UNIVERSITÄT BERN

Faculty of Economics and Social Sciences

Department of Economics

A Large Deviation Approach to the Measurement of Mobility

Robert Aebi
Klaus Neusser
Peter Steiner

05-18

December 2005

DISCUSSION PAPERS

Schanzeneckstrasse 1
Postfach 8573
CH-3001 Bern, Switzerland
http://www.vwi.unibe.ch

A Large Deviation Approach to the Measurement of Mobility*

Robert Aebi† Klaus Neusser‡ Peter Steiner§

December 9, 2005

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

† Institut de Recherche Mathématique Avancée, Université Louis Pasteur, 7 rue René-Descartes, F-67084 Strasbourg Cedex, France

‡ Department of Economics, University of Bern, Schanzeneckstrasse 1, Postfach 8573, CH-3001 Bern, Switzerland

§ State Secretariat for Economic Affairs, seco, Effingerstrasse 1, CH-3003 Bern, Switzerland

Abstract

We propose an approach to measure the mobility immanent in regular 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

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 alternative approaches prevail in the literature (see the surveys by Fields and Ok [16] or Maasoumi [20]). Here we follow the spirit of Batholomew [2] and interpret mobility as movements between states.[1] We view these movements as realizations of a stochastic process which we take as the primitive for our approach to the measurement of mobility. In order to make our point, we restrict ourselves to time homogenous regular Markov processes defined on finite state spaces. This means that the process is characterized by an initial distribution and a primitive transition matrix.

The paper is motivated by the longstanding insight that the notion of mobility comprises different aspects: the extent to which the process leads to movements between states over time and the degree to which future states

[1] Alternative interpretations view mobility as equalizing opportunity (Bénabou and Ok [3]), welfare enhancing or similarly as inequality reducing (Atkinson [1], Dardanoni [10], or Maasoumi [20, 132]). From an empirical point of view, the stochastic dominance approach represents a promising alternative because it allows to implement different mobility concepts. Although all the approaches start from different views, they are not unrelated to each other. In subsections 2.2 and 2.4 we investigate some connections to our approach.

do not depend on the initial state.[2] Usually, the first aspect is measured on the basis of the equilibrium or invariant distribution of the stochastic process whilst the measurement of the second aspect is based on the eigenvalues of the underlying transition matrix (see Sommers and Conlisk [28]). This insight led us to classify mobility indices into equilibrium and convergence mobility indices. Most conventional mobility indices can actually be classified according to these two characteristics.

The aim of this paper is a methodological one as it provides a joint basis for equilibrium and convergence mobility indices. The starting point of the analysis consists in the specification of a mobility functional. This functional is defined on the Cartesian square product of the state space and represents just a rule of weighting movements between states. The expected value of this functional with respect to the invariant distribution of the underlying stochastic process then defines an equilibrium mobility index. Popular mobility indices, like Bartholomew's index or the index of unconditional probability of leaving the current class, can actually be represented in this way. An application of the Ergodic theorem then implies that the time average of the mobility functional converges to the corresponding equilibrium mobility index. We show that this time average satisfies a large deviation principle (LDP). This means that the probability that the time average exceeds the value of the equilibrium mobility index by some prescribed amount converges to zero at a constant exponential rate. This exponential rate then gives rise to a kind of convergence mobility index which we call period mobility. We

[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 [2] refers to these two aspects as measures of movements and measures of generation dependencies. Gottschalk and Spolaore [18] refer to these two aspects of mobility as "reversal" and "origin independence".

will show that this exponential rate can actually be computed from a specific relative entropy. In this way the specification of a mobility functional gives rise simultaneously to an equilibrium and a convergence mobility index. Thus the way to measure both aspects of mobility is no longer independent from each other, but reduced to the choice of a mobility functional.

Replacing the expectation in the computation of the equilibrium mobility index by the corresponding ensemble average (i.e. the average over the individuals in the population) shows that the mobility functional approach has much in common with the measurement of total mobility by "economic distances" as analyzed by Fields and Ok [15] and Mitra and Ok [23]. Indeed their axiomatic view can serve as a guide for the appropriate choice of a mobility functional. An aspect we do not cover here.

The approach via a mobility functional must be contrasted with an older, but important, strand of literature that defines mobility as a functional on the set of transition matrices. This literature proposes an axiomatic approach and postulates a set of desirable axioms for mobility indices (Shorrocks [27]). Geweke, Marshall and Zarkin [17] 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.

Although we do not investigate the implications of particular properties of mobility functionals, we nevertheless highlight the importance of so-called 2-decreasing mobility functionals. The corresponding equilibrium mobility indices turn out to be consistent with Shorrocks' [27] monotonicity axiom,

Conlisk's [9] weak D-criterion as well as with Dardanoni's [10] partial ordering in the case of monotone transition matrices with identical equilibrium distributions.

While the notion of the equilibrium mobility index is related to concepts discussed in the literature, the measurement of convergence mobility based on the large deviation principle is completely new. Although the specific LDP we derive in this paper can be regarded as 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 applications 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.

Our paper is organized as follows. Section 2 states the assumptions which must be fulfilled by the underlying stochastic process and reviews some of their most immediate implications. Next, we define the mobility functional and the associated equilibrium mobility index. We then show that the corresponding sample averages obey a strong law of large numbers and a central limit theorem. Finally, we draw some connections to the existing literature. Section 3 introduces the Large Deviations Principle and proves the core theorem. Section 4 defines our convergence mobility index, called period mobility index, and discusses some illustrative examples. Finally, section 5 discusses a number of conclusions.

[3] Hollander [19] presents an excellent introduction to the Theory of Large Deviations, see especially chapter 4 on "Large Deviations for Markov Sequences". See also Dembo and Zeitouni [13] who provide a general treatment of the subject and an application to finite state space Markov Chains in Chapter 3.

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 $\mathcal{E} = \{1, 2, \ldots, K\}$. 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 $\mathcal{E}^\infty$ 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 $\mathrm{tr}(P) > 0$, P becomes a primitive matrix (i.e. $\exists m \in \mathbb{N} : P^m \gg 0$;[5] Berman and Plemmons [4, 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. In addition, $\rho(P) = 1$ is a simple eigenvalue greater in magnitude than any other eigenvalue.[7] Thus $\lambda \in \sigma(P)$ implies that $\lambda = 1$ or that

[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 $\mathrm{tr}(P) > 0$ 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

$|\lambda| < 1$. 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 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 label them as convergence mobility indices. These indices measure the degree to which future states do not depend on the initial state. A list of the most commonly used indices is given in Table 1.

2.2 Definitions

In contrast to Shorrocks [27] or Geweke , Marshall and Zarkin [17], we do not define our mobility index directly on the set of transition matrices. Instead, more in the spirit of Bartholomew [2, 24-30], we base our concept on the valuation of movements between states where the valuation is represented by a mobility functional. This way of proceeding has one great advantage that the definitions of the mobility indices proposed below can be easily carried over to general stochastic processes.

