0.1406	0.1949	0.2075	0.1634	0.0838	0.0381
(0.0239)	(0.0190)	(0.0423)	(0.0930)	(0.1506)	(0.1951)	(0.2014)	(0.1579)	(0.0815)	(0.0353)
0.0266	0.0086	0.0179	0.0490	0.0878	0.1461	0.2236	0.2238	0.1430	0.0735
(0.0185)	(0.0087)	(0.0202)	(0.0582)	(0.0955)	(0.1485)	(0.2204)	(0.2196)	(0.1411)	(0.0691)
0.0223	0.0042	0.0094	0.0356	0.0667	0.1332	0.2140	0.2235	0.1695	0.1217
(0.0157)	(0.0043)	(0.0107)	(0.0425)	(0.0731)	(0.1364)	(0.2125)	(0.2210)	(0.1686)	(0.1153)
0.0029	0.0012	0.0087	0.0288	0.0535	0.1101	0.1131	0.1349	0.1603	0.3865
(0.0020)	(0.0013)	(0.0100)	(0.0348)	(0.0592)	(0.1139)	(0.1135)	(0.1347)	(0.1610)	(0.3697)

☐ Adjustment factors above 1.25   ☐ Adjustment factors below 0.8

hypothesized transition matrix in parenthesis

19 Because our method shows the best result if there are no zero-entries in the hypothesized transition matrices, we used 2-dimensional kernel density estimation to derive the PSID-model.

Figure 6: Schrödinger Multipliers for the PSID-Hypothesis
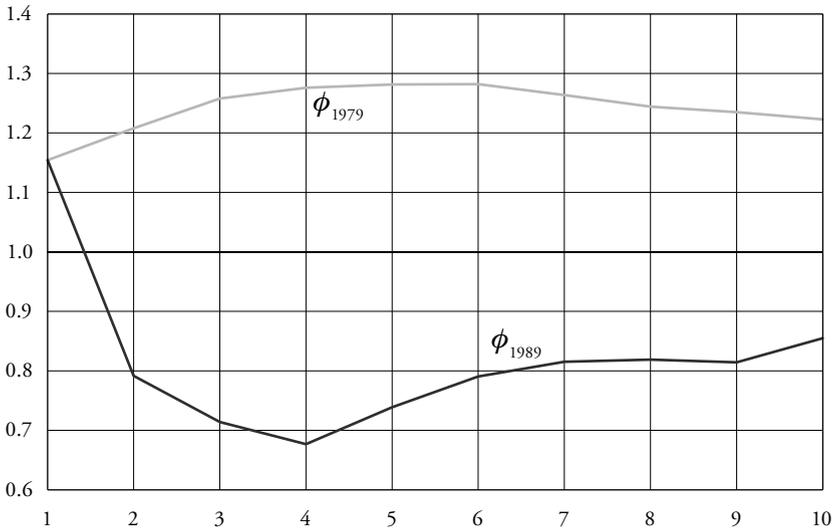
The Schrödinger multipliers in Figure 6 and the adjusted matrix reported in Table 4 indicate that the probabilities to stay in and to fall back to the lowest income class have to be increased considerably. Moreover, the probabilities to move up into higher income classes must be adjusted downwards for the low income classes. As for the previous specifications, the adjustment leads to a reduction in Prais' mobility index (see Table 5). Again, a look at the main diagonal shows that this decrease shows up because the rise in persistence in income classes 1 and 6 to 10 is larger than the rise in mobility in income classes 2 to 5. In contrast to the 3-band and the AR(1) hypotheses, the adjustment results in a rise in both equilibrium and convergence mobility, as measured by Bartholomew's index and the period mobility index, respectively.

Because the PSID-hypothesis is based on actual data ($n = 692$ individuals), we can test for the appropriateness of this hypothesis. We treat the cross-sectional distributions m_{1979} and m_{1989} as known and equal to the true (population) distribution. This is not an unrealistic assumption given the large sample size (see Table 1). The problem of adjusting the transition probabilities estimated from the PSID data can then be treated as the problem of estimating the cell probabilities of a contingency table for which the population marginal probabilities are known and fixed (IRELAND and KULLBACK 1968; AEBI, NEUSSER, and STEINER

Table 5: Summary Measures of the Four Hypotheses

	maximal mobility	3-band-hypothesis	AR(1)-hypothesis	PSID-hypothesis
Bartholomew's mobility index a				
hypothesized transition matrix	2.5957	1.5961	1.3835	1.6916
adjusted transition matrix	2.7750	1.5450	1.3317	1.7244
period mobility index b				
hypothesized transition matrix	0.4904	0.0814	0.0363	0.2224
adjusted transition matrix	0.5246	0.0876	0.0380	0.2394
Prais' mobility index c				
hypothesized transition matrix	1	0.8694	0.8369	0.8683
adjusted transition matrix	1	0.8546	0.8313	0.8612
relative entropy	0.0077	0.0110	0.0203	0.0083
test statistic (distributed as $\chi^2(9)$)				11.51[d]

a   Bartholomew's index is defined as $\sum_i \pi_i \sum_j p_{ij} |i - j|$ where π is the invariant distribution of $\mathbf{P}$.
b   The period mobility index is the convergence index that corresponds to the equilibrium index given here by Bartholomew's index (AEBI, NEUSSER, and STEINER 2006).
c   Prais' mobility index is defined as $K - tr(P)/(K - 1)$.
d   Indicates insignificance at the 1 percent level (critical value: 21.67).

