Diskussionsschriften

A Large Deviation Approach to the Measurement of Mobility

Robert Aebi
Klaus Neusser
Peter Steiner

02-20

December 2002

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

Version from 19.12.02

A Large Deviation Approach to the Measurement of Mobility

Robert AEBI, Klaus NEUSSER, and Peter STEINER

University of Bern

Abstract

We propose an approach to measure the mobility immanent in Markov processes. For this purpose, we distinguish between mobility in equilibrium and mobility associated with convergence towards equilibrium. The former aspect is measured as the expectation of a functional, defined on the Cartesian square product of the state space, with respect to the invariant distribution. Based on large deviations techniques, we show how the two aspects of mobility are related and how the second one can be characterized by a certain relative entropy. Finally, we show that some prominent mobility indices can be considered as special cases.

JEL classification: C22, J62

Keywords: mobility index, large deviations, relative entropy

Correspondence

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

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

A Large Deviation Approach to the Measurement of Mobility*

1. Introduction

The capacity or facility of movement from one state to another is an important characteristic of a stochastic process. It is therefore not surprising that there have been several attempts in the literature to capture this aspect in terms of a single so-called mobility index. As it turns out the notion of mobility is of a multifaceted nature so that different authors have approached it from different perspectives (see the surveys by Fields and Ok 1999 or Maasoumi 1998). Here we follow the spirit of Batholomew (1982) by interpreting mobility as movements between states.[1] The main advantage of approaching the notion of mobility from this angle is that it opens up interesting links to the estimation of mobility from actual data.

Following the traditional literature of mobility measurement, we base our analysis on time homogeneous Markov processes with finite state space. In this context, mobility indices are conventionally defined as a mapping from the set of transition matrices into the positive real numbers. In order to define meaningful indices, Shorrocks (1978) proposed an axiomatic approach and postulated a set of desirable axioms for mobility indices. Geweke, Marshall and Zarkin (1986) grouped these axioms into persistence-, convergence- and temporal aggregation criteria. Whereas several mobility indices are consistent with the persistence- and convergence criteria within a considerable class of transition matrices, none of them satisfies all three categories of criteria. Such inconsistencies had to be expected if one wants to condense a matrix into a single number. Obviously, different indices detect rather different aspects of mobility.

This paper is motivated by the insight that the notion of mobility is comprised of two aspects: mobility which occurs in equilibrium and mobility associated with convergence towards

---

* We gratefully acknowledge the comments by Martin Wagner and by seminar participants in Rauischolzhausen, Vienna, Berlin and Basel.

1 Alternative interpretations view mobility as equalizing opportunity (Bénabou and Ok 2001), welfare enhancing (Atkinson 1983; Dardanoni 1993), or as inequality reducing (Maasoumi 1998). Although these approaches start from different concepts, they are not unrelated to each other. In section 2.3 we will highlight some of these relations.

equilibrium.[2] Most conventional mobility indices can be classified according to these two aspects into equilibrium and convergence indices. The fact that some well-known indices are based on the equilibrium or invariant distribution whilst others are based on the eigenvalues of the underlying transition matrix serves as our guideline to discriminate between the two classes. Mathematically speaking, equilibrium mobility indices can be expressed as expected values of appropriately defined functionals (called *mobility functionals*) where the expectation is taken with respect to the invariant distribution of the Markovian process. For monotone transition matrices and for a certain class of mobility functionals, the so-defined equilibrium mobility indices turn out to be consistent with Shorrocks' (1978) monotonicity axiom, Conlisk's (1990) weak D-criterion as well as with Dardanoni's (1993) partial ordering.

Mobility indices in the second class are defined with respect to the spectrum of the underlying transition matrix. Because the eigenvalues are related to the speed at which the process converges towards its invariant distribution, we label these indices as *convergence mobility indices*. They can be interpreted as measuring the dependence of being in some state in the future on today's state (Conlisk and Sommers 1979).

In view of this distinction, empirical studies should distinguish between the two aspects of mobility and, ideally, measure both of them. Using conventional indices, there exist many combinations of indices of the two types. Thus it is easy to devise examples which demonstrate that different combinations lead to different results, making the interpretation of mobility arbitrary (see e.g. Dardanoni 1993).

Consequently, our approach proposes a pair of complementary indices where each single index measures one of the aspects of mobility. We call them *equilibrium and period mobility index*. Equilibrium mobility refers to that part of mobility which occurs when the process is in its equilibrium. Period mobility is related to that part of mobility which is associated with the convergence of the process towards equilibrium. As it is not directly based on the spectrum of the transition matrix, we refer to it as period instead of convergence mobility. In fact, the two indices cannot be chosen independently from each other as they are intimately related in our approach through the choice of a *mobility functional*. This is just a rule of weighting

2 In the sociologically oriented literature the first aspect is sometimes called pure or exchange mobility as the spot-distribution remains unchanged in equilibrium. Bartholomew (1982) refers to these two aspects as measures of movements and measures of generation dependences.

movements between states. It implies that, in contrast to Shorrocks, our mobility functional is not defined on the set of transition matrices but on the Cartesian square product of the state space. The choice of a mobility functional determines both the equilibrium and the period mobility index simultaneously. This way of proceeding therefore eliminates the arbitrariness inherent in the choice of a pair of conventional indices.

Our *equilibrium mobility index* is then just the expected value of an appropriately defined (mobility) functional where the expectation is taken with respect to the invariant distribution of the underlying Markov process. We show that popular mobility indices, like Bartholomew's index or the index of unconditional probability of leaving the current class, are actually of this form. An application of the Ergodic theorem then implies that the time average of the mobility functional on a sample path converges to our equilibrium mobility index. Using large deviation techniques, we prove that this convergence takes places at an exponential rate where the rate of convergence is given by a specific relative entropy. This rate of convergence then gives rise to our measure of *period mobility*. Thus the period mobility index is based on the speed at which the mobility functional averaged over time converges to the corresponding equilibrium mobility index as time goes to infinity. It can be interpreted as the conditional probability of observing a large deviation from equilibrium mobility (i.e. above average mobility) given such a large deviation occurred in the past.

Although the corresponding large deviations principle is a special case of a much more general theory, we nevertheless state and prove a complete version of it. This makes the paper self-contained and therefore easily accessible to non-specialists.[3] A full-fledged development of the large deviation principle also allows to fully adapt the theory to the application we have in mind and prepares the ground for the numerical computations. We think that this way to proceed enhances the interpretability and comparability of empirical applications.

Although we presented our theory in terms of finite-state Markov processes, all our concepts can be easily extended to more general stochastic processes. The reason for this easiness in generalization is that our mobility indices are not defined on the transition matrices themselves as in the conventional literature, but as the expectation of some functional defined

---

[3] Hollander (2000) presents an excellent introduction to the Theory of Large Deviation. The technically inclined reader is referred to Miller (1961), Iscoe, Ney, and Nummelin (1985), Ney and Nummelin (1987a), and Ney and Nummelin (1987b) for applications to Markov Chains and to Dembo and Zeitouni (1998) for a general treatment of the subject.

on a stochastic process. Thus it is straightforward, for example, to expand our concepts to ARMA processes which would relate the literature on mobility measurement to the modern econometrics literature on the estimation of income processes (see Alvarez, Browning, and Ejrnæs 2001 and the references therein).

2. Definitions and Properties of the Equilibrium Mobility Index

2.1. Preliminaries

Our analysis is based on a discrete-time stochastic process $\{X_t\}$, $t = 0, 1, 2, \ldots$, where the random variables X_t take values in a finite state space $E = \{1,2,\ldots,K\}$. The indices i and j always denote generic states running from 1 to K. For some arbitrary initial probability distribution μ at $t = 0$, we assume that $\{X_t\}$ is a Markov chain with a stationary $K \times K$ transition matrix $P = P(i,j)$. The measure induced by the Markov chain on the set of trajectories E^∞ is denoted by $\mathbf{P}_\mu$.[4] Following the literature on mobility indices, we assume that the transition matrix is irreducible. With the additional assumption that $tr(P) > 0$, P becomes a primitive matrix (i.e. $\exists$ m: $P^m >> 0$;[5] Berman and Plemmons 1994, Corrolary 2.2.28) which implies that the Markov chain $\mathbf{P}_\mu$ is regular.[6]

Thus there exists a unique invariant or ergodic probability distribution π. Moreover, $\lim_{T\to\infty} \mu' P^T = \pi'$ for any probability distribution μ or equivalently $\lim_{T\to\infty} P^T = P^\infty$ where P^∞ is a transition matrix whose rows are all equal to π'. Moreover, $\rho(P) = 1$ is a simple eigenvalue greater in magnitude than any other eigenvalue.[7] Thus $\lambda \in \sigma(P)$ implies that $\lambda = 1$ or that $|\lambda| < 1$. The speed of convergence of P^T towards P^∞ as T goes to infinity is therefore governed by those eigenvalues with moduli strictly smaller than one. In particular, one can show that the *asymptotic speed of convergence* is given by $-\log \delta(P)$ where $\delta(P)$ is the second largest

---

4 When there is no confusion, we omit the index referring to the initial distribution.

5 We adopt the following notation: $A \geq B$ if $A(i,j) \geq B(i,j)$ for all i and j; $A > B$ if $A \geq B$ and $A \neq B$; $A >> B$ if $A(i,j) > B(i,j)$ for all i and j. $\sigma(A)$ and $\rho(A)$ denote the spectrum and the spectral radius of A.

6 The assumption $tr(P) > 0$ is slightly more restrictive than is actually necessary. Its purpose is to avoid the discussion of uninteresting degenerate cases. Practically all arguments carry over to primitive matrices.

