Measuring Social Mobility as Unpredictability
Author(s): Simon C. Parker and Jonathan Rougier
Source: *Economica*, Vol. 68, No. 269 (Feb., 2001), pp. 63-76
Published by: Wiley on behalf of The London School of Economics and Political Science and The Suntory and Toyota International Centres for Economics and Related Disciplines
Stable URL: https://www.jstor.org/stable/3548881
Accessed: 19-12-2019 13:49 UTC

Economica (2001) **68**, 63–76

Measuring Social Mobility as Unpredictability

By SIMON C. PARKER and JONATHAN ROUGIER

University of Durham

Final version received 27 September 1999.

By associating mobility with the unpredictability of social states, new measures of social mobility may be constructed. We propose a family of three state-by-state and aggregate (scalar) predictability measures. The first set of measures is based on the transition matrix. The second uses a sampling approach and permits statistical testing of the hypothesis of perfect mobility, providing a new justification for the use of the χ^2 statistic. The third satisfies the demanding criterion of 'period consistency'. An empirical example demonstrates the usefulness of the new measures to complement existing ones in the literature.

INTRODUCTION

An enduring and popular medium for representing social mobility is the transition matrix, which describes the probabilities of persons moving from any one state to another state, or remaining where they are. Despite their popularity in applied work, there is still no commonly agreed definition or measure of social mobility based on transition matrices. This is due partly to an absence of unanimous agreement over what the word 'mobility' actually connotes. For example, Bartholomew (1982; 1996, ch. 5) distinguishes between mobility as distance moved between different states ('movement'), and mobility as the speed at which a social process changes over time ('dependence'). A number of measures have been proposed for both notions of mobility, most of which map transition matrices into scalar summary indices. Whereas the movement measures emphasize elements of the transition matrix that involve cases travelling over a large number of states, dependence measures regard as mobile those social structures that rapidly converge to their steady state.[1]

In fact, it is possible to suggest a third notion of mobility that is quite distinct from the other two. This is mobility as freedom of movement ('unpredictability'). Unpredictability is distinct from the other two notions of mobility, in that a social structure can adjust rapidly to its steady state with lots of movement of cases between states, yet can do so in a predetermined and predictable fashion.

An example might help to clarify the meaning of the unpredictability concept. Consider an occupational classification that includes the following two states: first-year apprentices and second-year apprentices. Individuals in the first state will almost invariably move to the second after one year; yet this is a very predictable movement, which says more about the way the occupational groups have been defined than about any genuine mobility. 'Movement' measures would rather misleadingly indicate mobility in this case, despite the predictability of the transition; and 'dependence' measures might also indicate mobility, depending on the other elements of the transition matrix.

The purpose of this paper is to develop new measures of social mobility based on the notion of mobility as unpredictability. As will be seen, unpredictability measures may give very different indications of social mobility to 'traditional' movement-based or dependence measures—having clear implications for policy-makers concerned about the degree of 'social mobility'. A family of predictability measures is proposed, consisting of three members: measures based on the transition matrix alone, measures permitting statistical inference about the predictability of a social structure, and measures designed to satisfy a demanding criterion called 'period consistency'. Each member enjoys some special properties not shared by the others, meriting their separate treatment. For example, the ability to test the hypothesis of perfect mobility may be especially attractive when transition matrices are estimated from small samples and/or noisy data. In this context it is interesting that one of the measures we propose is none other than the well known χ^2 statistic. Also, the period-consistent measures broaden the range of possible comparisons between transition matrices, provided that some particular technical requirements are satisfied.

Throughout the paper, the discussion relates purely to the standard transition matrix framework based on a discrete-time first-order Markov process with discrete states.[2] The discussion is general in that the concept of mobility is not confined to any particular context, but is equally applicable to social, occupational, intergenerational, geographical and income mobility processes, among others. Of course, the applicability of the Markov frame work may be limited in practice, for example if the transition matrix is not time-invariant as assumed.[3]

The paper is organized as follows. Section I distinguishes between different types of mobility, and states the defining characteristics of state-by-state and scalar predictability measures. Section II presents a set of new measures, and Section III discusses measures that satisfy period consistency. Section IV illustrates the new measures with a transition matrix adapted from Harbury and Hitchens's (1979) work on British intergenerational wealth mobility. Section V concludes.

I. Types of Mobility and Mobility Measures

Suppose each member of the population can be classified into exactly one of the k^2 ordered tuples (i, j), denoting 'from i to j'. The proportion of those members starting in i that finish in j is given by the element P_{ij} of the population matrix P. Row i of P, the proportionate destinations of members starting in i, is denoted P_i. We commence with two definitions based upon P. The first distinguishes between predictability and movement notions of mobility, and the second defines precisely the properties a predictability measure should possess.

