ON THE NON-STATIONARITY OF GERMAN INCOME MOBILITY*
(and some observations on poverty dynamics)

Christian Schluter
London School of Economics and Political Science

Discussion Paper
No.DARP/30
April 1997

The Toyota Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020-7955 6678

---

* I am grateful to Ramses Abul Naga, Tim Besley, Frank Cowell, Stephen Jenkins, and John Micklewright for comments on earlier drafts of this paper.

Abstract

The intra-distributional mobility of German income dynamics is analysed using GSOEP. Transition probabilities are found to be time-varying. The tested models comprise various mixed Markov chains in discrete time and a non-stationary mover-stayer model is proposed. In order to explain the observed mobility profiles, we concentrate on one important class – the poor – instead of the entire transition matrix. Various poverty duration models accommodating unobserved population heterogeneity and duration dependence are examined.

Keywords: Intra-distributional mobility, Markovian models, time-varying transition probabilities, poverty.

JEL Nos.: C41, D31, I32.

Contact address: Christian Schluter, STICERD, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, email: **c.schluter@lse.ac.uk**

1 Introduction

A useful taxonomy for the analysis of distributional dynamics is the explicit distinction between shape dynamics - referring to the changing external shape of the distribution - and intra-distributional mobility. These two dimensions are orthogonal, require different types of data, and correspond to different sets of economic questions. An examination of the shape dynamics is a purely cross-sectional exercise, which implies that the set of addressable economic issues is quite limited (such as income polarisation, cross-sectional inequality or the incidence of poverty).

However, the increasing availability of longitudinal datasets has given a strong impetus to policy-related research based on the analysis of income histories. For instance, the extent of intra-distributional mobility is important for the design of welfare programmes. As most canonical models of the income or earnings process such as permanent income and life-cycle models are indeed dynamic, so should welfare assessments be. Lifetime equity depends on the extent of movement up or down the distribution. The above taxonomy can be fruitfully applied here because the analysis of the shape dynamics is inadequate for the problem of (lifetime) welfare assessments[1]. This purely

[1]See Deaton (1992) for a critical review of the relation between income and consumption (and therefore welfare). Although the simplest life-cycle model predicts that "the temporal pattern of life-cycle income " does not determine "the evolution of consumption" so that "anticipated changes in consumption have no effect on consumption", empirical evidence reveals that, contrary to this prediction, "consumption tracks income closely over the life-cycle" (pp.

cross-sectional analysis cannot distinguish between such diametrically opposed worlds in which income positions are retained or permuted. It would rank the static economy exhibiting perfect persistence to possess the same level of welfare as the very mobile economy. But such a ranking would not conform to the value judgment of most people. Friedman (1962), for instance, considers "two societies that have the same distribution of annual income. In one there is great mobility and change so that the position of particular families in the income hierarchy varies widely from year to year. In the other, there is great rigidity so that each family stays in the same position year after year. Clearly, in any meaningful sense, the second would be the more unequal society" (p.171) - as the inequalities in permanent income are greater. A society may be better equipped to deal with short term fluctuations and a high degree of mobility - for instance consumption may be smoothed if credit and insurance markets are perfect- than with long term poverty and persistent social exclusion.

In this paper the German Socio-Economic Panel (GSOEP) is used to examine income mobility for the German case. Such an analysis is insightful since the German model is often offered as a counter-paradigm to Anglo-Saxon models. The US is characterised by a dynamic labour market, a small welfare state, and income large inequalities. By contrast, the German labour market is inflexible, the welfare state is large, and the resulting income inequality is low. According to common prejudice Germany can be caricatured as an economy is stasis. Below, we test the common prejudice that income mobility is

26, 161).

low. The characterisation of German income mobility is made in two stages. The first stage is descriptive and follows an established literature which represents social processes by Markov models [2] (and a new mover-stayer model is proposed). Instead of estimating a representative agent model, we model the evolution of the entire income distribution. But do such popular models give an adequate representation ? The second stage goes beyond mere description and attempts to explain the observed mobility.

This discussion of the "empirics for distributions" (Quah (1996)) is illustrated for the German case for the period 1983 to 1989 in Figures 1 to 3. Kernel density estimators (see Silverman (1986)) are a natural tool for such a preliminary and exploratory analysis. As Figure 1 demonstrates, the shape of the net income distribution has hardly changed in this period [3]. This almost unchanged cross-sectional inequality appears to confirm the common prejudice. Figure 2 depicts a typical example of an estimate of the joint distribution of incomes in two consecutive years. This distribution is unimodal and its

---

[2] See, for instance, Champernowne (1973), McCall (1971), the references in the eponymous paper by Singer and Spilerman (1974), Geweke, Marshall, and Zarkin (1986). Quah has applied the above taxonomy in a series of papers in the context of the debate about the (non)convergence amongst countries. See, for instance, Quah (1996).

[3] The changes in the shape of the distribution are minor when compared to the vast changes experienced in the UK (Cowell, Jenkins, and Litchfield (1996)) or the US (Burkhauser, Crews, Daly, and Jenkins (1995)) in the 1980s. In both cases, a dramatic polarisation has taken place; nearly unimodal shapes have turned into twin-peaks as the middle class occupies a sinking valley between them. As Jenkins (1995) observes: "The shift away from the middle class in both directions is strong evidence that the 'middle class' was shrunk, however one defines the middle." Fears of a shrinking German middle class are currently unfounded.

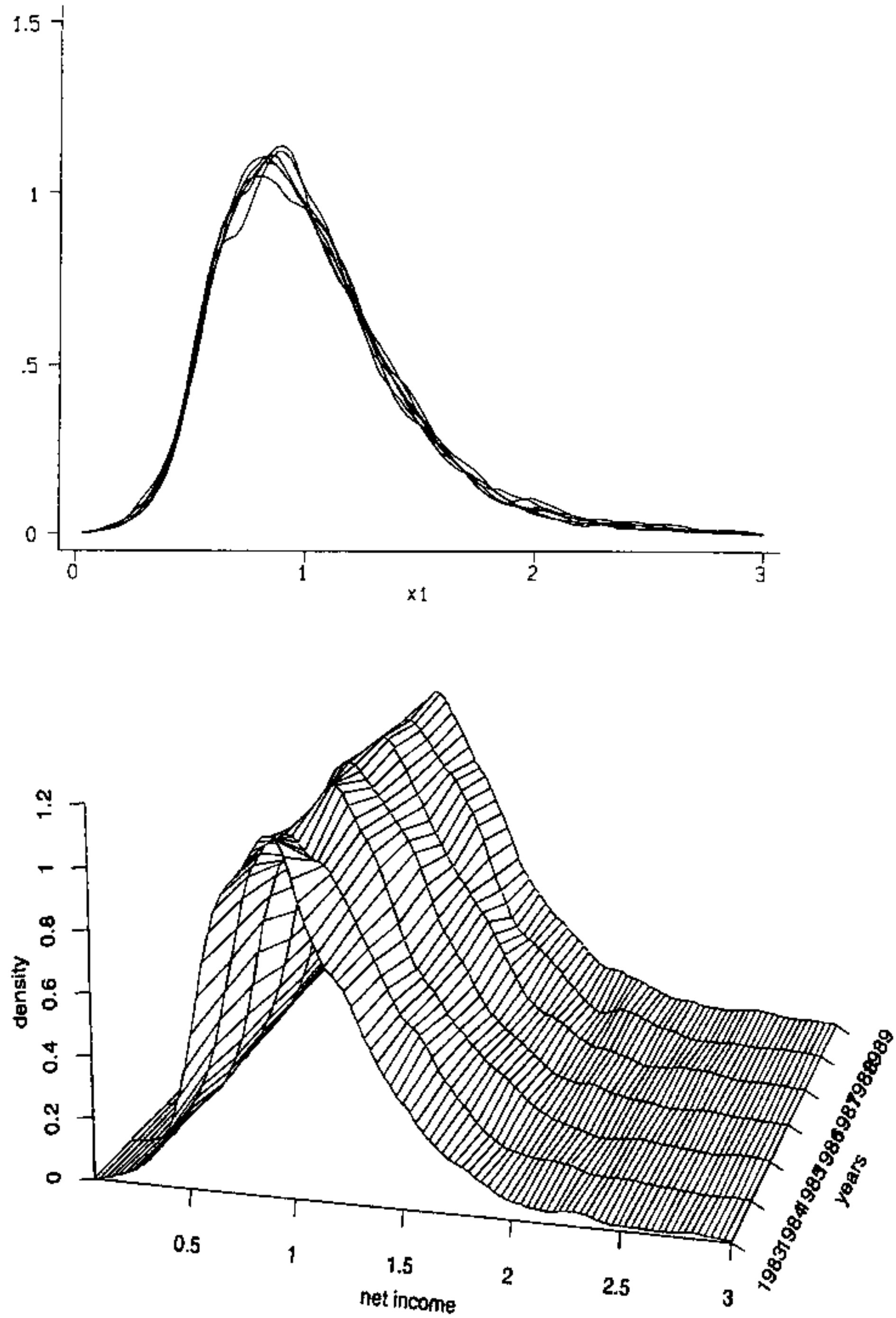
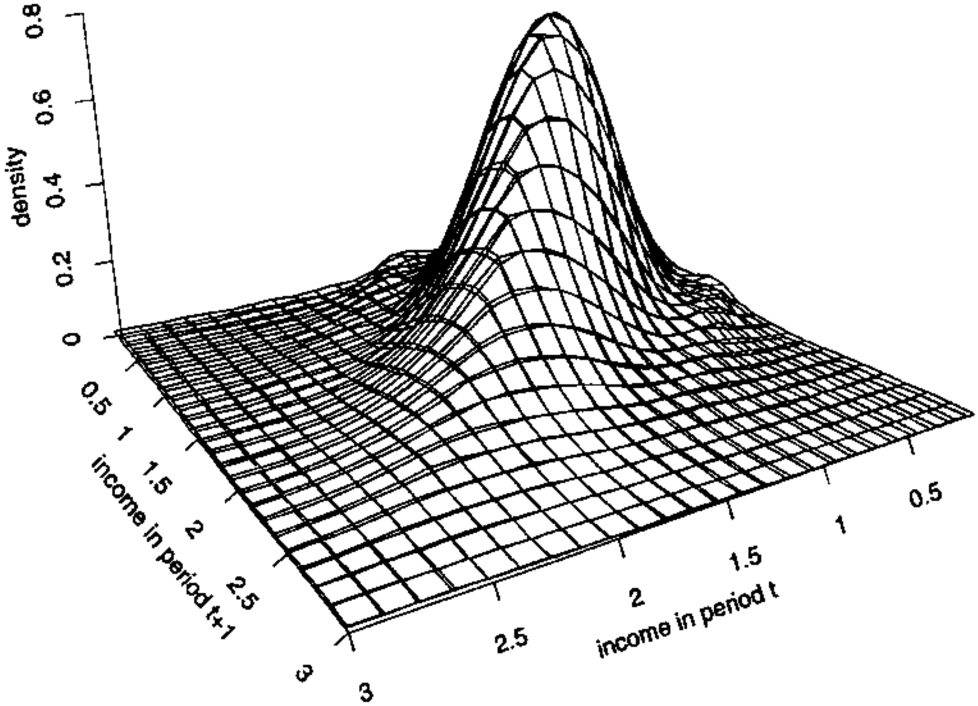