Definition 1. *A* mobility functional f *is a nonnegative functional on* $\mathcal{E} \times \mathcal{E}$

chains (for example Berman and Plemmons [4], Norris [25] or Stroock [30])

[8] The asymptotic speed of convergence is defined as $-\log \alpha$ with $\alpha = \sup_\mu \lim_{T\to\infty} \|\mu' P^T - \pi'\|^{\frac{1}{T}}$ where the supremum is taken over all initial distributions μ (Berman and Plemmons [4, 172]). It can be shown that the asymptotic speed of convergence equals $-\log \delta(P)$ in our case (Berman and Plemmons [4, 199]). Sommers and Conlisk [28] proposed $\delta(P)$ as a measure of immobility, respectively $1 - \delta(P)$ as a measure of mobility.

such that

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

The mobility functional therefore attaches positive values to movements from one state to another and zero when no movement occurs. Thus the mobility functional provides some kind of "economic distance" between states. Although f may define a metric on $\mathcal{E}$, definition 1 does not impose this requirement: in particular neither the triangle inequality nor the symmetry of f must hold. Upward movements can be valued differently from downward movements. Note also that movements toward states which are "farther away" need not receive higher values. From the Markovian viewpoint of equilibrium and convergence mobility a generalization to functionals f defined on higher powers than two of the state space $\mathcal{E}$ is not indicated. In fact, in the equilibrium described by the stationary probability distribution π, the Markov chain $\mathbf{P}_\mu$ 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 $\mathcal{E} \times \mathcal{E}$ and any irreducible transition matrix P with its unique invariant distribution π,*

$$M_f^e(P) = \sum_{i \in \mathcal{E}} \pi(i) \sum_{j \in \mathcal{E}} P(i,j) f(i,j) \tag{2.1}$$

is called the equilibrium f-mobility index *of P. For any two irreducible Markov chains $\mathbf{P}_{\mu_1}$ and $\mathbf{Q}_{\mu_2}$, we say that $\mathbf{P}_{\mu_1}$ is* more mobile than $\mathbf{Q}_{\mu_2}$ with respect to f, *denoted by $P \succeq_e Q$, if and only if $M_f^e(P) \geq M_f^e(Q)$.*

The definition can be written more compactly as $M_f^e(P) = \text{tr}(P'\text{diag}(\pi)f)$ 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 $M_f^e(P)$ is divided by $a_{\max}$, a number which depends only on f (see section 3.2). It is easy to see that the equilibrium f-mobility of the identity matrix I_K equals zero, i.e. $M_f^e(I_K) = 0$, so that the index fulfills Shorrocks [27, 1015] Immobility axiom. As we restrict ourselves to irreducible transition matrices with $\text{tr}(P) > 0$ (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 irreducible transition matrices with $\text{tr}(P) > 0$ 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 [10], Fields and Ok [16], Maasoumi [20]).

The definition of the equilibrium mobility index encompasses several specifications encountered in the literature. Consider first, the power functional: $f(i,j) = |i-j|^\alpha, \alpha \geq 1$. For $\alpha = 1$, the equilibrium mobility index specializes to Bartholomew's index:[10]

$$M_f^e(P) = \sum_{i \in \mathcal{E}} \pi(i) \sum_{j \in \mathcal{E}} P(i,j)|i-j|. \tag{2.2}$$

Another interesting choice for the mobility functional is $f(i,j) = 1 - \delta(i,j)$ where $\delta(i,j)$ denotes Kronecker's delta. This results in the index of unconditional probability of leaving the current class which is nothing but

[9] The diag(x) operator transforms any K-vector x into a $K \times K$ diagonal matrix with x on the diagonal.

[10] Bartholomew [2] scaled this index by $\frac{1}{K-1}$ to confine it to the interval $(0,1)$.

the expected number of class changes:[11]

$$M_f^e(P) = \sum_{i\in\mathcal{E}} \pi(i)(1 - P(i,i)) = \sum_{i\in\mathcal{E}} \pi(i) \sum_{j\in\mathcal{E}} P(i,j)(1 - \delta(i,j)). \qquad (2.3)$$

The above mobility functionals actually define metrics on the state space $\mathcal{E}$: 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 measurement of equilibrium mobility as the expected value of a mobility functional lies in the spirit of Fields and Ok [15] and Mitra and Ok [23]. To see this, suppose that the population consists of N individuals, then replacing the expectations by the corresponding ensemble average (i.e. the average over all individuals) leads to the following measure of mobility between two periods:

$$\frac{1}{N} \sum_{i=1}^{N} f(x_i, y_i) \qquad (2.4)$$

where x_i and y_i denote the state of individual i in the first, respectively the second period. But this is nothing but the per capita version of "total absolute income mobility" where the distance function between $x = (x_1, \ldots, x_N)$ and $y = (y_1, \ldots, y_N)$, in their terminology, is just given by $d_N(x, y) = \sum_{i=1}^{N} f(x_i, y_i)$. The interest in this interpretation of the equilibrium mobility index is that the axioms proposed by Fields and Ok [15] and Mitra and Ok [23] for $d_N(x, y)$ restrict the set of possible mobility functionals. Indeed, if one views their axioms as compelling, the power mobility functional turns out to be the generic case with $\alpha = 1$ (Bartholomew's case) being of special importance.

[11] In the literature, this index is usually scaled by $\frac{K}{K-1}$.

2.3 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 3. *For any Markov process, $\{X_t\}$, defined on the state space $\mathcal{E}$ and a mobility functional f on $\mathcal{E} \times \mathcal{E}$, the time average S_T of $f(X_{t-1}, X_t)$*

$$S_T = \frac{1}{T} \sum_{t=1}^{T} f(X_{t-1}, X_t) \tag{2.5}$$

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

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 movements. Note that in the latter case the assumption $\mathrm{tr}(P) > 0$ precludes the degenerate situation that S_T is constant over all possible realizations. Given the regularity assumptions about the Markov chain, a strong law of large numbers (SLLN) holds.

Theorem 1 (SLLN)**.** *The empirical f-mobility converges to the following limit:*

$$\lim_{T\to\infty} S_T = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} f(X_{t-1}, X_t) = M_f^e(P) \tag{2.6}$$

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

Proof. This is just an application of the Ergodic theorem (see for example Stroock [30]) to a function f defined on two consecutive states. □

Thus one can use S_T to estimate the equilibrium f-mobility index directly from the sample paths without estimating in a prior step the transition matrix of the process. This immediate conclusion from SLLN is reinforced because a central limit theorem (CLT) also holds in this context.

Theorem 2 (CLT)**.** *Let $\{X_t\}$ be a stationary regular Markov chain with finite state space, then the empirical f-mobility satisfies the central limit theorem:*

$$\sqrt{T}\left(S_T - M_f^e(P)\right) \xrightarrow{\mathcal{D}} \mathrm{N}(0, \sigma^2) \tag{2.7}$$

for any mobility functional f. The variance σ^2 of the normal distribution is given by

$$\sigma^2 = \mathrm{var}(Y_t) + 2\sum_{j=1}^{\infty} \mathrm{cov}(Y_t, Y_{t+1}) > 0$$

where $Y_t = f(X_{t-1}, X_t)$ for $t = 1, 2, \ldots$.

Proof. $\{X_t\}$ is ϕ-mixing with mixing coefficients $\phi_m^{(X)}$ declining to zero exponentially fast, i.e. there exist positive constants c and ρ, $\rho < 1$, such that $\phi_m^{(X)} = c\rho^m$ (Billingsley [6, Example 2, 167-8]). Theorem 14.1 in Davidson [12, 210] implies that $Y_t = f(X_{t-1}, X_t)$ is also ϕ-mixing with mixing coefficients $\phi_m^{(Y)} \leq \phi_m^{(X)}$, $m > 1$. The CLT then follows from theorem 20.1 in Billingsley [6, 174] because $\sum_{m=1}^{\infty} \sqrt{\phi_m^{(Y)}} < \infty$. □

Note that the above theorem uses the additional assumption that $\{X_t\}$ is stationary. This is equivalent to the assumption that the initial distribution (i.e. the distribution of X_0) equals the unique invariant distribution π. Whereas the CLT assesses the probability that S_T differs from $M_f^e(P)$ by an amount of order $\frac{1}{\sqrt{T}}$, the large deviation approach, to which we will turn next, relates to events where S_T differs from $M_f^e(P)$ by an amount of order $\frac{1}{T}$. Such deviations may be termed "large". Although these events are "rare" and their probabilities vanish exponentially fast, the rate at which this decay takes place can be quantified. Moreover, this rate can be used to define a convergence mobility index.

2.4 Relations to Existing Criteria and Rankings

An important class of mobility functionals is given by *2-decreasing mobility functionals.*

Definition 4. *A mobility functional f on $\mathcal{E} \times \mathcal{E}$ is* 2-decreasing *if*

$$V(i,j) = f(i+1, j+1) - f(i+1, j) - f(i, j+1) + f(i,j) \leq 0 \qquad (2.8)$$

for all $i, j \in \{1, 2, \ldots, K-1\}$.

The inequality is strict if $i = j$. 2-decreasing functions are the two-dimensional analogues of non-increasing functions in one variable. $-V(i,j)$ can be interpreted as the area assigned by f to the rectangle with vertices $(i,j), (i+1,j), (i,j+1), (i+1,j+1)$ (see Nelsen [24]). The 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 for $\alpha \geq 1$ whereas the functional $f(i,j) = 1 - \delta(i,j)$ is not.

