MIGRATIONS OF HETEROGENEOUS POPULATION OF DRIVERS ACROSS CLASSES OF A BONUS-MALUS SYSTEM

BY

WOJCIECH OTTO
Department of Economics
University of Warsaw
00-241 Warsaw, Poland
Długa Str. 44/50
e-mail: wotto@wne.uw.edu.pl

ABSTRACT

Traditionally, migrations of a single driver across classes of the Bonus-Malus system are modeled as a simple homogeneous Markov chain. Generalization to the case when the modeled object is a population of drivers is straightforward provided all drivers are characterized by the same claim frequency. In practice however, drivers differ in this respect.

The article is devoted to present implications of departure from the assumption on homogeneity of population. At first the probabilistic model of the migration process adequate to the case of heterogeneous population of drivers is introduced. Then the analysis focuses on how much properties of the migration process depart from properties of the homogeneous Markov chain. Also, some consequences of heterogeneity that concern estimators of transition probabilities from empirical data are presented.

The approach presented in the paper might be applied in various areas of empirical research, especially when the process of migrations across the space of states is observed for a large heterogeneous population over a small number of time periods.

KEYWORDS

Markov chain, estimation of transition probabilities, bonus-malus system, population heterogeneity.

1. INTRODUCTION

It is usually assumed that migrations of an individual driver across classes of a bonus-malus (BM) system satisfy assumptions of the homogeneous Marcov Chain. Generalization of the model onto the case of a population of drivers is straightforward, provided the population is homogeneous. However, in practice populations of drivers are heterogeneous. Moreover, this heterogeneity is the raison d'être of BM systems, as their aim is to make bad drivers paying higher insurance premium than good drivers do. The aim of this paper is to study implications of the departure from the assumption on homogeneity of population.

Somehow similar issues arise on the ground of the so called „Movers-Stayers Model" [5], used in panel data studies on income mobility. The model aims at explaining empirical evidence of non-Markov properties of migrations of individuals between decile groups of the income distribution. It is assumed that within each of two subpopulations migrations follow (two different) homogeneous Marcov chains. However, we observe only number of members of the whole population migrating between decile groups in subsequent years, and we are unable to indentify properly who is a „stayer" and who is a „mover". The result is that the prediction of the position of an individual in the income hierarchy in a coming year could be substantially improved by taking into account her position in the number of past years, and not only in the last year.

In motor insurance, some effects of heterogeneity of population are well known, however, explicit model of the migration process across BM classes of the population as a whole has neither been formulated, nor systematically analysed. Lack of an adequate model might result in misinterpretations, especially when analysing empirical data on migrations of a population of drivers across classes of the BM system. Misinterpretations might come from thinking in terms of properties of a homogeneous Markov chain, whereas in fact the departure from the assumption that the population is homogeneous make some of these properties heavily distorted.

Section 2 presents basic assumptions of the model of the migration process with population heterogeneity, with continuous and discrete versions of the distribution of the claim frequency within the population of drivers.

Section 3 is devoted to the analysis based on the continuous version. This version is especially convenient to analyse long-run properties of the process, and the essence of stationarity as well as some characteristics of the rate of convergence to the stationary state.

Section 4 presents the analysis based on the discrete version. This is especially convenient when we focus on short-term fluctuations of fractions of drivers staying in BM classes. Also, the approach is more adequate when we focus on properties that depend on the number of drivers in the population. This is especially relevant when transition probabilities are to be estimated from empirical data, and we have to rely on properties of estimators that are asymptotic in respect of the number of drivers in the population, as the number of years of observations is small or very small.

Finally, section 5 presents summary of main conclusions.

2. BASIC ASSUMPTIONS OF THE MIGRATION'S MODEL

Basic assumptions concern an individual driver, a population of drivers, and a BM system.

About an individual driver we assume:

- A driver characterised by the value λ of the risk parameter Λ produces losses according to the Poisson process with intensity λ per year
- A driver claims to the insurer all losses that happen to him/her ($claims \equiv losses$)
- Appearance of a driver in a coming year in a given class depends exclusively on two factors: number of losses claimed during last year and class where he/she has been then

About the population of drivers we assume:

- Population is closed,
- Migration processes of individual drivers are independent.

Moreover, in a discrete variant of the model we assume that:

- Population consists of N drivers characterised by values $\lambda_1, \lambda_2, ..., \lambda_N$ of risk parameter, respectively.

The continuous variant assumes approximation of the discrete distribution $\lambda_1, \lambda_2, ..., \lambda_N$ by:

- a continuous density $f(\lambda)$.

In both variants it is assumed that the distribution is known whereas parameter λ for a particular driver is not.

About the BM system we assume:

- It consists of S classes with transition probabilities, which for a driver characterised by the value λ of parameter Λ form a matrix:
$$\mathbf{P}(\lambda) = \{p_{i,j}(\lambda)\}_{i,j=1,2,...,S},$$
- An insurer sets the BM system by positioning zeroes and non-zero elements of $\mathbf{P}(\lambda)$ corresponding to events of zero claims, 1 claim, 2 claims, 3 claims, etc.,
- A driver's response is his/her intensity parameter λ, and values of non-zero elements $\mathbf{P}(\lambda)$ that are functions of this parameter,
- Construction of the BM system ensures that $\mathbf{P}(\lambda)$ is ergodic for all $\lambda > 0$.