Definition 1 (Transition types of states). State i is *perfectly mobile* if $P_i = k^{-1}\mathbf{1}_k$. State i is $\mathcal{P}_i = e_i$, where e_i is the unit vector with a 1 in position i and 0 elsewhere. State i is *perfectly predictable* if $P_i \in \{e_1, ..., e_k\}$. The set of all possible P_i is denoted $\mathbb{P}_i$. In obvious notation, $\mathbb{P}_i^{PI} \subset \mathbb{P}_i^{PP} \subset \mathbb{P}_i$. We also denote by $\mathbb{P}$ the set of all possible P.

According to Definition 1, perfect mobility (PM) in a state implies that any case is equally likely to move to any of the k possible destination states.

Conversely, perfect immobility (PI) implies that any case remains in its starting state for ever.[4] Thus, PI in state i is characterized by a row of the transition matrix in which there is unity in column i and zero elements in all other columns. Perfect predictability (PP) differs from perfect immobility because unity occurring in *any* of the columns with zeros elsewhere implies that the transition is predetermined. It is in this sense that PP is a generalization, or superset, of PI.

Definition 2 (Predictability measure). A measure $m : \mathbb{P}_i \mapsto \mathbb{R}$ is a *predictability measure* if (i) it is symmetric in the individual probabilities, and (ii) transferring an infinitesimal quantity of probability mass within P_i from state j to state j' increases $m(P_i)$ if and only if $P_{ij} \leqslant P_{ij'}$.

Consider part (i) of Definition 2 first. For a symmetric function, $m(P_i) = m(\text{perm } \mathrm{P_i})$ for all $P_i \in \mathbb{P}_i$. Symmetry is inherent in measures of predictability: there is no notion of distance between states, so all states are treated equally. Thus, symmetry ensures that all members of $\mathbb{P}_i^{\mathrm{PP}}$ indicate the same (maximal) level of immobility, which is desirable. The equal treatment of states in turn ensures that the same function m can be applied to any one of the sets $\mathbb{P}_i (i = 1, ..., k)$. Furthermore, symmetric measures are invariant to the ordering of states in the transition matrix. This also happens to be a desirable property if the state ordering decision is essentially arbitrary. Part (ii) of the definition asserts that predictability measures must increase whenever the probability mass becomes more concentrated, and must never rise when it becomes less concentrated. Thus, predictability is synonymous with the concentration of the probability mass of P_i, leading, ultimately, to $P_i \in \mathbb{P}_i^{\mathrm{PP}}$.

By implying invariance to state orderings, symmetry introduces a fundamental distinction between predictability and movement-based measures. Whereas the former possess symmetry as an innate characteristic, the latter do not, since the notion of movement by its nature attributes intrinsic importance to the positioning of states in the transition matrix. Therefore all movement measures necessarily presuppose that there is a natural ordering of states in the transition matrix.[5] This supposition is tenable if states are arranged according to a meaningful and unambiguous metric, e.g. prosperity, or a well defined and one-dimensional index of social class. However, it is common in the field of applied social mobility research to encounter situations when natural social orderings do not exist, or are not comparable across samples (see Bartholomew 1996, ch. 5, for some specific examples). Many occupational or social class classifications are functions of several different underlying variables, so a single unambiguous ordering which holds over all the variables will not exist.[6] Even when a meaningful ordering does exist, a symmetric treatment of the states is not of itself objectionable, because even some movement-based measures do this (i.e. those based just on the diagonal elements of P), and because unpredictability rather than movement is the focus of interest here.

Consequently, $m(P_i)$ cannot be expected to satisfy some particular properties built into certain movement-based measures.[7] For example, Shorrocks (1978) defines *monotonicity* (MO) as the attribution of greater immobility to matrix P than equal-sized Q if $p_{ij} \leqslant q_{ij}$ $\forall i \neq j$, with strict inequality for at least one $i \neq j$. MO favours matrices with greater off-diagonal elements: this is because they imply greater movement away from initial states. In contrast, symmetry allocates no special status to the diagonal elements.[8]

As a second example, consider the redistribution of probability mass within a transition matrix, such that the unit row sum constraints of the matrix are preserved. There are several variants of this type of transformation in the literature. For example, Atkinson *et al.* (1992) define a *diagonalizing switch* (DS) as a change that, for two states r and s, increases the proportions of individuals remaining in these states at the expense of reducing the proportions moving *between* r and s. Cowell (1985) discusses a similar property, called *monotonicity in distance*, in a non-Markovian context.[9] Dardanoni (1993) defines *a dynamic Pigou–Dalton transfer* (DPD) as a change that increases the probability of upward movement for those in lower ('worst') initial states while increasing the probability of downward movement for those in higher ('better') initial states. Dardanoni shows that this is equivalent to reducing the covariance between individuals' initial and lifetime status. Movement mobility is reduced by a DS and increased by a DPD; but there are obviously no necessary implications for m. Under symmetry, all states are treated equally, so there is no interest in shifting probability mass between states, except for its implications in concentrating it in a few states (recall Definition 2).[10]