2-decreasing functionals are especially useful in connection with monotone transition matrices. These matrices attracted some attention because they have theoretically plausible properties and are supported empirically (Conlisk [9], Dardanoni [10], Dardanoni [11], Fields and Ok [16]). Monotone transition matrices are transition matrices where row $i+1$ stochastically dominates row i for all $i = 1, \ldots, K-1$. This is equivalent to $T^{-1}\mathbf{P}T \geq 0$ where T denotes the summation matrix.[12]

In order to isolate the pure mobility effect, we follow among others Dardanoni [10, 377] and consider only Markov chains with identical invariant distributions. This normalization corresponds to the standard practice of

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

holding constant the mean when comparing the inequality of income distributions or the riskiness of asset return distributions.

Based on Lemma 1, we then show that equilibrium mobility indices induced by 2-decreasing mobility functionals are coherent with Dardanoni's partial ordering of monotone transition matrices sharing the same invariant distribution (Dardanoni [10]). Moreover, Theorem 3 and 4 imply the consistency with the monotonicity axiom of Shorrocks [27] and the weak D-criterion of Conlisk [9], so that they satisfy all persistence criteria listed by Geweke, Marshall and Zarkin [17].

Lemma 1. *For any two irreducible transition matrices P and Q with the same invariant distribution π and any 2-decreasing mobility functional f,*

$$T'\mathrm{diag}(\pi)(P-Q)T \leq 0 \quad \textit{implies} \quad P \succeq_e Q.$$

Proof. Noting that $M_f^e(P) = \mathrm{tr}(P'\mathrm{diag}(\pi)f)$ where f denotes the matrix with elements $f(i,j)$ and using the properties of the trace operator, we get:

$$\begin{aligned} M_f^e(P) - M_f^e(Q) &= \mathrm{tr}\left((P-Q)'\mathrm{diag}(\pi)f\right) \\ &= \mathrm{tr}\left(\left(T'\mathrm{diag}(\pi)(P-Q)T\right)\left(T^{-1}f'T'^{-1}\right)\right) \end{aligned}$$

The fact that $\sum_i \pi(i)\left(P(i,j)-Q(i,j)\right) = 0$ for all j and that $\sum_j \left((P(i,j)-Q(i,j)\right) = 0$ for all i implies that $T'\mathrm{diag}(\pi)(P-Q)T$ can be expressed as

$$T'\mathrm{diag}(\pi)(P-Q)T = \begin{pmatrix} N & 0_{(K-1)\times 1} \\ 0_{1\times(K-1)} & 0 \end{pmatrix}.$$

Because $T'\mathrm{diag}(\pi)(P-Q)T \leq 0$ by assumption, $N \leq 0$. On the other hand

$$T^{-1}f'T'^{-1} = \begin{pmatrix} \mathrm{V}' & c \\ b' & 0 \end{pmatrix}$$

where b and c are nonnegative $K-1$ vectors. The $(K-1)\times(K-1)$ matrix V has typical elements:

$$\mathrm{V}(i,j) = f(i,j) - f(i,j+1) + f(i+1,j+1) - f(i+1,j) \leq 0$$

where the inequality follows from f being 2-decreasing. This finally leads to:

$$M_f^e(P) - M_f^e(Q) = \mathrm{tr}\left(\begin{pmatrix} N & 0_{(K-1)\times 1} \\ 0_{1\times(K-1)} & 0 \end{pmatrix}\begin{pmatrix} \mathrm{V}' & c \\ b' & 0 \end{pmatrix}\right) = \mathrm{tr}(N\mathrm{V}') \geq 0$$

which is equal to $P \succeq_e Q$ (see Definition 2). □

The implication goes only in one direction as we can give examples such that $P \succeq_e Q$ with $T'\mathrm{diag}(\pi)(P-Q)T$ not being nonpositive. Furthermore, perfect mobility matrices $\iota\pi'$, $\iota = (1,\ldots,1)'$, are maximal elements with respect to equilibrium mobility because $T'\mathrm{diag}(\pi)(\iota\pi' - P)T \leq 0$ for all monotone transition matrices P with stationary probability distribution π (Dardanoni [10, theorem 2]). Therefore $(\iota\pi') \succeq_e P$ whenever f is 2-decreasing.

Theorem 3. *If P and Q are both monotone 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 $P \succeq_e Q$ if the mobility functional f is 2-decreasing.*

Proof. The assumptions imply that $T'\mathrm{diag}(\pi)(P-Q)T \leq 0$ (Dardanoni [10, Appendix 2]). For 2-decreasing functionals, $P \succeq_e Q$ follows from Lemma 1. □

Theorem 4. *Let P and Q be two monotone transition matrices with the same invariant distribution π. If the upper left $(K-1)\times(K-1)$ matrices of $T^{-1}PT$ and $T^{-1}QT$ are denoted by $\Delta(P)$ and $\Delta(Q)$, respectively, then $\Delta(Q) \geq \Delta(P)$ implies $P \succeq_e Q$ if the mobility functional f is 2-decreasing.*

Proof. $\Delta(Q) \geq \Delta(P)$ implies $T'\text{diag}(\pi)(P-Q)T \leq 0$ (Dardanoni [10, Appendix 2]). For 2-decreasing functionals, $P \succeq_e Q$ follows from Lemma 1. $\square$

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 kind of convergence mobility index which we call period f-mobility index. This requires, however, additional concepts which we will now introduce.

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

$$H_T(\mathbf{Q}|\mathbf{P}) = \int \log \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} \mathrm{d}\mathbf{Q}.$$

The Radon-Nikodym derivative of $\mathbf{Q}$ *with respect to* $\mathbf{P}$ *on* A_T *is defined as*

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

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

$$h(Q|P) = \lim_{T\to\infty} \frac{1}{T} H_T(\mathbf{Q}|\mathbf{P}) = \sum_{i\in\mathcal{E}} \pi_Q(i) \sum_{j\in\mathcal{E}} Q(i,j) \log\left(\frac{Q(i,j)}{P(i,j)}\right).$$

The second equality above is, strictly speaking, not a definition but an implication of the Shannon-McMillan-Breiman theorem (see, for example, Billingsley [5, 129]). The relative entropy plays a key role in the theory of large deviations so that it seems useful to restate two of its properties.[13] If $\mathbf{Q}$ is absolutely continuous with respect to $\mathbf{P}$, respectively if $P(i,j) = 0$ implies $Q(i,j) = 0$, we have:

- $H_T(.|\mathbf{P})$ and $h(.|P)$ are finite and strictly convex functions on the corresponding set of probability measures, respectively on the set of transition matrices.

- $H_T(\mathbf{Q}|\mathbf{P}) \geq 0$ and $h(Q|P) \geq 0$ with equality if and only if $\mathbf{Q} = \mathbf{P}$, respectively $Q = P$.

Definition 6. *For a given mobility functional f and any $\beta \in \mathbb{R}$, the* Perron-Frobenius transform *of an irreducible transition matrix P, denoted by P_β, is defined by the matrix*

$$P_\beta(i,j) = \frac{A_\beta(i,j) r_\beta(j)}{\lambda(\beta) r_\beta(i)} \tag{3.1}$$

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 $\{P_\beta\} = \{P_\beta | \beta \in \mathbb{R}\}$ is called the exponential Perron-Frobenius family *of P.*

The Perron-Frobenius transform of P, P_β, is also called the twisted transition matrix.[14] Taking $\beta > 0$, the matrix A_β is obtained from P by inflating

[13] Note that our motivation for the introduction of the relative entropy into the discussion of mobility measurement is completely different than in Chakravarty [8] or Maasoumi and Zandvakili [21].

[14] Our Perron-Frobenius transform corresponds to the Cramér transform (Hollander [19, 7]).

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 normalize 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. The idea behind the twisted transition matrix is to distort the original transition matrix P via the Perron-Frobenius transformation up to the point where movements which were "large" under the original transition matrix become "normal" under the twisted transition matrix.