7 The proofs of these implications can be found in any standard textbook on Markov chains (for example Berman and Plemmons 1994; Norris 1997; Seneta 1981; Rosenblatt 1974)

modulus of the eigenvalues of P, i.e. $\delta(P) = \max\{|\lambda| : \lambda \in \sigma(P) \text{ and } \lambda \neq 1\}$.[8] The asymptotic speed of convergence or any other commonly used mobility index based on $\sigma(P)$ can thus be related to the speed of convergence of P^T towards P^∞. Consequently, we may label them as *convergence mobility indices*. A list of the most commonly used indices is given in Table 1. This type of indices can be interpreted as measuring the dependence of future states on the initial state.

These indices capture, however, only one aspect of mobility as the stochastic process will still generate movements even if the distribution has already reached its equilibrium (i.e. is equal to its invariant distribution). We therefore consider a second class of indices, called *equilibrium mobility indices*. These indices are defined as expected values of some (mobility) functionals where the expectation is taken with respect to the invariant distribution. A formal definition is given in the next section.

This suggests that mobility of a stochastic process should actually be measured by two indices: one measuring mobility in equilibrium and one measuring the speed of convergence. We will show how these two measures can be related. In particular, we will prove that the choice of a mobility functional determines simultaneously an equilibrium and a convergence mobility index. In this way the arbitrariness in the choice of mobility indices is reduced to the choice of a mobility functional.

2.2. Definitions

In contrast to Shorrocks (1978) or Geweke , Marshall and Zarkin (1986), we do not define our mobility index directly on the set of transition matrices. Instead, more in the spirit of Bartholomew (1982, 24-30), we base our concept on the notion of a mobility functional which valuates on movements between states. This way of proceeding has the great advantage that the definitions of mobility indices proposed below can be easily generalized to other

---

[8] The asymptotic speed of convergence is defined as $-\log\alpha$ with $\alpha = \sup_{\mu} \lim_{T\to\infty} \left\| \mu' P^T - \pi \right\|^{1/T}$ where the supremum is taken over all initial distributions μ (Berman and Plemmons 1994, 172). It can be shown that the asymptotic speed of convergence equals $-\log \delta(P)$ in our case (Berman and Plemmons 1994, 199). Sommers and Conlisk (1979) proposed $\delta(P)$ as a measure of immobility, respectively $1 - \delta(P)$ as a measure of mobility.

stochastic processes. In particular, our approach can be extended to ARMA processes so that it ties in very well to the modern econometrics literature on modeling income processes (see Alvarez, Browning, and Ejrnaes 2001 and the references therein).

The equilibrium mobility index, we propose, is just the expected value of a functional taken with respect to the 2-dimensional invariant (ergodic or equilibrium) probability distribution of the Markov process. It measures the amount of mobility when the process is in equilibrium. For any realization or sample path of the process, the time average with respect to the mobility functional, called empirical mobility, converges to the equilibrium mobility index. We then show that convergence is achieved at an exponential rate. Furthermore, we give an explicit expression for this rate of convergence and show how it can be computed once a choice of a mobility functional has been made. This rate of convergence then gives rise to a convergence mobility index which we call period mobility index.

DEFINITION 1: A *mobility functional* f is a nonnegative functional on $E \times E$ such that

$$f(i,i) = 0 \quad \text{for all } i \in E \text{ and}$$
$$f(i,j) > 0 \quad \text{for all } i \text{ and } j \in E \text{ with } i \neq j$$

The mobility functional is therefore just a functional on $E \times E$ which attaches positive values or costs to movements from one state to another state and zero when no movement occurs. Thus the mobility functional provides some kind of "distance" between states. It is not necessarily a metric, in particular we do not require the symmetry of f: upward movements can be valued differently from downward movements. Note also that movements toward states which are "farther away" need not receive a higher value. From the viewpoint of equilibrium and period mobility a generalization to functionals f defined on higher powers than two of the state space E is not indicated. In fact, in the equilibrium described by the stationary probability distribution π, the Markov chain P_μ is entirely determined by its transition matrix P acting on the square of the state space.

Given a mobility functional, we then define the equilibrium mobility index as the expected value of this functional where the expectation is taken with respect to the invariant probability distribution:

DEFINITION 2: For any given mobility functional f on E×E and any primitive transition matrix P with its unique invariant distribution π,

$$M_f^e(P) = \sum_i \pi(i) \sum_j P(i,j) f(i,j)$$

is called the *equilibrium f-mobility index* of P.

The definition can be written more compactly as $M_f^e(P) = \mathrm{tr}(P' \mathrm{diag}(\pi) f)$ where f denotes the matrix with elements f(i,j).[9] The properties of f guarantee that $M_f^e(P) \geq 0$, but the index is not restricted to be smaller or equal than one. The normalization of the index to the interval [0,1] can be achieved if $M_f^e(.)$ is divided by a_{max}, a number which depends only on f.[10] It is easy to see that $M_f^e(I_K) = 0$ so that the index fulfills Shorrocks (1978, 1015) Immobility axiom. As we restrict ourselves to primitive transition matrices (which does not include I_K), the equilibrium mobility index is always strictly greater than zero. Hence the Strong Immobility axiom is fulfilled on the union of the set of primitive transition matrices and $\{I_K\}$. Because the equilibrium index measures mobility in a situation where the probability distribution remains unchanged over time (i.e. remains equal to π), it measures what is called pure exchange mobility in the sociologically oriented literature (see Dardanoni 1993; Fields and Ok 1999; Maasoumi 1998).

The definition of the equilibrium mobility index encompasses several specifications encountered in the literature. Consider first, the power functional: $f(i,j) = |i-j|^\alpha$. For $\alpha = 1$, the equilibrium mobility index specializes to Bartholomew's index:[11]

$$M_B^e(P) = \sum_i \pi(i) \sum_j P(i,j) |i-j|$$

The specification $\alpha = 2$ is also of some interest because it makes the mobility functional especially well suited for the analysis of real valued processes such as ARMA processes.

---

9 The diag(.) operator transforms any K×1 vector x into a K×K diagonal matrix with x on the diagonal.

10 $a_{max} = a_{max}(Q)$ where Q is any positive transition matrix (see section 3.2).

11 Bartholomew (1982) scaled this index by 1/(K–1) to confine it to the interval (0,1).

Another interesting choice for the mobility functional is f(i,j) = 1 – δ(i,j) where δ(i, j) denotes Kronecker's delta. This results in the index of unconditional probability of leaving the current class which is nothing but the expected number of class changes:[12]

$$M_U^e(P) = \sum_i \pi(i)(1 - P(i,i)) = \sum_i \pi(i) \sum_j P(i,j)\big(1 - \delta(i,j)\big).$$

The above functionals define actually a metric on the state space E: they are non-negative, symmetric, equal to zero if and only if the arguments coincide, and they satisfy the triangle inequality. While in the case of Bartholomew's index the functional expresses the ordinary distance between states i and j, the functional corresponding to the index of leaving the current class is known in topology as the trivial metric.

The definition of f is much too general and consequently does not impose enough structure on equilibrium mobility indices which would result in interesting properties. We therefore examine more closely the special class of, so-called, 2-decreasing mobility functionals. As it turns out, this class provides interesting connections to existing criteria and partial orderings of transition matrices.

DEFINITION 3: A mobility functional f on E×E is *2-decreasing* if

$$V(i,j) = f(i+1,j+1) - f(i+1,j) - f(i,j+1) + f(i,j) \le 0 \qquad \text{for all } i,j \in \{1,2,\ldots,K-1\}$$

Note that the inequality is strict if $i = j$. 2-decreasing functions can be considered as a two-dimensional analog of nonincreasing functions in one variable.[13] The above definition immediately implies that f(i+1,j) – f(i,j) and f(i,j+1) – f(i,j) are nonincreasing functions of j and i, respectively. The power functional is 2-decreasing provided $\alpha \ge 1$ whereas the functional f(i,j) = 1 – δ(i,j) is not.

2.3. Some Implications from 2-decreasing Mobility Functionals

Recently the class of monotone transition matrices has received special attention (Conlisk 1990; Dardanoni 1993; Dardanoni 1995; Fields and Ok 1999). Monotone transition matrices are transition matrices where row i+1 stochastically dominates row i for all i = 1, . . .,K-1.

---

12 In the literature, this index is usually scaled by K/(K–1).

13 See Nelsen (1999, 6). –V(i,j) can be interpreted as the area assigned by f to a rectangle with vertices (i,j), (i+1,j), (i,j+1), (i+1,j+1).

This condition can be written compactly as $T^{-1}PT \geq 0$ where T denotes the summation matrix.[14] It is argued that these matrices have theoretically plausible properties and are supported empirically.

PROPOSITION 1: For any two transition matrices P and Q with the same invariant distribution π and any 2-decreasing mobility functional f, $T' \operatorname{diag}(\pi)(P-Q)T \leq 0$ implies $M_f^e(P) \geq M_f^e(Q)$.

