SOEP Das Sozio-oekonomische Panel

427 2011

SOEPpapers
on Multidisciplinary Panel Data Research

SOEP — The German Socio-Economic Panel Study at DIW Berlin 427-2011

Asymptotic income distribution in the German Socio-Economic Panel (SOEP): Income inequality and Darwinian fitness

Diego Montano

DIW BERLIN

SOEPpapers on Multidisciplinary Panel Data Research
at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

The SOEPpapers are available at
http://www.diw.de/soeppapers

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | soeppapers@diw.de

Asymptotic income distribution in the German Socio-Economic Panel (SOEP)

Income inequality and Darwinian fitness

Diego Montano*

Abstract

In this paper, the relation between income inequality and population growth is analized from a Darwinian perspective. A Markov chain population growth model is presented and estimated using data from the German Socio-Economic Panel (SOEP). We estimate both population growth rates and steady-state income distribution for males and females. The results are compatible with the traditional age-based population growth models of demography, in so far as these are actually irreducible, positive-recurrent Markov chains. It is concluded that, from the perspective of individuals and family lines, income inequality may improve reproductive fitness of high-income individuals, and foster adaptive reproductive strategies for a given income level.
Keywords: Steady-state income distribution – Intrinsic growth rates – Markov chains – Population projection – Reproductive fitness

1 Introduction

The motivation of this paper is to analyze the relation between income inequality and population growth in a modern industrialized country from a Darwinian perspective. Notice that, since the times of Malthus, theoretical models of the relation between economic ressources and population dynamics have presupposed (maybe implicitly) a dependence structure given by a certain functional form, for instance, a linear or nonlinear dependence (e.g. Day et al 1989). The task of such theoretical explanation models is to estimate fertility rates given a certain amount of economic ressources. However, at least since Easterlin's work on the role of cohort size for fertility decisions, it has been clear that the relation between economic ressources and fertility is not absolute, but depends also on the population and income distribution (see Macunovich 1998 for a review). In other words, there is also a reference dependence that modulates the relation between economic

*Universität Düsseldorf, Philosophische Fakultät. Email: diego.montano@uni-duesseldorf.de

ressource and fertility, so that it is not unusual to find empirical evidence of positive as well as negative correlations between economic ressources and fertility (e.g. Myrskylä et al 2009). The most remarkable examples are probably the demographic transitions that have ocurred since the late 19th century in different regions of the world. In this paper, it is argued that distributional inequality is a fundamental element for the interpretation of the relation between economic ressources and population dynamics. Moreover, it is argued that Darwinian fitness maximization, i.e. the reproductive success of individuals in a given reference population (Stearns and Hoekstra 2005), can be thought of as an appropriate theoretical model for a partial explanation of the mechanism behind the persistent income inequality throughout human societies. In order to show this, it would be necessary to assess the reproductive behavior and success of family lines of a given economic ressource level across several generations. For historical populations there is evidence of Darwinian fitness maximization (e.g. Clark and Hamilton 2006, Josephson 1993, Voland 1990). However, fitness estimation of current populations must be based on asymptotic properties of population growth (Stearns 1992). The usual fitness measure is the so called intrinsic population growth rate r, which is calculated usually with classical demographic methods based on the age distribution of the population (see Keyfitz 1977). In order to investigate the relation between economic ressources and population growth, it is necessary to classify the population in terms of the distribution of certain socio-economic characteristics of individuals (see particularly Chu (1998) for a complete systematic approach, and Cheng and Chu (1999) for an empirical study with Taiwan data). In this paper, the income distribution of the population proxies its ressource distribution, and, at the same time, serves as basis for the calculation of the intrinsic growth rates r. From a stochastic perspective, this is possible because population growth is modeled on the assumptions of the Markov chain theory. Even though the Perron-Frobenius theory of nonnegative matrices gives the necessary and sufficient conditions to estimate population projection matrices and the intrinsic growth rates r from available demographic data, it can be shown that the validity of the Perron-Frobenius theory in classical demography is actually derived from the assumptions of the Markov-chain theory (for mathematical proofs see Meyer 2000). Thus, a Markov-chain estimation of population growth should be equivalent to the known demographic estimation methods of population projection. As a consequence, it is possible also to estimate fitness payoffs by evaluating the intrisic growth rates and the asymptotic income distribution of the population.

This paper is divided as follows. In section 2, the Markov chain population growth model is presented, and its assumptions and plausibility are discussed. In sections 3 and 4, the methods of estimation, the data and results are presented, respectively. In section 5 some conclusions are discussed.

2 Assumptions of a Markov chain population growth model

It is a well-established result that the steady-state age distribution of a population closed to migration is given by the Perron vector π satisfying the equation (Meyer 2000: 684; Keyfitz and Caswell 2005

$$L\pi = r\pi, \quad \lim_{t\to\infty} p(t) = \pi > 0, \quad ||\pi||_1 = 1 \tag{1}$$

L represents the Leslie matrix, r the intrinsic population growth rate, and $p(t)$ the age-group distribution at time t. The relation between the intrinsic growth rate r and the greatest eigenvalue λ_1 of L is given by:

$$\lambda_1 = e^{r\tau} \tag{2}$$

It should be remarked here that the projection matrix L is composed of the survival probabilities and fertility rates of a given population between times t and $t-1$. It is assumed also that mortality and fertility rates remain constant, so that (asymptotically) the proportion of individuals of a given population in each age group is the result of a certain mortality and fertility regime acting for a sufficient long period of time. The existance of an asymptotic age distribution π is secured by the theorems of the Perron-Frobenius theory of nonnegative matrices. Although the assumptions underlying equation (1) are unrealistic, they can help us evaluate the long-run impact of certain mortality and fertility rates on the age distribution of the population. Thus, in the analysis of the steady-state age distribution in classical demography, we are not interested in predicting population size in the future, but rather in assessing the long-term effects of period-specific intrinsic population growth rates on the age distribution.

