Testing Shape Restrictions on the Steady-State Distribution of a Finite Markov Chain.

Alexei Onatski*
Department of Economics, Harvard University
aonatski@kuznets.fas.harvard.edu

This draft: March 18, 2000.
JEL classification (C49)

*I thank James Stock, David Laibson, Jack Porter, Vladislav Karguine, and all participants of Harvard Econometrics lunch seminar for helpful discussions and suggestions.

Abstract

This paper develops a theory of the likelihood ratio test on the shape of the steady-state distribution of a finite Markov chain. A variety of assumptions about the shape of the steady-state distribution can be represented by a system of linear inequality restrictions on the steady-state probabilities. I show that, for given pseudo-true transition probabilities, asymptotic distribution of the test of linear inequality restrictions on the steady-state probabilities is equal to a mixture of the chi-squared distributions with different degrees of freedom. I use this result together with the finite Markov chain structure of the data generating process to find the exact asymptotic distribution of the global test of the monotonicity of the steady-state distribution. In doing this, I assume that the only non-zero transition probabilities are those on the three main diagonals of the transition matrix.

1. Introduction

Modeling social and economic processes as finite Markov chains is a widespread practice in empirical studies. Such models have been extensively used to explain patterns of social and occupational mobility, voting behavior, income distribution dynamics, distribution of firm sizes, success or failure of different marketing strategies, patterns of countries' convergence, and evolution of cities. For a review of related literature one can see Bartholomew (1973) and Shorrocks (1978).

Those studies that used Markov chains as their working model often estimated the corresponding steady-state distribution to predict the equilibrium state or to summarize tendencies in the past development of the underlying process. For a recent example, Quah (1993) divides countries into several groups according to their income relative to the world's average income, estimates the probabilities of transitions between the groups, and finds the corresponding steady-state distribution. The estimated steady-state distribution has two peaks with many rich countries, many poor countries, and relatively few middle income countries in between. This result has been interpreted as evidence supporting a hypothesis of world's polarization into two clubs, a club of rich and a club of poor.

Surprisingly, despite the importance of conclusions based on the estimated shape of the steady-state distribution, assessing statistical significance of the estimates received little attention. Usually, the form of the distribution is judged by visual inspection of the corresponding histogram. Unfortunately, as is well known, the steady-state probabilities are very sensitive to the underlying transition probabilities. Therefore, the estimated shape of the distribution might be a very fragile function of the data.

This paper develops a theory of the likelihood ratio test on the shape of the steady-state distribution of a finite Markov chain. A variety of assumptions about the shape of the steady-state distribution such as n-modality, degree of polarization, monotonicity, or degeneracy can be represented by a system of linear inequality restrictions on the steady-state probabilities. I show that, for given pseudo-true transition probabilities, asymptotic distribution of the test of linear inequality restrictions on the steady-state probabilities is equal to a mixture of the chi-squared distributions. I use this result together with the finite Markov chain structure of the data generating process to find the exact asymptotic distribution of the global test of the monotonicity[1] of the steady-state distribution. In

[1] A distribution is said to be monotone if the corresponding histogram has monotonically increasing or decreasing bars.

doing this, I assume that the only non-zero transition probabilities are those on the three main diagonals of the transition matrix.

Large literature on testing inequality restrictions is related to this work. Asymptotic distribution of the likelihood ratio test of linear inequality restrictions on the mean of a multivariate normal vector was studied in Kudo (1963), Perlman (1969), and Shapiro (1985). They show that the asymptotic distribution has a form of a mixture of the chi-squared distributions with different degrees of freedom. Wolak (1989,1991) analyzes tests of linear and non-linear inequality constraints in non-linear econometric models. Though very general, his results are not directly applicable to the problem at hand because he requires i.i.d. property for the data.[2]

An important early study that can be used to link Wolak's analysis to the multivariable normal literature is Chernoff (1954). He studies asymptotic properties of the likelihood ratio test when the true parameter value is on the boundary of both null and alternative hypotheses. Chernoff shows that asymptotically the problem reduces to a special case of testing linear inequality restrictions on the mean of a normal random variable with the covariance matrix equal to the inverse of the information matrix for the initial problem. This paper extends Chernoff's result to a finite Markov chain data generating process and uses this extension to find the asymptotic distribution of the likelihood ratio for given pseudo-true transition probabilities.

Unfortunately, the asymptotic distribution is not invariant to the choice of the pseudo-true transition probabilities. This is the same problem that makes Wolak (1989, 1991) conclude that tests of inequality restrictions in non-linear models are intrinsically local. I show that in one case, interesting for empirical research, the problem can be overcomed, that is, the critical values of the global test can be found. In particular, I analyze the monotonicity hypothesis about the steady-state distribution and assume that the only non-zero transition probabilities are those on the three main diagonals of the transition matrix.