PROOF: Using the assumption that P and Q share the same invariant distribution, we get:

$$M_f^e(P) - M_f^e(Q) = \operatorname{tr}((P-Q)' \operatorname{diag}(\pi) f) = \operatorname{tr}\left((T' \operatorname{diag}(\pi)(P-Q)T)\left(T^{-1} f' T'^{-1}\right)\right).$$

The matrices in parenthesis have a special structure: $T' \operatorname{diag}(\pi)(P-Q)T = \begin{pmatrix} N & 0_{(K-1)\times 1} \\ 0_{1\times(K-1)} & 0 \end{pmatrix}$ with $N \leq 0$ by assumption and $\left(T^{-1} f' T'^{-1}\right) = \begin{pmatrix} V' & c \\ b & 0 \end{pmatrix}$ where the (K-1)×(K-1) matrix V is a nonpositive matrix with typical elements $V(i,j) = f(i,j) - f(i,j+1) + f(i+1,j+1) - f(i+1,j)$ and where b' and c are nonnegative K-1 vectors. $V \leq 0$ because f is 2-decreasing. This then implies

$$M_f^e(P) - M_f^e(Q) = \operatorname{tr}\left(\begin{pmatrix} N & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} V' & c \\ b & 0 \end{pmatrix}\right) = \operatorname{tr}(NV') \geq 0.$$

q.e.d.

Note that the implication goes only in one direction as we can give examples such that $M_f^e(P) \geq M_f^e(Q)$ with $T' \operatorname{diag}(\pi)(P-Q)T > 0$. Proposition 1 implies that the equilibrium mobility index induced by a 2-decreasing mobility functional is coherent with Dardanoni's (1993) partial ordering of monotone mobility matrices sharing the same invariant distribution π. Furthermore, in this class the perfect mobility matrix $\iota\pi'$, $\iota = (1, \ldots, 1)'$, is a maximal element with respect to equilibrium mobility because $T' \operatorname{diag}(\pi)(\iota\pi' - P)T \leq 0$ for all monotone transition matrices P with stationary probability distribution π (Dardanoni 1993, Theorem 2) and therefore $M_f^e(\iota\pi') \geq M_f^e(P)$ where f is 2-decreasing.

---

[14] The summation matrix T is an upper triangular matrix with all elements on the diagonal and above equal to one. Its inverse T^{-1} is the matrix with ones on the diagonal, minus ones on the first superdiagonal and zeros elsewhere.

COROLLARY 1: If P and Q are two transition matrices with the same invariant distribution π such that $P(i,j) \geq Q(i,j)$ for all $i \neq j$ and $P(i,j) > Q(i,j)$ for some $i \neq j$, then $T'\,\mathrm{diag}(\pi)\,(P-Q)\,T \leq 0$ which implies $M_f^e(P) \geq M_f^e(Q)$ if the mobility functional f is 2-decreasing.

PROOF: see Dardanoni (1993, Appendix 2) for the first assertion. The second follows from proposition 1.

COROLLARY 2: Consider two transition matrices P and Q with the same invariant distribution π and denote by $\Delta(P)$ and $\Delta(Q)$ the upper left $(K-1)\times(K-1)$ matrix of $T^{-1}PT$ and $T^{-1}QT$, respectively, then $\Delta(P) \geq \Delta(Q)$ implies $T'\,\mathrm{diag}(\pi)\,(P-Q)\,T \leq 0$ which implies $M_f^e(P) \geq M_f^e(Q)$ if the mobility functional f is 2-decreasing.

Proof: see Dardanoni (1993, Appendix 2) for the first assertion. The second follows from proposition 1.

These corollaries show that in the case of monotone mobility matrices, our equilibrium mobility index with a 2-decreasing mobility functional is coherent with Conlisk's (1990) weak D-criterion as well as with Shorrocks' (1978) monotonicity axiom and thus satisfies all persistence criteria (see Geweke, Marshall and Zarkin (1986)).

2.4. Empirical Mobility

The empirical counterpart to the equilibrium mobility index is just the time average over consecutive $f(X_{t-1},X_t)$'s. We call this average the empirical f-mobility.

DEFINITION 4: For any Markov process, $\{X_t\}$, defined on the state space E and a mobility functional f on E×E, the time average S_T of $f(X_{t-1},X_t)$ on $\{0, 1, \ldots, T\}$

$$S_T = \frac{1}{T}\sum_{t=1}^{T} f(X_{t-1},X_t)$$

is called the *empirical f-mobility* up to period T.

Note that the assumption $\mathrm{tr}(P) > 0$ precludes the degenerate case that S_T is constant over all possible realizations.

THEOREM 1: The empirical f-mobility converges to the following limit:

$$\lim_{T\to\infty} S_T = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} f(X_{t-1}, X_t) = \sum_i \pi(i) \sum_j P(i,j) f(i,j) = M_f^e(P) \qquad \mathbf{P}_{\boldsymbol{\mu}}\text{–a.s.}$$

for every initial distribution μ and any primitive transition matrix P.

PROOF: This is just an application of the Ergodic theorem (see Seneta 1981 chap. 4; Rosenblatt 1974, chap V,b) to a function f(.,.) defined on two consecutive states.

In case of Bartholomew´s functional, the empirical f- mobility is just the average number of class changes. In case of the index of leaving the current class, it is the average number of changes of the current class. The above Theorem thus implies that one can estimate the equilibrium f-mobility index directly from sample paths without estimating in a prior step the transition matrix of the process as would be necessary in the traditional approach.

3. Large Deviations of Mobility Functionals

3.1. The Perron-Frobenius transformation

In this section we establish that the tail probabilities of the distribution of empirical f-mobility converge to zero at an exponential rate. The derivation of this result and the explicit expression of the rate of convergence will then serve as the key tools in the analysis of convergence mobility. This analysis will then naturally lead to a convergence mobility index, called period f-mobility index. This requires, however, additional concepts which we will now introduce.

DEFINITION 5: Let **P** and **Q** be two regular Markov chains with corresponding transition matrices P and Q and invariant distributions π_P and π_Q. If **Q** is absolutely continuous with respect to **P** (or equivalently, if P(i,j) = 0 implies Q(i,j) = 0), the *relative entropy of **Q** with respect to **P** up to period T*, $H_T(\mathbf{Q} \mid \mathbf{P})$, is defined on the σ-algebra $A_T = \sigma(X_t\,, 0 \le t \le T)$ by

$$H_T(\mathbf{Q} \mid \mathbf{P}) = \int \log \left.\frac{d\mathbf{Q}}{d\mathbf{P}}\right|_{A_T} d\mathbf{Q}\ .$$

The Radon-Nikodym derivative of **Q** with respect to **P** on A_T is defined as

$$\left.\frac{d\mathbf{Q}}{d\mathbf{P}}\right|_{A_T} = \frac{\pi_Q(X_0)\,Q(X_0,X_1)\cdot\ldots\cdot Q(X_{T-1},X_T)}{\pi_P(X_0)\,P(X_0,X_1)\cdot\ldots\cdot P(X_{T-1},X_T)} \qquad \mathbf{Q}\big|_{A_T} - \text{a.s.}$$

Moreover, *the specific relative entropy of the transition matrix Q with respect to P per period-unit*, h(Q | P), is defined as

$$h(Q \mid P) = \sum_i \pi_Q(i)\, H_1\big(Q(i,.)\big\|P(i,.)\big) = \sum_i \pi_Q(i) \sum_j Q(i,j) \log \frac{Q(i,j)}{P(i,j)}$$

DEFINITION 6: For a given mobility functional f and any $\beta \in \Re$, the *Perron-Frobenius transformation* of an irreducible transition matrix P, denoted by $P_\beta = T_\beta(P)$, is defined by the matrix

$$P_\beta(i,j) = \frac{A_\beta(i,j)\, r_\beta(j)}{\lambda(\beta)\, r_\beta(i)}$$

where $A_\beta(i,j) = P(i,j)\exp(\beta f(i,j))$ and where $r_\beta \neq 0$ is a right eigenvector associated with $\lambda(\beta)$, the largest positive eigenvalue of A_β. The set of matrices $\{P_\beta\} = \{P_\beta \mid \beta \in \Re\}$ is called the *exponential Perron-Frobenius family of* P.

The Perron-Frobenius transform of P, P_β, is also called the twisted transition matrix. Taking $\beta > 0$, the matrix A_β is obtained from P by inflating those entries of P which have a corresponding positive value of f(i,j). The higher the value of the corresponding f(i,j), the stronger the inflation of P(i,j). The diagonal elements P(i,i) remain unchanged because f(i,i) = 0. As A_β is not a transition matrix anymore, we renormalize it to obtain the transition matrix P_β. From the construction it is intuitively clear that the twisted transition matrix, as long as $\beta > 0$, is more mobile than the original one. Moreover, as β increases, the equilibrium mobility index of P_β increases. Before we provide exact proofs of these assertions, we establish that the Perron-Frobenius transform is well defined for any irreducible transition matrix P.

PROPOSITION 2: For any irreducible transition matrix P, A_β and the Perron-Frobenius transform of P, P_β, both defined in Definition 6, have the following properties:

(i) A_β is irreducible. Thus $\lambda(\beta)$ is a simple eigenvalue equal to $\rho(A_\beta)$. To this eigenvalue correspond a left and a right eigenvector, ℓ_β and r_β respectively, such that $\ell_\beta >> 0$, $r_\beta >> 0$, and $\ell'_\beta r_\beta = 1$. If P is primitive then A_β is also primitive.

(ii) $P_\beta = R_\beta^{-1} \dfrac{A_\beta}{\lambda(\beta)} R_\beta > 0$ with $R_\beta = \mathrm{diag}(r_\beta)$ is an irreducible stochastic matrix with unique invariant distribution π_β equal to $R_\beta \ell_\beta$. If P is primitive then P_β is also primitive.

(iii) If P is primitive, $\lim_{T\to\infty} P_\beta^T = \lim_{T\to\infty} R_\beta^{-1} \left(\dfrac{A_\beta}{\lambda(\beta)}\right)^T R_\beta = \iota\pi'_\beta$ where $\iota = (1, \dots, 1)'$. Or, equivalently, $A_\beta^{(T)}(i,j) = \lambda(\beta)^T r_\beta(i)\ell_\beta(j)\left[1 + O\left(\delta_\beta^T\right)\right]$ with $0 < \delta_\beta < 1$.[15]

PROOF: These are standard results based on the Perron-Frobenius theorem and can be found, for example, in Seneta (1981) or Berman and Plemmons (1994).