Definition 2 sets out the essential characteristics of a predictability measure at the state level. However, it may be desirable for a practical measure to possess some additional properties. One such property is normalization to some well defined finite interval on the real line, with extreme values being given by PM and PP defined above. A second is invariance to the sample size, in cases where it is not possible to collect data from the entire population. It may also be desirable to map the whole of the transition matrix into an aggregate scalar measure, $M : \mathbb{P} \mapsto \mathbb{R}$, in such a way that M is clearly related to the state-by-state measure m introduced in Definition 2. An obvious source of candidates for M is the set of symmeteric functions that are monotonic increasing in their individual arguments, as it then follows that M will also be a predictability measure in the spirit of Definition 2. We will address these issues in the following section.

II. Two Sets of Predictability Measures

This section presents two sets of predictability measures: those based upon the population transition matrix, and those based upon a random sample from the population. The latter permit statistical inference about the predictability of the population transition matrix. In both cases we start by proposing a state-by-state measure, before generalizing it to an all-states-together, or 'aggregate', measure. This constructive approach allows us to exploit the property of 'period consistency', which is discussed in Section III.

Measures based on the population transition matrix

For situations in which it is possible to collect data on every member of the population we propose the following measure. Not only does this measure satisfy the predictability criteria, but it is also straightforward to verify that it is normalized to the unit interval, taking a minimum of 0 for $P_i \in \mathbb{P}_i^{\mathrm{PM}}$ and a maximum of 1 for all $P_i \in \mathbb{P}_i^{\mathrm{PP}}$.

Proposition 1 (Population state-by-state measure). The measure

$$m(P_i) := \frac{k \sum_{j=1}^{k} P_{ij}^2 - 1}{k - 1}$$

is a predictability measure according to Definition 2.

Proof. Symmetry is obvious. Let P_i' be P_i following the transfer of an infinitesimal amount of probability mass δ from state j to state $j' \neq j$. Then it is straightforward to show that

$$m(P_i) - m(P_i') = \frac{2k\delta}{k-1} [P_{ij} - (P_{ij'} + \delta)],$$

confirming that $m(P_i)$ as defined increases if and only if $P_{ij} \leqslant P_{ij'}$, as required. □

For an aggregate measure on $P \in \mathbb{P}$, we would suggest the sum of the state-by-state measures

$$(1) \qquad M(P) = \sum_{i=1}^{k} m(P_i) = \frac{k}{k-1} \left(\sum_{i,j} P_{ij}^2 - 1 \right),$$

for which $0 \leqslant M(P) \leqslant k$, with the two extremes corresponding to $P \in \mathbb{P}^{\mathrm{PM}}$ and $P \in \mathbb{P}^{\mathrm{PP}}$, as before. The aggregate function $M(P)$ clearly enjoys the same predictability properties as $m(P_i)$.

Measures based upon samples

Measures of mobility that fulfil the criteria set out in Definition 2 presuppose that the whole population has been surveyed, as they are based on P. While there are situations in which this is possible, it is more likely that the data represent a finite sample of size n from a generally much larger population. Denote the sample, represented by transition frequencies, as X, and define

$$N = X\mathbf{1}_k, \qquad \hat{P} = (\mathrm{diag}(N))^{-1}X$$

to be, respectively, the numbers starting in each of the k states, and the sample transition matrix. We caution against the much-used practice of simply substituting $\hat{P}$ for P in population-based measures such as those we have described in the first subsection above. This discards the potentially useful information contained in N. We propose instead a second set of measures based explicitly within a sampling framework.

Assuming that X is a random sample, the distribution of X_i given n_i (component i of N) is multinomial:

$$(2) \qquad X_i \mid n_i \sim \mathrm{Mu}_k(P_i, n_i).$$

Consequently the asymptotic distribution of $\hat{P}_i$ is Gaussian:

$$\lim_{n_i \to \infty} \hat{P}_i \mid n_i \sim \mathrm{N}_k(\mu(P_i), n_i^{-1}\Sigma(P_i)), \tag{3}$$

where $\mu(P_i) = P_i$ and $\Sigma(P_i) = \mathrm{diag}(P_i) - P_i P_i^T$, from which it follows that the normal distribution in (3) is a singular distribution with $k-1$ degrees of freedom. Hence we have the usual chi-squared result,

$$n_i(\hat{P}_i - \mu(P_i))^T \Sigma^-(P_i)(\hat{P}_i - P_i) \overset{\text{asy}}{\sim} \chi^2_{k-1}, \tag{4}$$

where $\Sigma^-(P_i)$ is the Moore–Penrose generalized inverse of $\Sigma(P_i)$.[11]

We can use this sampling framework to define a sample-based predictability measure that will allow us to attach statistical significance to deviations from some benchmark value. As established by Definition 1, the two extreme cases are perfect mobility and perfect predictability. But, whereas the former has a unique characterization, the latter does not; therefore we compare our sample transition matrices against the benchmark of perfect mobility, but in the direction of perfect predictability. To put this another way, we require a predictability measure for $\hat{P}_i$ with a sampling distribution under the null hypothesis $P_i \in \mathbb{P}_i^{\mathrm{PM}}$. Our suggested sample-based measure is closely related to our population-based measure, $m(P_i)$.