The triple-diagonal assumption is not new in the literature. Shorrocks (1976) makes this assumption when studying income mobility, Champernowne (1953) uses similar assumption trying to explain Pareto-type income distributions, Adel-

[2] Wolak (1989) explicitly assumes i.i.d, data, whereas Wolak (1991) inherits this assumption from the 1989 paper by relying on its results. Andrews (1998, 1999) studies tests of a point null hypothesis in a very general framework, allowing in particular for a Markov chain data generating process. This paper, however, considers testing composite nulls versus composite alternatives.

man (1958) estimates a triple-diagonal transition probability matrix for evolution of concentration of firms in steel industry, Quah (1993) estimates a triple-diagonal transition matrix explaining countries' income distribution dynamics. The assumption implies a simple relationship between the steady-state and the transition probabilities.

The monotonicity hypothesis has considerable empirical interest. For example, for Quah's (1993) countries' income distribution analysis, the hypothesis corresponds to world becoming either all-rich or all-poor. An application of the results of this paper to analysis of countries' income distribution dynamics is currently in progress. Some preliminary findings are described in Kremer, Onatski, and Stock (2000). Interestingly, the authors cannot reject a hypothesis that world is becoming all-rich using the global monotonicity test described in this paper.

The rest of the paper is organized as follows. Section 2 describes the Markov chain framework and formulates general hypotheses about the shape of the steady-state distribution. Extension of Chernoff's results to the Markov chain data generating process is the subject of section 3. Section 4 contains results about the asymptotic distribution of the global likelihood ratio test of the shape restrictions. Section 5 concludes.

2. Shape restrictions on the steady-state distribution

The framework of my analysis is as follows. I consider N agents (these might be countries, individuals, firms, etc.) categorized between K groups for the time period from $t = 0$ to $t = T$. The evolution of each agent's position among the groups is governed by a first order Markov process. That is, an agent's membership in a particular group depends only on its classification in the previous period.

Denote the probability of transiting from group i to group j in one period as p_{ij} and the whole matrix[3] of the transition probabilities as P. We assume that the matrix P is ergodic so that there exists a unique steady-state distribution of the agents among the groups. Denote the steady-state probability of being in group i as π_i and the vector of these probabilities as π. Then π must satisfy three conditions:

$$\pi' P = \pi', \ \pi' J = 1, \text{ and } \pi \geq 0, \tag{2.1}$$

where J is a $K \times 1$ vector of ones. The first condition is the definition of the

[3] In the following P also denotes a vector of $K \times (K-1)$ independent parameters of the transition matrix. Whether the vector or the matrix is used should be clear from the context.

steady state. If a distribution of the agents among the groups is in the steady state currently then after one period it is still in the steady state. The second and the third conditions insure that π is a probability distribution.

The object of my study is the shape of the steady-state distribution, π. More precisely, I consider the following hypothesis about π :

$$H : A\pi \geq 0, \tag{2.2}$$

where A is a $C \times K$ matrix and C is the number of constraints. These simple linear inequality constraints have enough flexibility to represent a variety of assumptions about the shape of the steady-state distribution. Below I formulate several special cases of the above hypothesis that have clear interpretation in terms of the shape restrictions.

One interesting hypothesis restricts π_i to be monotonically non-decreasing function of i or, more generally:

$$H_1 : \pi_{k+1} \geq \delta_k \pi_k, \ \ k = 1, ..., K-1, \tag{2.3}$$

where δ_k are some positive numbers greater than or equal to 1. Clearly, the steady-state distribution satisfying these restrictions has a single peak at the extreme group, K. The importance of the peak is regulated by δ_k: the greater δ_k the more pronounced the peak. If, say: all δ_k are equal to 2 then the steady-state size of the largest group is at least 2^{K-1} times larger than that of the smallest group. To formulate a hypothesis that π_i are monotonically non-increasing in i one needs to change the inequality sign in H_1 and take δ_k to be positive numbers less than or equal to 1.

Another hypothesis of the form (2.2) having clear "shape interpretation" is that of n-peakedness of π. I call a distribution n-peaked if the corresponding histogram has n peaks. More formally, let $j_1, j_2, ..., j_r$ be an increasing sequence of integers with $j_1 = 0$ and $j_r = K + 1$ and such that π_k, as a function of $k \in [j_i, j_{i+1}]$, is non-decreasing if i is odd and strictly decreasing if i is even.[4] Then the distribution π is n-peaked with $n = (r + 1)/2$. Note that r is an odd number, so n is an integer. The peaks of the distribution are situated at groups $j_2, j_4, ..., j_{r-1}$. A slightly more general hypothesis than that of n-peakedness allows one to regulate importance of the peaks in the same manner as H_1 allows one to regulate importance of the peak at group K. This hypothesis can be fomulated

[4] I define π_0 and π_{K+1} to be equal to 0.

as follows:

$$\begin{aligned} H_2 &: \pi_{k+1} \geq \delta_k \pi_k, \text{ for } k \in [j_i, j_{i+1}), \ i\text{-odd and} \\ \pi_k &> \delta_k \pi_{k+1}, \text{ for } k \in [j_i, j_{i+1}), \ i\text{-even,} \end{aligned}$$

where δ_k are some positive numbers greater than or equal to 1.