However, the steady-state age distribution calculated in classical demography relies actually on the asumptions of the Markov theory, even though the asymptotic properties of Markov chains are also more easily analyzed within the theoretical framework of the Perron-Frobenius theory of nonnegative matrices. In order to make this clear, we define a Markov chain as follows (Bhattacharya and Waymire 1990, Bhat 1984). Let N_t be a random variable observed at time $t > 0$. Let S be a discrete set of n real numbers and let N_t take values only in the set S. The sequence $\{N_t : t \geq 0\} = \{N_t\}$ is called a stochastic process with index set $t \in [0, \infty)$, and denotes a collection of random variables on a probability space $(\Omega, \mathcal{F}, P)$. The set S is the so called state space, and denotes here a classification scheme of individuals in population $\mathcal{N}$ (i.e. every N_t has the same range $\{S_1, \ldots, S_n\}$). We will think of N_t as a classification variable, which assigns each individual of population $\mathcal{N}$ to a certain class S_i. In classical demography, the state space consists usually of age intervals, and N_t is a random variable that counts the number of individuals in each age group S_i. If we look at the age distribution F of population $\mathcal{N}$ at time t_n, it is clear that the number of individuals in each state depends on the frequencies of individuals in each state, and on period-specific fertility and mortalitiy rates at previous times t_k, for $k < n$. Consider the time set $(t_1, \ldots, t_n)$. The joint distribution $P(N_1, \dots, N_n)$ of the process $\{N_t\}$ at these time points cannot be simplified by assuming independent observations. Instead of calculating the joint probability of the stochastic process, it is assumed that we can describe the process by defining the (discrete) conditional transition distribution function P between times $n \geq m$ as

$$P_{ij}^{(m,n)} = P(N_n = S_j | N_m = S_i) \tag{3}$$

However, it is often impractical (or even impossible) to calculate the conditional distribution for large time sets. Instead, as in the case of the Leslie matrix, it is assumed that the transitional probabilities $P_{ij}^{(m,n)}$ of the stochastic process $\{N_t\}$ depend only on the last time point $t-1$ for every t (stationarity assumption). A stationary Markov chain is a stochastic process that satisfies the following Markov assumption of the transition probabilities $P(N_n = S_j | N_m = S_i)$ for all t

$$p_{ij} \equiv P(N_{t+1} = S_j | N_t = S_{it}, \dots, N_0 = S_{i0}) = P(N_{t+1} = S_j | N_t = S_{it}) \tag{4}$$

The corresponding matrix $\boldsymbol{P}_{n \times n}$ of transition probabilities p_{ij} is nonnegative. The rows of $\boldsymbol{P}$ sum up obviously to 1. The matrix $\boldsymbol{P}$ is called a stochastic matrix. We will call the vector $\boldsymbol{p}^T = (p_1, \dots, p_n), \sum_i p_{ij} = 1, i = 1, \dots, n$, the probability distribution vector of population $\mathcal{N}$ (Meyer 2000: 688). In other words, $\boldsymbol{p}$ gives the conditional probabilities of moving to state j given states i. However, the convergence of $\boldsymbol{p}$ to a limiting distribution vector π is not given for every Markov chain. In the case of the Leslie matrix, there exists such a limiting distribution vector because L is the transition matrix of an irreducible positive-recurrent Markov chain. Because of this property, the Leslie matrix is primitive (Meyer 2000: 693) and, according to the Perron-Frobenius theory, it has a positive eigenvalue λ_1 that is greater than any other eigenvalue (Minc 1988: 47)[1]. Therefore, we have that every (primitive) Leslie matrix is a Markov chain, and likewise, irreducible positive-recurrent Markov chains can serve as population growth models. Nontheless, because we can easily identify irreducible positive-recurrent Markov chains by looking at its corresponding graph, we will not discuss the (somewhat difficult) exact definition of irreducible positive-recurrent Markov chains (see details in Bhattacharya and Waymire 1990: chap. 2, Bhat 1984: chap. 3 and 4; Guttorp 1995). Irreducible and positive-recurrent chains can be thought of as processes, in which individuals can move freely from one state to the other with a certain nonzero probability. Because individuals are not "trapped" in a specific state (i.e. there is no absorbing state), we can expect for $t \to \infty$ a stable distribution of individuals across all states given some stationary transition probabilities.