Figure 1: Shape dynamics of the net income distribution. Net income is normalised at the contemporaneous mean. The kernel density estimator uses the Epanechnikov kernel and the bandwidth is chosen using cross-validation methods.



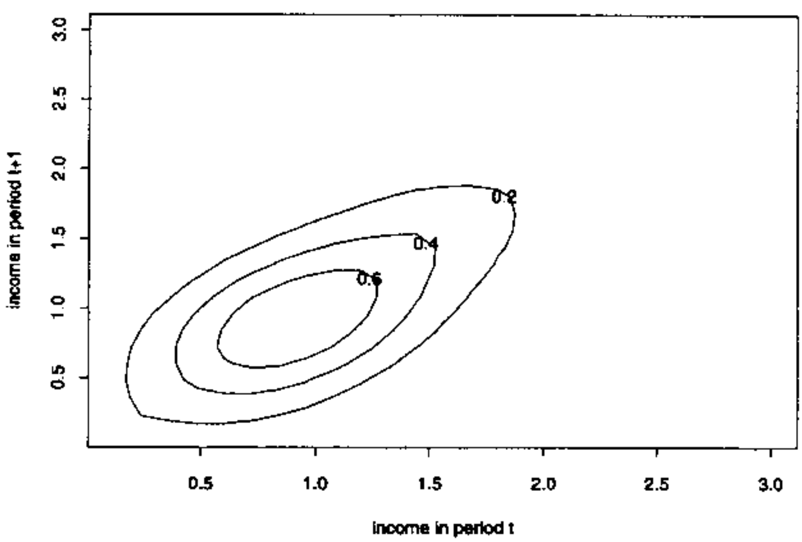


Figure 2: Upper panel: density estimates of two joint distributions. Lower panel: contour plot of a typical density estimate. Income is normalised at the contemporaneous median.

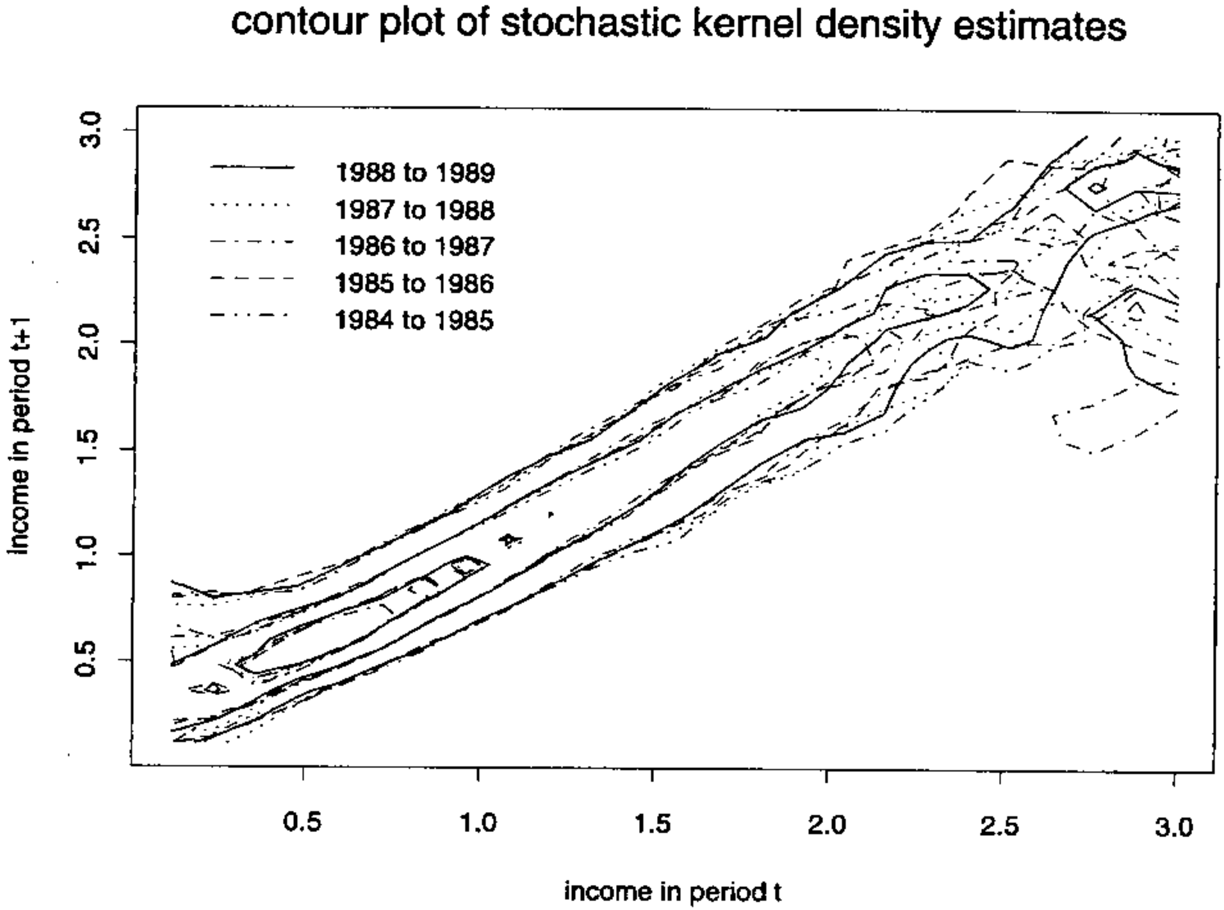


Figure 3: contour plots of the stochastic kernel density estimates 1984-1985 to 1988-1989. Income is normalised at the contemporaneous median.

tion heterogeneity, a (new) non-stationary mover-stayer model [5]. We pursue the twin-track approach to mobility which exploits the important complementarity between stochastic kernel density estimates and transition matrices as the strength of one tool compensates the weakness of the other. Although the mobility analysis is based on transition matrices, the income partition is not only chosen in an economically meaningful way, but also happens to coincide with the natural partition suggested by Figure 3. On the other hand, transition matrices are a powerful tool for making rigorous statistical inferences. The next section attempts to explain the observed mobility. The method pursued here is to concentrate on one important income state instead of the entire transition matrix, and the chosen income state is poverty -the income state about which most policy makers are concerned and for whose alleviation considerable resources are deployed. Persistent social exclusion is recognised as a grave problem facing any society. Moreover, this group has experienced the most dramatic changes in mobility. Section 5.1 explains the chances of escaping or descending into poverty by means of a Markov model which accounts both for the non-stationarity of the data and the heterogeneity of the population. A second class of models, analysed in section 5.2, comprises duration models. Section 6 concludes.