Proposition 2 (Sample state-by-state measure). The measure

$$r(X_i) = n_i\left(k \sum_{j=1}^{k} \hat{P}_{ij}^2 - 1\right) = n_i(k-1)m(\hat{P}_i) \tag{5}$$

is a predictability measure with the property $r(X_i) \mid P_i \in \mathbb{P}_i^{\mathrm{PM}} \overset{\text{asy}}{\sim} \chi^2_{k-1}$, where the distribution is asymptotic with respect to n_i.

Proof. That $r(X_i)$ is the sample-equivalent of a predictability measure follows directly from its relation to $m(\hat{P}_i)$. To prove that $r(X_i)$ has the required distribution, rewrite it in the equivalent form

$$r(X_i) = n_i(\hat{P}_i - \mu)^T \Sigma^-(\hat{P}_i - \mu),$$

where

$$\mu = k^{-1}\mathbf{1}_k \qquad \Sigma_{ij}^- := \begin{cases} k-1 & i=j \\ -1 & i \neq j. \end{cases}$$

Note the general asymptotic result given in (4), substituting $k^{-1}\mathbf{1}_k$ for P_i in $\mu(P_i)$ and $\Sigma(P_i)$. It only remains to show that Σ^- is indeed the generalized inverse of

$$\Sigma(k^{-1}\mathbf{1}_k) = k^{-2}\begin{pmatrix} k-1 & -1 & \dots & -1 \\ -1 & k-1 & \dots & -1 \\ \vdots & \vdots & & \vdots \\ -1 & -1 & \dots & k-1 \end{pmatrix},$$

which can be verified by direct computation (see e.g. Searle 1982, p. 212 *et seq.*). □

Interestingly, and perhaps not surprisingly, the measure $r(X_i)$ proposed in Proposition 2 can be re-expressed as the much used chi-squared 'goodness-of-fit' statistic,

$$r(X_i) = \sum_{j=1}^{k} \frac{(X_{ij} - n_i/k)^2}{n_i/k}, \tag{6}$$

where n_i/k is the number expected in each state implied by the hypothesis $P_i \in \mathbb{P}_i^{\mathrm{PM}}$. Moreover, by the additive property of independent chi-squared distributions, the aggregate measure

$$R(X) = \sum_{i=1}^{k} r(X_i) \tag{7}$$

has a $\chi^2_{k(k-1)}$ distribution under the null hypothesis of $P \in \mathbb{P}^{\mathrm{PM}}$. Note that there are $k(k-1)$ degrees of freedom, as we are only conditioning upon the row sums, N.

Thus, Proposition 2 provides a new interpretation of the chi-squared statistic when applied to sample transition matrices: it is a *predictability* measure, respecting the two properties described in Definition 2, for which the null hypothesis is the extreme case where predictability of the population is at a minimum.

One of the features of our measure $r(X_i)$ is that it is increasing in n_i. This of course is an implication of its derivation from a sampling framework: when comparing two states with the same sample transition probabilities, the state with the larger number of cases will have a smaller p-value for $r(X_i)$, reflecting the fact that with more data the evidence against PM is stronger.[12] But where the difference in the number of cases is not important for the analysis (e.g. where the investigator wishes to utilize measures that are not sensitive to n_i), different states can be compared directly using $m(\hat{P}_i)$ instead. There is one case, however, where the sampling framework is crucial, which is the comparison of states in different transition structures. This raises the problem that the number of states in the transition structures may not be the same, so that the scaling of the measure by k becomes important. By quoting the p-values of $r(X_i)$ the different values of k are accounted for through the different sampling distributions.

III. MEASURES THAT SATISFY 'PERIOD CONSISTENCY'

A general class of measures

We have constructed measures $M(P)$ and $R(X)$ that evaluate the predictability of transition structures, either of the population or of a random sample from that population. It would be highly desirable if these measures, or something similar, satisfied the property of 'period consistency' (Shorrocks 1978).

Definition 3 (Period consistency). A function $U : \mathbb{P} \mapsto \mathbb{R}$ is period consistent if, for all $P, Q \in \mathbb{P}$,

$$U(P^s) \geqslant U(Q^s) \Leftrightarrow U(P) \geqslant U(Q), \qquad s = 1, 2, \ldots$$

If a mobility measure is period-consistent (PC), then it is possible to compare directly transition structures that span different periods of time. As Shorrocks points out, PC is a very demanding property, which is not possessed by any conventional mobility measure.[13] Nor is PC possessed by our two measures $M(P)$ or $R(X)$. However, our approach of formulating aggregate measures on the basis of state-by-state measures suggests the following procedure for creating PC measures of mobility.