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 1. *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) *$P_\beta = R_\beta^{-1} \frac{A_\beta}{\lambda(\beta)} R_\beta$ 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,*

$$\lim_{T\to\infty} P_\beta^T = \lim_{T\to\infty} R_\beta^{-1}\left(\frac{A_\beta}{\lambda(\beta)}\right)^T R_\beta = \iota\pi'_\beta$$

where $\iota = (1,\ldots,1)'$. Or, equivalently, $A_\beta^{(T)}(i,j) = \lambda(\beta)^T r_\beta(i)\ell_\beta(j)\left[1+\mathrm{O}(\delta_\beta^T)\right]$ with $0<\delta_\beta<1$.[15]

Proof. These are standard results based on the Perron-Frobenius theorem and can be found, for example, in Berman and Plemmons [4]. □

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 2. *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 $\mathbb{R}$.*

(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)

$$\frac{\lambda'(\beta)}{\lambda(\beta)} = \sum_{i\in\mathcal{E}} \pi_{P_\beta}(i) \sum_{j\in\mathcal{E}} P_\beta(i,j) f(i,j) = M_f^e(P_\beta)$$

where π_{P_β} is the invariant probability distribution of P_β. In particular,

$$\frac{\lambda'(0)}{\lambda(0)} = \lambda'(0) = \sum_{i\in\mathcal{E}} \pi_P(i) \sum_{j\in\mathcal{E}} P(i,j) f(i,j) = M_f^e(P)$$

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

Proof. See appendix. □

Note that the assumption $\mathrm{tr}(P) > 0$ 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 $\frac{\lambda'(\beta)}{\lambda(\beta)}$, being the derivative of the convex function $\log(\lambda(\beta))$, is an increasing function.

3.2 Maximal Deviation

For the implementation of our approach it is important to characterize, for a given mobility functional f, 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 $V_1, \ldots, V_K$ 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 $\Pi = \{V_{i_0}, V_{i_1}, \ldots, V_{i_N}\} = \{i_0, i_1, \ldots, i_N\}$ such that $P(i_{n-1}, i_n) \neq 0$ for all $n = 1, \ldots, N$. In analogy to the definition of the empirical f-mobility, we assign to each path $\Pi = i_0, i_1, \ldots, i_N$ a number $s = s(\Pi)$ as follows:

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

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

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

where a circuit is a path $i_0, i_1, \ldots, i_N$ such that $i_1, \ldots, i_N$ are distinct but $i_0 = i_N$. The maximum must be achieved by a circuit of length $2 \leq N \leq K$ because $f(i,i) = 0$ for all i. It is clear that the value of $a_{\max}(P)$ does not

[16] See Berman and Plemmons [4] for further details.

depend on the value of the positive transition probabilities, but only on the positions of the zero entries. Thus equivalent transition matrices must necessarily have the same $a_{\max}(P)$.[17] In particular, all positive transition matrices P, i.e. $P \gg 0$, have the same $a_{\max} = a_{\max}(P) \geq a_{\max}(Q)$ where Q is any other transition matrix. Thus, there exists a maximal $a_{\max}$ that depends only on the mobility functional f 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 results are not new but can be deduced from a general theory (see Dembo and Zeitouni [13] or Hollander [19]). Although our setting fulfills all assumptions of this general theory, we have chosen a bottom up strategy because this general theory is not specific enough to be readily implemented. As we stress computational aspects and the possibility of empirical applications, we state and prove a version of the large deviation theorem which is self-contained and fully adapted to the application we have in mind.

Proposition 3. *For any irreducible transition matrix P with $\operatorname{tr}(P) > 0$, the Legendre-Fenchel transform $I(a)$ of $\log \lambda(\beta)$ is given for any threshold $a \in \left(M_f^e(P), a_{\max}(P)\right)$ by*

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

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

Proof. See appendix. □

[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 $P(i,j) = 0$.

Theorem 5. *For any threshold* $a \in \left(M_f^e(P), a_{\max}(P)\right)$*, there exists a unique* $\beta(a) \in \mathbb{R}$ *and a Perron-Frobenius transform of* P*,* $P_\beta(a)$*, such that*

(i)

$$\begin{aligned}\lim_{T\to\infty} \frac{1}{T} \log \mathbb{P}\left\{ S_T = \frac{1}{T}\sum_{t=1}^{T} f(X_{t-1}, X_t) \geq a \right\} &= -I(a) \\ &= -\sup_{\beta\in\mathbb{R}} \left(a\beta - \log\lambda(\beta)\right) \\ &= -h(P_{\beta(a)}|P)\end{aligned}$$

(ii)

$$M_f^e(P_{\beta(a)}) = \sum_{i\in\mathcal{E}} \pi_{P_{\beta(a)}} \sum_{j\in\mathcal{E}} P_{\beta(a)}(i,j) f(i,j) = a$$

Proof. See appendix. □

Note that the assumption $\mathrm{tr}(P) > 0$ guarantees that there always exists a non-trivial threshold $a > M_f^e(P)$. This Theorem shows that the tail probabilities of the distribution of S_T decline exponentially fast towards zero. For large T, the exponential 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 $\beta(a)$ and therefore the distortion of P is chosen such that the equilibrium f-mobility index of the twisted transition matrix, $P_{\beta(a)}$, equals a.

Consider two positive transition matrices P and Q with the same equilibrium mobility index M_f^e. It seems plausible to view the transition matrix P as being more mobile than Q if the event $\left\{S_T \geq a \text{ for } a > M_f^e\right\}$ is more probable under P than under Q. For large T, this is, according to Theorem 5, 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 an equilibrium mobility index equal

to the threshold a is larger for Q than for P.[18] 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.

4 Period mobility and examples

4.1 Period Mobility Index

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

Definition 7. *Given a threshold* $a \in \left(M_f^e(P), a_{\max}(P)\right)$*, the* period *f-mobility index,* $M_f^p(P|a)$ *, is defined as*

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

where $P_{\beta(a)}$ *is the Perron-Frobenius transform of* P *with the property* $M_f^e(P_{\beta(a)}) = a$ *(see Theorem 5).*

Straightforward arguments show that our period mobility index $M_f^p(P|a)$ is nothing but the asymptotic probability for T to infinity of consecutive deviations above threshold a from one period to the next:

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

This interpretation justifies the name period mobility index. Since the index corresponds to a probability, it automatically lies between 0 and 1. Values near 0 correspond to low mobility whereas values near 1 correspond to high period mobility. The main purpose of mobility indices is to compare stochastic processes with respect to their mobility.

[18] Steiner [29, Section 9.2.2.1] discusses a generalized form of Theorem 5 which treats also events $\left\{S_T \leq a \text{ for } 0 < a < M_f^e\right\}$.

Definition 8. *Given two regular Markov processes having transition matrices P and Q with* $\mathrm{tr}(P)$ *and* $\mathrm{tr}(Q) > 0$*, P is defined to be* strictly more mobile *with respect to period f-mobility at γ than Q, denoted by $P \succ_p Q$ at γ, if*

$$M_f^p(P|a(\gamma, P)) > M_f^p(Q|a(\gamma, Q)), \qquad \textit{for}\, \gamma \in (0, 1).$$

To any number $\gamma \in (0, 1)$ and any irreducible transition matrix P, the function $a(\gamma, P)$ associates a threshold a according to the following rule:

$$a(\gamma, P) = M_f^e(P) + \gamma \left(a_{\max}(P) - M_f^e(P)\right).$$

P is uniformly more mobile than Q if the above inequality holds for all γ.