From (i) we see that r_β cannot have a zero coordinate. Thus a division by zero in the definition of P_β is impossible. (ii) implies that the Perron-Frobenius transformation defines an operator on the set of irreducible (primitive) transition matrices. We next summarize the properties of $\lambda(\beta)$.

PROPOSITION 3: For any irreducible transition matrix P with $\mathrm{tr}(P) > 0$, $\lambda(\beta)$, as defined in Definition 6, has the following properties:

(i) The domain of $\lambda(\beta)$ is $\Re$.

(ii) $\lambda(0) = 1$.

(iii) $\lambda(\beta)$ is strictly increasing.

(iv) $\lambda(\beta)$ is analytic.

(v) $\lambda(\beta)$ and $\log \lambda(\beta)$ are strictly convex.

(vi) $\dfrac{\lambda'(\beta)}{\lambda(\beta)} = \sum_i \pi_{P_\beta}(i) \sum_j P_\beta(i,j) f(i,j) = M_f^e(P_\beta)$ where π_{P_β} is the invariant probability distribution of P_β. In particular, $\dfrac{\lambda'(0)}{\lambda(0)} = \lambda'(0) = \sum_i \pi_P(i) \sum_j P(i,j) f(i,j) = M_f^e(P)$.

[15] We denote $A_\beta^{(T)}(i,j)$ as the (i,j)-th element of the Matrix A_β^T.

PROOF: see appendix.

Note that the assumption $tr(P) > 0$ ensures that $\log \lambda(\beta)$ cannot be linear and is therefore strictly convex and not just convex. Assuming P to be a primitive matrix is not sufficient as shown by some counterexamples. From (v) and (vi) we see that $M_f^e(P_\beta)$ increases in β because $\lambda'(\beta)/\lambda(\beta)$ being the derivative of the convex function $\log \lambda(\beta)$ is an increasing function.

3.2. *Maximal Large Deviation*

We then want to characterize the maximal empirical mobility, denoted by $a_{max}(P)$, which can be achieved with positive probability. For this purpose, consider the directed graph associated to the matrix P. [16] This graph consists of the vertices $V_1, ..., V_K$ where an edge leads from V_i to V_j if and only if $P(i,j) \neq 0$. A path Π of length N in this graph is then just a sequence $\Pi = \{V_{i_0}, V_{i_1}, \cdots, V_{i_N}\} = \{i_0, i_1, \cdots, i_N\}$ such that $P(i_{n-1}, i_n) \neq 0$ for all n = 1, ..., N. In analogy to the definition of the empirical f-mobility, we assign to each path $\Pi = \{i_0, i_1, ..., i_N\}$ a number $s = s(\Pi)$ as follows:

$$s = s(\Pi) = s(\{i_0, i_1, \cdots, i_N\}) = \frac{1}{N}\sum_{n=1}^{N} f(i_{n-1}, i_n).$$

It is easily checked that the maximal value of s over all paths, $a_{max}(P)$, is given by

$$a_{max}(P) = \max_{\{\text{all circuits}\}} \frac{1}{N}\sum_{n=1}^{N} f(i_{n-1}, i_n) \quad < \infty$$

where a circuit is a path $\{i_0, i_1, ..., i_N\}$ such that $i_1, ..., i_N$ are distinct but $i_0 = i_N$. The maximum must be achieved by a circuit of length $N \geq 2$ because we assume that $f(i,i) = 0$ for all i. It is clear that the value of a_{max} does not depend on the actual probabilities in P, but only on the position of zero entries. Thus equivalent transition matrices must necessarily have the same $a_{max}(P)$.[17] In particular, all positive transition matrices P, i.e. $P >> 0$, have the same $a_{max} = a_{max}(P) \geq a_{max}(Q)$ where Q is any other transition matrix. Thus, there exists a maximal

[16] See Berman and Plemmons (1994) for further details.

[17] Two transition matrices P and Q are equivalent if and only if $P(i,j) = 0$ implies $Q(i,j) = 0$ and $Q(i,j) = 0$ implies $P(i,j) = 0$.

a_{max} that depends on the mobility functional f only and that equals $a_{max}(P)$, where P can be any positive transition matrix.

3.3. *The Large Deviation Principle*

We are now in a position to state our main theorem. At this point, we want to emphasize again that the mathematical result is actually not new as it can be deduced as a special case from a general theory (Iscoe, Ney, and Nummelin 1985; Ney and Nummelin 1987a, 1987b; Dembo and Zeitouni 1998). As we stress the application and computational aspects, we nevertheless state and prove a version of the large deviation theorem which is self-contained and fully adapted to the usage we have in mind.

PROPOSITION 4: For any irreducible transition matrix P with $tr(P) > 0$, the Legendre-Fenchel transform I(a) of $\log \lambda(\beta)$ is given for any $a \in \left(M_f^e(P), a_{max}(P)\right)$ by

$$I(a) = -\inf_{\beta\in\Re}\left(\log\lambda(\beta) - a\beta\right) = \sup_{\beta\in\Re}\left(a\beta - \log\lambda(\beta)\right) = a\beta(a) - \log\lambda(\beta(a))$$

where $\beta(a)$ is positive, finite, and unique.

PROOF: see appendix.

THEOREM 2: For any $a \in \left(M_f^e(P), a_{max}(P)\right)$, there exists a unique $\beta(a) \in \Re$ and a Perron-Frobenius transform of P, $P_{\beta(a)} = T_{\beta(a)}(P)$, such that

(i) $$\lim_{T\to\infty}\frac{1}{T}\log P\left\{S_T = \frac{1}{T}\sum_{t=1}^{T} f(X_{t-1}, X_t) \geq a\right\} = -I(a) = -\sup_{\beta\in\Re}\left(\beta a - \log\lambda(\beta)\right) = -h(P_{\beta(a)} \mid P)$$

(ii) $$M_f^e\left(P_{\beta(a)}\right) = \sum_i \pi_{P_{\beta(a)}}(i)\sum_j P_{\beta(a)}(i,j)f(i,j) = a\,.$$

PROOF: see appendix

Note that the assumption $tr(P) > 0$ guarantees that $a_{max}(P) > M_f^e(P)$ so that there always exists a non-trivial large deviation level $a > M_f^e(P)$. This Theorem shows that the tail probabilities of the distribution of S_T, the empirical f-mobility (see Definition 4), decline exponentially fast towards zero. For large T, the speed of convergence approaches a constant equal to the relative entropy of the Perron-Frobenius transform of P, $P_{\beta(a)}$, with respect to P. The larger $h(P_{\beta(a)}|P)$ the quicker this convergence takes place. The second part of the Theorem shows that, depending on the fixed large deviation level a $\in (M_f^e(P), a_{max}(P))$, $\beta(a)$ and therefore the distortion of P is chosen such that the twisted transition matrix, $P_{\beta(a)}$, has exactly an equilibrium mobility index of a.

Suppose that we want to compare the speed of convergence of two transition matrices P and Q and suppose further, for the sake of the argument, that they have the same equilibrium mobility index. Then it seems plausible to view the transition matrix P as being more mobile than Q if the event $\{S_T \geq a\}$ is more probable under P than under Q. For large T, this is equivalent to saying that $h(Q_{\beta(a)}|Q)$ is larger than $h(P_{\beta(a)}|P)$ which means that the distortion necessary to achieve the given large deviation level a is larger for Q than for P. This reasoning leads in the next section to the definition of a convergence mobility index associated with f which we call period f-mobility index.

Although we elaborated our theory in the context of finite state Markov processes, it is straightforward to extend our concepts to general state spaces and to more general Markov processes. Such extensions, however, require a highly technical and analytical machinery to establish even the most basic theorems. Moreover, some restrictions have to be placed on the Markov process to guarantee that the Legendre-Fenchel transform or respectively the convex conjugate of $\log \lambda(\beta)$ are finite on some interval. These restrictions typically take the form of a uniform "recurrence" condition (Kim and David 1979, assumption A4; Iscoe, Ney, and Nummelin 1985, 377). Lemma 1 in the appendix shows that such a condition is indeed automatically satisfied in the context of a finite state space.

4. Period mobility and examples

4.1. Period Mobility Index

Based on the reasoning from the previous section, we propose to define a convergence mobility index associated with f which we call period mobility index as follows:

DEFINITION 7: Given a large deviation level a, $M_f^e(P) < a < a_{max}(P)$, the *period f-mobility index*, $M_f^p(P \mid a)$, is defined as

$$M_f^p(P \mid a) = \exp\{-h(P_{\beta(a)} \mid P)\}$$

where $P_{\beta(a)}$ is the Perron-Frobenius transform of P with the property $M_f^e\left(P_{\beta(a)}\right) = a$ (see Theorem 2). By convention the index equals zero if $M_f^e(P) = a_{max}(P)$.

Straightforward arguments show that our period mobility index $M_f^p(P \mid a)$ is nothing but the asymptotical probability for T to infinity of continued large deviations above level a within one unit of time:

$$M_f^p(P \mid a) = \lim_{T \to \infty} P\left\{S_{T+1} \geq a \,\middle|\, S_T \geq a\right\} .$$

Since the index corresponds to a probability, it lies automatically between 0 and 1. Values near 0 mean low, values near 1 mean high period mobility.