3. MODEL WITH CONTINUOUS DISTRIBUTION OF THE RISK PARAMETER Λ

Distribution of drivers by parameter $\Lambda > 0$ in the whole population is given by density $f(\lambda)$. Joint distribution by classes of the BM system and by Λ is given by the vector of densities:
$$\mathbf{f}^{(t)}(\lambda) := \left[f_1^{(t)}(\lambda) \quad f_2^{(t)}(\lambda) \quad ... \quad f_S^{(t)}(\lambda) \right], \qquad t = 1,2,3,....$$
Due to well known properties of the stochastic matrix the following recursive formula holds:
$$\underset{\lambda>0}{\forall} \mathbf{f}^{(t+1)}(\lambda) = \mathbf{f}^{(t)}(\lambda)\mathbf{P}(\lambda), \qquad t = 1,2,3,...,$$
as well as the following implication is true:
$$\underset{\lambda>0}{\forall} \sum_{i=1}^{S} f_i^{(1)}(\lambda) = f(\lambda) \Rightarrow \underset{t=2,3,...}{\forall} \underset{\lambda>0}{\forall} \sum_{i=1}^{S} f_i^{(t)}(\lambda) = f(\lambda).$$
As a result, taking $f_1^{(1)}(\lambda),...,f_S^{(1)}(\lambda)$ such that their sum equals $f(\lambda)$ ensures that:
$$\sum_{i=1}^{S} \int_0^\infty f_i^{(t)}(\lambda) d\lambda = 1 \qquad t = 1,2,3,...$$
Thus the coherent definition of the probability $f_i^{(t)}$ that a driver stays in class *i* in year *t* reads:

$$f_i^{(t)} := \int_0^\infty f_i^{(t)}(\lambda)d\lambda, \quad i = 1,2,...,S.$$

Due to ergodicity of $\mathbf{P}(\lambda)$ for all $\lambda > 0$ vectors $\mathbf{f}^{(t)}(\lambda)$ converge for $t \to \infty$ to the limit:

$$\mathbf{f}(\lambda) = [f_1(\lambda) \quad f_2(\lambda) \quad ... \quad f_S(\lambda)],$$

which integrated over the interval $(0, \infty)$ renders the limiting vector of probabilities:

$$\mathbf{f} = [f_1 \quad f_2 \quad ... \quad f_S].$$

Convergence takes place for any admissible starting density vector $\mathbf{f}^{(1)}(\lambda)$. For example, an admissible starting vector could consist of elements that equal:

$$f_i^{(1)}(\lambda) := w_i f(\lambda), \quad i = 1,2,...,S,$$

such that weights w_i are non-negative and sum up to one. This represents the case when at the start of the system fractions of drivers allocated to classes are predefined, but allocation of individual drivers to classes is random (does not depend on risk parameter values).

3.1. Implications: variable transition probabilities, and the essence of stationarity

By definition, $p_{i,j}(\lambda)$ is a conditional probability of staying next year in class *j* under the condition, that this year a driver stay in class *i* and that her risk parameter Λ equals λ. In order to make this probability concerning a driver randomly drawn form the population, we have to determine the year in question. Once this is done, the probability of transition from class *i* to in year *t* to class *j* in the next year (unconditional in respect of Λ) equals:

$$p_{i,j}^{(t)} = \frac{1}{f_i^{(t)}} \int_0^\infty p_{i,j}(\lambda) f_i^{(t)}(\lambda) d(\lambda),$$

and varies as long as $f_i^{(t)}(\lambda)$ changes with $t = 1,2,...$. However, for $t \to \infty$ densities $f_i^{(t)}(\lambda)$ converge to $f_i(\lambda)$, so that transition probabilities $p_{i,j}^{(t)}$ converge to their limits $p_{i,j}$ as well. Using matrix notation $\mathbf{P}^{(t)} := \{p_{i,j}^{(t)}\}_{i,j=1,...,S}$ this can be written as:

$$\mathbf{P}^{(t)} \to \mathbf{P}.$$

Of course, limits $\mathbf{P}$ and $\mathbf{f}$ of the sequences $\mathbf{P}^{(t)}$ and $\mathbf{f}^{(t)}$ satisfy a system of equations:

$$\mathbf{f} = \mathbf{fP}.$$

However, the similarity to the case of a standard Marcov chain is misleading. The essence of stationarity now is that the joint distribution (in respect of classes and Λ) does not change. As for a counterexample we may consider allocation of drivers in the first year such that:

- $\int_0^\infty f_i^{(1)}(\lambda)d\lambda = \int_0^\infty f_i(\lambda)d\lambda$ for all $i = 1,2,...,S$, but:
- there exists a measurable set $A \subset (0, \infty)$ such that $\int_A f_i^{(1)}(\lambda)d\lambda \neq \int_A f_i(\lambda)d\lambda$ for some $i \in \{1,2,...,S\}$.

Despite marginal distribution in respect of classes at $t = 1$ equals its own limit, transition probabilities $p_{i,j}^{(t)}$ vary for $t = 2,3,...$ until stabilisation of density vector $\mathbf{f}^{(t)}$ is attained.

3.2. Implications: second largest eigenvalue looses its interpretation

Under the case of homogenous population the second largest (in terms of absolute value) eigenvalue of the transition probability matrix characterises the rate of convergence of the distribution on the space of classes to the stationary distribution. This is no more true in the case of heterogeneous population. In order to show that, let us consider an example.