As the ranking with respect to period mobility may depend on γ (see section 4.2), the choice of the threshold can become crucial. In order to motivate the method proposed in the definition above, we restrict ourselves to equivalent transition matrices P and Q. They have the property that $a_{\max}(P) = a_{\max}(Q)$. Consider now the following two different cases:

case 1 ($M_f^e(P) = M_f^e(Q)$): In this situation both transition matrices have identical intervals from which the threshold can be chosen: $\left(M_f^e(P), a_{\max}(P)\right) = \left(M_f^e(Q), a_{\max}(Q)\right)$. Thus $a(\gamma, P) = a(\gamma, Q)$ for all $\gamma \in (0, 1)$ so that the resulting threshold is the same for both matrices in absolute terms.

case 2: Suppose without loss of generality that $M_f^e(P) < M_f^e(Q)$. In this case, the ranges to chose the threshold are no longer identical for both matrices. Thresholds in the interval $\left(M_f^e(P), M_f^e(Q)\right)$ are only feasible for matrix P. It therefore makes no sense to compare these matrices at the same threshold. However, it seems appropriate to compare them at identical relative distances above their corresponding equilibrium indices. This is just what the function $a(\gamma, P)$ does.

Although our rule for assigning a threshold may be considered ad hoc, it has the virtue that the applied researcher can fix a value for γ independently of the transition matrices under consideration. In addition, our rule can also be applied to transition matrices which are not equivalent.

4.2 Examples

We are now in a position to illustrate our approach. We do this on the basis of the Bartholomew-functional $f(i,j) = |i-j|$ and the following six transition matrices:

$$P_1 = \begin{pmatrix} 0.60 & 0.35 & 0.05 \\ 0.35 & 0.40 & 0.25 \\ 0.05 & 0.25 & 0.70 \end{pmatrix} \qquad P_2 = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.2 & 0.7 \end{pmatrix}$$

$$P_3 = \begin{pmatrix} 0.600 & 0.399 & 0.001 \\ 0.301 & 0.400 & 0.299 \\ 0.099 & 0.201 & 0.700 \end{pmatrix} \qquad P_{\text{x}} = \begin{pmatrix} 0.40 & 0.55 & 0.05 \\ 0.55 & 0.40 & 0.05 \\ 0.05 & 0.05 & 0.90 \end{pmatrix}$$

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

The first two transition matrices, P_1 and P_2, have been introduced by Dardanoni [10]. The third matrix P_3 is a positive analogue to the third matrix used in Dardanoni's examples.[19] Dardanoni used these matrices to document the inconsistency between alternative mobility indices. In the following, we call these matrices the Dardanoni-matrices. They share the particularity that their Bartholomew mobility index is the same. P_1 and P_3

[19] The original third matrix by Dardanoni [10] had a zero-entry in position (1,3). We substituted this matrix by a positive analogue P_3 in order to compare positive matrices only.

even have the same index of unconditional probability of leaving the current class as well as the same Prais and eigenvalue index.

The transition matrix P_{x} was chosen to demonstrate that the ranking of transition matrices according to period mobility may depend on the value chosen for the threshold. P_{x} shares the same Bartholomew index with the Dardanoni-matrices. In addition, P_{x} also shares the same values for the index of leaving the current class as well as the Prais and the eigenvalue index with P_1 and P_3. 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 that all six transition matrices share the same invariant distribution $(1/3, 1/3, 1/3)'$. Table 2 summarizes the characteristics of all transition matrices.

A straightforward computation shows that we have the following inequalities with respect to equilibrium mobility:

$$M_f^e(P_{\mathrm{ident}}) < M_f^e(P_1) = M_f^e(P_2) = M_f^e(P_3) = M_f^e(P_{\mathrm{x}}) < M_f^e(P_{\mathrm{mobile}})$$

Thus, according to the criterium of equilibrium mobility, P_{ident} represents the least mobile process whereas P_{mobile} represents the most mobile process. The other four processes have index values between these two but cannot be distinguished in terms of equilibrium mobility.

As all matrices have strictly positive entries, their $a_{\max}$ is the same and equals 2. A circuit which achieves $a_{\max}$ is $\{1, 3, 1\}$. Since the Dardanoni-matrices and P_{x} also share the same equilibrium mobility index, comparisons of period mobility in relative and absolute terms are identical (case 1 in

subsection 4.1). However, if these matrices are to be compared with P_{mobile} or P_{ident} only a relative perspective makes sense (case 2 in subsection 4.1).

In Figure 1 we have plotted our period mobility index as a function of γ.[20] This figure shows that the ranking

$$P_{\text{ident}} \prec_p P_3 \prec_p P_1 \prec_p P_2 \prec_p P_{\text{mobile}}$$

is independent of γ and therefore uniform. Table 2 reports the actual values of the index for $\gamma = 0.5$. This means that we measure period mobility at a threshold halfway between $a_{\max}$ and the value of the equilibrium mobility index.

While it is impossible to distinguish Dardanoni's matrices with respect to equilibrium mobility, the matrices are somewhat different concerning their convergence mobility. If one likes to capture both aspects of mobility, the resulting rankings were up to now completely arbitrary and depend heavily on the choice of combination of equilibrium and conventional convergence indices.[21] The virtue of our approach is that it reduces this arbitrariness to the choice of a mobility functional f. 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.

Although it was possible to rank the Dardanoni matrices, P_{ident}, and P_{mobile} uniformly in terms of period mobility such a situation cannot be ex-

[20] The numerical implementation is straightforward and is based on the results presented in Proposition 3 and Theorem 5. As the function to be optimized is strictly convex and possesses a unique supremum, the actual computations are free from numerical complications. MATLAB routines are available from the authors.

[21] For example, the combination of the Batholomew index with the Prais index leads to a different ranking than the combination of the Batholomew index with the second largest eigenvalue index.

pected to prevail under all circumstances. Consider, for example, the transition matrices P_2 and P_{x}. Figure 2 plots their period mobility indices as a function of γ. For $0 < \gamma < 0.2884$, $P_{\mathrm{x}} \succ_p P_2$ holds whereas for $0.2884 < \gamma < 1$ the reverse is true, i.e. $P_{\mathrm{x}} \prec_p P_2$. Since $a_{\max}(P_2) = a_{\max}(P_{\mathrm{x}}) = 2$ and $M_f^e(P_2) = M_f^e(P_{\mathrm{x}})$, $a(\gamma, P_2) = a(\gamma, P_{\mathrm{x}})$ for all $\gamma \in [0, 1]$ so that relative and absolute comparisons yield identical results. The value $\gamma = 0.2884$ corresponds to a threshold $a = 0.9089$. In order to get an intuition of the dependence of period mobility ranking with respect to γ, compare thresholds below and above 0.9089.

- For $a \in \big(M_f^e(P), 0.9089\big)$, the probabilities of consecutive large deviations of empirical mobility are higher for transition matrix P_{x} because P_{x} shows less weight on its main diagonal and is thus less persistent than P_2. Note in this respect the comparatively high transition probabilities $P_{\mathrm{x}}(1, 2)$ and $P_{\mathrm{x}}(2, 1)$ which receive weight 1 by the Bartholomew functional.

- For $a \in (0.9089, 2)$, the chances for consecutive large deviations are now higher for transition matrix P_2 for two reasons. First, moving to adjacent states does not boost empirical mobility further because such movements receive only weight 1. Thus the high transition probabilities $P_{\mathrm{x}}(1, 2)$ and $P_{\mathrm{x}}(2, 1)$ don't help anymore to keep the probability for consecutive large deviations of empirical mobility at high levels. Second, matrix P_2 shows higher probabilities of larger movements, i.e. transitions $1, 3$ and $3, 1$, which receive weight 2 by the Bartholomew functional. Therefore, such movements are necessary to keep the probability of continued large deviations at high levels. The higher the chosen threshold a, the more important the higher probabilities of transitions $1, 3$ and $3, 1$ of matrix P_2 become.

Although this example shows that there is no guarantee for a uniform ranking with respect to period mobility, it also instructs us to examine the whole plot of the period mobility index as a function of γ (as in figures 1 and 2) as this plot provides useful information which enhances the understanding and interpretation of mobility.[22]

5 Conclusions

This paper has shown how the choice of a mobility functional simultaneously determines 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, interesting relations to the existing literature are opened up. The period mobility index is related to the speed at which the tail probabilities of the empirical mobility converge to zero. For a given deviation from equilibrium mobility, this convergence takes place at an exponential rate which can be expressed as the relative entropy of the twisted with respect to the original transition matrix. This exponential rate then leads to the definition of the period mobility index. As the numerical computations are easily implemented, we suggest to report both, the value of the equilibrium mobility index and the plot of the period mobility index as a function of γ. This conveys information on both aspects of mobility in an efficient manner.