Still another potentially interesting hypothesis about the shape of the steady-state distribution is one about degree of polarization of π. Intuitively, a distribution is the most polarized when it is concentrated in the two most distant groups. A distribution is the least polarized when it is concentrated in a single group. According to this intuition the following hypothesis could be interpreted as a restriction on the degree of polarization of π:[5]

$$H_3 : \min_{i \in \{1,K\}} \pi_i - \delta \max_{j \in \{2,\ldots,K-1\}} \pi_j \geq 0,$$

where δ is greater than or equal to 1. Hypothesis H_3 can be represented as a system of $2(K-2)$ inequalities and therefore has form (2.2).

One of the traditional choices of the test of linear inequality restrictions is the likelihood ratio test. I follow this tradition and concentrate on testing the shape restrictions with a likelihood ratio test of size 0.05. To define the likelihood ratio statistic I introduce some additional notations. Denote the total number of observed transitions from group i to group j as N_{ij}, and the number of agents initially in group i as N_i. Define f_i as the probability of being initially in group i, and f as the vector of the initial probabilities, f_i. I assume that f does not depend on the transition probabilities, P. Given this, the likelihood function of the data X is

$$L(X|P,f) = \prod_{r=1}^{K} f_r^{N_r} \prod_{i,j=1}^{K} p_{ij}^{N_{ij}}$$

and the maximum likelihood estimates of p_{ij} and f_r are $\hat{p}_{ij} = N_{ij}/\sum_{j=1}^{K} N_{ij}$ and $\hat{f}_r = N_r/\sum_{k=1}^{K} N_k$ respectively.

---

[5] A measure of polarization that satisfies this intuition and several other axioms is the coefficient of polarization, $\varkappa$, proposed by Esteban and Ray (1994):

$$\varkappa(\pi) = R \sum_{i,j=1}^{K} \pi_i^{1+\alpha} \pi_j \left| i - j \right|,$$

where α is greater than zero and R is a constant chosen so that $\varkappa$ is between zero and one. However, the hypothesis $\varkappa(\pi) \geq \varkappa_0$ cannot be represented in the form (2.2).

Let us denote the set of transition probabilities, P, satisfying hypothesis (2.2) as Π and the set of discrete probability distributions having no more than K masses as S. Note that conditions (2.1) imply the following convenient formula for π:[6]

$$\pi = \left[(P - I)(P - I)' + JJ' \right]^{-1} J, \tag{2.4}$$

where I is a $K \times K$ identity matrix and J is a $K \times 1$ vector of ones. Thus, the set Π consists of those P that satisfy conditions (2.4) and (2.2). Given these notations, the likelihood ratio statistic is defined as

$$\lambda(X) = -2 \sup_{P \in \Pi, f \in S} \log \left(\frac{L(X|P, f)}{L(X|\hat{P}, \hat{f})} \right).$$

The likelihood ratio test rejects H in favor of its complementary alternative if λ is greater than a critical value c, defined so that the following equality holds:

$$\sup_{P_0 \in \Pi, f_0 \in S} \Pr(\lambda(X) \geq c) = 0.05,$$

where P_0 is the pseudo-true matrix of the transition probabilities and f_0 is the vector of the pseudo-true initial probabilities.

Computation of λ is a relatively easy constrained maximization task. It could be done using standard maximization programs such as `fmincon` routine in Matlab. A much harder task is finding the correct finite sample critical value, c. I briefly discuss difficulties of finite sample computations at the end of the paper. In what follows I concentrate on finding asymptotic approximation to the exact finite sample critical value, c.

3. Asymptotic reduction of the shape restrictions to a simpler hypothesis

Wilks' (1938) classical result on the asymptotic distribution of the likelihood ratio (LR) does not hold in the framework described in the previous section because the transition probabilities, P, are allowed to lie on the boundary of the null set Π. The asymptotic theory of the likelihood ratio test when parameters might lie on the boundary of both the null and the alternative sets was originated in the work by Chernoff (1954). He proved that under suitable regularity conditions

[6] See, for example, Guilbaud (1976).

the general testing problem could be reduced to a test of a related hypothesis about mean of multivariable normal distribution when a single observation from the distribution is available. This reduction facilitates dramatically analysis of the likelihood ratio test.

Below I will extend Chernoff's (1954) results to the Markov processes to show that the asymptotic distribution of the likelihood ratio statistic corresponding to hypothesis (2.2) is the same as that corresponding to a simple hypothesis about the mean of a normal distribution. I now introduce some helpful definitions and new notations.

Definition 3.1. *A set C is positively homogeneous if $p \in C$ implies $ap \in C$ for $a > 0$.*

Definition 3.2. *A set Ψ is approximated by a positively homogeneous set C_Ψ at point 0 if*

$$\inf_{x \in C_\Psi} |x - y| = o(|y|) \text{ for } y \in \Psi \text{ and } \inf_{y \in \Psi} |x - y| = o(|x|) \text{ for } x \in C_\Psi.$$

Denote the constrained maximum likelihood estimate of transition probabilities satisfying H as $\hat{P}_H$. We can reparametrize the problem so that the true parameter P_0 is zero. Indeed, whatever P_0 is, one can define new parameters $\tilde{P} = P - P_0$ such that $\tilde{P}_0 = 0$. Denote C_Π and C_Π^c the positively homogeneous sets approximating Π and Π^c respectively.[7] The following proposition is true.