Proposition 3 (PC aggregate dual). If P is a regular transition matrix in a time-homogeneous Markovian process, then any state-by-state measure $v(P_i)$ has a PC aggregate dual, $V^*(P) = v(\Pi(P))$, which is period-consistent, where $\Pi(P)$ is the equilibrium distribution of P (i.e. the standardized eigenvector of the largest eigenvalue of P).

Proof. The proposition can be verified by observing that, under the given conditions, $V^*(P)$ possesses the stronger property

$$V^*(P^s) \equiv v(\Pi(P^s)) = v(\Pi(P)) \equiv V(P),$$

which follows from $\Pi(P^s) = \Pi(P)$, giving rise to what we might term *strong period consistency*. It is clearly the only aggregate measure that can satisfy this property. □

Proposition 3 shows that the PC aggregate dual as defined above not only possesses period consistency, but also satisfies an attractive stronger version of this invariance property, whereby the time period leaves unchanged not only the orderings of mobility measures, but also the values of the individual measures themselves. As stated in the proof of the proposition, this arises because the measure is defined on the equilibrium distribution of the transition matrix, provided that this equilibrium distribution exists.

One drawback to a PC-aggregate dual is that the mapping from transition matrix to mobility measure is 'many-to-one', as can be illustrated by the two-state bi-stochastic transition matrix

$$P = \begin{pmatrix} \theta & 1-\theta \\ 1-\theta & \theta \end{pmatrix} \qquad 0 \leqslant \theta \leqslant 1.$$

The case of $\theta = 1/2$ corresponds to PM, while the extreme cases of both $\theta = 0$ and $\theta = 1$ correspond to PP. But whatever the value of θ, P has the same equilibrium vector $\Pi(P) = (1/2, 1/2)$, and would therefore be adjudged to be PM. *This highlights the importance of using the PC aggregate dual only in conjunction with the state-by-state measures, as the latter would be sensitive to the various values of θ.* Disagreement between the typical state-by-state value and the aggregate value then becomes a diagnostic that the transition matrix has an interesting and possibly unusual structure; it would also suggest that an aggregate measure defined on P or X would probably be preferable to one based on $\Pi(P)$, even though the former are not PC.[14]

PC-aggregate measures of predictability

We turn now to our predictability measures, $m(P_i)$ and $r(X_i)$. There are two possibilities. The first is where P is known or where the differences in the numbers of cases in the different states is considered to be immaterial (e.g. where the investigator wishes to utilize measures that are not sensitive to N). Then $M^*(P) = m(\Pi(P))$ and $M^*(\hat{P}) = m(\Pi(\hat{P}))$ are PC aggregate measures in cases where the conditions of Proposition 3 are thought to hold. In cases where these conditions do not hold, the PC property is not relevant, and the investigator should stick with the aggregate measures already described.

The other possibility is where P is not known and the difference in the numbers of cases in the different states is considered to be important. What is required is a sample-based PC measure. We require, in effect, the sampling distribution of some simple transformation of $m(\Pi(\hat{P})) \mid N$. Unfortunately, this is a problem, because the sampling distribution of eigenvectors is notoriously intractable. However, the following approximation is reasonable in situations to be discussed below.

Proposition 4 (PC-aggregate sample measure). The measure

$$R_1^*(X) = \bar{n}k(k-1)m(\Pi_1)$$

where $\Pi_1 = k^{-1}\hat{P}^T\mathbf{1}_k$ and $\bar{n}$ is the harmonic mean of $\{n_1, \ldots, n_k\}$, is a first-order approximation to the PC-aggregate measure based upon $m(\Pi(\hat{P})) \mid N$, with the sampling distribution

$$R_1^*(X) \mid P \in \mathbb{P}^{PM} \overset{\text{asy}}{\sim} \chi^2_{k-1},$$

where the limit is taken as $\min\{n_1, \ldots, n_k\} \rightarrow \infty$.

Proof. To see that Π_1 is a first-order approximation to $\Pi(P)$, note the general result that

$$\lim_{s \rightarrow \infty} \Pi_s = k^{-1}(P^s)^T\mathbf{1}_k = \Pi(P),$$

where we then replace P with its consistent ML estimator $\hat{P}$.