[5]This appears to be a natural research agenda. Indeed, after completing of this paper, I read the following suggestions in Atkinson, Bourguignon, and Morrisson (1992): "..the assumption .. may be relaxed by considering a second order Markov process... Alternatively, we may relax the assumption .. of population homogeneity. One route by which this may be done is the mover/stayer model.." (p.17)

2 The data

In this paper the German panel dataset GSOEP is used in its incarnation as the "Equivalent Datafile". Comprising the years 1983 to 1989, the latter is a subset of the former and not only contains its principal income variable but also includes some derived variables, the most important being post-tax post-benefit income. Since GSOEP proper was described in detail in Schluter (1996a), a brief outline should suffice. Two income concepts are used. The elements of annual gross (pre-tax pre-benefit) income are raw data but need to be aggregated. However, the Equivalent Datafile conveniently supplies an estimate of annual household post-tax post-benefit income, which is derived from the gross income data by means of a tax-benefit simulation. This income variable is computed as the sum of total family income from earnings, asset flows, private and public transfers, the imputed rental value of owner occupied housing, and a tax simulation is applied[6].

In order to take account of scale economies within the household, income was equivalised using the OECD equivalent scales[7].Finally, incomes were standardised at 1991 prices. The data remained unweighted for the sub-

[6]Unfortunately, the data providers have not yet conducted an external validation study as in the case of the PSID (see, for instance, (Pischke 1995) or Bound and Krueger (1989)) in order to assess the severity of measurement error. However, several checks are run by the data provider to ensure the high quality of the data. In the absence of any concrete evidence, we do not model the error process.

[7]Disposable income was divided by household size raised to the power 0.5. This choice of equivalence scales had been made for two reasons. First, Burkhauser, Merz, and Smeeding (1994) show that the German Social Assistance scale implies scale economies which are too low. Second, the use of the OECD scale, being the standard scale for datasets included in the LIS project, facilitates first ad hoc international comparisons.

sequent estimation procedures. The sample examined in this paper was selected by keeping only persons with a complete income record for the years 1984 to 1990. This selection procedure resulted in 9022 observations.

Subsequently, the income data is analysed by means of transition matrices. Four income groups were defined with respect to the contemporaneous median (a statistic which is robust against outliers). The poverty line is set (arbitrarily) at 0.5 times median income. Modest incomes are equivalised incomes between 0.5 and 1 times the median. Middle incomes are between 1 and 1.5 times the median. Finally, high incomes are those above 1.5 times the median. The choice of these income groups is inherently arbitrary, but the relative definition of poverty applied in this paper has become standard practise for European countries. In fact, this partition is also suggested by Figure 3.

3 Pure Markov models in discrete time

First order Markov chains are very popular in the theoretical and the applied literature perhaps because future income distributions can be easily predicted by iterating a well-behaved transition matrix and the stationary distribution can be computed as the eigenvector of the eigenvalue one. This section explores to what extent standard Markovian models can explain the observed income transitions -i.e. whether such a representation is an acceptable abstraction or rather wishful thinking. Competing models are juxtaposed, and the following sequence of tests is conducted: non-stationary and stationary Markov chains of the same order are tested against each other.

Then, a non-stationary first order chain is tested against a non-stationary second order chain. These tests were first developed by Anderson and Goodman (1957).

3.1 First order Markov chains

Let $P(t) = [p_{ij}(t)]$ be an mxm transition matrix where $p_{ij}(t)$ denotes the conditional probability of moving to state j in the current period, given that state i was occupied in the preceding period. The chain is observed up to time T at time points $t = 1, 2, .., T$. If the chain is stationary $p_{ij}(t_1) = p_{ij}(t_2) = p_{ij}, \quad \forall t_1, t_2$. Let $N(t) = [n_{ij}(t)]$ denote the associated matrix of actual transition counts. The transition probabilities $p_{ij}(t)$ need to be estimated. Since these are multinomially distributed, their maximum likelihood estimator $\hat{p}_{ij}(t)$ can be derived by maximising the likelihood function, conditional on the initial distribution

$$\log L = \sum_t \sum_i \sum_j n_{ij}(t) \cdot \log p_{ij}(t)$$

subject to $P(t)$ being a stochastic matrix. The Lagrangean for this programme is

$$\mathcal{L} = \sum_t \sum_i \sum_j n_{ij}(t) \log p_{ij}(t) - \sum_t \sum_i \lambda_{ti} \left(\sum_j p_{ij}(t) - 1 \right)$$

The first order condition implies $n_{ij}(t) = \lambda_{ti} \cdot \hat{p}_{ij}(t)$. Summing out j yields the Lagrangean multipliers $\lambda_{ti} = \sum_j n_{ij}$, which upon substitution gives the estimator

$$\hat{p}_{ij}(t) = \frac{n_{ij}(t)}{\sum_j n_{ij}(t)}$$

being a simple frequency count. For the stationary model, a similar calculation gives the maximum likelihood estimator [8]

$$\tilde{p}_{ij} = \frac{\sum_{t=1}^{T} n_{ij}(t)}{\sum_{t=1}^{T}\sum_j n_{ij}(t)} = \frac{\sum_{t=1}^{T} n_{ij}(t)}{\sum_{t=1}^{T} n_{i+}(t)}$$

where $\sum_j n_{ij}(t) = n_{i+}(t)$ for notational convenience. The frequency counts for each period are just averaged.

Given these functions, tests for non-stationarity can be easily implemented. Let the null hypothesis be that the transition probabilities are stationary, i.e. $\mathcal{H}_0$: $p_{ij}(t) = p_{ij} \quad \forall t, i, j$. Using the respective likelihood functions, the likelihood ratio is

$$\log \lambda = \sum_t \sum_i \sum_j n_{ij}(t) \cdot \{\log \tilde{p}_{ij} - \log \hat{p}_{ij}(t)\}$$

and $-2\log\lambda$ is asymptotically distributed as χ^2 with $(T-1)m(m-1)$ degrees of freedom[9]. If the null hypothesis is true, an asymptotically equivalent test is based

[8] Anderson and Goodman (1957) show its asymptotic sampling distribution to be normal.

[9] See Rao (1973). The number of degrees of freedom of the asymptotic χ^2-distribution equals the number of linearly independent restrictions. Note that P is a stochastic matrix.

on the similarity between transition matrices and contingency tables. The well-known χ^2-test then gives the test statistic

$$X^2 = \sum_i X_i^2 = \sum_i \sum_t \sum_j n_{i+} \cdot \frac{\tilde{p}_{ij} - \hat{p}_{ij}(t)}{\tilde{p}_{ij}}$$

which is asymptotically distributed as χ^2 with $m(m-1)(T-1)$ degrees of freedom. However, Anderson and Goodman (1957) show that if the null hypothesis is not true, the power of the χ^2 -test can be different from the power of the likelihood test. Thus both tests should be performed.

It is not surprising that both tests confirm the greater explanatory power of the non-stationary model. The test statistics evaluate to $-2\log\lambda = 336.3$ and $X^2 = 341.4$.

One convenient way to represent the data and to bring out the non-stationarity is to compute a mobility index, such as (Shorrocks 1978) index defined as $\mu(P) = (m - tr(P))/(m-1)$ where tr is the trace of the matrix [10]. It is the inverse of the harmonic mean of expected

[10] Schluter (1996b) shows that this index is asymptotically normally distributed. This follows simply from noting that the n_{ij} follow a multinomial law. The asymptotic normality of the index is then a consequence of the central limit theorem, $\sqrt{n}vec(\hat{P}' - P') \rightarrow N(0,V)$, and some simple manipulations. A different argument is based on the delta method (Rao (1973)) which the can be applied to most standard mobility indices to demonstrate their asymptotic normality. The first order Taylor expansion of $\mu(\hat{P})$ about $\mu(P)$ is $\mu(\hat{P}) = \mu(P) + DM(P)(vec(\hat{P}' - P'))$ where $DM(P) := \partial DM(P)/\partial vec(P')'$. Since $\sqrt{n}vec(\hat{P}' - P') \rightarrow N(0,V)$ it follows that $\sqrt{n}(\mu(\hat{P}) - \mu(P)) \rightarrow N(0,\Sigma)$

durations of remaining in a given part of the cross section distribution. The higher the index, the lower is the persistence or the greater is the mobility. Its time series is depicted in Figure 4. Other popular indices yield similar diagrams. In contrast to the seeming stability of the shape dynamics, mobility dynamics behave differently. The lack of action at the surface conceals substantial movements beneath it. In fact, there is a downwards trend which implies a consistent fall of income mobility over the years except for the last year. Comparing the confidence intervals suggests that the differences in the values are statistically significant. However, the index aggregates quite a lot of information and it is useful to examine what drives the index. Since the Shorrocks mobility index is the inverse of the harmonic mean of expected durations of remaining in a given part of the cross-section distribution, the lower panel of the figure depicts the time series of the staying probabilities. For the three richest income groups, these have a tendency to rise, but not monotonically and some movements are in opposite directions. By contrast, the lowest income group - the poor - experience a dramatic increase in immobility, but there is also a sharp fall in the last year.