Proposition 3.3. *If $\hat{P}_H$ is consistent, then, the asymptotic distribution of λ when $N \to \infty$ $(T \to \infty)$ is the same as it would be for the likelihood ratio test of $P \in C_\Pi$ against $P \in C_\Pi^c$ based on one observation from a population with distribution $N(P, J^{-1})$, where J^{-1}/N (respectively J^{-1}/T) is the asymptotic variance-covariance matrix of the unrestricted maximum likelihood estimates.*

Proof: In the case when $N \to \infty$ we can consider our data as N i.i.d. observations of a sequence of states of length T. The distribution of each particular observation is multinomial with probability of a sequence $\{x_1, x_2, ..., x_T\}$ equal to $f_{x_1} p_{x_1 x_2} p_{x_2 x_3} ... p_{x_{T-1} x_T}$. It is then straightforward to check Chernoff's (1954) Condition $\Re$ that imply the above theorem.

In the case when $T \to \infty$ we can consider our data as T observations of a single Markov chain with vector states $\{x_i\} = \{x_1, x_2, ..., x_N\}$ and probability

[7] Here and elsewhere in the paper the superscript c denotes the complementary set.

of transition between states $\{x_i\}$ and $\{y_i\}$ equal to $p_{x_1y_1}p_{x_2y_2}...p_{x_Ny_N}$. It is easy to generalize Chernoff's (1954) result to Markov chain processes by substituting Billingsley's (1961) Conditions 1.1 and 1.2 instead of Condition $\Re$ and otherwise leaving Chernoff's proof unchanged. Checking Conditions 1.1 and 1.2 for the Markov process at hand is straightforward.□

To use Proposition 3.3 we need to prove that $\hat{P}_H$ is consistent. The consistency follows from the fact that the likelihood function is concave and has enough curvature. More formally, the following proposition holds.

Proposition 3.4. *Let P_0 belong to a closed set Ψ and let $\hat{P}_\Psi$ denote a constrained maximum likelihood estimate of P when it is allowed to vary in Ψ. Then $\hat{P}_\Psi$ is consistent.*

Proof: Let $l(X,P)$ denote $\frac{1}{n}$ of the logarithm of the likelihood function $L(X|P,f)$. Consider the Taylor expansion of $l(X,p)$ at the unrestricted maximum likelihood estimate of P, $\hat{P}$:

$$l(X,P) = l(X,\hat{P}) + \frac{1}{2}(P-\hat{P})'\frac{\partial^2 l(X,\tilde{P})}{\partial P^2}(P-\hat{P}),$$

where $\tilde{P}$ is some suitable vector of transition probabilities. Substitute $\hat{P}_\Psi$ and P_0 instead of P, and subtract one resulting equality from the other. We get:

$$l(X,\hat{P}_\Psi)-l(X,P_0) = \frac{1}{2}(\hat{P}_\Psi-\hat{P})'\frac{\partial^2 l(X,\tilde{P})}{\partial P^2}(\hat{P}_\Psi-\hat{P})-\frac{1}{2}(P_0-\hat{P})'\frac{\partial^2 l(X,\tilde{\tilde{P}})}{\partial P^2}(P_0-\hat{P})$$

The left hand side of the equality is non-negative, the second term in the right hand side tends to zero in probability. Therefore, because the second derivative of $l(X,P)$ is negative-definite and uniformly bounded from zero, $\hat{P}_\Psi$ tends to $\hat{P}$ in probability. But $\hat{P}$ is consistent, therefore $\hat{P}_\Psi$ is consistent.□

To summarize, Proposition 3.3 reduces the problem of testing the shape hypothesis (2.2) to that of testing a simpler hypothesis about the mean of a multivariate normal distribution. In the next section we use this result to develop asymptotic theory of the likelihood ratio test of the shape restrictions.

4. Asymptotic distribution of the likelihood ratio test

To use Proposition 3.3 we need to answer two questions. First, how to describe the set C_Π? Second, how to test a hypothesis that the mean of a multivariate normal distribution lies in C_Π? I consider these two questions in turn.

The answer on the first question is as follows. If P_0 lies inside Π, then C_Π is the whole space, $R^{K(K-1)}$, and asymptotically we cannot reject hypothesis (2.2). Similarly, if P_0 lies outside Π, then C_Π is empty and we reject (2.2). The only case left is the one in which P_0 lies on the boundary of Π. In this case, the first order approximation to Π at P_0 is the set of all those transition probabilities satisfying

$$A\frac{\partial}{\partial P}\pi(P_0)(P-P_0) \geq 0 \tag{4.1}$$

The above set is a polyhedral cone and it is obviously a positively homogeneous set (we consider the case when $P_0 = 0$). It is possible to find a closed form expression for matrix $R \equiv A\frac{\partial}{\partial P}\pi(P_0)$. Indeed, denote the fundamental matrix of the Markov chain, $(I - P + J\pi')^{-1}$, as Z. Then, the following lemma is true.