Example 1. Consider a simple BM system called "no claim discount", where each driver after a year with claims goes to class 1, and after no-claim year falls one class down (provided there are still classes down the current one). Assume $S=3$. Thus a driver with the value λ of risk parameter Λ follows the Marcov Chain characterised by the transition probability matrix:

$$\mathbf{P}(\lambda)=\begin{bmatrix} q(\lambda) & p(\lambda) & 0 \\ q(\lambda) & 0 & p(\lambda) \\ q(\lambda) & 0 & p(\lambda) \end{bmatrix}, \qquad \text{where } p(\lambda)=\exp(-\lambda), \text{ and } q(\lambda)=1-p(\lambda).$$

For any $\lambda>0$ a limiting distribution on the space of classes equals:

$$\begin{bmatrix} q(\lambda) & q(\lambda)p(\lambda) & p^2(\lambda) \end{bmatrix}.$$

For any $\lambda>0$ matrix $\mathbf{P}(\lambda)$ has eigenvalues $\{\rho_1, \rho_2, \rho_3\} \approx \{1, 0, 0\}$, which is in concordance with the fact that drivers attain their limiting distributions in two steps (at $t=3$). Let us assume now that the parameter Λ in the population of drivers is exponentially distributed with expectation equal 1/5. Then the following results can be easily obtained:

- probability of zero-claims in a year for a representative driver equals $\overline{p}=5/6$,
- joint distribution $\mathbf{f}^{(t)}(\lambda)$ for $t=3,4,...$ equals the stationary distribution $\mathbf{f}(\lambda)$:
  $$\mathbf{f}(\lambda)=\begin{bmatrix} (1-e^{-\lambda})5e^{-5\lambda} & e^{-\lambda}(1-e^{-\lambda})5e^{-5\lambda} & e^{-2\lambda}5e^{-5\lambda} \end{bmatrix},$$
- stationary marginal distribution in respect of classes equals:
  $$\mathbf{f}=\begin{bmatrix} (5/6) & (5/6-5/7) & (5/7) \end{bmatrix}.$$
- matrix $\mathbf{P}$ characterising transitions from year $t \in \{3,4,5,...\}$ to the next year equals:
  $$\mathbf{P}=\begin{bmatrix} 2/7 & 5/7 & 0 \\ 1/4 & 0 & 3/4 \\ 1/8 & 0 & 7/8 \end{bmatrix},$$
  with eigenvalues $\{\rho_1, \rho_2, \rho_3\} \approx \{1, 0.390, -0.229\}$.

Under the standard Marcov chain such a result means that convergence takes place in an infinite number of steps, with the limiting rate of convergence equal 0.39 (see [2]). In our case it is misleading indeed, as the population attains stationary state as early as for $t=3$.

4. MODEL WITH DISCRETE DISTRIBUTION OF THE RISK PARAMETER Λ

Our further analysis is focused on characteristics of the counting vector:

- $\mathbf{N}^{(t)} := \begin{bmatrix} N_1^{(t)} & N_2^{(t)} & ... & N_s^{(t)} \end{bmatrix}$, such that:
- $N_i^{(t)} := N_{i,1}^{(t)} + N_{i,2}^{(t)} + ... + N_{i,S}^{(t)}$ is a number of drivers staying in class *i* in year *t*,
- $N_{i,j}^{(t)}$ is a number of drivers migrating from class *i* in year *t* to class *j* in the next year.

As the population size N is a finite number, it is now convenient to switch to the discrete model. Driver number k with parameter λ_k is now characterised by:

- Transition probability matrix $\mathbf{P}_k := \mathbf{P}(\lambda_k)$ with elements $p_{i,j}(k)$
- Sequence $\{\mathbf{h}_k^{(t)}\}_{t=1}^{\infty}$ of vectors $\mathbf{h}_k^{(t)} := [h_{1,k}^{(t)} \quad h_{2,k}^{(t)} \quad \ldots \quad h_{S,k}^{(t)}]$ of probabilities of staying in year t in classes $i = 1,2,\ldots,S$.

Assuming now a starting vector $\mathbf{h}_k^{(1)}$ we obtain recursively next terms of the sequence:

$$\mathbf{h}_k^{(t)} = \mathbf{h}_k^{(1)}\mathbf{P}_k^{t-1}, \; t = 2,3,\ldots,$$

which under reasonably designed BM system converges to the limit $\mathbf{h}_k$ for $k = 1,2,\ldots,N$.

Let us consider now random drawing (with equal probabilities) of a driver from the population. Migration process of such a driver are characterised by:

- probability vectors $\mathbf{h}^{(t)} = [h_1^{(t)} \quad h_2^{(t)} \quad \ldots \quad h_S^{(t)}]$ such that $\mathbf{h}^{(t)} \to \mathbf{h}$, where:

  $$\mathbf{h}^{(t)} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{h}_k^{(t)}, \text{ and } \mathbf{h} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{h}_k,$$

- transition probability matrices $\mathbf{P}^{(t)} = \{p_{i,j}^{(t)}\}_{i,j=1,\ldots,S}$ such that $\mathbf{P}^{(t)} \to \mathbf{P}$, where:

  $$p_{i,j}^{(t)} = \frac{1}{Nh_i^{(t)}}\sum_{k=1}^{N} p_{i,j}(k)h_{i,k}^{(t)}, \text{ and } p_{i,j} = \frac{1}{Nh_i}\sum_{k=1}^{N} p_{i,j}(k)h_{i,k}.$$

As well as in the case of the continuous model matrix $\mathbf{P}$ and vector $\mathbf{h}$ are linked by the system of equations $\mathbf{h} = \mathbf{hP}$.

Further on we focus on the analysis of the unconditional distribution of the vector $\mathbf{N}^{(t)}$ and conditional distribution of this vector given $\mathbf{N}^{(t-1)}$. Terms “unconditional” and “conditional” are rather conventional, as $\mathbf{N}^{(t)}$ depends also on starting conditions. The simplification is justified by focussing on the case of large t, when that dependence disappears.