To find the sampling distribution of $R_1^*(X)$, write Π_1 as the equivalent form $k^{-1}(\hat{P}_1 + \cdots + \hat{P}_k)$, remembering that the vector P_i is defined as row i of P. Given n_i, each P_i has an asymptotic Gaussian distribution with mean μ and variance $(n_i)^{-1}\Sigma$, where μ and Σ were given in Proposition 2; furthermore, given N, the P_i are mutually independent. It follows that

$$\Pi_1 \mid N, P \in \mathbb{P}^{\text{PM}} \overset{\text{asy}}{\sim} \text{N}_k(\mu, (k\bar{n})^{-1}\Sigma).$$

To finish the proof, note that $R_1^*(X)$ can be written as the quadratic form:

$$R_1^*(X) = k\bar{n}(\Pi_1 - \mu)^T\Sigma^-(\Pi_1 - \mu),$$

in exactly the same manner as in Proposition 2. □

The proposed measure $R_1^*(X)$ is a reasonable approximation to the PC-aggregate measure whenever Π_1 is a reasonable approximation to $\Pi(P)$. This will be the case when (i) the least of the components of N is large, in which case $\hat{P}$ will be close to P, and (ii) when $\Pi_1 \approx \Pi(\hat{P})$, which it is always possible to assess directly given $\hat{P}$. In other circumstances it may be possible to determine the sampling distribution of $m(\Pi(\hat{P})) \mid N$ by simulation.

IV. Empirical Illustration

In this section the new measures are illustrated using a transition matrix constructed from Harbury and Hitchens's (1979) study of UK intergenerational wealth mobility. This particular example provides a succinct and clear distinction between the different concepts of mobility.

Harbury and Hitchens were interested in discovering whether the terminal wealth left by sons was related to the terminal wealth left by their fathers. Following an earlier approach of Wedgwood (1929), they drew random samples of sons and fathers who could be traced using probate records. Both 'forward tracing' (tracing sons of fathers who died in a given year) and 'backward tracing' (tracing fathers of sons who died in a given year) were used. Table 1 presents a sample transition matrix, X, based on backward tracing of fathers of sons who died in 1973. It is adapted from table 3.2 of Harbury and Hitchens (1979), and contains information on transitions between five wealth states.[15] All wealth values are given in constant prices, and so are comparable between fathers and sons.

Two points should be borne in mind when interpreting these data. First, the absence of information about individuals with terminal wealth below £10,000 (which is partly a consequence of the structure of death duties at the time of the sample) means that inference relates only to the 'moderately well off', and not to the UK resident population as a whole. Second, the data describe gross estates left by fathers, not net inheritances received by sons. Bequests and inheritances will often differ not only because of death duties, but also because more than one sibling shares an estate. Table 1 shows that the modal destination of sons whose fathers belonged to the very top wealth class is the class below.

Table 1
The Intergenerational Wealth Sample Transition Matrix

Fathers' terminal wealth group (TWG) £'000	Sons' TWG (1)	(2)	(3)	(4)	(5)	N
1 10– 25	16	12	10	13	2	53
2 25– 50	7	8	7	10	0	32
3 50–100	4	9	6	13	0	32
4 100–500	1	5	5	22	2	35
5 500+	0	1	1	13	1	16

Source: adapted from table 3.2 of Harbury and Hitchens (1979).

The transition matrix in Table 1 was used to compute movement, dependence and predictability measures of mobility. For movement we use Bartholomew's (1982) measure,

$$b(P_i) = \sum_{j=1}^{k} P_{ij} \mid i - j \mid \qquad B(P) = \sum_{i=1}^{k} \Pi_i(P) b(P_i),$$

and Shorrocks's (1978) measure,

$$s(P_i) = \frac{k}{k-1}(1 - P_{ii}) \qquad S(P) = k^{-1} \sum_{i=1}^{k} s(P_i),$$

where in both cases we have inferred the state-by-state measures from the aggregate measures.[16] We note that, as the wealth classes are of different widths, we may prefer symmetric measures such as $s(P_i)$ or our own $m(P_i)$ to measures such as $b(P_i)$ because the latter ascribes, implicitly, the same width to each state.

Table 2 presents the state-by-state and aggregate measures. The $m(\hat{P}_i)$ and $r(X_i)$ values indicate that predictability is greater for the high-wealth groups and smaller for the low-wealth groups. At a significance level of 5%, the null hypothesis of perfect mobility cannot be rejected for the second lowest wealth group. However, it can be rejected for all other wealth groups, with the top two wealth groups, 4 and 5, being particularly predictable. Interestingly, both Bartholomew's and Shorrocks's measures suggest that 4 is an immobile state and that 5 is a mobile state, because for both states most of the cases end in state 4. This highlights very clearly the difference between movement and predictability.