Lemma 4.1. *The derivative of k-th component of the steady-state vector, π, with respect to p_{ij} is equal to*

$$\frac{d\pi_k}{dp_{ij}} = (z_{jk} - z_{ik})\pi_i$$

Proof: see Appendix.□

We now turn to the second question. In the following we assume that P_0 lies on the boundary of Π. The LR statistic for the test of $R(P - P_0) \geq 0$ when P is multivariate normal distributed is usually referred to as chi-bar-squared, $\bar{\chi}^2$, statistic. The properties of this statistic and the related ones were studied extensively. Kudo (1963) shows that if R is an identity matrix then $\bar{\chi}^2$ is distributed as a mixture of chi-squared distributions

$$\Pr(\bar{\chi}^2 \geq z^2) = \sum_{i=0}^{q} \omega_i \Pr(\chi_i^2 \geq z^2), \tag{4.2}$$

where q is the number of constraints and the weights, ω_i, are nonnegative and sum to one. The restriction on R to be an identity matrix is without loss of generality. Indeed, any non-identity matrix R can be absorbed into the variance-covariance matrix of P and vice versa.

Thus, the 95% quantile, $c(P_0)$, of the LR statistic given the pseudo-true transition probabilities, P_0, can be found as a solution to the equation

$$\sum_{i=0}^{q} \omega_i \Pr(\chi_i^2 \geq c(P_0)) = 0.05.$$

The critical value of the LR test, c, is equal to $c(P^*)$, where P^* is the worst possible transition probabilities from the null set. That is

$$c = \sup_{P_0 \in \Pi} c(P_0). \tag{4.3}$$

The problem of finding c is a difficult one. The difficulty stems from the fact that both the variance-covariance matrix of P and the matrix of constraints, R, nontrivially depend on the choice of P_0. This dependence implies that there is no simple way to find the worst possible P_0 where supremum in (4.3) is attained. For example, similar to Wolak's (1991) results, it is not generally true that all inequality constraints are binding at the worst possible P_0. Therefore, in general the critical value could only be found by numerical maximization of the 95% quantiles over the whole set of possible pseudo-true transition probabilities.

However, in one special case we are able to find the critical value analytically. The hypothesis we would like to test is the monotonicity hypothesis (2.3). Let transition probability matrix have the triple-diagonal structure. Precisely, assume that $p_{ij} \neq 0$ if and only if $|i - j| \leq 1$. The last assumption implies a particularly simple relationship between transition probabilities and the steady-state probabilities:

$$\frac{\pi_i}{\pi_{i+1}} = \frac{p_{i+1,i}}{p_{i,i+1}}, \; i = 1, ..., K - 1. \tag{4.4}$$

To see this, note that after one transition, the probability of being in the first group equals the probability of initially being in the first group and remaining there, plus the probability of initially being in group 2 and transiting to group 1. Thus, in steady state, $\pi_1 = \pi_1(1 - p_{12}) + \pi_2 p_{21}$. Simplifying yields $\frac{\pi_1}{\pi_2} = \frac{p_{21}}{p_{12}}$, and the remaining equalities in (4.4) follow by induction.

The following proposition is true.

Proposition 4.2. *Assume that the triple diagonal condition (4.4) holds. Assume also that T goes to infinity. Then the likelihood ratio statistics, λ, for testing monotonicity hypotheis H_1 against the complementary alternative is asymptotically stochastically dominated by $\bar{\chi}^2$ random variable with distribution*

$$\Pr(\bar{\chi}^2 \geq c^2) = \frac{1}{2^{K-1}} \sum_{i=0}^{K-1} \binom{K-1}{i} \Pr(\chi_i^2 \geq c^2)$$

where χ_i^2 is a chi-squared variable with i degrees of freedom, $\chi_0^2 \equiv 0$, and $\binom{K-1}{i}$ is the binomial coefficient. Moreover, there exists a sequence of pseudo-true P_0,

$\{P_{0n}\}$, *all satisfying* H_1 *such that the LR test statistic converges in distribution to* $\bar{\chi}^2$.[8]

To prove the proposition we first need to establish the following lemma.

Lemma 4.3. *Assume that we have one observation,* x*, from* $N(P, I)$*, a multivariate normal distribution with mean* P *and identity variance-covariance matrix. Assume further that the true value of* P *is* 0 *and let* λ_1 *and* λ_2 *be LR statistics for the test that the mean,* P*, lies in the sets* C_1*,* C_2 *respectively. If* $C_1 \subseteq C_2$ *then* $\lambda_1 \geq \lambda_2$*. If in addition* C_2 *uniformly converges to* C_1 *in any bounded subset of the whole space then* $\lambda_2 \to \lambda_1$ *in distribution.*

Proof:

The likelihood ratio statistics λ_1 (λ_2) is equal to half the standard Euclidean distance from x to C_1 (C_2). The distance from x to C_1 is no less than that from x to C_2. Therefore $\lambda_1 \geq \lambda_2$.