1999). In this case, the sample size n equals the number of observations from the PSID data, in our case 692. With this interpretation, the value of the test statistic (21) becomes 11.51 (see Table 5) which does not lead to a rejection of the hypothesis that the PSID model is compatible with the cross-sectional observations. Although the PSID model is not rejected, we still use the adjusted transition matrix for further computations because the adjusted matrix is at least as close to the true dynamics as the hypothesis.

As a final exercise we use the adjusted transition matrices to project the 1989 income distribution ten years into the future. The corresponding distributions in 1999, m_{1999}, and the implied invariant distributions are reported in Table 6. For each distribution we also computed Atkinson's inequality index, A_ε, and the generalized entropy index, GE_α. The maximum mobility transition matrix maps m_{1989} into itself so that the distribution is expected to remain unchanged.

Table 6: Projected and Steady State Distributions

Income class	observed distributions		projected distributions					
			adjusted 3-band model		adjusted AR(1) model		adjusted PSID model	
	m_{1979}	m_{1989}	m_{1999}	SS	m_{1999}	SS	m_{1999}	SS
1	8.23	10.14	11.69	15.77	11.67	15.98	11.03	11.83
2	4.81	5.64	6.21	7.67	6.16	7.60	6.02	6.37
3	7.13	7.82	7.98	8.46	8.16	9.18	8.07	8.31
4	10.27	10.50	10.30	10.00	10.49	10.77	10.52	10.59
5	14.73	14.19	13.68	12.73	13.73	13.05	13.92	13.79
6	18.23	16.63	15.92	14.52	15.78	14.20	16.08	15.70
7	16.89	14.97	14.32	12.92	14.18	12.33	14.38	13.93
8	11.71	10.67	10.25	9.19	10.25	8.75	10.31	9.98
9	5.46	5.82	5.74	5.15	5.76	4.89	5.74	5.56
10	2.54	3.62	3.91	3.58	3.82	3.26	3.93	3.93
$A\varepsilon = 0.5$	0.0944	0.1080	0.1148	0.1271	0.1148	0.1266	0.1129	0.1160
$GE\alpha = 2$	0.1993	0.2350	0.2522	0.2873	0.2526	0.2911	0.2483	0.2575

m_{1999} projected income distribution in 1999 based on the adjusted transition matrix with initial income distribution m_{1989}.
SS   steady state or invariant distribution of the adjusted transition matrix.
$A_{\varepsilon = 0.5}$ Atkinson's inequality index with $\varepsilon = 0.5$.
$GE_{\alpha = 2}$ generalized entropy index with $\alpha = 2$.

The adjusted 3-band hypothesis implies a further increase in inequality. Note the increase of the size of the lowest three income classes as well as the increase in the top income class. In the long run, inequality is expected to increase even further although the size of the top income class shrinks below its 1989 level. This is due to the large increase of the size of the lowest income classes. According to the adjusted 3-band hypothesis, we should see a large increase in the lower tail of the income distribution. The projections of the adjusted AR(1) model are quite similar to those of the adjusted 3-band hypothesis, both in the short and in the long run. The two hypotheses are remarkably similar concerning the direction as well as the magnitude of the change. The adjusted PSID model produces similar results. Inequality is predicted to increase in the short as well as in the

long-run. However, the raise in both inequality indices is considerably less pronounced. In contrast to the 3-band and the AR(1) hypothesis, the top income class is keeping its share in the long run.

4. Conclusion

This paper presents a method to estimate or adjust transition matrices using just cross-sectional observations at two points in time. The method has been applied to explain the development of the US income distribution, in particular the movement of the middle income classes. Irrespective of the prior specification, most of the mass corresponding to the middle income classes shifted downwards. These developments led to increased inequality which is expected to continue in the short as well as in the long-run. However, because we adjusted the 1979 data with the mean of the 1989 data to control for economic growth, our projections must be interpreted in relative terms. Therefore, our results do not contradict the possibility that economic growth could "lift all boats". But there is a clear tendency towards segregation in the US society.

Although the theory based models of income dynamics used in this paper are pretty simple, it is interesting to note that after applying our method to adjust the initial theory based models, the resulting projections point in the same direction as the conclusions based on the empirically derived PSID-model. In order to make theory based models a serious alternative to empirical hypotheses, not just in qualitative but also in quantitative analysis, further investigations into theoretic models of the underlying income dynamics are necessary.

For empirically derived hypotheses, the combination of cross-sectional observations with longitudinal data is a promising research strategy. It allows taking advantage of the large sample sizes typically available in cross-sectional data without disregarding the information of individual income dynamics available from panel data. Furthermore, income dynamics directly derived from panel data can be biased, because on average, individuals earn more when growing older, and because data attrition is a serious problem. As our method corrects such flaws, it is well suited to improve the estimation of income processes from panel data.