3.2 Second order Markov chains

The methods of the preceding paragraphs extend in a very natural manner to second order Markov chains. $P(t) = [p_{ijk}(t)]$ denotes the conditional probability of being in state k at time t, given states i and j at times $t-2$ and

where $\Sigma = DM(P)VDM(P)'$. See Trede (1995) for explicit derivations.

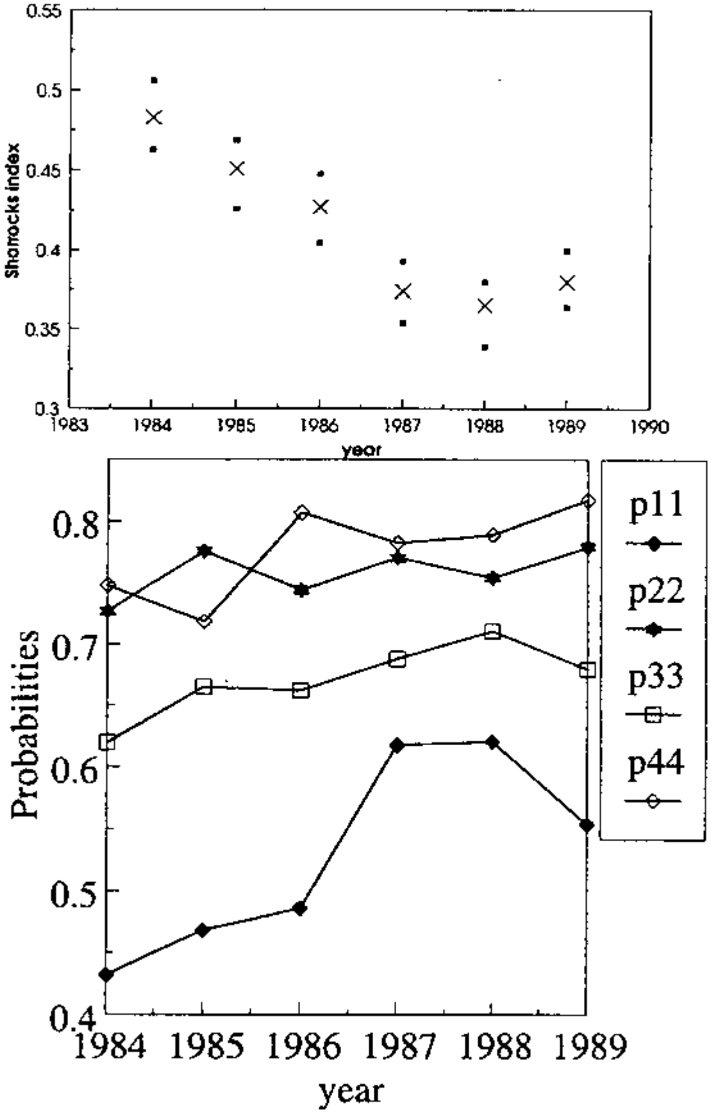


Figure 4: Upper Panel: Time series of the Shorrocks mobility index, bracketed by estimated 10% confidence intervals. The confidence intervals were computed using bootstrapping and (Efron 1987)'s BC_α method. The transition matrices were defined on the weighted data. Lower Panel: Time series of staying probabilities.

$t-1$ respectively. Again, the transition probabilities are estimated by maximising the log-likelihood function, and similar likelihood ratio and χ^2- tests apply.

The tests suggest once again that the non-stationary (second order) model has a greater explanatory power than the stationary model. The test statistics evaluate to $-2\log\lambda = 561.4$ and $\chi^2 = 545.7$.

3.3 First order against second order Markov chains

The theoretical results of the two preceding sections can be combined in order to test which order of the non-stationary model has the greater explanatory power. If the null hypothesis is that a first order non-stationary model is applicable, $p_{1jk} = p_{2jk} = \ldots = p_{mjk} = p_{jk}, \forall j, k$, the likelihood ratio becomes

$$\log\lambda = \sum_t \sum_i \sum_j \sum_k n_{ijk}(t) \cdot \{\log\hat{p}_{jk} - \log\hat{p}_{ijk}\}$$

$-2\log\lambda$ being asymptotically distributed as χ^2 with Tm^2 degrees of freedom.

The statistic evaluates to $-2\log\lambda = 2,955.8$, being very significant evidence against the null hypothesis. In consequence, the memory of the process governing income transitions extends over more than one period. Despite its popularity in the theoretical and the applied literature the (first order) Markov assumption has to be rejected.

4 Mixed Markovian models in discrete time

Pure Markovian models are popular both in the theoretical as well as in the empirical literature because of their mathematical structure. Yet, as the previous section demonstrated, they do not fit the data too well. One principal assumption underlying their estimation is the 'homogeneity of persons': individuals are the same except for their income. This assumption is likely to be flawed. Indeed, results derived elsewhere suggest that the population is very heterogeneous. Schluter (1996a), for instance, depicts for several socio-economic groups income distributions which differ dramatically. In contrast to this observable heterogeneity, some latent variable may be important. A particular type of unobservable heterogeneity is treated next.

The next level of complexity is achieved by mixing independent Markovian models, the easiest of which is the following mover-stayer model. An unobservable fraction of the population stays with certainty in its income group for all periods, whilst the evolution of incomes of everyone else, the movers, is determined by a non-degenerate first order Markov chain. The pure Markovian model is nested within this richer structure, since stayers may not be present. This nesting gives rise to a natural test, a likelihood ratio test, by means of which to discriminate between these two models.

4.1 A mover-stayer model : the stationary case

Although Goodman (1961) presents an extensive mathematical treatment of this model, his estimators, proposed without derivation, fail to be maximum likelihood estimators. These are supplied in Frydman (1984) where transitions are stationary.[11] A non-stationary model is proposed below.

Let the unobserved fraction of the population who are stayers in income group i be denoted by s_i. The income of movers $(1 - s_i)$ evolves according to the stationary first order Markov chain with mxm transition matrix $M = [m_{ij}]$. The composite process thus evolves according to $P(t) = SI + (I - S)M^t$ where S is a diagonal matrix with entries s_i. Let $n_i\,(t)$ denote the number of persons in state i at time t, n_i the number of persons staying in state i during the entire period of observation, $n_{ik} = \sum_t n_{ik}\,(t)$ the total number of transitions from state i to state k, $n_i^* = \sum_t n_i\,(t-1)$, and n the total number of persons. The log-likelihood function conditional on the initial distribution can be factorised thus

$$\begin{aligned} \log L(s, M) \;=\; & \sum_i n_i \log\left[s_i + (1 - s_i)m_{ii}^T\right] \qquad (1) \\ & + \sum_i \left[n_i\,(0) - n_i\right] \log\,(1 - s_i) + \left[n_{ii} - Tn_i\right] \log m_{ii} \\ & + \sum_i \sum_{k \neq i} n_{ik} \log m_{ik} \end{aligned}$$

The last summation pertains only to transition between unequal states, and thus concerns only movers. As re-

[11] McCall (1971) applies the mover-stayer model to the issue of earnings mobility. He simplifies the estimator proposed in Goodman (1961) by letting $T \longrightarrow \infty$ despite the fact that his empirical time series is very short.

gards the first sum, a person may remain in income class i for two reasons: either he is a stayer with probability s_i, or with probability $(1 - s_i)$ he is a mover but remains in that state for T consecutive periods with probability m_{ii}^T. The second term captures movers returning to their initial state, who have at least once left it.
The maximisation strategy is to resubstitute solutions from the first order conditions into the objective function[12]. Eventually the size of the equations system is reduced to the number of income classes, and the equations for $\hat{m}_{ii}$ can be solved numerically. The estimators of the off-diagonal elements $\hat{m}_{ij}$ are then computed recursively

$$\hat{m}_{ij} = n_{ij}(1 - \hat{m}_{ii} - \sum_{\substack{k=1 \\ k \neq i}}^{j-1} \hat{m}_{ik}) / \sum_{\substack{k=j \\ k \neq i}} n_{ik}$$

The estimates for the stayers are[13]

$$\hat{s}_i = \frac{n_i - n_i\left(0\right)\hat{m}_{ii}^T}{n_i\left(0\right)\left(1 - \hat{m}_{ii}^T\right)} = 1 - \frac{n_i\left(0\right) - n_i}{n_i\left(0\right)\left(1 - \hat{m}_{ii}^T\right)} \qquad (2)$$

[12]Amemiya (1985) proposes an alternative method of estimation. Two equation systems are established by considering transitions within two periods, viz. $P = SI + (I - S)M$ and $P^{(2)} = SI + (I - S)M^2$. P and $P^{(2)}$ can be consistently estimated by the maximum likelihood estimators presented in the preceding sections. Resubstituting these yields $2m(m-1)$ equations in m^2 unknowns. However, Frydman's estimation strategy is more parsimonious.