Towards the end of Table 2 we give the aggregate measures, although after observing the variation in the state-by-state measures we may well be cautious about summarizing the whole matrix in an aggregate value. The statistic $R(X)$ is highly significant, which is not surprising given the values of the state-by-state measure. Of more interest is the PC aggregate measure $R_1^*(X)$. From the final two columns of the table we can see that the approximation Π_1 is very

TABLE 2
MOBILITY MEASURES[a]

TWG	$b(\hat{P}_i)$	$s(\hat{P}_i)$	$m(\hat{P}_i)$	$r(X_i)$	$\Pi(\hat{P})$	Π_1
1 10 – 25	1.49	0.87	0.05	10.49*	0.11	0.13
2 25 – 50	1.06	0.94	0.07	8.94	0.19	0.19
3 50 – 100	0.94	1.02	0.12	15.19**	0.17	0.16
4 100 – 500	0.57	0.46	0.30	42.00**	0.49	0.48
5 500+	1.12	1.17	0.59	37.75**	0.03	0.03
Aggregate	0.85	0.89	1.13	114.37**		
$R_1^*(X)$				82.11**		

[a] Measures as defined in the text.
* indicates rejection of null hypothesis of perfect mobility at the 5% Type I error level; ** at 1% or less.

similar to $\Pi(\hat{P})$, and, as there are a reasonable number of cases starting in each of the five states, we may be confident that we have a period-consistent measure of predictability which indicates that P is very significantly different from perfect mobility.

The two measures $B(\hat{P})$ and $S(\hat{P})$ appear at first to give an inconsistent picture, because they measure different aspects of mobility. Bartholomew's B lies in the range 0 to $k-1$, although the upper limit could be attained only by a non-regular periodic process. This makes the observed value of 0.85 seem rather low. Shorrocks's S lies in the range 0 to 1, which makes 0.89 seem rather high.[17] Taken together, these two measures suggest that cases tend to move away from their initial state, but not very far. A similar difficulty of interpretation applies to the Sommers–Conlisk (1979) dependence measure $|\lambda_2|$, where λ is the vector of eigenvalues of $\hat{P}$, which measures the speed of convergence to the steady state and lies in the range 0 to 1. In our case this value is 0.39, which would seem to be indeterminate. The similarity of Π_1 and $\Pi(\hat{P})$ is a more compelling demonstration that convergence will be fairly quick, as this similarity depends upon the full eigenstructure of $\hat{P}$.

V. Conclusion

This paper has considered alternative concepts of mobility and issues in measuring it using discrete transition matrices. A new family of measures was proposed, based on the concept of mobility as unpredictability of movement, in contrast to the traditional treatment of mobility as distance travelled between states or the speed of convergence to equilibrium. The new measures can be used to measure predictability on both a state-by-state and an aggregate (i.e. whole-matrix) basis. The first set of measures is based on the transition matrix. The second set uses a sampling approach and permits statistical testing of the hypothesis of perfect mobility, providing a new justification for the use of the χ^2 statistic. The third set satisfies the demanding criterion of 'period consistency', including a measure that also follows a well-known sampling distribution to a first-order approximation. The approximation was found to be a good one in an empirical example based on UK intergenerational wealth data; this example also demonstrated the usefulness of the new measures to complement existing ones in the literature.

We hope that future researchers will find it useful to supplement the traditional measures of mobility with the new predictability measures proposed here. Apart from their advantages summarized above, the new measures can be expected to facilitate a more comprehensive understanding of the innate mobility of social structures.

ACKNOWLEDGMENTS

We would like to thank two anonymous referees of this Journal, David Bartholomew, and seminar participants at St Andrews, Brunel, Hull and the ESRC Research Centre on Micro-Social Change at Essex for helpful comments. The usual disclaimer applies.