4.1. Characteristics of the unconditional distribution of $\mathbf{N}^{(t)}$

Expected value of the counting vector $\mathbf{N}^{(t)}$ equals:

$$\mathrm{E}(\mathbf{N}^{(t)}) = N\mathbf{h}^{(t)}.$$

Due to independence assumption the covariance matrix is just a simple sum:

$$\mathrm{cov}(\mathbf{N}^{(t)},\mathbf{N}^{(t)}) = \sum_{k=1}^{N}\begin{bmatrix} h_{1,k}^{(t)}(1-h_{1,k}^{(t)}) & -h_{1,k}^{(t)}h_{2,k}^{(t)} & \ldots & -h_{1,k}^{(t)}h_{s,k}^{(t)} \\ -h_{2,k}^{(t)}h_{1,k}^{(t)} & h_{2,k}^{(t)}(1-h_{2,k}^{(t)}) & \ldots & -h_{2,k}^{(t)}h_{s,k}^{(t)} \\ \ldots & \ldots & \ldots & \ldots \\ -h_{s,k}^{(t)}h_{1,k}^{(t)} & -h_{s,k}^{(t)}h_{2,k}^{(t)} & \ldots & h_{s,k}^{(t)}(1-h_{s,k}^{(t)}) \end{bmatrix}.$$

The above matrix can be decomposed into a difference of two matrices:

$$\mathrm{cov}(\mathbf{N}^{(t)},\mathbf{N}^{(t)}) = N\{diag(\mathbf{h}^{(t)}) - (\mathbf{h}^{(t)})'\mathbf{h}^{(t)}\} - \sum_{k=1}^{N}(\mathbf{h}_k^{(t)} - \mathbf{h}^{(t)})'(\mathbf{h}_k^{(t)} - \mathbf{h}^{(t)}),$$

Where:

- First matrix appearing on the RHS is a covariance matrix adequate for the case of homogeneous population where $\mathbf{h}_1^{(t)} = \mathbf{h}_2^{(t)} = ... = \mathbf{h}_N^{(t)} = \mathbf{h}^{(t)}$.
- Second matrix measures heterogeneity within the set of vectors $\mathbf{h}_1^{(t)}, \mathbf{h}_2^{(t)}, ..., \mathbf{h}_N^{(t)}$ that comes from heterogeneity within the set $\{\lambda_1, \lambda_2, ..., \lambda_N\}$. It is in fact a *N*-multiple of the covariance matrix of the random vector $\mathbf{h}_K^{(t)}$ being a function of random variable *K* that takes values from the set $\{1,2,...,N\}$ with the same probability 1/*N*.

Taking the following notations for the case of $t \to \infty$:

- $\mathbf{V_N} := \lim_{t\to\infty} \mathrm{cov}(\mathbf{N}^{(t)}, \mathbf{N}^{(t)})$,
- $\mathbf{V_h} := \frac{1}{N}\sum_{k=1}^{N}(\mathbf{h}_k - \mathbf{h})'(\mathbf{h}_k - \mathbf{h})$,

we can present the above decomposition in matrix notation:

$$\mathbf{V_N} = N\{diag(\mathbf{h}) - \mathbf{h}'\mathbf{h} - \mathbf{V_h}\},$$

where the last term on the RHS stands for the effect of heterogeneity of population of drivers.

4.2. Characteristics of the conditional distribution of $\mathbf{N}^{(t)}$ given $\mathbf{N}^{(t-1)}$

The dynamics of the vector $\mathbf{N}^{(t)}$ can be represented by the system of equations:

$$\mathbf{N}^{(t)} = \mathbf{N}^{(t-1)}\mathbf{P} + \boldsymbol{\varepsilon}^{(t)}.$$

In the case of homogeneous population the first term on the RHS is just the conditional expectation $\mathrm{E}(\mathbf{N}^{(t)} | \mathbf{N}^{(t-1)})$. Therefore the random term satisfy the condition $\mathrm{E}(\boldsymbol{\varepsilon}^{(t)} | N^{(t-1)}) = 0$, and only the conditional covariance matrix $\mathrm{cov}(\boldsymbol{\varepsilon}^{(t)}, \boldsymbol{\varepsilon}^{(t)} | N^{(t-1)})$ depends on $\mathbf{N}^{(t-1)}$. This allows to interpret the system of equations as linear regression model. In particular, the estimation of transition probabilities from aggregate time series data is based on this interpretation (see [3]).

In the case of heterogeneous population matrix $\mathbf{P}^{(t)}$ differs from the limit **P**. However, we can still pose the question, whether the dynamics of the vector $\mathbf{N}^{(t)}$ can be described analogously as in the case of homogeneous population, at least for large *t*. The answer is negative, as it turns out that $\mathrm{E}(\mathbf{N}^{(t)} | \mathbf{N}^{(t-1)})$ is a non-linear function of the vector $\mathbf{N}^{(t-1)}$. This can be easily shown in the simple case when the BM system assumes that the only way to enter the class *j* is to be last year in class *i*. We can focus then on the conditional expectation $\mathrm{E}(N_j^{(t)} | N_i^{(t-1)})$, and consider (for simplicity) the extreme case when $N_i^{(t-1)} = N$ and the almost-extreme opposite case of $N_i^{(t-1)} = 1$. Direct calculations render the following results:

$\mathrm{E}(N_j^{(t)} | N_i^{(t-1)} = N) = \sum_{k=1}^{N} p_{i,j}(k)$, which generally differs from $Np_{i,j}^{(t-1)}$, and:

$\mathrm{E}(N_j^{(t)} | N_i^{(t-1)} = 1) = \dfrac{\sum_{k=1}^{N} p_{i,j}(k) h_{i,k}^{(t-1)} \left(1 - h_{i,k}^{(t-1)}\right)^{-1}}{\sum_{k=1}^{N} h_{i,k}^{(t-1)} \left(1 - h_{i,k}^{(t-1)}\right)^{-1}}$, which differs from $p_{i,j}^{(t-1)}$.