[13]This estimator may become useless if the observation period T is small, since $\hat{s}_i$ may become negative as Frydman failed to impose a non-negativity constraint. This problem decays with m_{ii}^T as T increases.

the last term being the ratio of the observed to the expected number of persons who make a least one transition from state i during T periods.

Since the pure Markovian model is nested within the mover-stayer model, their relative performance can be assessed using a likelihood ratio test. Let the null hypothesis be that the pure model is appropriate ($s_i = 0, \forall i$). The maximised likelihood of the pure model is $\log(L_{s=0}) = \sum_i n_i(0) \log(n_i(0) - n) + \sum_{i,k} n_{ik} \log(n_{ik}/n_i^*)$. Denote the ratio of the likelihoods by λ, $-2\log\lambda$ is distributed as χ^2 with m degrees of freedom.[14] Can the statistical significance of an individual s_i be tested ? If T and $\hat{m}_{ii}$ are such that $\hat{m}_{ii}^T$ is negligible, the estimator for s_i simplifies to $\hat{s}_{i=} n_i/n_i(0)$, being the fraction of persons initially in state i who remain there for all consecutive periods. In this case Goodman (1961)'s argument applies. Let $\tilde{p}_{ij}$ denote the estimator of the stationary first order Markov chain derived in the previous section. The test is based on a comparison between $\hat{s}$ and its expected value $\tilde{p}_{ii}^T$. $(\hat{s} - \tilde{p}_{ii})$ is normally distributed with mean zero and a variance which can be consistently estimated by $\hat{\sigma}^2 = \tilde{p}_{ii}^T(1 - \tilde{p}_{ii}^T)/n_i(0) - n\tilde{p}_{ii}^{2T-1}(1 - \tilde{p}_{ii}^T)/\bar{n}_i$, where $\bar{n}_i = \sum_t n_i(t)/T$. Under the null hypothesis $s_i = 0$, and $X_i^2 = (s - \tilde{p}_{ii}^T)/\hat{\sigma}^2$ is distributed as χ^2 with one degree of freedom.

The estimated transitions matrix is not reported here for the sake of brevity. Compared to the pure model,

[14]See Rao (1973). The number of degrees of freedom of the asymptotic χ^2-distribution equals the number of linearly independent restrictions. There are m restrictions imposed, viz. $\{s_i = 0\}_{i=1}^m$.

probability mass has been redistributed away from the main diagonal. The movers are thus more mobile than the pure model suggests. The stayers fractions are estimated to be $\hat{s} = (0.1; 0.224; 0.04; 0.14)$. Testing the competing models, the likelihood ratio test confirms the greater explanatory power of the mover-stayer model

$(-2 \log \lambda = 23{,}147.8)$. These estimate have a profound implication, since the first income group is occupied by the poor (whose income falls short of the contemporaneous poverty line). Income mobility is sufficiently high so that most persons are able to escape poverty at least temporarily. Yet, a statistically significant 10%[15] of those deemed in poverty at the beginning of the observation period constitute a hard-core of poverty[16]- remaining poor with certainty.

These results, of course, have to be taken with a pinch of salt, as the previous section suggested that income transitions are non-stationary. The problem caused by time-varying transition probabilities is addressed in the next section.

4.2 A mover-stayer model: the non-stationary case

The previous model can be generalised so that non-stationarity in income transitions can be admitted. Let movers transit according to the non-stationary first order Markov chain $M(t)$. In consequence, the composite process evolves according to $P(t) = SI + (I - S)\prod_{\tau=1}^{t} M(\tau)$. Analogous to equation (1), the likelihood function can be written as

[15]Applying the above test to the estimate of S_1, the estimate of the hard-core of poverty, yields a statistically significant result ($\chi_1^2 = 3,766.8$).

[16]Labelled by McCall (1971) the "back-wash hypothesis".

$$\log L(s, M(1), ..., M(T)) \tag{3}$$

$$= \sum_i n_i \log \left(s_i + (1 - s_i) \prod_\tau m_{ii}(\tau) \right) \tag{4}$$
$$+ \sum_i (n_i(0) - n_i) \log (1 - s_i)$$
$$+ \sum_i \sum_\tau (n_{ii}(\tau) - n_i) \log m_{ii}(\tau)$$
$$+ \sum_i \sum_\tau \sum_{k \neq i} n_{ik}(\tau) \log m_{ik}(\tau)$$

Observe that the stationary case of equation (1) is nested within equation (3). Maximising this with respect to s_i yields the estimator

$$s_i = \frac{n_i - n_i(0) \prod_\tau m_{ii}(\tau)}{n_i(0) (1 - \prod_\tau m_{ii}(\tau))} \tag{5}$$

(compare to (2).) Resubstituting this into equation (3) yields

$$\log L(\hat{s}, M(1), ..., M(T))$$
$$= c - \sum_i (n_i(0) - n_i) \log \left(1 - \prod_\tau m_{ii}(\tau) \right)$$

$$+\sum_i \sum_\tau (n_{ii}(\tau) - n_i) \log m_{ii}(\tau)$$
$$+\sum_i \sum_\tau \sum_{k \neq i} n_{ik}(\tau) \log m_{ik}(\tau)$$

where c denotes a constant. The Lagrangean of this problem is

$$\begin{aligned}\mathfrak{L} &= \log L(\hat{s}, M(1), ..., M(T)) \\ &\quad - \sum_i \sum_\tau \lambda_i(t) \left(\sum_k m_{ik}(t) - 1 \right)\end{aligned}$$

Maximising this with respect to $m_{ii}(t)$ and $m_{ik}(t)$, summing these and solving out the Lagrange multipliers $\lambda_i(t)$ yields a non-linear equations system for $m_{ii}(t)$

$$\begin{aligned}n_{ii}(t) - n_i &= (n_{i+}(t) - n_i)\, m_{ii}(t) \\ &\quad + \left[n_i(0) - n_i \frac{\prod_\tau m_{ii}(\tau)}{1 - \prod_\tau m_{ii}(\tau)} \right] m_{ii}(t)\end{aligned}$$

This is a non-linear system comprising T equations in T unknowns with solution $\hat{m}_{ii} = (\hat{m}_{ii}(1), ..., \hat{m}_{ii}(T)) \in [0,1]^T$. It is solved numerically using the multidimensional Newton's method (see Press, Teukolsky, Vetterling, and Flannery (1992)). The solutions to $\hat{m}_{ik}(t)$ and $\hat{s}_i$ are then computed recursively.

Tests of hypotheses can be implemented following the methods outlined in the previous sections.

This estimation procedure resulted in the following estimates for the stayers $\hat{s} = (0.11; 0.33; 0.13; 0.31)$. Compared to the stationary model, the estimates of the fraction of the poor has not changed, whilst the other estimates all have increased. As regards the mover probabilities, the entries of $M(t)$ follow the changes suggested by the movement of the Shorrock mobility index. Probability mass is moved onto the main diagonal as time passes, suggesting that incomes have become more immobile.

5 Poverty re-examined

The aim of this section is to go beyond the descriptive Markov models of the preceding sections and to attempt to explain the observed mobility profiles. Instead of analysing the entire transition matrix, we concentrate on one important income state - poverty - and analyse the processes governing the movements into and out of poverty. This state is not only important from a welfare point of view, but, as Figures 3 and 4 have shown, the probability of remaining poor in two consecutive periods has exhibited the largest changes. Two types of models are examined.

5.1 A Markov model with observed heterogeneity

This section examines a two state Markov model with exogenous variables as proposed in Boskin and Nold (1975) and further discussed in Amemiya (1985). Person i may be in either of two states: either he is in poverty at time

t, $y_i(t) = 1$, or he is not $y_i(t) = 0$. The probability of being in poverty conditional on the preceding state is $\Pr(y_i(t) = 1|y_i(t-1)) = F(\beta' x_i(t) + \gamma' x_i(t) y_i(t-1))$ where $F(.)$ is a distribution function with corresponding density f. Thus, the model is a generalised first order Markov model, in which the exogenous variables $x_i(t)$ exhibit non-stationarity and heterogeneity amongst persons. This formulation nests within it a variety of observationally equivalent models, depending on the parametrisation of γ. For instance, setting $\gamma = -(\alpha + \beta)$ and if f is symmetric, the model has the following interpretation. The (conditional) probability of person i entering poverty is determined as $p^i_{01}(t) = F(\beta' x_i(t))$, whereas the (conditional) probability of escaping poverty is $p^i_{10}(t) = F(\alpha' x_i(t))$. Thus the profile of a representative person entering poverty is stipulated to be different from that of a representative person escaping poverty.