2.1 The model

The Markov chain model presented in this paper is based on the directed life cycle graph depicted in figure 1. Our state space consists of five interconnected states (or nodes), which in this case represent income classes (see Keyfitz and Caswell 2005 for details on the

[1] Actually, the Leslie matrix is the result of ignoring those age groups with (practically) zero fertility. Otherwise, the corresponding Markov chain would not be an irreducible positive-recurrent chain.

theory of graphs in demography). Income mobility is possible only between generations, i.e. income mobility is given as the difference between the income classes of parents and children. However, children can move from their initial income class to any other class of the chain with transition probability p_{ij}. At each state, individuals can reproduce with state-specific reproduction rates $F_i, i = 1, \dots 5$. From figure 1, it is clear that for each pair of nodes there is a sequence of directed ledges leading from one node to the other. This property is called strong connectivity. Moreover, because the greatest common divisor of the set of all lenghts from one node to the other is one, the Markov chain is irreducible and positive recurrent, and the distribution vector $\boldsymbol{p}$ converges to the stationary distribution vector π of equation (1) (Berman 1984; Heller et al 1978: 72-74, 111). If we let $\boldsymbol{P}$ denote the transition matrix, whose elements are the transition probabilities p_{ij}, we can estimate the limiting income distribution vector π by solving the equation system:

$$(I - \boldsymbol{P}^T)\pi = 0 \qquad \text{and} \qquad \sum_{i}^{J} \pi_i = 1. \tag{5}$$

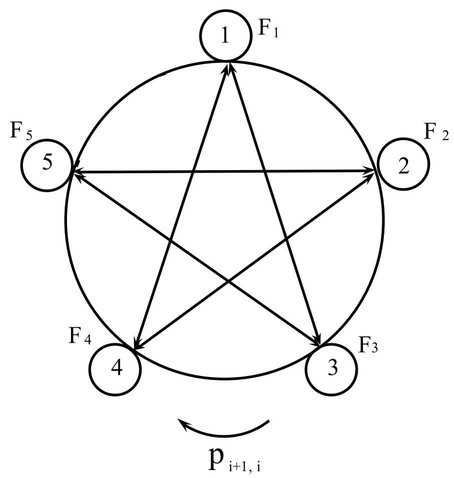

Figure 1: Life cycle graph of the population growth model based on income distribution.

This population growth model is equivalent to the Markov chain models proposed in the studies of occupational mobility, and has thus some important assumptions (and therefore weaknesses). According to Fararo, we have the following assumptions (Fararo 1978: chap. 16):

- *Assumption 1*: The population forms a set of family lines, so that only one son or one daughter of a given father or mother is consider in tracing the line.

- *Assumption 2*: The intergenerational transition probabilities from one income class to another satisfies the Markov assumption with time steps of 24 years, or one generation. Moreover, the corresponding Markov chain is stationary. Note here that this assumption actually means that the period-specific transition probabilities would continue to be valid for the rest of the process time of the Markov chain. Because real populations are not expected to have such stable income transition rates, this assumption is, as we said above, the basis for a ceteris paribus population projection, just as the intrinsic growth rates of the age-based demographic model assume that some constant fertility and mortality rates act for a very long period of time (Keyfitz and Caswell 2005: 128 ff.)

- *Assumption 3*: Families trace out their income position paths independently.

- *Assumption 4*: Families remain on the same income position between generations, and contribute new individuals only to their own income group, i.e. income mobility is possible only in discrete time steps of one generation.

With these assumptions and figure 1, we can define the following population projection matrix $\boldsymbol{P}$, which is equivalent to the Leslie matrix L in equation (1):

$$\boldsymbol{P} = \begin{pmatrix} p_{11}F_1 & p_{12} & \cdots & p_{15} \\ p_{21} & p_{22}F_2 & \cdots & p_{25} \\ \vdots & & & \vdots \\ p_{51} & \cdots & & p_{55}F_5 \end{pmatrix} \tag{6}$$

Thus, by estimating the projection matrix $\boldsymbol{P}$, it is possible to estimate the limiting distribution vector π in equation (5), and the intrinsic growth rates given in equation (2).

3 Data and methods

We estimated the transition probabilities p_{ij} with data from the German Socio-Economic Panel (SOEP) (see Wagner et al 2007 for a general description of SOEP). The SOEP is an ongoing representative panel survey of private households in the Federal Republic of Germany. It started in 1984, and includes several socio-economic variables at the household and individual level. Every person of a household is followed up, once he/she becomes 16 years old, even if children of respondents leave the parental household. Therefore, it is possible to reconstruct family relationships across generations. Given the model of section 2.1, the stochastic matrices of transition probabilities were calculated from the West German SOEP samples by using the household pre-government income in euro adjusted with the price index of 2006. The method was the following. The weighted mean for males and women was estimated for each year, and each person was classified to a certain relative income position according to the classification scheme given in table 1 (see Grabka and Krause 2008; Eurostat 2010 for more details). This was done separatedly for men and women. Thus, a new variable was defined that indicated the income class of a person in relation to the mean income of the whole male or female population for each year, respectively. The corresponding transition matrix, which gives the probability of transitions between income classes between two generations, is called mean transition matrix (Formby et al 2003: 190). Even though the household equivalized median income is the official measurement instrument to assess income distribution (and income poverty), the pre-government income before taxation was utilized here as a measurement of household economic output. In this way, it was intented to isolate the effects of governmental intervention on the population growth rates, and on the asymptotic income distribution. Note that the income classes do not have the same size, in so far as it was intended to estimate the inequality of groups based only on their economic output.

Table 1: Relative income position - Classification scheme.