Thus in both cases (provided population is heterogeneous) $\mathrm{E}(N_j^{(t)} | N_i^{(t-1)}) \neq N_i^{(t-1)} p_{i,j}^{(t-1)}$.

Non-linearity of the function $\mathrm{E}(\mathbf{N}^{(t)}|\mathbf{N}^{(t-1)})$ implies that $\lim_{t\to\infty}\mathrm{E}(\boldsymbol{\varepsilon}^{(t)}|\mathbf{N}^{(t-1)}) \neq 0$. Only the weaker condition $\lim_{t\to\infty}\mathrm{E}(\boldsymbol{\varepsilon}^{(t)}) = 0$ is satisfied, but it is insufficient to interpret the analysed set of equations in terms of a more or less standard linear regression model. To study the problem in greater details we can decompose the vector of random terms into two components:

$$\boldsymbol{\varepsilon}^{(t)} = \left\{\mathbf{N}^{(t)} - \mathrm{E}(\mathbf{N}^{(t)}|\mathbf{N}^{(t-1)})\right\} + \left\{\mathrm{E}(\mathbf{N}^{(t)}|\mathbf{N}^{(t-1)}) - \mathbf{N}^{(t-1)}\mathbf{P}\right\},$$

Where the first component represents fluctuations around the conditional expected value, whereas the second one represents linearization error. It turns out that the share of covariance matrix of the second component in the total covariance matrix of the vector $\boldsymbol{\varepsilon}^{(t)}$ does not disappear even when both t and N tend to infinity. This conclusion is illustrated in more details on an example in the Appendix.

4.3. Estimating transition probabilities when data on transitions are available

When number of transitions $N_{i,j}^{(t)}$ as well as their sums $N_i^{(t)}$ are observed, then transition probabilities could be estimated on the basis of observations from a single year. A natural estimator of the transition probability $p_{i,j}^{(t)}$ can be defined then as:

$$\hat{p}_{i,j}^{(t)} := \frac{N_{i,j}^{(t)}}{N_i^{(t)}}, \text{ with some predefined constant for the case when } N_i^{(t)} = 0.$$

Using formulas for moments of the quotient of sample means (well known on the ground of sampling theory) we obtain approximation of its expectation:

$$\mathrm{E}(\hat{p}_{i,j}^{(t)}) \approx p_{i,j}^{(t)} + \frac{1}{\left(Nh_i^{(t)}\right)^2}\sum_{k=1}^{N}\left(h_{i,k}^{(t)}\right)^2\left(p_{i,j}(k) - p_{i,j}^{(t)}\right),$$

with bias proportional to $1/N$. There is no bias when the population is homogeneous, as in this case the second component disappears. The approximated variance is given by the formula:

$$\mathrm{var}(\hat{p}_{i,j}^{(t)}) \approx \frac{1}{Nh_i^{(t)}}p_{i,j}^{(t)}(1 - p_{i,j}^{(t)}) - \frac{1}{N^2\left(h_i^{(t)}\right)^2}\sum_{k=1}^{N}\left(h_{i,k}^{(t)}\right)^2\left(p_{i,j}(k) - p_{i,j}^{(t)}\right)^2.$$

The first component is the variance of the classical estimator of the conditional probability of some event when the condition is satisfied for an individual observation with probability $h_i^{(t)}$. The second component (negative) represents the efficiency gain due to heterogeneity of population. Both components are of order $1/N$. Thus the efficiency gain does not disappear with increasing population size N. The same remark could be made for covariances. For the case of transitions from two different classes the formula reads:

$$\mathrm{cov}(\hat{p}_{i,j}^{(t)}, \hat{p}_{m,n}^{(t)}) \approx -\frac{1}{N^2h_i^{(t)}h_m^{(t)}}\sum_{k=1}^{N}h_{i,k}^{(t)}h_{m,k}^{(t)}\left(p_{i,j}(k) - p_{i,j}^{(t)}\right)\left(p_{m,n}(k) - p_{m,n}^{(t)}\right), \qquad i \neq m.$$

In this case covariance reduces to zero when population is homogeneous. It is not the case for the covariance of estimators of probability of transition from the same class to two different classes. In this case the formula reads:

$$\mathrm{cov}(\hat{p}_{i,j}^{(t)}, \hat{p}_{i,n}^{(t)}) \approx -\frac{1}{Nh_i^{(t)}}p_{i,j}^{(t)}p_{i,n}^{(t)} - \frac{1}{\left(Nh_i^{(t)}\right)^2}\sum_{k=1}^{N}\left(h_{i,k}^{(t)}\right)^2\left(p_{i,j}(k) - p_{i,j}^{(t)}\right)\left(p_{i,n}(k) - p_{i,n}^{(t)}\right),$$

$$j \neq n.$$

The first term comes from the competing character of transitions from the same class to two different classes. This is the second term only that represents the effect of heterogeneity of population.

In all the above formulae errors of approximation are of order N^{-2}. An illustrative example concerning this error and the scope of the efficiency gain of estimators resulting from heterogeneity of population is presented in the Appendix.

5. CONCLUSIONS

Major implications of the heterogeneity of the population that make the migration process different from the simple generalisation of the Marcov Chain are the following:

- Transition probabilities vary in time for small t (*nothing really new*)
- Different properties of the process $\mathbf{N}^{(t)}$ even for large t:
    - Probability of loss depending on class where a driver stays (*known, although often interpreted as being due to other factors than heterogeneity*),
    - Loss of interpretation of the second largest eigenvalue of probability transition matrix **P** (*new result*),
    - Non-linearity of function $\mathrm{E}(\mathbf{N}^{(t)} \mid \mathbf{N}^{(t-1)})$ (*new facts and interpretations*),
    - Efficiency gain of estimators of transition probabilities (*new result*)

Lack of awareness of these effects may lead to misinterpretations of various anomalies that might be observed in empirical research. These anomalies are quite common and may come from the so-called hunger for bonus phenomenon as well as from various factors that change parameters of the migration process in subsequent calendar years (see [4], [6]).