The log-likelihood function can be written as

$$\begin{aligned} & \log L(\alpha, \beta) \\ = & \sum_i \sum_t y_i(t) \log F(\beta' x_i(t) + \gamma' x_i(t) y_i(t-1)) \\ & +(1 - y_i(t)) \log [1 - F(\beta' x_i(t) + \gamma' x_i(t) y_i(t-1))] \end{aligned}$$

The distribution function is chosen to be logistic $F(x) = e^x/(1+e^x)$, so that the objective function is globally concave and the estimation step reduces to estimating a standard logit model. The maximisation strategy is to employ the iterative method of scoring separately for each parameter. The MLE is consistent and asymptotically normal (see Amemiya (1985)). Note also that the indices $(i;t)$ can be treated as a single index. Thus, although the time series is relatively short but the cross section is large, the sample can be considered to be large.

The sample was chosen to contain only persons above the age of 20 in order to focus on the causes of poverty, a step which reduces the size of the sample to 6266 observations. The regressors comprise: indicators for employment status, disability, and household size in a given year [17], nationality, the age, and education level (measured in years) of the person in the year 1984. The importance of these variables is not surprising given the results of a static analysis in Schluter (1996a) who estimates the income distributions for various partitions of the sample using kernel density estimators.

[17] Bane and Ellwood (1985), for instance, emphasise the importance of the household formation process as a determinant of poverty in the case of the US.

Table 1: Maximum likelihood estimates for poverty model

unweighted model	variables (standard errors)							chi-squared	log-likelihood	N
parameter	nationality	age84	hhsize	education84	disability	unemployment	constant			
β	.886 (.076)	-.027 (.002)	-.304 (.02)	-.131 (.02)	.437 (.08)	1.53 (.077)	-1.24 (.29)	716.48	-4,174.4	35,574
γ	.268 (.11)	.006 (.003)	-.137 (.036)	-.085 (.027)	.203 (.11)	.64 (.11)	.57 (.38)	114.3	-1,337.9	2,022
α=-(β+γ)	-1.154	.021	.441	.216	-.64	-2.17	.67			

weighted model	(robust standard errors)									
β	.859 (.1)	-.038 (.003)	-.55 (.05)	-.14 (.027)	.363 (.11)	1.75 (.124)	-.034 (.418)		-4,340.1	35,574
γ	.43 (.156)	.008 (.004)	-.15 (.053)	-.103 (.033)	.08 (.15)	.727 (.169)	.75 (.45)		-1,302.5	2,022
α	-.1.289	.03	.697	.24	-.443	-2.47	-.72			

Notes: The variables are defined as follows: nationality is an indicator set equal to one for foreigners; age84 is the age of a person in year 1984 on the survey date; the education level is measured by years of education up to year 1984; hhsize refers to the size of the household at time t disability at time t is an indicator equal to one for disabled persons; unemployment at time t equals one when unemployed.

Chi-squared refers to the likelihood ratio test, the null hypothesis being that only the constant term has explanatory power.

Robust standard errors were computed following the methods proposed in Huber (1967).

Both weighted and unweighted data are used and Table 1 collects the estimation results. The results contain some surprises. As regards the unweighted data, the probability of escaping poverty is higher when a person is a German, is well educated, healthy and ends unemployment spells quickly. This last ability is the most decisive and the relative size of the parameter estimate is perhaps astonishing because of the general presumption that the entry into poverty should be uncoupled from the entry into unemployment as plicy makers generally claim that German social insurance benefits are sufficiently generous. Informal insurance between members of a household and precautionary savings should also dampen the effect of unemployment. Despite this, the labour market performance is the most important, limiting the scope for the welfare system to alleviate the incidence of poverty (this point is reinforced below as we find evidence of negative duration dependence). More formally, the relative importance of the variables can be assessed by computing an elasticity such as $\eta_{i;j} := (\partial p^i_{01}(t)/\partial x^i_j(t))/(x^i_j(t)/p^i_{01}(t)) = \beta_j(1 - p^i_{01}(t))x^i_j(t)$ which approximates the effect of a change in a discrete variable x_j for person i. The effect of becoming unemployed is dramatic: $\eta_{i;unemployed} = 1.53(1 - p^i_{01}(t))$ (but this effect diminishes as $\beta' x^i(t)$ increases). Unemployment and nationality are of even greater importance for those escaping poverty.

The coefficient for nationality is large but this may be due to oversampling foreigners. When the data is weighted, the coefficient on nationality is expected to fall because foreigners were oversampled. Surprisingly, the coefficient is only slightly lower, but the employment status coefficient is markedly higher. This implies a different profile for persons slipping into poverty. For Germans, the principal reason appears to be unemployment, whilst it seems to be low earnings for foreigners.

5.2 Semi-Markov processes: non-stationary duration models

The discrete time models of the preceding sections have to confronted the time aggregation problem, highlighted in Singer and Spilerman (1976), caused by the absence of a natural time unit: income transitions do not happen at the end of regularly spaced intervals which coincide with those of the panel survey. As a consequence, parameter estimates cannot be interpreted as structural information. In this case it is more appropriate to fit a continuous time model. But this strategy gives rise to two problems. First, the model needs to be formulated in such a way that the actual discrete time observation is embeddable within the continuous time model. A nice set of necessary and sufficient conditions has yet not been found. [18]

The second problem is caused by the particular data under scrutiny, viz. their non-stationarity. Whilst it is impracticable to estimate a general continuous time Markov chain, researchers have pursued two avenues. Singer and Spilerman (1976) discuss (but do not estimate) a mixture model in which transitions follow a stationary

[18] Geweke, Marshall, and Zarkin (1986), for instance, present a calculation to test the embeddability of a discrete first order stationary Markov chain within a stationary continuous time model. See also their references for the embeddability problem.

Markov chain but waiting times between transitions may vary with time. A second possibility and the strategy pursued below is to focus on one economically meaningful state, such as poverty, and to estimate a parametrised duration model.

This section presents a standard duration model as outlined in Cox and Oakes (1984) and follows suggestions of Amemiya (1985). Person i may be in either of two states: either he is poor or he is not. The time in poverty T, i.e. the length of the poverty spell, is a random variable with distribution F and associated density f. If the population is heterogeneous, these may differ across persons, written as F_i. It is convenient to work with the hazard rate $\lambda_i(t) := f_i(t) / [1 - F_i(t)]$ where $\lambda_i(t)\Delta t$ has a probabilistic interpretation: it is the probability that, given the person has not left poverty in the time interval$(0, t)$, he will do so the next moment, i.e. in $(t, t+\Delta t)$. A basic assumption of the continuous time model is reminiscent of Poisson processes, since the probability that a person changes her income state more than once in a small time interval $(t, t+\Delta t)$ is negligible. $\lambda_i(t)$ may vary with time. The duration function F can then be written as

$$F_i(t) = 1 - \exp\{-\int_0^t \lambda_i(z)\, dz\} \qquad (5)$$

If person i completes J poverty spells of individual length $t_{i,j}$ the contribution to the likelihood function is $\prod_{j=1}^{J} f_i(t_{i,j})$. However, the estimation problem is complicated by the fact that person i may have censored spells. A spell at the end of the panel t^* is right-censored and thus incomplete if the person cannot be observed to leave that state, leading to the contribution $1 - F_i(t^*)$ to the likelihood function. A spell is left-censored if person i is in poverty at the beginning of the panel, and may have been in this state for a long time. Amemiya (1985) shows that the contribution to the likelihood function then is $[1 - F_i(t)] / \int s f_i(s) ds$.[19] For the sample under scrutiny Table 2 collects information on the incidence and duration of poverty spells, and the extent of censoring.

The problem, of course, is how to parametrise the hazard rate $\lambda_i(t)$. A parametrisation, popular in the econometrics of labour turnover, is a Cox proportional hazard rate $\lambda_i(t) = h(t) \exp(\beta' x_i(t))$, where $x_i(t)$ is a vector of time-varying exogenous variables. Following the previous Markov model, $x_i(t)$ includes two different

[19]In some applications, such as duration models of criminal recedivism or fertility, the probability of eventual "failure" is less than one; some censored observations will never "fail". If the survival function is thus defective for some persons, Schmidt and Witte (1989) propose to use a split population model, which parametrises Pr{never fail}$=1 - G(\alpha' z_i) = 1 - 1/(1 + \exp(\alpha' z_i))$, where z_i is a vector of explanatory variables. The likelihood function then needs to be adjusted accordingly.