X = Income, $\bar{X}$ = Mean income	
Income class	**Classification intervals**
Lower - Position 1	$X \leq 0.6\bar{X}$
Low - Position 2	$0.6\bar{X} < X \leq 0.8\bar{X}$
Middle - Position 3	$0.8\bar{X} < X \leq \bar{X}$
High - Position 4	$\bar{X} < X \leq 1.5\bar{X}$
Upper - Position 5	$1.5\bar{X} \leq X$

If we let Y be the random variable for the income position of the father or mother, and X the random variable for the income position of the son or daughter, we have to estimate the conditional probabilities $P(X = i|Y = j)$. In order to estimate these transition probabilities from and to income positions between generations, it was necessary to identify family lines and cohorts of parents and children, and to build the correspoding two dimensional contingency tables. Although SOEP offers sample weights at the individual and household level, which are based on selection and staying probabilies, and on the marginal distribution of some socio-economic traits of the population (Rendtel 1995; Pannenberg et al 2005), it was not clear how such sample weights would apply for family lines, which are actually the observational units. The task was to build a weighted contigency table of the income position of children conditioned on the income position of the parents. Sampling weights were needed in order to reduce the standard errors of the transition probabilities, since the sample sizes of family lines were quite small (n = 427, n = 289 for men and women, respectively). Nevertheless, because the sampling weights of children are not the same as those for parents, i.e. the contingency table cannot be extrapolated to the marginal distribution of the population by only one set of sampling weights, say those of children or of parents alone, it was necessary to estimate the transition probabilites of the stochastic matrix by using the discrete formulation of Bayes theorem. The conditional probabilites $P(Y = j|X = i)$, i.e. the probability of a father or mother of income position j given a son or a daughter of income position i, were estimated with the sample weights of the parents. The prior probabilites $P(X = i)$, i.e. the distribution of the relative income position of the children generation, were estimated using the sample weights of the children. Thus, we could estimate the transition probabilities $P(X=i|Y=j)$, as if we were using indirectly some sort of sample weights for the family lines (for more estimation procedures, and for a comparison of transition matrices estimates for Germany see Formby et al 2003; and for a general overview Bartholomew 1973: chap. 2).

Addtionally, in order to estimate the conditional probabilies $P(Y=j|X=i)$ with the sampling weights of the parents, it was necessary to assume that children and parents form only two different generations. This assumption, as discussed by T. Fararo (Fararo 1978: 445), can have serious consequences for the estimation of the transition probabilities, in so far as the birth years of children and parents scatter over the time interval chosen to define the generations. This scattering implies that the associated Markov chain steps do not coordinate to a single step for all family lines, if the system is not in equilibrium. Moreover, because of sample attrition in panel studies, the reconstruction of family lines in SOEP was limited regarding sample size. In order to overcome to some extent these limitations, the choice of "generations" was based on two criteria: first, by finding a sample year, in which equilibrium conditions could be assumed, and second, by choosing the greatest sample. We chose the cohorts 1922-1950 for parents, and the cohorts 1960-1975 for children. All cohorts were observed in the year 2000. Only the relative income position of fathers were considered here, because for those children born between 1960 and 1975 the economic output of fathers is a better proxy for the household economic output in West Germany. This was due to the fact that fathers had a more active participation in the labor market as mothers (see e.g. Mayer and Hillmert 2004). The year 2000 was chosen as observation time, because we could define for that year a relative big sample, for which it was plausible to assume that both fathers (aged 50-78) and children (aged 25-40) had attained their "final" relative income position. Thus, we intended to reduce coherence problems at each time step in the proposed Markov chain. However, it should be noticed here that, in particular for children aged 25-30, there might be some individuals for whom the equilibrium assumption may not be yet valid.

4 Results

Tables 2 and 3 reproduce the estimates of the stochastic matrices of intergenerational income mobility for men and women in the year 2000 using Bayes theorem. The limiting distribution vector is given in the last row of each transition matrix, and it was calculated by equation (5). Moreover, if we denote with T the number or generations spent in income class i, we can calculate the average number of generations $\hat{E}(T)$, and its estimated variance $\hat{\sigma}^2(T)$ using equations (7) derived by S. J. Prais (Prais 1955). The results are reproduced in table 4.

$$\hat{E}(T) = 1/(1-p_{ii}), \qquad \hat{\sigma}^2(T) = p_{ii}/(1-p_{ii}) \tag{7}$$

If we compare the estimates of the limiting distribution in tables 2 and 3 with the standard deviations of the average number of generations in each income class in table 4, we can identify gender-specific income characteristics. Men seem to show a greater upward social mobility as women, and a longer average mean time in each income class. From a socio-economic point of view, women's income distribution seems to have a greater variance and a stronger income polarization given as a tendency toward the highest or the lowest relative income positions (columns 2 to 4 in table 3). This results in a greater average time spent in income classes 2 and 5. Income polarization for women can be expected already by comparing the transition probabilities of men and women between income positions 1 and 5. In general, women of the children cohorts seem to have a greater probability of both staying in the lowest income position, or of reaching the highest income position. On the other hand, men seem to have been moving upwards rather homogenously from the lowest tail of the distribution toward the higher relative income positions. Thus, their average number of generations seems to follow a more or less monoton increase with each successive income position. Family lines of men can expect to remain in the highest income position for about 3.2 generations, the longest time spent in any income class (table 2).

Table 2: Transition matrix of intergenerational income mobility and asymptotic income class distribution vector. Men. Source: SOEP, own calculations.