6. APPENDIX

6.1. Illustration of issues presented in subsection 4.2.

We consider the same no claim discount system with 3 classes as in section 3. However, we assume now that the population size N is an even number with one half of population consisting of good drivers and the other half of bad ones. Respective transition probability matrices are:

$$\mathbf{P}_G = \begin{bmatrix} q_G & p_G & 0 \\ q_G & 0 & p_G \\ q_G & 0 & p_G \end{bmatrix}, \quad \text{and} \quad \mathbf{P}_B = \begin{bmatrix} q_B & p_B & 0 \\ q_B & 0 & p_B \\ q_B & 0 & p_B \end{bmatrix},$$

where:

- Good drivers are characterised by $p_G = 0.9$, and $q_G = 0.1$,
- Bad drivers are characterised by $p_B = 0.6$, $q_B = 0.4$.

Thus limiting probability vectors (attained as early as for $t = 3$) equal:

$$\mathbf{h}_G = \begin{bmatrix} q_G & q_G p_G & p_G^2 \end{bmatrix} = \begin{bmatrix} 0.10 & 0.09 & 0.81 \end{bmatrix},$$
$$\mathbf{h}_B = \begin{bmatrix} q_B & q_B p_B & p_B^2 \end{bmatrix} = \begin{bmatrix} 0.40 & 0.24 & 0.36 \end{bmatrix}.$$

A randomly drawn driver from that population is characterised by the limiting vector:

$$\mathbf{h} = (\mathbf{h}_G + \mathbf{h}_B)/2 = [0.25 \quad 0.165 \quad 0.585],$$

And by the limiting transition probability matrix:

$$\mathbf{P} = \begin{bmatrix} q_{(1)} & p_{(1)} & 0 \\ q_{(2)} & 0 & p_{(2)} \\ q_{(3)} & 0 & p_{(3)} \end{bmatrix} \approx \begin{bmatrix} 0.34 & 0.66 & 0 \\ 0.3182 & 0 & 0.6818 \\ 0.1923 & 0 & 0.8077 \end{bmatrix}.$$

Surely, the average no-claim probability $\bar{p}$ and class-specific limiting probabilities $p_{(1)}$, $p_{(2)}$, and $p_{(3)}$ satisfy the equation:

$$p_{(1)}h_1 + p_{(2)}h_2 + p_{(3)}h_3 = \bar{p} = 0.75 .$$

We focus now on the conditional expectation $\mathrm{E}(N_2^{(t+1)} \big| N_1^{(t)})$, selected from the vector $\mathrm{E}(\mathbf{N}^{(t+1)} \big| \mathbf{N}^{(t)})$ just for simplicity, as a direct transition to class 2 is possible only from class 1. Let us now define a function g: $\{1,2,...,N\} \to (0,1)$ such that:

- $mg(m) = \mathrm{E}(N_2^{(t+1)} \big| N_1^{(t)} = m)$ for $m = 1,2,...,N$

Of course, in the case of homogeneous population $g(m) \equiv \bar{p}$. In our case, however, the function is related to the class-specific probability of no-claim $p_{(1)}$ by the equation:

$$p_{(1)} = \frac{\lim\limits_{t\to\infty} \sum_{k=1}^{N} g(k) k \Pr(N_1^{(t)} = k)}{\lim\limits_{t\to\infty} \sum_{k=1}^{N} k \Pr(N_1^{(t)} = k)},$$

Where the RHS is just a ratio of expected values $\mathrm{E}(N_2^{(t+1)})$ and $\mathrm{E}(N_1^{(t)})$. At the same time the above formula clearly reveal the fact that $p_{(1)}$ is just a weighted average of values of $g(m)$. These values for $m = 1,2,3,...,N$ have to be calculated from a general formula, which takes into account the expected structure of the class 1 in year t, provided there are m drivers in this class at the moment. The formula reads:

$$g(m) = \frac{\sum_{j=\max\{0,m-N/2\}}^{\min\{m,N/2\}} [jp_A + (m-j)p_B] h_{1,A}^j h_{1,B}^{m-j} (1-h_{1,A})^{N/2-j} (1-h_{1,B})^{N/2-m+j}}{m \sum_{j=\max\{0,m-N/2\}}^{\min\{m,N/2\}} h_{1,A}^j h_{1,B}^{m-j} (1-h_{1,A})^{N/2-j} (1-h_{1,B})^{N/2-m+j}}, \qquad m = 1,...,N .$$