The above definition makes clear that our period mobility index depends on the choice of the large deviation level a, the underlying equilibrium mobility index as well as a_{max}. We propose to consider large deviation levels a such that the difference between a and $M_f^e(P)$ is a fraction γ of the difference between the maximal large deviation level, $a_{max}(P)$, and the long-run equilibrium level $M_f^e(P)$:

$$a = a(\gamma) = M_f^e(P) + \gamma\left(a_{max}(P) - M_f^e(P)\right) \qquad \text{where } 0 < \gamma < 1.$$

Therefore by construction, the value of γ can be chosen independently of P. As we compare different transition matrices at the same fraction γ of their respective maximal possible large deviation from equilibrium mobility, we take a relative point of view. This is so because a large deviation level of e.g. $\gamma = 0.5$ leads only to the same a, if the matrices under consideration share the same equilibrium index as well as the same a_{max}.

4.2. Examples

We are now in a position to illustrate our approach. Throughout this illustration, we used the Bartholomew-functional $f = |i-j|$ (see section 2.2) to determine the respective equilibrium and period f-mobility indices. For this purpose we selected the following five test transition matrices:

$$P_1 = \begin{pmatrix} 0.6 & 0.35 & 0.05 \\ 0.35 & 0.4 & 0.25 \\ 0.05 & 0.25 & 0.7 \end{pmatrix} \qquad P_2 = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.2 & 0.7 \end{pmatrix} \qquad P_3 = \begin{pmatrix} 0.6 & 0.399 & 0.001 \\ 0.301 & 0.4 & 0.299 \\ 0.099 & 0.201 & 0.7 \end{pmatrix}$$

$$P_{\text{mobile}} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \qquad P_{\text{ident}} = \begin{pmatrix} 0.998 & 0.001 & 0.001 \\ 0.001 & 0.998 & 0.001 \\ 0.001 & 0.001 & 0.998 \end{pmatrix}$$

The first two transition matrices, P_1 and P_2, have been introduced by Dardanoni (1993). The third matrix P_3 is a positive analogon to the third, non-positive matrix used in Dardanoni's examples.[18] Dardanoni used these matrices to document the inconsistency between alternative mobility indices. In the following, we call those matrices the Dardanoni-matrices. They share the particularity that their Bartholomew mobility index is the same. P_1 and P_3 even have the same index of unconditional probability of leaving the current class as well as the same Prais and eigenvalue index. Because these matrices share the same Bartholomew index as well as the same maximal level of large deviations a_{max}, this leads to a comparison of their period mobility indices in absolute terms. In general, it is obvious from the formula suggested to compute the large deviation level $a = a(\gamma)$ that for positive matrices sharing the same equilibrium index the comparison of period mobility indices is the same whether one takes an absolute or a relative view.

Besides the Dardanoni-matrices, we examined two other test transition matrices. The matrix P_{mobile} has rows equal to its invariant distribution, (1/3, 1/3, 1/3)′. Transition matrices with equal rows are commonly described as perfectly mobile because the probability of moving to any class is independent of the state initially occupied. Finally, the matrix P_{ident} denotes a transition matrix close to the identity matrix and is thus considered as representing a Markov process with high persistence, that is with a low probability to move to a different state.

Note further that all transition matrices share the same invariant distribution (1/3, 1/3, 1/3)′. As all matrices have strictly positive entries, their a_{max} is the same and equal to two. A circuit which achieves a_{max} is {1,3,1}. Table 2 summarizes the characteristics of all test transition matrices.

---

18 The original third matrix by Dardanoni (1993) had a zero-entry in position (1,3). We substituted this matrix by its positive analogue P_3 in order to compare positive matrices only.

In the following discussion we denote by $\prec_e$ and $\prec_p$ the relations "is strictly less mobile than" where the indices e and p indicate if the relation is defined with respect to the equilibrium or the period mobility index. The equilibrium mobility index, in this case Bartholomew's mobility index, leads to the following ranking: $P_{ident} \prec_e P_1 \approx_e P_2 \approx_e P_3 \prec_e P_{mobile}$ where $\approx_e$ indicates equivalence concerning equilibrium mobility (see Table 2). Thus according to the equilibrium mobility index, P_{ident} is the least mobile process followed by P_1, P_2, and P_3 which are equivalent. P_{mobile} turns out to be the most mobile process. This ranking thus conforms to our intuition.

For any γ, $0 < \gamma < 1$, with corresponding large deviation level $a = a(\gamma)$, we compute numerically the supremum of $a\beta$ - log $\lambda(\beta)$ over $\beta \in \Re$ (see Proposition 4). If the supremum is attained at $\beta(a)$, we calculate the twisted transition matrix $P_{\beta(a)}$, as defined in Definition 6. Theorem 2 then implies that this twisted transition matrix satisfies the relation $M_f^e(P_{\beta(a)}) = a$. As the function to be optimized is strictly convex and possesses a unique supremum, the actual computations are free from numerical complications.[19] In Figure 1 we have plotted our period mobility index as a function of γ. This figure shows that the ranking $P_{ident} \prec_p P_3 \prec_p P_1 \prec_p P_2 \prec_p P_{mobile}$ is independent of γ. In Table 2 we report the actual values of the index for the arbitrary value $\gamma = 0.5$. This means that we propose to measure period mobility at a large deviation level halfway between a_{max} and the value of the equilibrium mobility index. The period mobility index leads to a ranking in which P_{mobile} is the most mobile process whereas P_{ident} is the least mobile process. That is, independent of the large deviation level a under consideration, chances for further movement and thus further mobility are much higher for P_{mobile} than for the process P_{ident}. Looking at the Dardanoni-matrices P_1, P_2 and P_3, we are now in a position to make a valid statement concerning the two aspects of mobility. Whereas the matrices do not differ concerning equilibrium mobility, it is possible to rank these matrices with respect to period mobility. Table 2 shows that for a level γ of 50 % of the maximal possible range of large deviations, P_2 clearly is the most mobile process whereas P_3 is least mobile. In our view, this kind of analysis clearly enhances the depth and the quality of statements concerning mobility of Markov processes.

---

19 MATLAB routines are available from the authors.

Table 2 gives an overview of the results for all convergence indices mentioned in this paper. To strengthen our point, we report in Table 3 the rankings of the three Dardanoni-matrices. In order to distinguish between the conventional convergence indices and our period mobility index, we use the symbols $\approx_c$ and $\prec_c$ to indicate "is equivalent", respectively "is strictly less mobile", according to the corresponding conventional convergence index.

While it's impossible to rank Dardanoni's matrices with the equilibrium index, the matrices are different concerning their convergence mobility and can thus be ranked with respect to all convergence indices. If one likes to capture both aspects of mobility, the resulting rankings are completely arbitrary and depend heavily on the choice of combination of equilibrium and convergence indices.

The virtue of our approach is that it reduces this arbitrariness to the choice of a mobility functional. Moreover, we think that the specification of a mobility functional is straightforward given a particular application in mind. This decision then determines the pair of indices which captures both equilibrium and convergence mobility.

5. Conclusions

This paper has shown how the choice of a mobility functional can simultaneously determine an equilibrium and a period mobility index. The equilibrium mobility index was defined as the expected value of the mobility functional evaluated with respect to the invariant distribution. By restricting the class of mobility functionals to so-called 2-decreasing functionals, the equilibrium mobility index is consistent with Dardanoni's (1993) partial ordering of monotone mobility matrices sharing the same invariant distribution. The period mobility index is related to the speed at which the tail probabilities of the empirical mobility converge to zero. Given a large deviation from equilibrium mobility, this convergence takes place at an exponential rate which can be expressed as the relative entropy of a twisted relative to the original transition matrix. We also showed how this relative entropy can be computed using the Legendre-Fenchel transform of the logged eigenvalues of some matrix family. As the numerical computations are easily implemented we suggest to report both the equilibrium and the period mobility index. This guarantees that both aspects of mobility are taken into account.

The measurement of mobility has thus been reduced to the specification of a mobility functional. This way of proceeding presents several advantages. First, the weighting of movements between states by a mobility functional seems to us a natural starting point which facilitates the interpretation and evaluation of mobility. Second, the arbitrariness inherent in the measurement of both aspects of mobility with conventional indices is reduced. Third, the approach can be readily extended to more general state spaces and Markov processes. For example, taking the power functional with $\alpha = 2$, both mobility indices can be readily extended to ARMA processes. In this way, our approach links the literature on mobility measurement to the modern panel data estimation techniques of income processes. Finally, our approach also suggests a way to estimate directly both aspects of mobility so that the effort of fitting transition matrices or ARMA models to the data can be avoided.

Appendix: Proofs

LEMMA 1: Any primitive transition matrix P and any functional f on E×E satisfy the following UNIFORM RECURRENCE CONDITION (R): There exists a positive integer m such that

$$\eta = \min_i \sum_j P\{X_m = j, S_m > a | X_0 = i\} > 0 \quad \text{for } a \in \left(M_f^e(P), a_{\max}(P)\right).$$

PROOF: We will show that for all i and j there exists a path Π which leads from i to j such that s(Π) > a. Let Π^* be a circuit such that $s(\Pi^*) = s(\{i_0^*, i_1^*, \cdots, i_{N^*}^*\}) = a_{\max}(P)$. As P is primitive, there exists an integer m_1 such that for all i we can find a path Π_1 which leads from i to i_0^* in m_1 steps. Similarly, we can find for any j a path Π_2 which leads from $i_{N^*}^*$ to j in m_1 steps. We can then construct a path $\Pi = \left\{\Pi_1, \underbrace{\Pi^*, \cdots, \Pi^*}_{\text{q times}}, \Pi_2\right\}$ which leads from i to i_0^*, passes q times through the circuit Π^*, and finally reaches j. The functional f assigns to this path the value:

$$s(\Pi) = \frac{m_1}{2m_1 + qN^*} s(\Pi_1) + \frac{qN^*}{2m_1 + qN^*} a_{\max}(P) + \frac{m_1}{2m_1 + qN^*} s(\Pi_2).$$

For q going to infinity, the first and the last term in this expression go to zero whereas the second term approaches $a_{\max}(P)$. As $a < a_{\max}(P)$, $s(\Pi) > a$ if we choose q large enough. Although q still depends on i and j, we can choose q^* as the maximum of all q's over all i and j. The integer m is then defined as $m = 2m_1 + q^*N^*$.

q.e.d.