Fathers / Sons	Position 1	Position 2	Position 3	Position 4	Position 5
P(X = k |Y = 1)	0.3140	0.0812	0.1569	0.2659	0.1820
P(X = k |Y = 2)	0.0379	0.6378	0.1279	0.0790	0.1174
P(X = k |Y = 3)	0.0778	0.1068	0.6462	0.1003	0.0689
P(X = k |Y = 4)	0.1464	0.0684	0.0282	0.6070	0.1501
P(X = k |Y = 5)	0.0940	0.0823	0.0411	0.0884	0.6941
Asymptotic distribution	0.12	0.19	0.17	0.23	0.29

Table 3: Transition matrix of intergenerational income mobility and asymptotic income class distribution vector. Women. Source: SOEP, own calculations.

Fathers / Daughters	Position 1	Position 2	Position 3	Position 4	Position 5
P(X = k |Y = 1)	0.2405	0.1734	0.0967	0.2149	0.2745
P(X = k |Y = 2)	0.2105	0.4826	0.0549	0.0511	0.2010
P(X = k |Y = 3)	0.2135	0.0802	0.2710	0.3143	0.1211
P(X = k |Y = 4)	0.1198	0.0460	0.1566	0.4165	0.2611
P(X = k |Y = 5)	0.1802	0.0514	0.0449	0.2074	0.5162
Asymptotic distribution	0.18	0.13	0.11	0.25	0.32

Even though the estimation of the asymptotic distribution of income positions was based on weighted contingency tables, it was necessary to compare changes in the household income distribution for male and female cohorts 1960-1975 with the male cohorts

Table 4: Average number of generations T spent in a specific income class $\hat{E}(T)$, and its estimated standard error $\hat{\sigma}(T)$. Source: SOEP, own calculations.

	Men		Women	
Density	E(T)	Std(T)	E(T)	Std(T)
P(X = k |Y = 1)	1.45	0.81	1.31	0.64
P(X = k |Y = 2)	2.76	2.20	1.93	1.34
P(X = k |Y = 3)	2.82	2.27	1.37	0.71
P(X = k |Y = 4)	2.54	1.98	1.71	1.10
P(X = k |Y = 5)	3.26	2.72	2.06	1.48

1922-1950 for the whole West German SOEP samples. The idea behind this comparison was to assess, whether the intergenerational transition probabilities and their asymptotic distribution agreed to a more general distribution shift for those cohorts in the whole sample, or whether they were only related to those specific family lines. It can be expected that the transition probabilites of familiy lines give not only information on the dependence structure of intergenerational income class mobility, but also information on cohort effects affecting the whole population. These effects might be caused to some extent by macro-structural changes affecting the socio-economic conditions of West German households (e.g. labor-market shifts, educational expansion, womens' empowerment, and so on). For this reason, we estimated the weighted relative income distributions as described in Handcock and Morris (1999) for male cohorts 1922-1950 in the year 1984 in relation to the income distribution of male and female cohorts 1960-1975 in the year 2008[2].

The results are depicted in figures 2 and 3. The curves describe the relative income distribution of two cohort groups across income deciles. The reference group in figure 2 are cohorts of men born between 1922 and 1950. The comparison groups are males (continuous line) and women (dotted line) born between 1960 and 1975. The upper horizontal axis gives the original values in euro corresponding to the income deciles of the abscissa. On the other side, the reference group in figure 3 are cohorts of men born between 1960 and 1975. The comparison group consists of women born in the same period. The curves in each figure become concave, if there are more observations of the reference group in a given income decile. Conversely, if there are more observations of the comparison group in a given decile, the curves become convex. At the intersection of the curve of the relative income distribution with the horizontal line at $y = 1$, we get the same number of observations in each cohort for a given income decile. As expected from the estimated transition probabilities, males of cohorts 1960-1975 in figure 2 show rather a continuous upward mobility trend. Males of the "children" generation are almost twice as likely as men of the "fathers" generation of being in the ninth decile of the household income distribution, and, at the same time, they seem less likely to be in the lower deciles.

[2]For each of the samples of male cohorts 1922-1950 and 1960-1975 one outlier was simply removed.

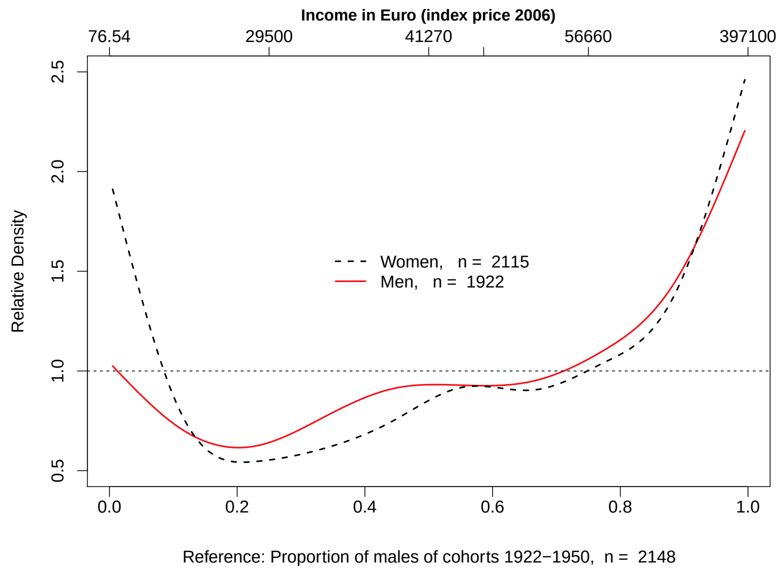

Figure 2: Weighted relative income density of male cohorts 1922-1950 in comparison with male and female cohorts 1960-1975 for each income decile. Source: SOEP, own calculations.