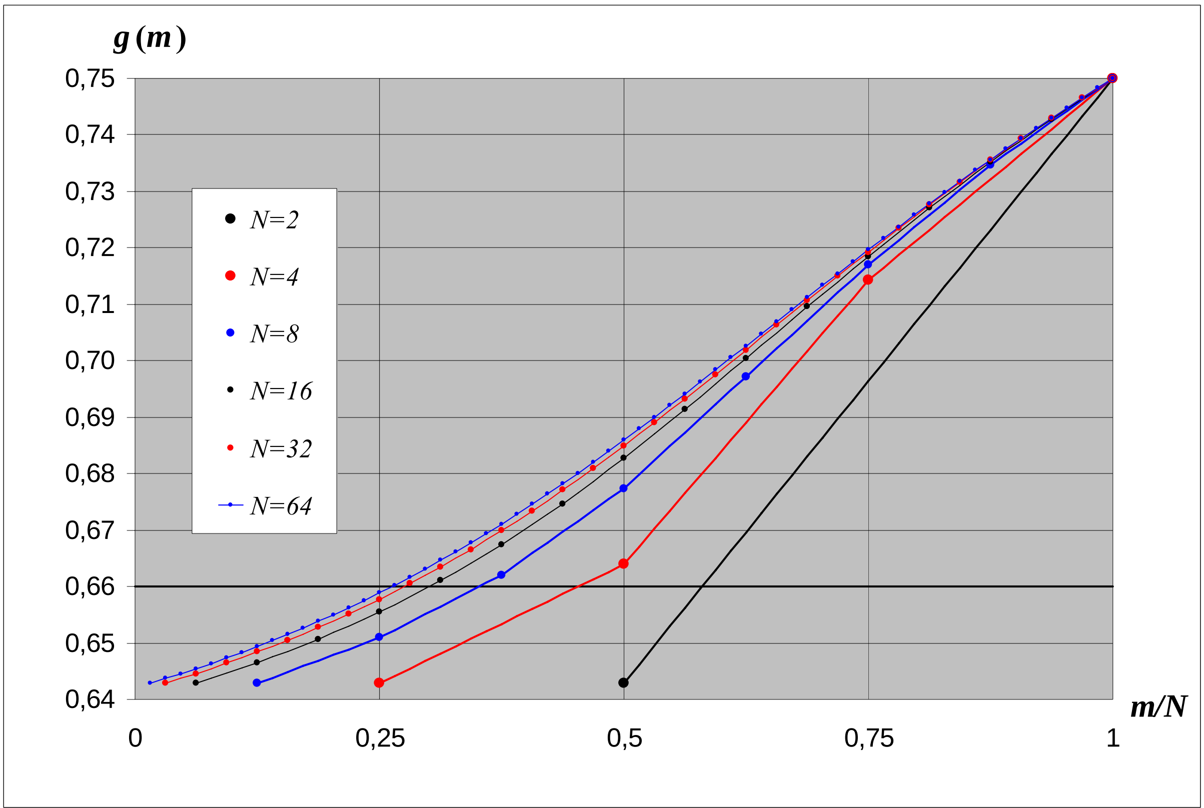
Calculations, cumbersome in general, simplify in the extreme cases to the form:

$$g(N) = \bar{p} = 0.75 ,$$

$$g(1) = \frac{p_G h_{1,G}(1-h_{1,G})^{-1} + p_B h_{1,B}(1-h_{1,B})^{-1}}{h_{1,G}(1-h_{1,G})^{-1} + h_{1,B}(1-h_{1,B})^{-1}} = \frac{9}{14} \approx 0.643 .$$

The general formula has been used to calculate all medium values of the function g for populations of size $N = 2, 4, 8, 16, 64$. Results are depicted on graph 1.

Graph 1. Function $g(m)$ for selected population sizes *N*.



Apart of the obvious conclusion that function $g(m)$ depends significantly on *m*, other conclusions that could be drawn on the basis of these results are the following:

- g increases monotonically between the extremes $g(1) \approx 0.643$ and $g(N) = 0.75$,
- for $m = \mathrm{E}(N_1^{(t)})$ that equals one quarter of the population the value $g(m)$ seems to tend (as the population size N increases) to $p_{(1)} = 0.66$
- around the point $m = \mathrm{E}(N_1^{(t)})$ the slope of the function $g(m)$ seems to tend also to some limiting value as $N \to \infty$.

Further insight into the issue of the difference between nonlinear function $\mathrm{E}(N_2^{(t+1)} \big| N_1^{(t)})$ and its linear approximation $N_1^{(t)} p_{(1)}$ can be gained by the following decomposition of the total variance $\mathrm{var}(N_2^{(t+1)})$:

- $\mathrm{var}(N_1^{(t)} p_{(1)})$ - the regression component of variance based on linear approximation
- $\mathrm{var}\{\mathrm{E}(N_2^{(t+1)} \big| N_1^{(t)})\} - \mathrm{var}(N_1^{(t)} p_{(1)})$ - effect of linearization error
- $\mathrm{E}\{\mathrm{var}(N_2^{(t+1)} \big| N_1^{(t)})\}$ - effect of fluctuations around the true conditional expectation

Results of calculations (made on the same basis as results depicted on Graph 1) are summarised in Table 1 below. Presented results confirm that the relative share of the linearization error in the total variance decreases at first, but tends to stabilise at the level (probably) above 3.6% for large *N*.

Table 1. Variance of the random variable $N_2^{(t+1)}$ and its components

No.	N	2	4	8	16	32	64
(1)	$\mathrm{var}(N_2^{(t+1)})$	0.264	0.527	1.057	2.114	4.229	8.458
(2)	$\mathrm{var}(N_1^{(t)} p_{(1)})$	0.144	0.288	0.575	1.150	2.300	4.600
(3)	$\mathrm{var}\{\mathrm{E}(N_2^{(t+1)} \mid N_1^{(t)})\}$	0.155	0.308	0.614	1.228	2.455	4.910
(4)	(3) – (2)	0.011	0.020	0.039	0.078	0.155	0.310
(5)	(2)/(1)	54.39%	54.39%	54.39%	54.39%	54.39%	54.39%
(6)	(4)/(1)	4.133%	3.832%	3.727%	3.688%	3.671%	3.663%
(7)	(3)/(1)	41.48%	41.78%	41.88%	41.92%	41.94%	41.95%