The measurement of mobility thus reduces to the specification of a mobility functional. This way of proceeding presents several advantages. First,

[22] The problem is similar to the case of no first-order stochastic dominance in Fields, Leary and Ok [14]. Like in this paper, we propose to examine the mobility ranking over the whole range (over all $\gamma \in (0, 1)$ in our case).

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. Finally, the method is very flexible as it can be readily extended to more general state spaces and Markov processes.

A Proofs

Lemma 2. *Any primitive transition matrix P and any functional f on $E \times E$ satisfy the following uniform recurrence condition (R): There exists a positive integer m such that*

$$\eta = \min_{i \in \mathcal{E}} \sum_{j \in \mathcal{E}} \mathbb{P}(X_m = j, S_m > a | X_0 = i) > 0 \qquad \text{for } a \in \left(M_f^e(P), a_{\max}(P)\right). \tag{A.1}$$

Proof. We will show for all i and j, that there exists a path Π which leads from i to j such that $s(\Pi) > a$. Let Π^* be a circuit such that $s(\Pi^*) = s\left(\{i_0^*, i_1^*, \ldots, i_N^*\}\right) = a_{\max}(P)$. 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 $\Pi = \left\{ \Pi_1, \underbrace{\Pi^*, \ldots, \Pi^*}_{q \text{ times}}, \Pi_2 \right\}$ 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:

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

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 $m = 2m_1 + q^*N^*$. □

Lemma 3. *The moment generating function $M_T(\beta)$ of $\sum_{t=1}^T f(X_{t-1}, X_t)$ equals*

$$M_T(\beta) = \mathbb{E}_{\mathbf{P}} \left(\exp\left(\beta \sum_{t=1}^T f(X_{t-1}, X_t) \right) \right) = \sum_{i \in \mathcal{E}} \mu(i) \sum_{j \in \mathcal{E}} A_\beta^{(T)}(i, j) \tag{A.3}$$

Proof.

$$
\begin{aligned}
M_T(\beta) &= \mathbb{E}_{\mathbf{P}}\left(\exp\left(\beta\sum_{t=1}^{T} f(X_{t-1},X_t)\right)\right)\\
&= \sum_{x_0}\sum_{x_1}\cdots\sum_{x_T}\exp\left\{\beta\sum_{t=1}^{T} f(x_{t-1},x_t)\right\}\\
&\qquad\qquad \times P(x_{T-1},x_T)\times\ldots\times P(x_0,x_1)\times\mu(x_0)\\
&= \sum_{x_0}\sum_{x_1}\cdots\left(\sum_{x_T}\exp\left\{\beta f(x_{T-1},x_T)\right\}\times P(x_{T-1},x_T)\right)\\
&\quad\times\exp\left\{\beta\sum_{t=1}^{T-1} f(x_{t-1},x_t)\right\}\times P(x_{T-2},x_{T-1})\times\ldots\times P(x_0,x_1)\times\mu(x_0)\\
&= \sum_{x_0}\sum_{x_1}\cdots\sum_{x_{T-1}} A_\beta(1)(x_{T-1})\exp\left\{\beta\sum_{t=1}^{T-1} f(x_{t-1},x_t)\right\}\\
&\qquad\qquad \times P(x_{T-2},x_{T-1})\times\ldots\times P(x_0,x_1)\times\mu(x_0)
\end{aligned}
$$

where $A_\beta(1)(x_{T-1})$ denotes the x_{T-1}-th element of the vector of row sums. Proceeding further in this manner, one finally gets[23]

$$M_T(\beta)=\sum_{x_0} A_\beta^{(T)}(1)(x_0)\mu(x_0)=\sum_{i\in\mathcal{E}}\mu(i)\sum_{j\in\mathcal{E}} A_\beta^{(T)}(i,j).$$

□

Proof. Proof of proposition 2

(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 [4, corollary 1.3.29]). Thus $\lambda(\beta)$ is strictly increasing which proves (iii).

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

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

If $M_T(\beta)$ denotes the moment generating function of $\sum_{t=1}^{T} f(X_{t-1}, X_t)$, lemma 3 implies that

$$M_T(\beta) = \sum_{i \in \mathcal{E}} \mu(i) \sum_{j \in \mathcal{E}} A_\beta^{(T)}(i,j) \tag{A.4}$$

$$= \lambda(\beta)^T \sum_{i \in \mathcal{E}} \mu(i) \sum_{j \in \mathcal{E}} r_\beta(i) \ell_\beta(j) \left[1 + O(\delta_\beta^T)\right] \tag{A.5}$$

where the second equality follows from Proposition 1. This implies that $\lim_{T\to\infty} [M_T(\beta)]^{1/T} = \lambda(\beta)$ because $\left\{ \sum_{i \in \mathcal{E}} \mu(i) \sum_{j \in \mathcal{E}} r_\beta(i) \ell_\beta(j) \left[1 + O(\delta_\beta^T)\right] \right\}^{1/T}$ 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 $\mathbb{R}$ (Billingsley [7, 148]) for every T. As $\lambda(\beta)$ and $\log \lambda(\beta)$ are the pointwise limits of convex functions, they are convex (Rockafellar [26, 90]).

The two functions cannot be linear on some proper subinterval of $\mathbb{R}$ 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 infinity as β goes to infinity and because on the other hand $\lambda(\beta)$ and $\log \lambda(\beta)$ are bounded from below by $\max_{i \in \mathcal{E}} P(i,i) > 0$, respectively by $\max_{i \in \mathcal{E}} \log P(i,i) > -\infty$ as $\mathrm{tr}(P) > 0$. $\lambda(\beta)$ 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 (see Magnus and Neudecker [22, 161-162])

$$\lambda'(\beta) = \frac{\mathrm{d}\lambda(\beta)}{\mathrm{d}\beta} = \ell_\beta' \frac{\mathrm{d}A_\beta}{\mathrm{d}\beta} r_\beta \tag{A.6}$$

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 1,

we obtain:

$$\lambda'(\beta) = \lambda(\beta) \sum_{i \in \mathcal{E}} \ell_\beta(i) r_\beta(i) \sum_{j \in \mathcal{E}} \frac{P(i,j) \mathrm{e}^{\beta f(i,j)} r_\beta(j)}{\lambda(\beta) r_\beta(i)} f(i,j) \tag{A.7}$$

$$= \lambda(\beta) \sum_{i \in \mathcal{E}} \pi_{P_\beta}(i) \sum_{j \in \mathcal{E}} P_\beta(i,j) f(i,j). \tag{A.8}$$

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

Proof. Proof of Proposition 3

Instead of A_β consider the matrix $e^{-a\beta} A_\beta$. This matrix has maximal eigenvalue $e^{-a\beta}\lambda(\beta)$. Applying Proposition 2 to $e^{-a\beta} A_\beta$ implies that $g(\beta) = e^{-a\beta}\lambda(\beta)$ is a differentiable strictly convex function of β with derivative equal to

$$g'(\beta) = \frac{\mathrm{d} e^{-a\beta}\lambda(\beta)}{\mathrm{d}\beta} = e^{-a\beta}[-a\lambda(\beta) + \lambda'(\beta)]. \tag{A.9}$$

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 and we have

$$\lambda^m(\beta) = \lambda^m(\beta) \sum_{j \in \mathcal{E}} \varphi_\beta(j) = \sum_{j \in \mathcal{E}} \sum_{i \in \mathcal{E}} \varphi_\beta(i) A_\beta^{(m)}(i,j) = \sum_{i \in \mathcal{E}} \varphi_\beta(i) \sum_{j \in \mathcal{E}} A_\beta^{(m)}(i,j).$$