On the other side, the relative distribution of female cohorts in comparison to the males of cohorts 1922-1950 in figure 2 suggest that there is a stronger income polarization for females. They are overrepresented in the first and ninth decile of the income distribution in comparison with the "fathers" generation, even though they are also less likely to be in the intervall between the second and seventh income deciles. In figure 3 the relative income distribution between male and female children of cohorts 1960-1975 is plotted. The U-shape of the relative distribution confirms the relative stronger income polarization of women, who are not only almost twice as likely as men of being in the first income decile, but also they are overrepresented in the ninth decile. Therefore, these results suggest that men of cohorts 1960-1975 have been moving rather uniformly toward higher household income positions, while women of the same cohorts have been moving toward the opposite extremes of the income distribution. This results in a stronger income polarization for females. In order to analyze the relation between the income distribution of those same cohorts and population growth, the income model in figure 1 will be estimated using fertility data of SOEP.

This model was calculated using the transition probabilities given in tables 2 and 3. The fertility rates of each income position were calculated for men and women in the years 1984 and 2008. The samples were based on West Germans aged 20 to 45. Following the estimates of Kroll and Lampert (2008), income specific mortality risks were taken into account for both men and women. Because the projection matrix in equation (6) satisfies equation (1) as well, we could estimate the intrinsic growth rates for the period 1984-2008 by applying equation (2), and setting $\tau = 24$. For that purpose, we multiplied the main diagonal of the each transition matrix with the income-specific fertility rates and

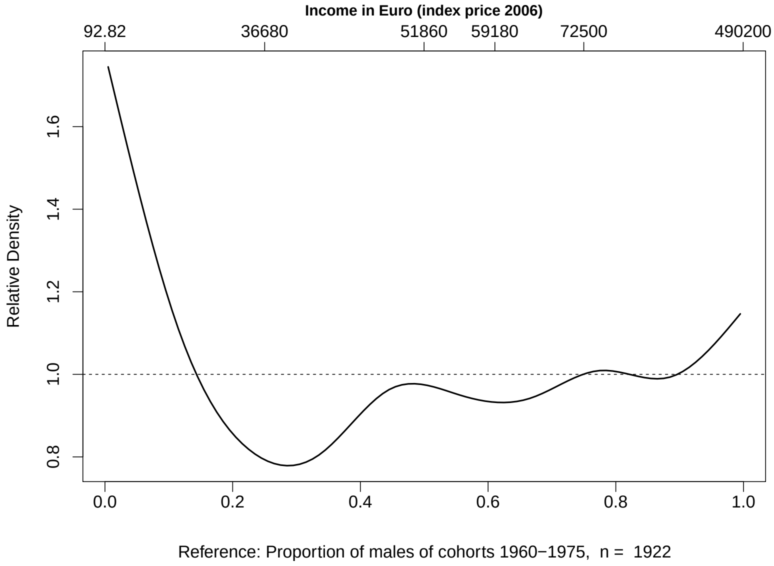

Figure 3: Weighted relative income density of male cohorts 1960-1975 in comparison with female cohorts 1960-1975 for each income decile. Source: SOEP, own calculations.

mortality risks reproduced in tables 6 and 7. The number of children under age 18 in the household was used as a proxy for fertility for both men and women. This variable seemed to be the adequate proxy for cross-gender comparative fertility analysis, because information on male fertility in SOEP is available only since 2001. Under the assumption of stationarity of the corresponding Markov chain, the same transition probabilities were used for estimating the impact of fertility changes on the intrisic growth rate r for every income position in the given years. The estimation of the limiting distribution vector π_0 given the initial income distributions in 1984 and 2008 can be easily calculated by the formula (Keyfitz and Caswell 2005: 161)

$$\pi_0 = \lim_{t \to \infty} \frac{p(t)}{\lambda_1^t} = w^T p_0 w \tag{8}$$

where w is the eigenvector of unit length corresponding to the dominant eigenvalue λ_1 of the projection matrix $\boldsymbol{P}$, and p_0 is the initial income class distribution vector. In order to assess the validity of r calculated by the income model, we calculated also the intrinsic growth rates by the traditional methods of the age-based population growth model. We used the abriged life tables of the years 1980 and 2008 supplied by the German Statistical Bureau (Statistisches Bundesamt 2009). The estimation methods utilized are fully described in Keyfitz (1977). Table 5 reproduces both estimates of the intrinsic growth rates of the population aged 20 to 45 in West Germany between 1984 and 2008.

The estimates of the age model can be considered very close estimates of the true parameters, in so far as they are based on a sample, which is approximately three orders of magnitude greater than the SOEP samples. Nevertheless, the estimates of the intrinsic

Table 5: Intrinsic growth rates from the income-based and age-based population growth models. λ_1 is the greatest eigenvalue of the projection matrix $\boldsymbol{P}$ of the income model. Source: SOEP, own calculations.