In the current model, the problem is minor, since this criticism could at most be applied to the old, living on social benefits. However, the density estimates reported in Schluter (1996a) show that poverty is not a predominant old age phenomenon. Moreover, a poor old pensioner could alter her income state by entering the household of her children. The problem, however, is not completely absent given the previous results of the mover-stayer model.

length [years]	numbers of spells	left censored	right censored
1	893	194	145
2	246	62	63
3	113	19	53
4	64	16	27
5	33	6	23
6	30	10	20
7	51	51	51

Table 2: Poverty spells: incidence, duration, and censoring

processes: an unemployment process and a household formation process which traces the evolution of the size of the household. Note that the parameter β does not vary with the number of spells. The baseline hazard rate $h(t)$ captures duration dependence of the poverty process.

The parameters are estimated by maximising the (partial) likelihood function[20], but in order to simplify the estimation problem left censored spells were deleted. In order to evaluate the integral in (5) with discrete data, the exogenous variables were assumed to remain constant during the interval between observations.

The estimation results on the unweighted data are reported in Table 3.

Both the nationality and the age variable are not significant. An increased household size increases the poverty hazard. But most important, confirming the evidence of

[20]See Cox (1975) or Lancaster (1990) chapter 9 for a discussion of the partial likelihood function.

variables	ML estimates	standard errors
unemployment status	-.236	.0858
household size	.065	.029
nationality	-.1556	.091
age in 1984	-.0009	.0027

Table 3: The continuous-time Cox poverty hazard model

the preceding section, is the employment process. Being unemployed reduces the hazard of leaving poverty.

Furthermore, the plot of the baseline hazard rate $h(t)$ is derived by setting the parameter values β to zero. Inspecting the (not provided) plot of the non-parametric estimate of the baseline hazard rate, it increases at first, but then falls monotonically. Thus medium and long term poverty profiles differ with the latter exhibiting negative duration dependence.(Cf. also the non-parametric estimates of the discrete duration model.) [21]. Thus, the longer the poverty spell, the less likely is the person to escape from it. However, these findings must be considered tentative in the light of a result due to Heckman and Singer (1984). They have demonstrated that variable selection is a grave problem since "uncontrolled unobservables bias estimated hazard rates towards negative

[21] $h(t)$ is often assumed to be Weibull, $h(t) = \alpha t^{\alpha-1}$, since a Weibull specification leads to a non-constant hazard rate (but nests within it the exponential distribution which exhibits a lack of memory). Duration dependence is negative (positive) if $\alpha < 1$ $(\alpha > 1)$. Fitting a Weibull distribution leads to an estimate of $\hat{\alpha} = 0.833 < 1$, confirming the conjectured negative duration dependence.

duration dependence". This follows since more mobile persons leave the less mobile persons behind, creating the appearance of stronger negative duration dependence than actually exists.

Unobservable heterogeneity can be modelled by introducing a mixing distribution, so that the hazard rate is perturbed by an unobservable random variable V. Following Lancaster (1979), let v be iid from a Gamma$(1, \eta)$ distribution with variance η^{-1}, assumed to mimic the unobservables. Thus, the hazard rate for person i becomes

$$\lambda_i(t) = v_i \mu_i(t) \text{ where } \mu_i(t) = \alpha t^{\alpha-1} \exp\left(\beta' x_i(t)\right) \quad (6)$$

This specification leads to conditional distributions $F_i(t|v)$ and the unobservable v needs to be integrated out. This yields the unconditional distribution $F^*(t) = E_v(F_i(t|v)) = 1-[1+z(t)/\eta]^{-\eta}$ and density $f^*(t) = \mu(t)[1+z(t)/\eta]^{-(1+\eta)}$ where $z(t) = \int_0^t \mu(s)ds$. The maximum likelihood estimation of the parameter vector (β, η) is carried out using the $EM-$algorithm (see the appendix for a description). However, the resulting estimate of the variance of the mixing Gamma distribution, η^{-1}, is already very high on only the uncensored data, $\eta^{-1} = 14$. This implies that a Gamma mixing model, popular in the literature, is inappropriate in the present context.

It may be argued that the continuous time model is misspecified in that discrete-time data have inappropriately been treated as if they were continuous. Does a discrete-time model have different implications ?

The theory outlined above extends in a natural manner to the discrete-time case. For instance, the hazard rate now has the interpretation $\lambda_i(t) = \Pr\{T_i = t | T_i \geq t; x_i(t)\}$. As pointed out by Allison (1982) and reiterated by Jenkins (1996), estimation of this model is straightforward. Making the unit of analysis the spell month and thus reorganising the data, the likelihood function for the discrete-time duration model can be rewritten in a form which is standard in the analysis of a binary variable. Two parametrisations of the hazard rate are examined. First, the complementary log-log hazard rate $\lambda_i(t) = 1 - \exp\{-\exp\{h(t) + \beta' x_i(t)\}\}$ is chosen, since it is the counterpart of the underlying continuous time proportional hazard model examined above. But since there is no reason why hazard rates should be proportional, the second parametrisation is the logistic hazard rate $\lambda_i(t) = 1/(1 + \exp\{-h(t) - \beta' x_i(t)\})$.

The results of the estimation are reported in Table 4. The selected variables are the same as in the previous models. Duration dependence is captured by the baseline hazard $h(t)$, which is estimated non-parametrically by a sequence of dummies. The results of the two parametrisations are very similar. This should not be too surprising, since it is well known that the logistic model converges to the proportional hazard model as the hazard rate converges to zero. Once again, poverty spells of the long-term poor exhibit negative duration dependence. The hazard of leaving poverty is lower for foreign nationals, and the household formation process is neither important nor very significant. Finding employment is the principal way of escaping from poverty.

What are the determinants of re-entering poverty ?

model	c. log-log		logistic	
	MLE	SE	MLE	robust SE
duration=2 years	.138	.102	.169	.124
duration=3 years	-.163	.159	-.181	.185
duration=4 years	-.549	.232	-.64	.263
duration≥5 years	-.639	.504	-.73	.553
employment status	.407	.0818	.495	.099
household size	.0656	.0283	.0828	.034
nationality	-.161	.089	-.202	.105
disabled	-.044	.092	-.048	.108
education in 1984	.03	.021	.0377	.0268
age in 1984	.0014	.0027	.0018	.0032

Table 4: The discrete-time duration models of the hazard of leaving poverty

Applying a duration model to this issue is problematic, since sample sizes are small: there are 1050 single spells out of poverty which followed a poverty spell of which 70% are right censored. So the subsequent statistical analysis has to be regarded as tentative. However, this data structure suggests that for economically mobile persons (the movers in section 4) poverty is a predominantly transitory and rare event, which once overcome is unlikely to be experienced again.

The hazard rate $\lambda_i(t)$ now captures the probability of person i re-entering poverty at time t. The results of estimating the model with the two hazard parametrisations are reported in Table 5. The selected variables are those of the previous models. The estimates show the expected strong negative duration dependence: the longer the spell out of poverty, the less likely is the person to experience poverty again. The surprise, however, is that although the coefficients on all other explanatory variables have signs consistent with the previous results, they are not statistically significant. Moreover, the size of the employment coefficient is very small. This duration model is thus inadequate for analysing the probabilities of re-entering poverty.

How do these findings relate to results found by other researchers for other countries such as the US ? The results are, in many ways, similar to those of Bane and Ellwood (1985). Using the PSID for the years 1970 to 1982, they find that most of those who become poor will have only a short stay in poverty, whilst the stock of the poor is predominantly composed of the long-term poor. The hazard of leaving poverty also exhibits negative duration dependence (although these are computed ignoring

model	c. log-log		logistic	
	MLE	SE	MLE	robust SE
duration=2 years	-.522	.139	-.676	.176
duration=3 years	-.88	.223	-1.1	.263
duration=4 years	-1.11	.279	-1.364	.317
duration≥5 years	-5.12	.326	-5.47	.33
employment status	-.039	.129	-.064	.166
household size	-.043	.045	-.059	.058
nationality	.07	.14	.0999	.185
disabled	.144	.135	.188	.177
education in 1984	-.005	.04	-.0076	.053
age in 1984	-.005	.04	-.004	.005

Table 5: The discrete-time duration models of the hazard of re-entering poverty

observable and unobservable population heterogeneity). Using cross-tabulation techniques, they find that earnings changes explain 75% of all poverty spell endings, but this figure is dramatically lower for beginning spells. This result is mirrored in the German case by the importance of the (un)employment process.

The household formation process is found to be of lesser importance than in the US. However, this process is modelled only crudely here as a change in the size of the household, whereas Bane and Ellwood examine separately the various possibilities such as the birth of a child, the wife becoming household head or escaping poverty through marriage or the departure of children from the household. Thus, changes in the size of the household subsume possible events with opposite effects on the poverty status which might explain the small estimated coefficient. Since the current study analyses annual income data, the caveat of Ruggles and Williams (1989) applies, who, using monthly data, find that the typical poverty spell is much shorter than would be anticipated using annual data as 2/3 succeed in escaping poverty before 12 months.[22] In this case, annual data combines multiple short spells into one long spell, giving the impression of longer duration dependence.