The (i,j)-th element of A_β^m can be written as $A_\beta^{(m)}(i,j) = \sum_\nu \mathbb{P}\left\{X_m = j, S_m = \nu | X_0 = i\right\} e^{\beta m \nu}$ where ν runs over all positive values of S_m. As Lemma 2 implies that P and f satisfy the Uniform Recurrence Condition (R), we get:

$$\lambda^m(\beta) = \sum_{i \in \mathcal{E}} \varphi_\beta(i) \sum_{j \in \mathcal{E}} \sum_\nu \mathbb{P}\left\{X_m = j, S_m = \nu | X_0 = i\right\} e^{\beta m \nu}$$

$$\geq \eta \sum_{\nu > a} e^{\beta m \nu}$$

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 $(0, +\infty)$. □

Proof. Proof of Theorem 5: The following lemma is needed before proceeding to the proof □

Lemma 4. *Let P be a primitive transition matrix. For any $\beta \in \mathbb{R}$ and any transition matrix Q which is absolutely continuous with respect to P, the following decomposition holds:*

$$h(Q|P) = h(Q|P_\beta) + \beta \sum_{i\in\mathcal{E}} \pi_Q(i) \sum_{j\in\mathcal{E}} Q(i,j) f(i,j) - \log \lambda(\beta)$$

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

Proof. As $\mathbf{Q}$ is absolutely continuous with respect to $\mathbf{P}$, and $\mathbf{P}$ and $\mathbf{P}_\beta$ are equivalent, $\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\Big|_{A_T} = \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}_\beta}\frac{\mathrm{d}\mathbf{P}_\beta}{\mathrm{d}\mathbf{P}}\Big|_{A_T}$.[24] The definition of the relative entropy then leads to: $H_T(\mathbf{Q}|\mathbf{P}) = H_T(\mathbf{Q}|\mathbf{P}_\beta) + \int \log \frac{\mathrm{d}\mathbf{P}_\beta}{\mathrm{d}\mathbf{P}}\Big|_{A_T} d\mathbf{Q}$.

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

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

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

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

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

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:

$$\begin{aligned}
\int \log \frac{\mathrm{d}\mathbf{P}_\beta}{\mathrm{d}\mathbf{P}}\bigg|_{A_T} d\mathbf{Q} &= \sum_{i \in \mathcal{E}} \pi_Q(i) \log \frac{\pi_{P_\beta}(i)}{\pi_P(i)} \\
&\quad + \beta T \sum_{i \in \mathcal{E}} \pi_Q(i) \sum_{j \in \mathcal{E}} Q(i, j) f(i, j) - T \log \lambda(\beta)
\end{aligned}$$

where we have substituted i for x_0 and x_{t-1} and j for x_t. An application of the ergodic theorem (see Stroock [30]) shows that $h(Q|P) = \lim_{T\to\infty} \frac{1}{T} H_T(\mathbf{Q}|\mathbf{P})$ and $h(Q|P) = \lim_{T\to\infty} \frac{1}{T} \log \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\Big|_{A_T}$ $\mathbf{Q}|_{A_t}$-a.s. This result is also known as the Shannon-MacMillan-Breiman theorem. From this we get

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

For $a \in \left(M_f^e(P), a_{\max}(P)\right)$, Propositions 2 and 3 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 $\min_{Q\in\Pi_a} h(Q|P) = h(P_{\beta(a)}|P)$. □

The proof of Theorem 5 proceeds in two steps. First we derive an upper and a lower bound and then show that they converge to the same limit.

Proof. Upper bound.

The application of Chebycheffs inequality to the function $g(x) = e^{T\beta x}$ with $\beta > 0$, sometimes called the exponential overbound lemma, leads to

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

where $\mathbb{E}_{\mathbf{P}}$ is the expectation with respect $\mathbf{P}_\mu$. Lemma 3 implies that the expectation equals

$$\mathbb{E}_{\mathbf{P}} \left[e^{\beta \sum_{t=1}^{T} f(X_{t-1}, X_t)} \right] = \sum_{x_0\in\mathcal{E}} A_\beta^{(T)}(1)(x_0) \mu(x_0).$$

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

$$\begin{aligned}\frac{1}{T}\log\mathbb{P}\left\{\frac{1}{T}\sum_{t=1}^{T}f(X_{t-1},X_t)\geq a\right\} &\leq -\beta a+\frac{1}{T}\log\left(\sum_{x_0\in\mathcal{E}}A_\beta^{(T)}(1)(x_0)\mu(x_0)\right)\\ &\leq -\beta a+\log\lambda(\beta)\\ &\qquad+\frac{1}{T}\log\left(\sum_{x_0\in\mathcal{E}}\frac{A_\beta^{(T)}(1)(x_0)}{\lambda(\beta)^T}\mu(x_0)\right)\\ &\leq -\beta a+\log\lambda(\beta)+o\left(\frac{1}{T}\right)\end{aligned}$$

The last step follows from Proposition 1 by observing that $\left(\frac{A_\beta}{\lambda_\beta}\right)^T$ converges to $\ell_\beta r'_\beta \gg 0$. Taking the limit with respect to T implies:

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

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

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

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

Proof. Lower bound.

Take any $\mathbf{Q}$ such that $\sum_{i\in\mathcal{E}}\pi_Q(i)\sum_{j\in\mathcal{E}}Q(i,j)f(i,j)=a$. 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

$$B = \left\{ \left. \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} \right|_{A_T} \leq e^{T(h(Q|P)+\delta)} \right\} = \left\{ e^{-T(h(Q|P)+\delta)} \left. \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} \right|_{A_T} \leq 1 \right\}$$
$$= \left\{ \frac{1}{T} \log \left. \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} \right|_{A_T} \leq h(Q|P) + \delta \right\}.$$

Then:

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

By assumption the event A occurs almost surely under $\mathbf{Q}$ for $T \to \infty$ as a consequence of ergodicity. $\mathbb{P}_{\mathrm{Q}}\{B\}$ is controlled by the Shannon-MacMillan-Breiman theorem. Thus $\mathbb{P}_{\mathrm{Q}}\{A \cap B\}$ converges to 1 in probability. This implies that

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

Setting $Q = P_{\beta(a)}$ with $a = \sum_{i \in \mathcal{E}} \pi_{P_{\beta(a)}}(i) \sum_{j \in \mathcal{E}} P_{\beta(a)}(i,j) f(i,j)$ and taking limits with respect to T, we finally get:

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

□

References

[1] A. B. Atkinson, *The measurement of economic mobility*, in Social Justice and Public Policy, A. B. Atkinson, ed., Cambridge Massachusetts, 1983, MIT Press, pp. 61–75.

[2] D. J. Bartholomew, *Stochastic Models for Social Processes*, John Wiley & Sons, Chichester, 3rd ed., 1983.

[3] R. Bénabou and E. A. Ok, *Mobility as progressivity: Ranking income processes according to equality of opportunity*, Working Paper W8431, NBER, 2001.

[4] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, vol. 9, Society for Industrial and Applied Mathematics, 1994.

[5] P. Billingsley, *Ergodic Theory and Information*, John Wiley & Sons, New York, 1965.

[6] ——, *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968.

[7] ——, *Probability and Measure*, John Wiley & Sons, New York, 3rd ed., 1995.

[8] S. R. Chakravarty, *A note on the measurement of mobility*, Economics Letters, 48 (1995), pp. 33–36.

[9] J. Conlisk, *Monotone mobility matrices*, Journal of Mathematical Sociology, 15 (1990), pp. 173–191.

[10] V. Dardanoni, *Measuring social mobility*, Journal of Economic Theory, 61 (1993), pp. 372–394.

[11] ——, *Income distribution dynamics: Monotone markov chains make light work*, Social Choice and Welfare, 12 (1995), pp. 181–192.

[12] J. Davidson, *Stochastic Limit Theory*, Oxford University Press, Oxford, 1994.

[13] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Spinger-Verlag, New York, 2nd ed., 1998.

[14] G. S. Fields, J. B. Leary, and E. A. Ok, *Stochastic dominance in mobility analysis*, Economics Letters, 75 (2002), pp. 333–339.

[15] G. S. Fields and E. A. Ok, *The meaning and measurement of income mobility*, Journal of Economic Theory, 71 (1996), pp. 349–377.

[16] ——, *The measurement of income mobility: An introduction to the literature*, in Handbook on Income Inequality Measurement, J. Silber, ed., Boston, 1999, Kluwer Academic Publishers, pp. 557–596.