Year	λ_1	Income model r	Age model r
		Men	
1984	0.5957	-0.0216	-0.0120
2008	0.5791	-0.0228	-0.0121
		Women	
1984	0.7621	-0.0113	-0.0132
2008	0.7474	-0.0121	-0.0135

growth rates based on the SOEP samples are of the same order of magnitude and sign as those of the age-based model. It should be noticed that, specially for men, there seems to be a much greater estimation error, which might be caused by the use of the number of children in the household as a proxy for fertility. This seems plausible, since children in West Germany remain almost always in the household of the mother after separation or divorce of the parents. For all single, divorced and separated fathers this would imply an underestimation of fertility, which in turn causes a greater decline in the estimated intrinsic growth rates in the male model. In any case, it is clear that both sexes have had since 1984 a negative growth rate.

However, the estimates of the asymptotic distribution calculated by equation (8) in tables 6 and 7 suggest that income positions 1 and 4 are decreasing faster, both for men and women, while the relative proportions of the income classes 2, 3 and 5 increase with and without the assumption of differential mortality risks. This increase of the asymptotic proportion of the highest income class is at first sight surprising, if it is taken into account that both its fertility rates and its observed initial proportions are lowest (with the exception of the fertility rates of income position 1). Even if it is assumed that there are no statistically significant differential mortality risks among income classes[3], there is, in general, a tendency of increase in the asymptotic proportion of the highest income class for both sexes (see tables 6 and 7). The increase of income positions 2, 3 and 5 would imply, therefore, that the negative intrinsic growth rates would reflect rather (i) the decrease of income positions 1 and 4, (ii) the upward mobility reflected in the transition probabilities discussed above, and (iii) a higher reproductive payoff of income positions 2 and 3. As can be inferred from the life cycle graph in figure 1, higher transition probabilities toward higher income classes would increase the proportion of these classes under the Markov assumption given any initial distribution, because individuals from generation to generation would have a greater probability of arriving at higher income classes, and, eventually, of arriving at income position 5. Accordingly, this would cause

[3]However, the assumption that income classes show statistically significant mortality risks (and in general overall health differences) has been supported by various empirical findings, see e.g. Scholz and Thoelke 2002; Lampert and Kurth 2007; Breyer and Marcus 2010.

a shift of the asymptotic distribution or, equivalently, an increase of the asymptotic proportion of the highest income class and a decrease of the lowest. The increase of income positions 2 and 3, on the other hand, relies heavily on their higher fertility, as can be inferred from the estimates of the corresponding tables, in so far as fertility may increase the "surviving" probabilities of those income positions.

Table 6: Income class fertility rates, observed income class proportion, and limiting distribution vector π of the income-based population growth model for men. Relative mortality risks taken from Kroll and Lampert (2008). West Germany. Source: SOEP, own calculations.

			Men		
	Position 1	Position 2	Position 3	Position 4	Position 5
Relative mortality risk	1.61	1.19	1.01	0.84	0.60
			1984		
Fertility	0.36	0.54	0.50	0.47	0.41
Observed proportion	0.22	0.19	0.18	0.27	0.14
π with risk	0.24	0.17	0.18	0.21	0.20
π without risk	0.23	0.21	0.20	0.20	0.16
Eigenvector	0.54	0.38	0.39	0.47	0.44
			2008		
Fertility	0.27	0.45	0.47	0.43	0.38
Observed proportion	0.28	0.15	0.15	0.26	0.16
π with risk	0.23	0.21	0.20	0.20	0.16
π without risk	0.23	0.19	0.20	0.20	0.17
Eigenvector	0.51	0.46	0.44	0.44	0.36

5 Concluding remarks

The estimates of the limiting distribution vector π suggest that, on the long-run, there is a strong income mobility toward higher income positions. However, the average number of generations spent in each income class increases across income classes for both men and women. The estimates of the population growth rates, and of the limiting distribution vector given income-specific fertility rates suggest, on the other hand, that higher fertility rates of the lower income classes can increase the probability that the next generations can attain a higher income class. Thus, the longer "waiting times" of upward social mobility can be compensated with higher fertility rates[4]. On the contrary, lower fertility rates for the higher income family lines seem to increase their fitness. These results point to different reproductive strategies for lower and higher income classes that, on the long run, may increase their chances of reproductive success given certain socio-economic conditions of social mobility and income inequality. In this sense, each family line of a

[4]Notice also that lower-income classes can also reduce fertility if social mobility rates compensate to some extent the longer waiting times of upward social mobility (see e.g. Caldwell et al 1999).

Table 7: Income class fertility rates, observed income class proportion, and limiting distribution vector π of the income-based population growth model for women. Relative mortality risks taken from Kroll and Lampert (2008). West Germany. Source: SOEP, own calculations.

			Women		
	Position 1	Position 2	Position 3	Position 4	Position 5
Relative mortality risk	1.57	0.94	0.92	0.72	0.65
			1984		
Fertility	0.44	0.57	0.58	0.49	0.46
Observed proportion	0.23	0.18	0.18	0.26	0.15
π with risk	0.21	0.18	0.22	0.20	0.19
π without risk	0.21	0.19	0.22	0.19	0.18
Eigenvector	0.46	0.40	0.50	0.45	0.42
			2008		
Fertility	0.37	0.48	0.48	0.48	0.42
Observed proportion	0.31	0.14	0.14	0.23	0.17
π with risk	0.21	0.18	0.22	0.21	0.19
π without risk	0.21	0.19	0.22	0.20	0.18
Eigenvector	0.46	0.39	0.50	0.46	0.42

given income class would try to enhance their Darwinian fitness by adapting to certain social mobility and income distribution conditions. The reproductive success of lower or higher fertility would depend on the relative socio-economic and mortality risks of individuals, which eventually lead to different, but maybe (asymptotically) equivalent reproductive strategies.