References

Adelman, Irma, Samuel Morley, Christoph Schenzler, and Matthew Warning (1994), "Estimating Income Mobility from Census Data", *Journal of Policy Modeling*, 16, pp. 187–213.

Aebi, Robert and Masao Nagasawa (1992), "Large Deviations and the propagation of Chaos for Schrödinger Processes", *Probability Theory and Related Fields*, 94, pp. 53–68.

Aebi, Robert (1996), "Schrödinger's Time-Reversal of Natural Laws", *The Mathematical Intelligencer*, 18, pp. 62–67.

Aebi, Robert (1997), "Contingency Tables with Prescribed Marginals", *Statistical Papers*, 38, pp. 219–229.

Aebi, Robert, Klaus Neusser, and Peter Steiner (1999), "Evaluating Theories of Income Dynamics: A Probabilistic Approach", Discussion Paper 99-05, University of Berne.

Aebi, Robert, Klaus Neusser, and Peter Steiner (2006), "A Large Deviation Approach to the Measurement of Mobility", *Swiss Journal of Economics and Statistics*, 142 (2), pp. 195–222.

Bartholomew, David J. (1982), *Stochastic Models of Social Processes* (3rd ed.), Chichester.

Boyarsky, Abraham, and Pawel Góra (1997), *Laws of Chaos*, Boston.

Burkhauser, Richard V., Amy Crews Cutts, Mary C. Daly, and Stephen P. Jenkins (1999), "Testing the Significance of Income Distribution Changes over the 1980s Business Cycle: A Cross-National Comparison", *Journal of Applied Econometrics*, 14, pp. 253–272.

Champernowne, David G. (1953), "A Model of Income Distribution", *Economic Journal*, 63, pp. 318–351.

Csiszár, Imre. (1975), "I-Divergence Geometry of Probability Distributions and Minimization Problems", *The Annals of Probability*, 3, pp. 146–158.

Deming, W. Edwards and Frederick 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, pp. 427–444.

Ellis, Richard S. (1985), *Entropy, Large Deviations, and Statistical Mechanics*, New York.

Golan, Amos, George G. Judge, and Douglas Miller (1996), *Maximum Entropy Econometrics*, Chichester.

Haberman, Shelby J. (1984), "Adjustment by Minimum Discriminant Information", *Annals of Statistics*, 12, pp. 971–988.

Ireland, C. Terrence, and Solomon Kullback (1968), "Contingency Tables with Given Marginals", *Biometrika*, 55, pp. 179–188.

Kalbfleisch, John D., and Jerald F. Lawless (1984), "Least-Squares Estimation of Transition Probabilities from Aggregate Data", *Canadian Journal of Statistics*, 12, pp. 169–182.

Kullback, Solomon (1959), *Information Theory and Statistics*, New York.

Lanford, Oscar E. (1973), "Entropy and Equilibrium States in Classical Statistical Mechanics", *Lecture Notes in Physics*, 20, pp. 1–113.

Lee, Tsoung-Chao, George G. Judge, and Arnold Zellner (1970), *Estimating the Parameters of the Markov Model from Aggregate Time Series Data*, Amsterdam.

Prais, Sig J. (1955), "Measuring Social Mobility", *Journal of the Royal Statistical Society*, 188, pp. 56–66.

Schrödinger, Erwin (1931), "Über die Umkehrung der Naturgesetze", *Sitzungsberichte der Preussischen Akademie der Wissenschaften, physikalisch-mathematische Klasse*, pp. 144–153.

Silverman, Bernard W. (1986), *Density Estimation for Statistics and Data Analysis*, London.

Sinkhorn, Richard (1964), "A Relationship between Arbitrary Positive Matrices and Doubly Stochastic Matrices", *The Annals of Mathematical Statistics*, 35, pp. 876–879.

Sinkhorn, Richard (1967), "Diagonal Equivalence to Matrices with Prescribed Row and Column Sums", *The American Mathematical Monthly*, 74, pp. 402–405.

Smith, John H. (1947), "Estimation of Linear Functions of Cell Proportions", *Annals of Mathematical Statistics*, 13, pp. 166–178.

Steiner, Peter (2004), *Anwendungen von Konzepten grosser Abweichungen auf Fragestellungen der Einkommensdynamik*, Berlin.

Tauchen, George (1986), "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions", *Economics Letters*, 20, pp. 177–181.

White, Halbert (1994), *Estimation, Inference and Specification Analysis*, Cambridge.

SUMMARY

Based on a relative entropy approach, this paper proposes a method to estimate or update transition matrices using just cross-sectional observations at two points in time. The method is then applied to explain the development of the US income distribution. Starting from three hypothesized transition matrices and a transition matrix estimated from the PSID data, we show how these matrices must be adjusted in the light of the cross-sectional information. Finally, we explore the consequences of these updated transition matrices for the future development of the US income distribution.