NOTES

1. For examples of movement measures, see Bartholomew (1996); an example of a dependence measure is the modulus of the second largest eigenvalue of the transition matrix, suggested by Sommers and Conlisk (1979). See Bartholomew (1996, ch. 5) for an extended discussion of dependence, which is also referred to as 'social inheritance'.
2. Alternative approaches include measuring movement by differences between vectors of observations at different times (Fields and Ok 1996); and measuring dependence by correlation coefficients (Sommers and Conlisk 1979).
3. See Shorrocks (1976), Atkinson *et al.* (1992) and Atoda and Tachibanaki (1991) for some international evidence against this assumption in the context of income mobility.
4. Both perfect mobility and PI are widely adopted as benchmark cases in the literature (Prais 1955; Bartholomew 1982).
5. Note that this excludes those measures based on lack of movement, i.e. those that use the diagonal elements of P only (see Bartholomew 1996, ch. 5).
6. For example, it is a matter of subjective choice whether farmers who own their farm and employ labourers should be placed adjacent to a 'managerial and supervisory' group in the frame of the transition matrix, or adjacent to 'semi-skilled manual workers'. The problem arises because there are conflicting aspects of 'occupational class'; i.e. there is no unique metric for directing the classification.
7. The use of the word 'symmetry' in the present context should not be confused with its other meaning of *anonymity*. For example, Dardanoni (1993) considers the importance of symmetry as anonymity with regard to social welfare rankings of different mobility structures; and Shorrocks (1993) considers symmetry/anonymity a desirable property of income inequality and mobility measures in a non-Markovian context.
8. Shorrocks (1978) observed that in general MO conflicts with PM. But because MO is not a relevant property for predictability measures, the latter will be unaffected by any inconsistency of this sort.
9. See also King (1983) and Chakravarty (1984) for mobility measures based on the 're-ranking' of individuals within income distributions. Re-ranking involves individuals switching places in the distribution, which is taken as prima facie evidence of movement mobility.
10. One implication of this is that predictability measures will not in general be coherent with the type of welfare orderings characterized by Dardanoni (1993), since the latter are based explicitly on a particular ('bad' to 'good') state ordering.
11. For a full discussion, see e.g. Mardia *et al.* (1979, p. 41).
12. It is interesting to compare in this respect the favoured non-Markovian income mobility measure of Fields and Ok (1993), which is also an increasing function of the sample (or population) size. As here, those authors also proposed a complementary per capita measure that controls for the sample size. We are grateful to an anonymous referee for emphasizing the importance of the sample size issue.
13. Shorrocks does suggest some measures that satisfy the less demanding condition of 'period invariance'. In obvious notation, period invariance is defined as $U_{PI}(P;\ T) = U_{PI}(P^s;\ sT)$, for $s = 2, 3, \ldots$.
14. We are grateful to an anonymous referee for emphasizing the points raised in this paragraph.
15. Adaptation of Harbury and Hitchens's tabulation was needed to group the data on a common basis for sons and fathers, and to calculate the numbers of fathers starting in each state. An additional table given by Harbury and Hitchens (table 3.7a) provided the latter information, but for only three states and with a smaller sample size.
16. Another important (income mobility) movement measure has been proposed by Fields and Ok (1996). However, these authors show that a version of their measure is related to Bartholomew's; the 'raw' Harbury–Hitchens data required for computing their non-Markovian measure is unavailable.
17. Other grounds for supposing 0.89 to indicate 'relatively high mobility' come from comparisons with S values reported in other studies. These include Atkinson *et al.* (1983) and Dearden *et al.* (1997), though these studies analysed income rather than wealth groups.

REFERENCES

ATKINSON, A., MAYNARD, A. and TRINDER, C. (1983). *Parents and Children: Incomes in Two Generations*. London: Heinemann.

ATKINSON, A., BOUGUIGNON, F. and MORISSON, C. (1992). *Empirical Studies of Earnings Mobility*. New York: Harwood.

ATODA, N. and TACHIBANAKI, T. (1991). Earnings distribution and inequality over time: education versus relative position and cohort. *International Economic Review*, **32**, 475–89.

BARTHOLOMEW, D. J. (1982). *Stochastic Models for Social Processes*, 3rd edn. Chichester: John Wiley.

—— (1996). *The Statistical Approach to Social Measurement*. San Diego: Academic Press.

CHAKRAVARTY, S. J. (1984). Normative indices for measuring social mobility. *Economics Letters*, **15**, 175–80.

COWELL, F. (1985). Measures of distributional change: an axiomatic approach. *Review of Economic Studies*, **52**, 135–51.

DARDANONI, V. (1993). Measuring social mobility. *Journal of Economic Theory*, **61**, 372–94.

DEARDEN, L., MACHIN, S. and REED, H. (1997). Intergenerational mobility in Britain. *Economic Journal*, **107**, 47–66.

FIELDS, G. S. and OK, E. A. (1996). The meaning and measurement of income mobility. *Journal of Economic Theory*, **71**, 349–77.

HARBURY, C. D. and HITCHENS, D. M. (1979). *Inheritance and Wealth Inequality in Britain*. London: Allen & Unwin.

KING, M. A. (1983). An index of inequality: with applications to horizontal equity and social mobility. *Econometrica*, **51**, 99–115.

MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. London: Academic Press.

PRAIS, S. J. (1955). Measuring social mobility. *Journal of the Royal Statistical Society*, **A118**, 56–66.

SEARLE, S. (1982). *Matrix Algebra Useful for Statistics*. New York: John Wiley.

SHORROCKS, A. F. (1976). Income mobility and the Markov assumption. *Economic Journal*, **86**, 556–78.

—— (1978). The measurement of mobility. *Econometrica*, **46**, 1013–24.

—— (1993). On the Hart measure of income mobility. In M. Casson and J. Creedy (eds.), *Industrial Concentration and Economic Inequality*. Cheltenham: Edward Elgar.

SOMMERS, P. M. and CONLISK, J. (1979). Eigenvalue immobility measures for Markov chains. *Journal of Mathematical Sociology*, **6**, 253–76.

WEDGWOOD, J. (1929). The influence of inheritance on the distribution of wealth: some new evidence. *Economic Journal*, **38**, 38–55.