There exist a ball, B, of large enough radius such that $\Pr(\lambda_i \leq l)$ is in an arbitrarily small neighborhood of $\Pr(\lambda_i \leq l, x \in B)$ for any l. This and the fact that C_2 converges to C_1 uniformly in B implies the second statement of the lemma.□

Now we turn to the proof of Proposition 4.2:

Let p be the vector of transition probabilities

$$p = (p_{12}, p_{21}, p_{23}, p_{32}, ..., p_{K-1,K}, p_{K,K-1})'.$$

The triple diagonal condition implies that vector p satisfies H_2 if and only if $Rp \geq 0$, where

$$R = \begin{bmatrix} 1 & -\delta_1 & 0 & 0 & ... & 0 & 0 \\ 0 & 0 & 1 & -\delta_2 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & ... & 1 & -\delta_{K-1} \end{bmatrix}.$$

The unrestricted maximum likelihood estimates of p_{ij} have asymptotic variances $p_{ij}(1-p_{ij})/TN\pi_i$ and covariances $-\delta_{ig}p_{ij}p_{gh}/TN\pi_i$. Let J^{-1} be the asymptotic variance-covariance matrix of the unrestricted maximum likelihood estimate

[8] The same proposition is true for the monotone non-increasing version of H_1.

of $\sqrt{T}p$. A linear transformation $\tilde{p} = J^{1/2}(p - p^0)$ changes the null set of H_1 into a positively homogeneous set

$$\Omega = \{\tilde{p} : RJ^{-1/2}\tilde{p} \geq -Rp^0\}$$

and the true transition probabilities, p^0, into $\tilde{p} = 0$.

Denote $S = \{s_1, s_2, ..., s_q\}$ the set of indexes of those constraints that bind at p^0. Let R_S be the matrix that consists of s_i'th rows of R. Then the positively homogeneous set, C_Ω, approximating Ω is

$$C_\Omega = \{\tilde{p} : R_S J^{-1/2}\tilde{p} \geq 0\}.$$

According to Proposition 3.3, the asymptotic distribution of the test H_1 given that p^0 is the vector of the true transition probabilities is the same as the asymptotic distribution of the test of the hypothesis that $R_S J^{-1/2}\tilde{p} \geq 0$ based on one observation from normal distribution $N(\tilde{p}, I)$ when the true $\tilde{p}$ is zero. The distribution $N(0, I)$ is invariant with respect to any orthogonal transformation, O. Therefore, the asymptotic distribution of the test H_1 is the same as the asymptotic distribution of the test of hypothesis that $R_S J^{-1/2} O\tilde{p} \geq 0$.

We now show that there exists an orthogonal transformation O, such that the set $\{\tilde{p} : R_S J^{-1/2} O\tilde{p} \geq 0\}$ includes the "positive orthant", $\mathcal{R}_+^{K-1}$, that consists of all vectors $\tilde{p}$ with positive first $K - 1$ entries.

Indeed, consider matrix $R_S J^{-1/2}$. Denote its i'th raw as x_i. Straightforward but somewhat lengthy algebraic manipulations show that the correlation between x_i and x_{i+1}, is zero if $s_{i+1} - s_i > 1$ and it is positive and equal to

$$\sqrt{\frac{p_{i,i-1}p_{i,i+1}}{(2 - (1 + \delta_{i-1})p_{i,i-1})\,(2 - (1 + 1/\delta_i)p_{i,i+1})}}$$

if $s_{i+1} - s_i = 1$.

Let y_i denote the orthogonal system of vectors obtained from x_i, $i = 1, ..., r$, by Gram-Schmidt orthogonalization procedure. That is

$$y_i = x_i - \sum_{j=1}^{i-1} \frac{(x_i y_j)}{(y_j y_j)} y_j.$$

Taking into account that $(x_i, x_j) = 0$ if $|i - j| > 1$, one can rewrite the above expression as

$$y_i = x_i - \frac{(x_i x_{i-1})}{(y_{i-1} y_{i-1})} y_{i-1}.$$

The last formula implies that the matrix $R_S J^{-1/2}$ can be represented in the form ZO^{-1}, where Z is a $q \times 2(K-1)$ matrix such that $z_{ij} \geq 0$ for any $j \leq q$ and $z_{ij} = 0$ if $j > q$, and O^{-1} is an orthogonal matrix with q first rows equal to $y_i/|y_i|$. This implies that any vector $\tilde{p}$ with first $K-1$ non-negative elements satisfies inequality $R_S J^{-1/2} O\tilde{p} \geq 0$. Note that O is an orthogonal transformation. Thus, the asymptotic distribution of the likelihood ratio statistic for the test of $R_S J^{-1/2}\tilde{p} \geq 0$ is equal to that of the LR statistic for the test of $R_S J^{-1/2} O\tilde{p} \geq 0$. According to Lemma 4.3 the LR statistic for the initial test is dominated by the LR statistic for the test of $\tilde{p}_1 \geq 0, ..., \tilde{p}_{K-1} \geq 0$.