LEMMA 2: The moment generating function $M_T(\beta)$ of $\sum_{t=1}^T f(X_{t-1}, X_t)$ equals

$$M_T(\beta) = E_{\mathbf{P}}\left(e^{\beta \sum_{t=1}^T f(X_{t-1}, X_t)}\right) = \sum_i \mu(i) \sum_j A_\beta^{(T)}(i, j)$$

PROOF:

$$M_T(\beta) = E_{\mathbf{P}}\left(e^{\beta\sum_{t=1}^{T} f(X_{t-1},X_t)}\right) = \sum_{x_0}\sum_{x_1}\cdots\sum_{x_T}\exp\left\{\beta\sum_{t=1}^{T} f(x_{t-1},x_t)\right\}\times P(x_{T-1},x_T)\times\cdots\times P(x_0,x_1)\times\mu(x_0)$$

$$= \sum_{x_0}\sum_{x_1}\cdots\left(\sum_{x_T}\exp\{\beta f(x_{T-1},x_T)\}\times P(x_{T-1},x_T)\right)$$

$$\times\exp\left\{\beta\sum_{t=1}^{T-1} f(x_{t-1},x_t)\right\}\times P(x_{T-2},x_{T-1})\times\cdots\times P(x_0,x_1)\times\mu(x_0)$$

$$= \sum_{x_0}\sum_{x_1}\cdots\sum_{x_{T-1}} A_\beta(1)(x_{T-1})\exp\left\{\beta\sum_{t=1}^{T-1} f(x_{t-1},x_t)\right\}\times P(x_{T-2},x_{T-1})\times\cdots\times P(x_0,x_1)\times\mu(x_0)$$

where $A_\beta(1)(x_{T-1})$ denotes the x_{T-1}-th element of the vector of row sums. Proceeding further in this way, one finally gets:[20]

$$M_T(\beta) = E_{\mathbf{P}}[e^{\beta\sum_{t=1}^{T} f(X_{t-1},X_t)}] = \sum_{x_0} A_\beta^{(T)}(1)(x_0)\,\mu(x_0) = \sum_i \mu(i)\sum_j A_\beta^{(T)}(i,j)$$

q.e.d.

PROOF OF PROPOSITION 3:

(i) and (ii) are obvious.

Because $f \geq 0$, $0 < A_\beta < A_{\beta'}$ if $\beta < \beta'$. This implies $\lambda(\beta) = \rho(A_\beta) < \rho(A_{\beta'}) = \lambda(\beta')$ because A_β is primitive and therefore irreducible (see Berman and Plemmons 1994, corollary 1.3.29). Thus $\lambda(\beta)$ is strictly increasing which proves (iii).

Because A_β is irreducible, $\lambda(\beta)$ is a simple root of the characteristic equation for A_β. The implicit function theorem (Dieudonné 1960) then implies that $\lambda(\beta)$ is analytic for all $\beta \in \Re$. This proves (iv).

If $M_T(\beta)$ denotes the moment generating function of $\sum_{t=1}^{T} f(X_{t-1},X_t)$, lemma 2 implies that

$$M_T(\beta) = \sum_i \mu(i)\sum_j A_\beta^{(T)}(i,j) = \lambda(\beta)^T \sum_i \mu(i)\sum_j r_\beta(i)\ell_\beta(j)\left[1 + O\left(\delta_\beta^T\right)\right]$$

where the second equality follows from Proposition 2. This implies that $\lim_{T\to\infty}[M_T(\beta)]^{1/T} = \lambda(\beta)$ because $\left\{\sum_i \mu(i)\sum_j r_\beta(i)\ell_\beta(j)\left[1 + O\left(\delta_\beta^T\right)\right]\right\}^{1/T}$ approaches one as $T \to \infty$.

As $[M_T(\beta)]^{1/T}$ is a moment generating function and $(1/T)\log[M_T(\beta)]$ is a cumulant generating function, these functions are convex on $\Re$ (Billingsley 1986, 144-5) for every

---

20 By $A_\beta^{(T)}(1)(x_0)$ we denote the sum of the x_0-th row of the matrix A_β^T.

T. As $\lambda(\beta)$ and log $\lambda(\beta)$ are the pointwise limits of convex functions, $\lambda(\beta)$ and log $\lambda(\beta)$ are convex (Rockafellar 1970, 90).

The two functions cannot be linear on some proper subinterval of $\Re$ because they are analytic. They cannot be linear over the whole real line either, because on the one hand $\lambda(\beta)$ and log $\lambda(\beta)$ diverge to ∞ as β goes to infinity and because on the other hand $\lambda(\beta)$ and log $\lambda(\beta)$ are bounded from below by $\max_i P(i,i) > 0$, respectively by $\max_i \log P(i,i) > -\infty$ as tr(P) > 0. $\lambda(\beta)$ and log $\lambda(\beta)$ must therefore be strictly convex functions which proves (v).

Because $\lambda(\beta)$ is a simple root of the characteristic equation for A_β, the differential of $\lambda(\beta)$ with respect to β is given by (Magnus and Neudecker 1988, 161-2)

$$\lambda'(\beta) = \frac{d\lambda(\beta)}{d\beta} = \ell'_\beta \frac{dA_\beta}{d\beta} r_\beta$$

where ℓ_β and r_β are left and right eigenvectors corresponding to $\lambda(\beta)$ normalized as $\ell_\beta' r_\beta = 1$. Using the properties of A_β and P_β listed in Proposition 2, we obtain:

$$\lambda'(\beta) = \lambda(\beta)\sum_i \ell_\beta(i) r_\beta(i) \sum_j \frac{P(i,j)e^{\beta f(i,j)}\, r_\beta(j)}{\lambda(\beta) r_\beta(i)} f(i,j) = \lambda(\beta)\sum_i \pi_{P_\beta}(i)\sum_j P_\beta(i,j) f(i,j).$$

This proves (vi). Thus f(.,.) has expectation $\lambda'(\beta)/\lambda(\beta)$ with respect to the invariant distribution.

q.e.d.

PROOF OF PROPOSITION 4:

Instead of A_β consider the matrix exp(-aβ) A_β. This matrix has maximal eigenvalue exp(-aβ) $\lambda(\beta)$. Applying Proposition 3 to exp(-aβ) A_β implies that that $g(\beta) = \exp(-a\beta)\,\lambda(\beta)$ is a differentiable strictly convex function of β with derivative equal to

$$g'(\beta) = \frac{d\, e^{-a\beta}\lambda(\beta)}{d\beta} = e^{-a\beta}[-a\lambda(\beta) + \lambda'(\beta)]$$

This derivative is negative at $\beta = 0$ for any $a \in \left(M_f^e(P), a_{max}(P)\right)$ because $-a\lambda(0) + \lambda'(0)$ $= -a + M_f^e(P) < 0$.

Let φ_β denote the left eigenvector of A_β corresponding to $\lambda(\beta)$ normalized as $\sum_i \varphi_\beta(i) = 1$.

This eigenvector has strictly positive coordinates because A_β is primitive. Then

$\lambda^m(\beta)=\lambda^m(\beta)\sum_j \varphi_\beta(j)=\sum_j\sum_i \varphi_\beta(i)A_\beta^{(m)}(i,j)=\sum_i \varphi_\beta(i)\sum_j A_\beta^{(m)}(i,j)$. The (i,j)-th element of A_β^m can be written as $A_\beta^{(m)}(i,j)=\sum_f P\{X_m=j, S_m=\nu \mid X_0=i\}e^{\beta m\nu}$ where ν runs over all positive values of S_m. As Lemma 1 implies that P and f satisfy the Uniform Recurrence Condition (R), we get:

$$\lambda^m(\beta)=\sum_i \varphi_\beta(i)\sum_j\sum_f P\{X_m=j,S_m=\nu|X_0=i\}e^{\beta m\nu}\geq \eta\sum_{\nu>a}e^{\beta m\nu}$$

which implies $\lambda(\beta)\geq \eta^{1/m}e^{\beta\nu}$, or equivalently $g(\beta)=e^{-a\beta}\lambda(\beta)\geq \eta^{1/m}e^{\beta(\nu-a)}$, for some $\nu > a$. This shows that, although $g'(0) < 0$, $g'(\beta)$ cannot be negative over the whole domain of β and that, for β large enough, $g'(\beta)$ must become positive. The strict convexity of $g(\beta)$ then ensures that $g(\beta)$ and therefore $\log(g(\beta)) = \log\lambda(\beta) - a\beta$ attain a unique infimum in the interval $(0,+\infty)$.

q.e.d.

PROOF OF THEOREM 2: Before proceeding to the proof we need the following Lemma.