---

[22]See also Blanc (1989) for a similar econometric approach in the context of single AFDC spells (amongst female household heads) in the US using monthly data. The principal focus of her analysis is duration dependence whose various parametrisations she compares with a non-parametric stepwise specification. In contrast to annual data which might combine multiple short spells, she finds evidence of only weak duration dependence but two distinct groups of beneficiaries.

6 Conclusion

Intra-distributional mobility is a very important dimension of income dynamics and merely extrapolating from the shape dynamics of the income distribution is likely to result in misleading judgments about lifetime welfare. In the German case, the lack of action at the surface conceals substantial movements beneath it. Indeed according to Friedman's criterion Germany has become a more unequal society because overall mobility has fallen. The largest changes have occured in the lowest income group, amongst the poor, whereas the middle income groups have experienced remarkably stable incomes. Contrary to common prejudice, however, Germany is a fairly mobile society. Most probabilities of remaining poor in two consecutive periods are below 0.55. The other income groups are less mobile, but all staying probabilities for the second richest income group are bounded by 0.7 and for the remaining income groups by 0.8.

Several statistical models based on transition matrices were estimated in order to provide a concise description of the mobility process and to draw rigorous statistical inferences. Although any discretisation of the state space is arbitrary, contour plots of the stochastic kernel density estimates -the continuous analogue of transition matrices-suggest a very natural and economically meaningfull partition. The transition probabilities vary with time and the process exhibits a memory which extends beyond one period. The (first order) Markov assumption is rejected. The mover-stayer models also suggest the importance of population heterogeneity.

In order to examine the economic determinants of the income process further and to go beyond the descriptive analysis, we have concentrated on one very important income state -poverty- instead of the entire transition matrix. Although different models were estimated - a Markov model with exogenous variables and several duration models - the principal findings are similar: unemployment is the principal determinant of poverty; in contrast to the US, the household formation process is only of minor importance, as are age and educational background. Poverty spells of the long-term poor exhibit negative duration dependence: the longer the poverty spell, the less likely is the person to escape poverty.

The current specifications of these models are very parsimonious and only one income state is examined in detail. The economic determinants of income mobility warrant further research.

7 Appendix: The EM-Algorithm

This section describes the EM-algorithm used for maximum likelihood estimation in the poverty hazard model with unobservable heterogeneity. For a more detailed description see Cox and Oakes (1984) or Lancaster (1990), upon which the following discussion is based.

The EM-algorithm consists of two principal steps, viz. taking an Expectation, and Maximising the objective function thereafter.

Let the random variable V with realisation v be iid with distribution function $G(.;\eta)$ and associated density $g(.;\eta)$, known up to a parameter vector η. This process generates the unobservable heterogeneity. Let T with realisation t denote the random variable waiting time, parametrised such that its conditional density is $f(t|v;\beta,\eta) = v\mu(t;\beta)\exp(-vz(t;\beta))$ where $z(t;\beta) = \int_0^t \mu(s)ds$. The log-likelihood of the joint distribution of V and T is

$$\begin{aligned}
\log L(\beta;\eta;t;v) &= \sum_{i=1}^{N}[\log f(t_i|v_i;\beta) + \log h(v_i|\eta) \\
&= \sum_{i=1}^{N}[\log v_i + \log \mu(t_i;\beta) - v_i z(t_i;\beta) + \log g(v_i;\eta)
\end{aligned}$$

The algorithm proceeds as follows:

1. From an initial guess $(\beta_n;\eta_n)$ calculate the log-likelihood function of the joint distribution of V and T.

2. Calculate its expected value using the initial guess

$$Q((\beta;\eta)\,(\beta_n;\eta_n)) = E(\log L(\beta;\eta);V|t,(\beta_n;\eta_n))$$

In the context of the present model, the following calculations are typical. Since $g(v)$ and $f(t|v)$ are known, $f(v|t)$ can be calculated. The terms such as $E(V|t)$ and $E(\log V|t)$ are then readily derived.

3. Maximise this with respect to $(\beta;\eta)$. The solutions to the first order conditions define the new iteration values $(\beta_{n+1};\eta_{n+1})$.

4. Continue to iterate until the value of the unconditional log-likelihood converges.

References

Allison, P. (1982). Discrete-time methods for the analysis of event histories. In S. Leinhardt (Ed.), *Sociological Methodology 1982.* Jossey-Bass.

Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press.

Anderson, T. and L. Goodman (1957). Statistical inference about Markov chains. *Annals of Mathematical Statistics*, 89–109.

Atkinson, A., F. Bourguignon, and C. Morrisson (1992). *Empirical Studies of Earnings Mobility.* Harwood Academic Publishers.

Bane, M. and D. Ellwood (1985). Slipping into and out of poverty: The dynamics of spells. *Journal of Human Resources.*

Blanc, R. (1989). Analysing the length of welfare spells. *Journal of Public Economics 39*, 245 – 273.

Boskin, M. and F. Nold (1975). A Markov model of turnover in aid to families with dependent children. *Journal of Human Resources 10*(4), 467–481.

Bound, J. and A. Krueger (1989). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right ? *NBER Working Paper No.2885.*

Burkhauser, R., A. Crews, M. Daly, and S. Jenkins (1995). Where in the world is the middle class ? A cross-national comparison of the shrinking middle class using kernel density estimates. *Syracuse University cross-national studies in aging Discussion Paper 26.*

Burkhauser, R., J. Merz, and T. Smeeding (1994). Relative inequality and poverty in Germany and the United States using alternative equivalence scales. *Discussion paper, Universitaet Lueneburg, Forschungsinstitut Freie Berufe.*

Champernowne, D. (1973). *The distribution of Incomes between persons.* Cambridge University Press.

Cowell, F., S. Jenkins, and J. Litchfield (1996). The changing shape of the UK income distribution: Kernel density estimates. In J. Hills (Ed.), *New Inequalities.* Cambridge University Press.

Cox, D. and D. Oakes (1984). *Analysis of Survival Data.* Chapman and Hall.

Deaton, A. (1992). *Understanding Consumption.* Clarendon Press, Oxford.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association 82*, 171 – 200.

Friedman, M. (1962). *Capitalism and Freedom.* Chicago, University of Chicago Press.

Frydman, H. (1984). Maximum likelihood in the mover-stayer model. *Journal of the American Statistical Association 79*(387), 632–638.

Geweke, J., R. Marshall, and G. Zarkin (1986). Mobility indices in continuous time Markov chains. *Econometrica 54*(6), 1407–1423.

Goodman, L. (1961). Statistical methods for the mover-stayer model. *Journal of the American Statistical Association 56*, 841–868.

Heckman, J. and B. Singer (1984). Econometric duration analysis. *Journal of Econometrics*, 32–132.

Huber, P. (1967). The behaviour of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1*, 221–233.

Jenkins, S. (1995). Did the middle class shrink during the 1980s? UK evidence from kernel density estimates. *Economics Letters*, 407–413.

Jenkins, S. (1996). Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics 57*(1), 129–137.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica* (4), 939–957.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data.* Econometric Society Monographs 17, Cambridge University Press.

McCall, J. (1971). A Markovian model of income dynamics. *Journal of the American Statistical Association 66*(335), 439–447.

Pischke, J.-S. (1995). Measurement error and earnings dynamics: Some estimates from the psid validation study. *Journal of Business and Economic Statistics 13*(3), 305–314.

Press, W., S. Teukolsky, W. Vetterling, and B. Flannery (1992). *Numerical Recipes in C.* Cambridge University Press.

Quah, D. (1996). Empirics for growth and distribution: stratefication, polarisation and convergence clubs. *LSE Department of Economics mimeo.*.

Rao, C. (1973). *Linear Statistical Inference and its Application.* John Wiley and Sons.

Ruggles and Williams (1989). Longitudinal measures of poverty: accounting for incomes and assets over time. *Review of Income and Wealth 35*, 225 – 243.

Schluter, C. (1996a). Income distribution and inequality in Germany - Evidence from panel data. *LSE DARP distributional analysis research programme discussion paper No. 16.*

Schluter, C. (1996b). Income mobility in Germany - Evidence from panel data. *LSE DARP distributional analysis research programme discussion paper No. 17.*

Schmidt, P. and A. Witte (1989). Predicting criminal recidivism using 'split population' survival time models. *Journal of Econometrics 40*(1), 141–159.

Shorrocks, A. (1978). The measurement of mobility. *Econometrica 46*, 1013 – 1024.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall.

Singer, B. and S. Spilerman (1974). The representation of social processes by Markov models. *American Journal of Sociology 81*(1), 1–54.

Singer, B. and S. Spilerman (1976). Some methodological issues in the analysis of longitudinal surveys. *Annals of Economic and Social Measurement* (5/4), 447–474.

Trede, M. (1995). Statistical inference in mobility measurement: Sex difference in earnings mobility. *University of Cologne, mimeo.*