As was discussed above, the distribution of the likelihood ratio statistic for the test of $\tilde{p}_1 \geq 0, ..., \tilde{p}_{K-1} \geq 0$ is a mixture of the chi-squared distributions (4.2). However, finding the weights, ω_i, of the mixture can be a non-trivial problem. In general, let $\mathcal{K}$ be a polyhedral cone with the vertex at zero. Suppose that we have one observation, x, from a n-variate normal distribution $N(a, I)$. Suppose further that we would like to test a hypothesis that a lies in $\mathcal{K}$ and let the pseudo-true a be zero. Denote by $\mathcal{K}^0$ the polar cone , that is

$$\mathcal{K}^0 = \{q \in R^n : \ q'p \leq 0 \text{ for all } p \in \mathcal{K}\}.$$

Each face[9], ϕ, of the cone $\mathcal{K}$ is accompained by the polar face ϕ^* of the polar cone $\mathcal{K}^0$. Let P_ϕ, P_{ϕ^*} be symmetric idempotent matrices giving the orthogonal projections onto the linear spaces generated by ϕ and ϕ^* respectively. Denote Π_ϕ and Π_{ϕ^*} the sets $\{v : P_\phi v \in \phi\}$ and $\{v : P_{\phi^*} v \in \phi^*\}$ respectively. Shapiro (1985, p.140) shows that

$$\omega_i = \sum \Pr(x \in \Pi_\phi) \Pr(x \in \Pi_{\phi^*})$$

where the sum is over all faces ϕ with dimension $n - i$.

For the present purpose the polyhedral cone $\mathcal{K}$ corresponds to the set of constraints $\tilde{p}_1 \geq 0, ..., \tilde{p}_{K-1} \geq 0$. Thus, $\mathcal{K}$ is simply the positive orthant of the $K-1$-dimensional Euclidean space. Therefore,

$$\Pr(\tilde{p} \in \Pi_\phi) \Pr(\tilde{p} \in \Pi_{\phi^*}) = \frac{1}{2^{K-1}}.$$

Furthermore, there exist $\begin{pmatrix} i \\ K-1 \end{pmatrix}$ i-dimensional faces of the positive orthant. Hence, the distribution of the worst possible chi-bar-squared statistic in our case

[9] A face of a polyhedral cone is an intersection of a boundary hyperplane with the cone. For further terminology related to polyhedral cones see Stoer and Witzgall (1970).

is

$$\Pr(\bar{\chi}^2 \geq c^2) = \frac{1}{2^{K-1}} \sum_{i=0}^{K-1} \begin{pmatrix} K-1 \\ i \end{pmatrix} \Pr(\chi_i^2 \geq c^2) \tag{4.5}$$

Now, if $p^0 \rightarrow 0$, that is, if the pseudo-true transition matrix tends to the identity matrix, then the correlation between x_i becomes weaker and weaker, so there exist an orthogonal transformation O such that the set $\left\{\tilde{p} : R_S J^{-1/2} O \tilde{p} \geq 0\right\}$ converges to $\mathcal{R}_+^{K-1}$ uniformly in any bounded neighborhood of zero. According to Lemma 4.3 there exists a sequence of p^0 such that the asymptotic distribution of the test H_1 converges to the distribution (4.5).□

An application of the above result to Quah's (1993) analysis of countries' income distribution is performed in Kremer, Onatski, and Stock (2000). The authors divide countries into five groups: those with less than 1/4 of the world average per capita income; those between 1/4 and 1/2 of world average income in that year; those between 1/2 world average income and world average income; those between 1 and 2 times world average income, and those with income greater than twice the world average. A country's membership in a particular group is assumed to depend only on its classification the previous year. Kremer, Onatski and Stock perform the global test of monotonicity of the long-run distribution of countries' income. They consider a simplified version of hypothesis H_1 with all δ_i equal to the same number, δ. A single peak at the rich end of the income range cannot he rejected for δ as large as 1.21. This result can be interpreted as saying that the data are consistent with a hypothesis that the world is eventually becoming all-rich. This is in contrast to the popular twin peaks hypothesis saying that in the long run the world will polarize into two clubs, a club of the poor and a club of the rich.

In this paper I concentrated on the asymptotic distribution of the likelihood ratio test. Such an approximation might work poorly in finite sample. This is because the steady-state probabilities are highly non-linear functions of the transition probabilities, which are the natural parameters of the likelihood function. The triple-diagonal assumption on the transition probability matrix alleviates this problem considerably. For example, given this assumption, both monotonicity and n-modality hypotheses correspond to linear inequality restrictions on the transition probabilities. In general, however, it is desirable to perform exact finite sample tests of the shape restrictions on the steady-state distribution.

A straightforward but computationally demanding way to find exact finite sample critical value, c, of the shape restrictions test is as follows. Denote the

95% quantile of the distribution of λ for arbitrarily pseudo-true $P \in \Pi$ and $f \in S$ as $c(P, f)$. We can find $c(P, f)$ for each P and f by simulating large number of the corresponding Markov chains, computing λ for each simulation, and approximating the 95% quantile by the quantile of the empirical distribution of λ. It is then possible in principle to find the global maximum of $c(P, f)$ over the set $\Pi \times S$ numerically. Unfortunately, not only the quantile function, $c(P, f)$, can have many local maxima, but also this function is very costly to estimate with high precision at any given point. Indeed, if the cumulative distribution function of λ happens to he almost flat at $c(P, f)$, virtually infinite number of simulations is needed to estimate $c(P, f)$ precisely enough. One way to deal with this difficulty is to allow the quantile function to be estimated with errors depending on the state P, f. However, maximization of functions measured with the state dependent and possibly non-zero mean noise is not well studied. This is an interesting topic for future research.