LEMMA 3: Let P be a primitive transition matrix. For any $\beta\in\Re$ and any transition matrix Q which is absolutely continuous with respect to P, the following decomposition holds:

$$h(Q\mid P)=h(Q\mid P_\beta)+\beta\sum_i \pi_Q(i)\sum_j Q(i,j)f(i,j)-\log\lambda(\beta)$$

where f is a functional on E×E, P_β is the Perron-Frobenius transform of P, and $\lambda(\beta)$ is the largest eigenvalue of A_β (see definition 6). Moreover, for $a\in\left(M_f^e(P),a_{max}(P)\right)$ define the set of transition matrices $\Pi_a=\left\{Q:\sum_i \pi_Q(i)\sum_j Q(i,j)f(i,j)=a\right\}$, then there exists a positive, finite and unique $\beta = \beta(a)$ such that $P_{\beta(a)}\in\Pi_a$ and $\min_{Q\in\Pi_a} h(Q\mid P)=h(P_{\beta(a)}\mid P)$.

PROOF: As $\mathbf{Q}$ is absolutely continuous with respect to $\mathbf{P}$, and $\mathbf{P}$ and $\mathbf{P}_\beta$ are equivalent, $\left.\frac{d\mathbf{Q}}{d\mathbf{P}}\right|_{A_T} = \left.\frac{d\mathbf{Q}}{d\mathbf{P}_\beta}\frac{d\mathbf{P}_\beta}{d\mathbf{P}}\right|_{A_T}$.[21] The definition of the relative entropy then leads to:

$$H_T(\mathbf{Q}|\mathbf{P}) = H_T(\mathbf{Q}|\mathbf{P}_\beta) + \int \log \left.\frac{d\mathbf{P}_\beta}{d\mathbf{P}}\right|_{A_T} d\mathbf{Q}.$$

For any given path $(x_0, x_1, \ldots, x_T)$, we have

$$\begin{aligned}
\log\frac{d\mathbf{P}_\beta}{d\mathbf{P}}(x_0,\cdots,x_T) &= \log\left(\frac{\pi_{P_\beta}(x_0)\,P_\beta(x_0,x_1)\cdots P_\beta(x_{T-1},x_T)}{\pi_P(x_0)\,P(x_0,x_1)\cdots P(x_{T-1},x_T)}\right)\\
&= \log\frac{\pi_{P_\beta}(x_0)}{\pi_P(x_0)} + \sum_{t=1}^{T}\log\left(\frac{P(x_{t-1},x_t)\,e^{\beta f(x_{t-1},x_t)}\,r(x_t)}{P(x_{t-1},x_t)\,\lambda(\beta)\,r(x_{t-1})}\right)\\
&= \log\frac{\pi_{P_\beta}(x_0)}{\pi_P(x_0)} + \beta\sum_{t=1}^{T} f(x_{t-1},x_t) + \sum_{t=1}^{T}\log r(x_t) - \sum_{t=1}^{T}\log r(x_{t-1}) - \sum_{t=1}^{T}\log\lambda(\beta)\\
&= \log\frac{\pi_{P_\beta}(x_0)}{\pi_P(x_0)} + \beta\sum_{t=1}^{T} f(x_{t-1},x_t) - T\log\lambda(\beta) + \log r(x_T) - \log r(x_0)
\end{aligned}$$

Taking expectations with respect to $\mathbf{Q}$ leads to:

$$\begin{aligned}
\int \log\left.\frac{d\mathbf{P}_\beta}{d\mathbf{P}}\right|_{A_T} d\mathbf{Q} = {} & \sum_{x_0=1}^{K}\pi_Q(x_0)\log\frac{\pi_{P_\beta}(x_0)}{\pi_P(x_0)} + \beta\sum_{t=1}^{T}\sum_{x_{t-1}=1}^{K}\pi_Q(x_{t-1})\sum_{x_t=1}^{K}Q(x_{t-1},x_t)f(x_{t-1},x_t)\\
& - T\log\lambda(\beta) + \sum_{x_T=1}^{K}\pi_Q(x_T)\log r(x_T) - \sum_{x_0=1}^{K}\pi_Q(x_0)\log r(x_0)
\end{aligned}$$

Because of the invariant distribution π_Q we have T times the same double sum in the second term and because the last two terms are equal, the above expression simplifies to:

$$\int \log\left.\frac{d\mathbf{P}_\beta}{d\mathbf{P}}\right|_{A_T} d\mathbf{Q} = \sum_{i=1}^{K}\pi_Q(i)\log\frac{\pi_{P_\beta}(i)}{\pi_P(i)} + \beta T\sum_{i=1}^{K}\pi_Q(i)\sum_{j=1}^{K}Q(i,j)f(i,j) - T\log\lambda(\beta)$$

where we have substituted i for x_0 and x_{t-1} and j for x_t.

21 The Markov processes $\mathbf{Q}$, $\mathbf{P}$, and $\mathbf{P}_\beta$ have initial distributions equal to their corresponding invariant distributions.

An application of the ergodic theorem (Rosenblatt 1974, chap. V,d) shows that $h(Q\,|\,P)=\lim_{T\to\infty}\frac{1}{T}H_T(\mathbf{Q}\,|\,\mathbf{P})$ and that $h(Q\,|\,P)=\lim_{T\to\infty}\frac{1}{T}\log\left.\frac{d\mathbf{Q}}{d\mathbf{P}}\right|_{A_T}$ $\mathbf{Q}|_{A_T}$ – a.s. This result is also known as the Shannon-MacMillan-Breiman theorem. From this we get

$$
\begin{aligned}
h(Q|P) &= \lim_{T\to\infty}\frac{1}{T}H_T(\mathbf{Q}|\mathbf{P})=\lim_{T\to\infty}\frac{1}{T}\left[H_T(\mathbf{Q}|\mathbf{P}_\beta)+\int\log\left.\frac{d\mathbf{P}_\beta}{d\mathbf{P}}\right|_{A_T}d\mathbf{Q}\right]\\
&=\lim_{T\to\infty}\frac{1}{T}\left[H_T(\mathbf{Q}|\mathbf{P}_\beta)+\sum_{i=1}^{K}\pi_Q(i)\log\frac{\pi_{P_\beta}(i)}{\pi_P(i)}+\beta T\sum_{i=1}^{K}\pi_Q(i)\sum_{j=1}^{K}Q(i,j)f(i,j)-T\log\lambda(\beta)\right]\\
&=h(Q|P_\beta)+\beta\sum_{i=1}^{K}\pi_Q(i)\sum_{j=1}^{K}Q(i,j)f(i,j)-\log\lambda(\beta)
\end{aligned}
$$

For $a\in\left(M_f^e(P),a_{max}(P)\right)$, Propositions 3 and 4 imply that there exists a finite and unique $\beta=\beta(a)>0$ such that $\sum_i\pi_{P_\beta}(i)\sum_j P_\beta(i,j)f(i,j)=a$. Therefore $P_{\beta(a)}\in\Pi_a$. Noting that $h(Q|P)\geq 0$ for all P and Q and that $h(Q|P)=0$ implies $Q=P$, this proves $\min_{Q\in\Pi_a}h(Q|P)=h(P_{\beta(a)}|P)$.

q.e.d.

The proof of Theorem 2 proceeds in two steps. First we present an upper bound then a lower bound and show that they converge to the same limit.

Upper bound. The application of Chebycheff's inequality to the function $g(x)=e^{T\beta x}$ with $\beta>0$, sometimes called the exponential overbound lemma, leads to

$$
P\left\{\frac{1}{T}\sum_{t=1}^{T}f(X_{t-1},X_t)\geq a\right\}\leq e^{-T\beta a}\,E_{\mathbf{P}}\left[e^{\beta\sum_{t=1}^{T}f(X_{t-1},X_t)}\right]
$$

where $E_{\mathbf{P}}$ is the expectation with respect $\mathbf{P}_{\boldsymbol{\mu}}$. Lemma 2 implies that the expectation equals

$$
E_{\mathbf{P}}\left(e^{\beta\sum_{t=1}^{T}f(X_{t-1},X_t)}\right)=\sum_{x_0}A_\beta^{(T)}(1)(x_0)\,\mu(x_0)
$$

Applying the log to Chebycheff's inequality and dividing by T yields:

$$\frac{1}{T}\log P\left\{\frac{1}{T}\sum_{t=1}^{T} f(X_{t-1},X_t)\geq a\right\}\leq -\beta a+\frac{1}{T}\log\left(\sum_{x_0} A_\beta^{(T)}(1)(x_0)\,\mu(x_0)\right)$$

$$\leq -\beta a+\log\lambda(\beta)+\frac{1}{T}\log\left(\sum_{x_0}\left(\frac{A_\beta^{(T)}(1)}{\lambda(\beta)^T}\right)(x_0)\,\mu(x_0)\right)$$

$$\leq -\beta a+\log\lambda(\beta)+o\left(\frac{1}{T}\right)$$

The last step follows from Proposition 2 by observing that $\left(A_\beta/\lambda_\beta\right)^T$ converges to $\ell_\beta r_\beta' >> 0$. Taking the limit with respect to T implies:

$$\lim_{T\to\infty}\frac{1}{T}\log P\left\{\frac{1}{T}\sum_{t=1}^{T} f(X_{t-1},X_t)\geq a\right\}\leq -\beta a+\log\lambda(\beta).$$

As this inequality holds for any $\beta > 0$, it must also hold for the infimum over $\beta > 0$:

$$\lim_{T\to\infty}\frac{1}{T}\log P\left\{\frac{1}{T}\sum_{t=1}^{T} f(X_{t-1},X_t)\geq a\right\}\leq \inf_{\beta\in\Re}\left(\log\lambda(\beta)-\beta a\right)=-\sup_{\beta\in\Re}\left(\beta a-\log\lambda(\beta)\right)=-h\left(P_{\beta(a)}\middle|P\right)$$

According to Proposition 4 the infimum over $\Re$ is attained in the interval $(0,+\infty)$. Thus we are allowed to take the infimum, respectively the supremum, over $\beta \in \Re$ and not just over $\beta > 0$. The last equality is a consequence of the decomposition in Lemma 3.