References

Bartholomew D (1973) Stochastic models for social processes. Wiley, London

Berman A (1984) Convexity, graph theory and non-negative matrices. Annals of Discrete Mathematics 20:55–59

Bhat N (1984) Elements of applied stochastic processes, 2nd edn. John Wiley, New York

Bhattacharya R, Waymire E (1990) Stochastic processes with applications. John Wiley, New York

Breyer F, Marcus J (2010) Income and longevity revisited: Do high-earning women live longer? SOEP-Discussion Papers 1037:1–25

Caldwell J, Khuda B, Caldwell B, Pieris I, Caldwell P (1999) The bangladesh fertility decline: An interpretation. Population and Development Review 25(1):67–84

Cheng R, Chu C (1999) Estimating the intergenerational income mobility matrix using pseudo panel data. In: Ranis G, Hu SC, Chu YP (eds) The political economy of Taiwan's development into the 21st century, vol 2, Edward Elgar Publishing, Southhampton, Mass., pp 227–245

Chu C (1998) Population dynamics. Oxford: Oxford University Press

Clark G, Hamilton G (2006) Survival of the richest: The malthusian mechanism in preindustrial england. Journal of Economic History 66(3):707–737

Day R, Kim KH, Macunovich D (1989) Complex demoeconomic dynamics. Journal of Population Economics 2:139–159

Eurostat (2010) Income poverty and material deprivation in european countries. DOI 10.2785/1144, URL http://epp.eurostat.ec.europa.eu

Fararo T (1978) Mathematical sociology. Robert Krieger Publishing Company, New York

Formby J, Smith J, Zheng B (2003) Mobility measurement, transition matrices and statistical inference. Journal of Econometrics 120:181–205

Grabka M, Krause P (2008) Einkommen und Armut von Familien und älteren Menschen. Wochenbericht DIW-Berlin 9:155–165

Guttorp P (1995) Stochastic modeling of scientific data. London: Chapman & Hall

Handcock M, Morris M (1999) Relative distribution methods in the social sciences. New York: Springer

Heller WD, Lindenberg H, Nuske M, Schriever KH (1978) Stochastische Systeme. Berlin: Walter de Gruyter

Josephson S (1993) Status, reproductive success, and marrying polygynously. Ethology and Sociobiology 14:391–396

Keyfitz N (1977) Introduction to the mathematics of population. Cambridge, Mass.: Addison-Wesley

Keyfitz N, Caswell J (2005) Applied mathematical demography. Springer, New York

Kroll LE, Lampert T (2008) Soziale Unterschiede in der Lebenserwartung - Möglichkeiten auf Basis des Sozio-oekonomischen Panels. SOEP-Papers 112:1–36

Lampert T, Kurth BM (2007) Sozialer Status und Gesundheit von Kindern und Jugendlichen. Ergebnisse des Kinder- und Jugendgesundheitssurveys (KiGGS). Deutsches Ärzteblatt 104(43):A 2944–9

Macunovich D (1998) Fertility and the easterlin hypothesis: an assessment of the literature. Journal of Population Economics 11:53–111

Mayer KU, Hillmert S (2004) Neue Flexibilitäten oder blockierte Gesellschaft? Sozialstruktur und Lebensverläufe in Deutschland 1960-2000. In: Kecskes RWM (ed) Angewandte Soziologie, VS Verlag für Sozialwissenschaften, pp 129–158

Meyer C (2000) Matrix analysis and applied linear algebra. Philadelphia: SIAM

Minc H (1988) Nonnegative matrices. New York: Wiley

Myrskylä M, Kohler HP, Billari F (2009) Advances in development reverse fertility declines. Nature 460(6):741–743

Pannenberg M, Pischner R, Rendtel U, Spiess M, Wagner G (2005) Sampling and weighting. In: Haisken-DeNew J, Frick J (eds) Desktop companion to the German Socio-Economic Panel (SOEP), DIW-Berlin, chap 5, pp 153–186

Prais SJ (1955) Measuring social mobility. Journal of the Royal Statistical Society Series A (General) 118(1):56–66

Rendtel U (1995) Lebenslagen im Wandel: Panelausfälle und Panelrepresentativität. Sozio-ökonomische Daten und Analysen für die Bundesrepublik Deutschland. Band 8, Campus Verlag, Frankfurt am Main

Scholz R, Thoelke H (2002) Lebenserwartung in Berlin 1986 bis 1996 - Trends und regionale Unterschiede. In: Cromm J, Scholz R (eds) Regionale Sterblichkeit in Deutschland, Göttingen: WiSoMed, pp 65–100

Statistisches Bundesamt (2009) Periodensterbetafeln für Deutschland. Accessed on 01.05.2010, URL `www.destatis.de`

Stearns S (1992) The evolution of life histories. Oxford University Press, New York

Stearns S, Hoekstra R (2005) Evolution: An introduction. Oxford University Press, Oxford

Voland E (1990) Differential reproductive success within the krümmhorn population (germany 18th and 19th centuries). Behavioral Ecology and Sociobiology 26:65–72

Wagner G, Frick J, Schupp J (2007) The German Socio-Economic Panel Study (SOEP) - Scope, evolution and enhancements. Journal of Applied Social Science Studies 127(1):139–169