5. Conclusion

This paper develops a framework for formal statistical analysis of the shape of the steady-state distribution of a finite Markov chain. It is shown that many reasonable shape hypotheses can be formulated in the form of simple linear inequality constraints on the steady-state probabilities. A likelihood ratio criterion is used to test the shape restrictions.

It is shown that the asymptotic distribution of the likelihood ratio statistic is the same as it would be for the test of a simple one-sided hypothesis about the mean of a particular multivariate normal distribution. This result is obtained by extending Chernoff's (1954) analysis to the case of a dependent data generating process.

Despite the fact that the distribution of the likelihood ratio statistic for the test of the one-sided hypothesis is well studied, it is difficult to find the critical value of the global likelihood ratio test. This is because the asymptotic distribution nontrivially depends on the assumed pseudo-true transition probabilities. I was able, however, to analytically find the exact asymptotic distribution of the global likelihood ratio test of the monotonicity restrictions on the steady-state probabilities. To obtain the asymptotic distribution I assume that the matrix of transition probabilities is triple-diagonal.

6. Appendix

Proof of Lemma 4.1:

Let

$$\tilde{P} = P + dp_{ij}e_ie_j' - dp_{ij}e_ie_i'$$

then one can check that

$$\tilde{\pi}' = \pi' \left(I - (\tilde{P} - P)Z\right)^{-1}$$

Using Bartlett's identity:

$$(A + ab')^{-1} = A^{-1} - \frac{A^{-1}ab'A^{-1}}{1 + b'A^{-1}a}$$

we get

$$\tilde{\pi}' = \pi' \left(I + \frac{e_ie_j'Zdp_{ij}}{1 - e_jZe_idp_{ij}}\right)$$

or

$$d\pi = \frac{Z'(e_j - e_i)e_i'dp_{ij}}{1 - (e_i - e_j)'Ze_idp_{ij}}\pi$$

dividing both sides by dp_{ij} and taking the limit when $dp_{ij} \to 0$ we get

$$\lim \frac{d\pi}{dp_{ij}} = Z'(e_j - e_i)e_i'dp_{ij}\pi$$

or, finally

$$\frac{d\pi_k}{dp_{ij}} = (z_{jk} - z_{ik})\pi_i$$

□

7. References

Adelman, I.G. 1958 "A Stochastic Analysis of the Size Distribution of Firms," *Journal of the American Statistical Association*, Vol. 53, pp. 893-904.

Andrews, D.W.K. 1998 "Hypothesis testing with a restricted parameter space," *Journal of Econometrics,* Vol. 84, pp. 155-199.

Andrews, D.W.K. 1999. "Testing when a Parameter Is on the Boundary of the maintained Hypothesis," *Cowles Foundation Discussion Paper.* No 1229.

Bartholomew, D.J. 1973 *Stochastic Models for Social Processes*, Second Edition. London: Wiley.

Billingsley, P. 1961 *Statistical Inference for Markov Processes*, The University of Chicago Press.

Champernowne, D.C. 1953 "A model of Income Distribution" *Economic Journal*, Vol. 63 pp. 318-351.

Chernoff, H. 1954. "On the distribution of the likelihood ratio," *The Annals of mathematical Statistics*, Vol. 25, No. 3, pp.573-578.

Esteban, J. and Debraj Ray. 1994. "On tile Measurement of Polarization," *Econometrica*, Vol. 62, No. 4, pp.819-851.

Guilbaud O. 1977. "Estimating the probability eigenvector and related characteristics of an ergodic transition matrix," *Scandinavian Journal of Statistics*, Vol. 4, pp. 105-112.

Kremer, M., A. Onatski, and J. Stock. 2000 "Searching for Prosperity" Harvard university, manuscript.

Kudo, A. 1963. "A multivariate analogue of the one-sided test," *Biometrika,* Vol.50, pp. 403-418.

Perlman, M.D. 1969. "One sided problems in multivariate analysis," *Annals of mathematical statistics*, Vol. 40, pp. 549-567.

Quah, Danny. 1993. "Empirical Cross-Section Dynamics in Economic Growth," *European Economic Review*, 37, pp.426-34.

Shapiro, A. 1985. "Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints," *Biometrika,* Vol. 72, No. 1, pp.133-144.

Shorrocks, A.F. 1976 "Income inobility and the Markov assumption," *Economic Journal*, Vol 86, pp. 566-578

Shorrocks, A. F. 1978. "The measurement of mobility", *Econometrica*, Vol. 46, No. 5, pp.1013-1024.

Stoer, J., Witzgall, C. 1970. *Convexity and Optimization in Finite Dimensions*, I. Berlin: Springer-Verlag.

Wilks, S.S. 1938. "The large sample distribution of the likelihood ratio for testing composite hypotheses," *Annals of Mathematical Statistics*, Vol. 9, p. 60

Wolak, F.A. 1989. "Local and global testing of linear and non-linear inequality constraints in non-linear econometric models," *Econometric Theory*, Vol. 5, pp. 1-35.

Wolak, F.A. 1991. "The local nature of hypothesis tests involving inequality constraints in non-linear models," *Econometrica,* Vol. 59, No. 4, pp. 981-996.