6.2. Illustration of issues presented in subsection 4.3.

We consider the same no claim discount system with 3 classes as in section 2. For simplicity, we drop the superscript (t) from our notations, as all data and estimators concern the year t and transitions from classes in year t to classes in the next year. At first, we define the ML estimator of the no-claim probability average in the population $\bar{p}$:

$$\hat{\bar{p}} := \frac{N_{1,2} + N_{2,3} + N_{3,3}}{N}.$$

The estimator is not interesting by itself, but we make use of it for defining coherently estimators of class-specific no-claim probabilities $p_{(1)}$, $p_{(2)}$, and $p_{(3)}$:

$$\hat{p}_{(1)} := \frac{N_{1,2}}{N_1} \text{ when } N_1 > 0, \quad \text{and} \quad \hat{p}_{(1)} := \hat{\bar{p}} \text{ when } N_1 = 0$$

$$\hat{p}_{(2)} := \frac{N_{2,3}}{N_2} \text{ when } N_2 > 0, \quad \text{and} \quad \hat{p}_{(2)} := \hat{\bar{p}} \text{ when } N_2 = 0$$

$$\hat{p}_{(3)} := \frac{N_{3,3}}{N_3} \text{ when } N_3 > 0, \quad \text{and} \quad \hat{p}_{(3)} := \hat{\bar{p}} \text{ when } N_3 = 0$$

Let us now take the same assumptions as in subsection 6.1. on population consisting of good and bad drivers. Using approximated formulas for expectations, variances, and covariances of estimators presented in section 3.3 we obtain the following results:

$$\left[\mathrm{E}(\hat{p}_{(1)} - p_{(1)}) \quad \mathrm{E}(\hat{p}_{(2)} - p_{(2)}) \quad \mathrm{E}(\hat{p}_{(3)} - p_{(3)})\right] \approx \frac{1}{N}\left[-0.0576 \quad -0.0541 \quad 0.0492\right],$$

$$\mathrm{cov}(\hat{p}_{(1)}, \hat{p}_{(2)}, \hat{p}_{(3)}) \approx \frac{1}{N}\begin{bmatrix} 0.8976 & 0 & 0 \\ 0 & 1.3148 & 0 \\ 0 & 0 & 0.2655 \end{bmatrix} - \frac{1}{N}\begin{bmatrix} 0.0092 & 0.0114 & 0.0123 \\ 0.0114 & 0.0142 & 0.0152 \\ 0.0123 & 0.0152 & 0.0163 \end{bmatrix}.$$

The first result is a pure effect of heterogeneity of population, as in case of homogeneous population the bias disappears. The efficiency gain appearing in the second result is moderate: variances of estimators $\hat{p}_{(1)}$ and $\hat{p}_{(2)}$ are reduced by about 1%, whereas variance of the estimator $\hat{p}_{(3)}$ is reduced by about 6%.

As it has been pointed out in section 4.3, formulas used to obtain these results are only approximations. In fact, during their derivation the power series expansion is used, and only terms of order 1 and of order N^{-1} are accounted for, while terms of order N^{-2}, N^{-3}, etc. are dropped. Also, the derivation neglects the term due to the event of empty class from which transitions take place, when the respective estimator $\hat{p}_{(i)}$ takes a value equal to $\hat{\bar{p}}$. Below, quality of these approximations is illustrated on the example of the bias of the estimator $\hat{p}_{(1)}$. It is because the exact expectation of this estimator can be assessed on the basis of calculations that have been done anyway for the purpose of analysis of the function $g(m)$, presented in subsection 6.1. The results are summarised in Table 2.

Table 2. Bias of the estimator $\hat{p}_{(1)}$ for small *N*, and its components.

Notations:

b_N - exact bias,

$\tilde{b}_N$ - bias as it is given by the approximate formula,

b_N^0 - component of b_N due to the event $N_1 = 0$,

b_N^1 - component of b_N due to the event $N_1 > 0$,

p^0 - probability of the event $N_1 = 0$

N	2	4	6	8	10	16	24	32	48	64
Nb_N	0.0900	0.0852	0.0537	0.0211	-0.0049	-0.0442	-0.0557	-0.0572	-0.0575	-0.0575
$N\tilde{b}_N$	-0.0576	-0.0576	-0.0576	-0.0576	-0.0576	-0.0576	-0.0576	-0.0576	-0.0576	-0.0576
$b_N / \tilde{b}_N$	-1.5625	-1.4792	-0.9330	-0.3670	0.0853	0.7681	0.9663	0.9930	0.9977	0.9984
Nb_N^0	0.0972	0.1050	0.0850	0.0612	0.0413	0.0104	0.0013	0.0002	0.0000	0.0000
Nb_N^1	-0.0072	-0.0198	-0.0313	-0.0401	-0.0462	-0.0547	-0.0570	-0.0573	-0.0575	-0.0575
p^0	0.5400	0.2916	0.1575	0.0850	0.0459	0.0072	0.0006	0.0001	0.0000	0.0000

Results confirm that the approximate formula works well even for populations of very moderate size.

REFERENCES

[1] *Billingsley P., Convergence of Probability Measures, John Wiley & Sons 1969*

[2] *Iosifescu M., Finite Marcov Processes and Their Applications, Dover Publications 2007*

[3] *Lee T. C., Judge G. G., Zellner A., Estimating the parameters of the Markov probability model from aggregate time series data, North-Holland, Amsterdam 1970.*

[4] *Lemaire J., Bonus-malus systems in automobile insurance, Kluwer, Boston 1995.*

[5] *McCall J. J., A Markovian model of income dynamics, Journal of the American Statistical Association, vol. 66, 1971*

[6] *Walhin J. F., Paris J., Using mixed Poisson processes in connection with bonus-malus systems, ASTIN Bulletin vol.29, 1999*