[17] J. Geweke, R. C. Marshall, and G. G. A. Zarkin, *Mobility indices in continuous time markov chains*, Econometrica, 54 (1986), pp. 1407–1423.

[18] P. Gottschalk and E. Spolaore, *On the evaluation of economic mobility*, Review of Economic Studies, 69 (2002), pp. 191–208.

[19] F. d. Hollander, *Large Deviations*, vol. 14 of Fields Institute Monographs, American Mathematical Society, Providence, Rhode Island, 2000.

[20] E. Maasoumi, *On mobility*, in Handbook of Applied Economic Statistics, A. Ullah and D. Giles, eds., New York, 1998, Marcel Dekker, pp. 119–173.

[21] E. Maasoumi and S. Zandvakili, *Generalized entropy measures of mobility for different sexes and income levels*, Journal of Econometrics, 43 (1990), pp. 121–133.

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

[23] T. Mitra and E. A. Ok, *The measurement of income mobility: A partial ordering approach*, Economic Theory, 12 (1998), pp. 77–102.

[24] R. B. Nelsen, *An Introduction to Copulas*, vol. 139 of Lecture Notes in Statistics, Springer-Verlag, New York, 1999.

[25] J. R. Norris, *Markov Chains*, Cambridge University Press, Cambridge, 1997.

[26] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.

[27] A. F. Shorrocks, *The measurement of mobility*, Econometrica, 46 (1978), pp. 1013–1034.

[28] P. M. Sommers and J. Conlisk, *Immobility measures for markov chains*, Journal of Mathematical Sociology, 6 (1979), pp. 253–276.

[29] P. Steiner, *Anwendungen von Konzepten grosser Abweichungen auf Fragestellungen der Einkommensdynamik*, dissertation.de - Verlag im Internet GmbH, Berlin, 2004.

[30] D. W. Stroock, *An Introduction to Markov Processes*, Springer-Verlag, Berlin, 2005.

Table 1: Some commonly used mobility indices

name of index	**definition**
equilibrium mobility indices	
Bartholomew's index	$\sum_{i=1}^{K} \pi(i) \sum_{j=1}^{K} P(i,j)\|i-j\|$
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	$\frac{K-\sum_{1=1}^{K}\|\lambda_i\|}{K-1}$
second largest eigenvalue index	$1-\delta(P)$
asymptotic speed of convergence	$-\log\delta(P)$
determinant index	$1-\det(P)$

P ... irreducible transition matrix

π ... invariant distribution of P

λ_i ... eigenvalues of P

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

Table 2: Characteristics and mobility indices for test transition matrices

	P_1	P_2	P_3	P_{x}	P_{mobile}	P_{ident}
equilibrium mobility index[a]	0.4667	0.4667	0.4667	0.4667	0.8889	0.0027
period mobility index[a] $\gamma = 0.5$	0.4554	0.5495	0.3799	0.4720	0.7841	0.0630
Prais' index[b]	0.65	0.60	0.65	0.65	1.00	0.003
eigenvalue index[b]	0.65	0.60	0.65	0.65	1.00	0.003
second largest eigenvalue index[b]	0.3854	0.4268	0.3994	0.1500	1.00	0.003
asymptotic speed of convergence[b]	0.4868	0.5565	0.5098	0.1625	$+\infty$	0.0030
determinant index[b]	0.9475	0.87	0.9403	0.8725	1.00	0.0060
$a_{\max}(P)$[a]	2	2	2	2	2	2
invariant distribution	$\frac{1}{3}(1,1,1)'$	$\frac{1}{3}(1,1,1)'$	$\frac{1}{3}(1,1,1)'$	$\frac{1}{3}(1,1,1)'$	$\frac{1}{3}(1,1,1)'$	$\frac{1}{3}(1,1,1)'$

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

b . . . For definitions see Table 1.

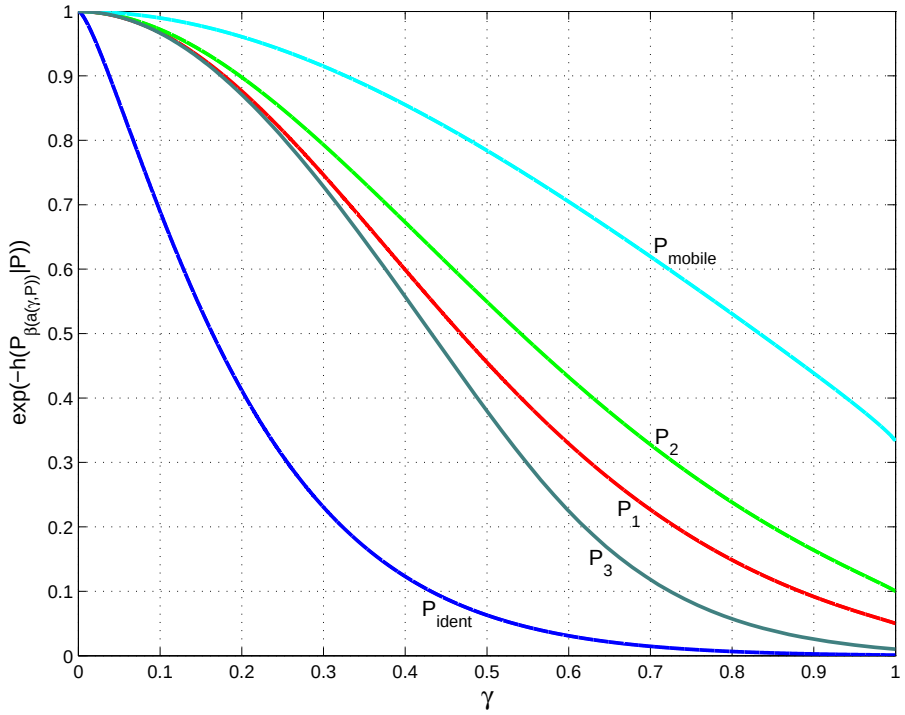

Figure 1: Period Mobility of Test Matrices (uniform ranking)

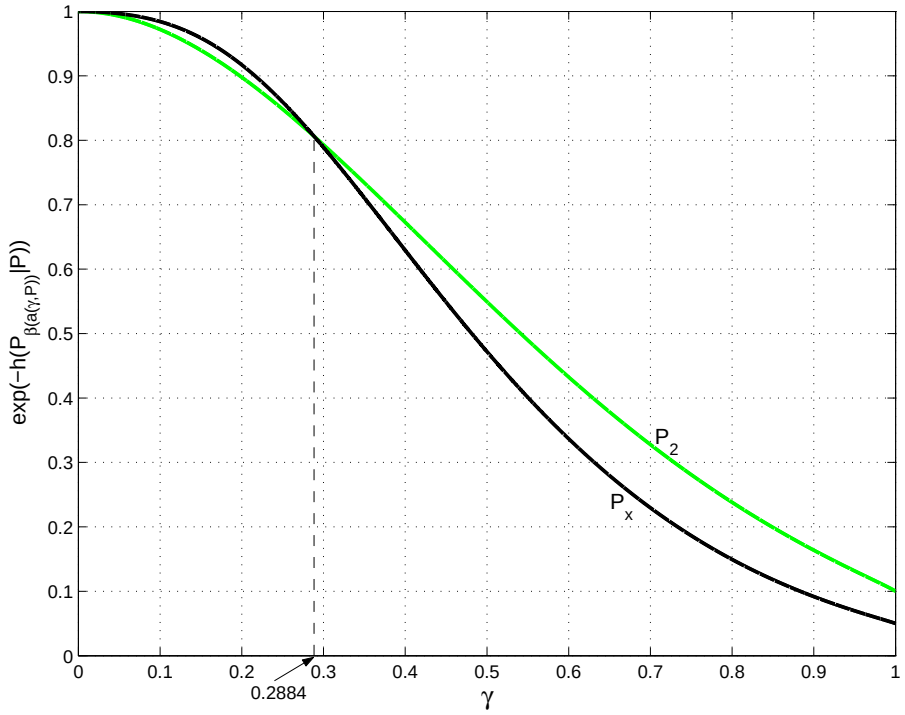

Figure 2: Period Mobility of Test Matrices (nonuniform ranking)