Lower bound. Take any **Q** such that $\sum_i \pi_Q(i)\sum_j Q(i,j)f(i,j)=a$.

For $\delta > 0$, consider the two events $A=\left\{\frac{1}{T}\sum_{t=1}^{T} f(X_{t-1},X_t)\geq a\right\}$ and $B=\left\{\frac{d\mathbf{Q}}{d\mathbf{P}}\Big|_{A_T}\leq e^{T(h(Q|P)+\delta)}\right\}=\left\{e^{-T(h(Q|P)+\delta)}\frac{d\mathbf{Q}}{d\mathbf{P}}\Big|_{A_T}\leq 1\right\}=\left\{\frac{1}{T}\log\frac{d\mathbf{Q}}{d\mathbf{P}}\Big|_{A_T}\leq h(Q|P)+\delta\right\}$. Then:

$$\mathbf{P}\{A\}\geq \mathbf{P}\{A\cap B\}=\int 1_{A\cap B}\,d\mathbf{P}\geq e^{-T(h(Q|P)+\delta)}\int 1_{A\cap B}\frac{d\mathbf{Q}}{d\mathbf{P}}\Big|_{A_T}d\mathbf{P}\geq e^{-T(h(Q|P)+\delta)}\mathbf{Q}\{A\cap B\}$$

By assumption the event A occurs almost surely under **Q** for $T \to \infty$ as a consequence of ergodicity. **Q**{B} is controlled by the Shannon-MacMillan-Breiman theorem (Rosenblatt 1974, chap. V,d). Thus **Q**{A∩B} converges to 1 in probability. This implies that

$$\frac{1}{T}\log\mathbf{P}\{\frac{1}{T}\sum_{t=1}^{T} f(X_{t-1},X_t)\geq a\}\geq -h(Q\,|\,P)-\delta+O\left(\frac{1}{T}\right).$$

Setting $Q = P_{\beta(a)}$ with $a = \sum_i \pi_{P_{\beta(a)}}(i) \sum_j P_{\beta(a)}(i,j) f(i,j)$ and taking limits with respect to T,

we finally get: $\lim_{T\to\infty} \frac{1}{T} \log P\{\frac{1}{T}\sum_{t=1}^{T} f(X_{t-1}, X_t) \geq a\} \geq -\inf_{\delta>0}\left(h(P_{\beta(a)} \mid P) + \delta\right) = -h(P_{\beta(a)} \mid P)$

q.e.d.

6. References

Alvarez, J., M. Browning, and M. Ejrnaes. 2001. Modelling Income Processes with Lots of Heterogeneity. Paper presented at ESEM 2001 in Lausanne, Switzerland.

Atkinson, A.B. 1983. The Measurement of Economic Mobility. In *Social Justice and Public Policy*, ed. A.B. Atkinson. Cambridge Massachusetts: MIT Press.

Bartholomew, D.J. 1982. *Stochastic Models for Social Processes*. 3rd Ed. Chichester: John Wiley & Sons.

Bénabou, R., and E.A. Ok. 2001. Mobility as Progressivity: Ranking Income Processes According to Equality of Opportunity. NBER Working Paper No. W8431.

Berman, A., and R.J. Plemmons. [1979] 1994. *Nonnegative Matrices in the Mathematical Sciences*. Philadelphia: Society for Industrial and Applied Mathematics, Classics in Applied Mathematics.

Billingsley, P. 1986. *Probability and Measure*. 2nd Ed. New York: John Wiley & Sons.

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

Conlisk, J., and P.M. Sommers. 1979. Eigenvalue Immobility Measures for Markov Chains. *Journal of Mathematical Sociology* 6: 253-76.

Dardanoni, V. 1993. Measuring Social Mobility. *Journal of Economic Theory*. 61: 372-394.

Dardanoni, V. 1995. Income Distribution Dynamics: Monotone Markov Chains Make Light Work. *Social Choice and Welfare* 12:181-92.

Dieudonné, J. 1960. *Foundations of Modern Analysis*. New York: Academic Press.

Dembo, A., and O. Zeitouni. 1998. *Large Deviations Techniques and Applications*. 2nd Ed. New York: Spinger-Verlag.

Fields, G.S., and E.A. Ok. 1999. The Measurement of Income Mobility: An Introduction to the Literature. In *Handbook on Income Inequality Measurement*, ed. J. Silber, 557-96. Boston: Kluwer Academic Publishers

Geweke, J, R.C. Marshall, and G.A. Zarkin. 1986. Mobility Indices in Continuous Time Markov Chains. *Econometrica* 54:1407-23.

Hollander, F. den. 2000. *Large Deviations.* Fields Institute Monographs 14. Providence, Rhode Island: American Mathematical Society.

Iscoe, I., P. Ney, and E. Nummelin. 1985. Large Deviations of Uniformly Recurrent Markov Additive Processes. *Advances in Applied Mathematics* 6:373-412.

Kim, G.-H., and H.T. David. 1979. Large Deviations of Functions of Markovian Transitions and Mathematical Programming Duality. *Annals of Probability* 7:874-81.

Maasoumi, E. 1998. On Mobility. In *Handbook of Applied Economic Statistics*, eds. A. Ullah and D. Giles, 119-73. New York: Marcel Dekker.

Magnus, J.R., and H. Neudecker. 1988. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester: John Wiley & Sons.

Miller, H.D. 1961. A Convexity Property in the Theory of Random Variables defined on a Finite Markov Chain. *Annales of Mathematical Statistics* 32:1260-70.

Nelsen, R.B. 1999. *An Introduction to Copulas*. Lecture Notes in Statistics 139. New York: Springer Verlag.

Ney, P., and E. Nummelin. 1987a. Markov Additive Processes I. Eigenvalue Properties and Limit Theorems. *Annals of Statistics* 15:561-92.

Ney, P., and E. Nummelin. 1987b. Markov Additive Processes II. Large Deviations. *Annals of Statistics* 15:593-609.

Norris, J.R. 1997. *Markov Chains*. Cambridge: Cambridge University Press.

Rockafellar, R.T. 1970. *Convex Analysis*. Princeton, New Jersey: Princeton University Press.

Rosenblatt, M. 1974. *Random Processes*. New York: Springer Graduate Texts in Mathematics.

Seneta, E. 1981. *Non-negative Matrices and Markov Chains*. New York: Springer Series in Statistics.

Shorrocks, A.F. 1978. The Measurement of Mobility. *Econometrica*. 46:1013-34

Table 1: Some commonly used mobility indices

equilibrium mobility indices	**Bartholomew's index**	$\sum_{i=1}^{K} \pi(i) \sum_{j=1}^{K} P(i,j)\lvert i-j \rvert$
	index of unconditional probability of leaving the current class	$\frac{K}{K-1} \sum_{i=1}^{K} \pi(i)(1-P(i,i))$
convergence mobility indices	**Prais' index**	$\frac{K - \mathrm{tr}(P)}{K-1}$
	eigenvalue index	$\frac{K - \sum_{i=1}^{K} \lvert \lambda_i \rvert}{K-1}$
	second largest eigenvalue index	$1-\delta(P)$
	asymptotic speed of convergence	$-\log \delta(P)$
	determinant index	$1-\lvert \det(P) \rvert$

P . . . . . . primitive transition matrix with invariant distribution π

λ_i . . . . . eigenvalues of P

$\delta(P) = \max\{\lvert\lambda\rvert : \lambda \in \sigma(P) \text{ and } \lambda \neq 1\}$

Table 2: characteristics and mobility indices for test transition matrices

		transition matrices				
		P_1	P_2	P_3	P_{mobile}	P_{ident}
	equilibrium mobility index [a]	0.4667	0.4667	0.4667	0.8889	0.00267
"convergence" indices	**period mobility index** [a] $\gamma = 0.5$	0.4554	0.5495	0.3799	0.7841	0.0630
	Prais' index [b]	0.65	0.6	0.65	1	0.003
	eigenvalue index [b]	0.65	0.6	0.65	1	0.003
	second largest eigenvalue index [b]	0.3854	0.4268	0.3994	1	0.003
	asymptotic speed of convergence [b]	0.4868	0.5565	0.5098	$+\infty$	0.00300
	determinant index [b]	0.9475	0.87	0.9403	1	0.00599
	$\mathbf{a_{max}(P)}$ [a]	2	2	2	2	2
	invariant distribution	$\left(\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}\right)'$	$\left(\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}\right)'$	$\left(\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}\right)'$	$\left(\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}\right)'$	$\left(\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}\right)'$

[a] Bartholomew mobility functional $f(i,j) = |i - j|$

[b] For definitions see table 1

Table 3: ranking the Dardanoni's matrices

mobility index	resulting ranking
Bartholomew-/equilibrium mobility index	$P_1 \approx_e P_2 \approx_e P_3$
period mobility index, $\gamma = 0.5$	$P_3 \prec_p P_1 \prec_p P_2$
Prais' index [a]	$P_2 \prec_c P_1 \approx_c P_3$
eigenvalue index [a]	$P_2 \prec_c P_1 \approx_c P_3$
second largest eigenvalue index [a]	$P_1 \prec_c P_3 \prec_c P_2$
asymptotic speed of convergence [a]	$P_1 \prec_c P_3 \prec_c P_2$
determinant index [a]	$P_2 \prec_c P_3 \prec_c P_1$

[a] For definitions see Table 1.

Figure 1: Period mobility of test matrices