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

wotto@wne.uw.edu.pl

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 the paper is to study implications of the departure from the assumption on homogeneity of population.

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 probably 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 may 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 several of these properties heavily distorted.

2. BASIC ASSUMPTIONS OF THE MIGRATION'S MODEL

- The BM system consists of S classes. Transitions of drivers between classes takes place once a year.
- A driver characterised by the value λ of risk parameter Λ produces losses according to the Poisson process with intensity λ per year, and claims all of them to the insurer.
- Transition probabilities for a driver with $\Lambda = \lambda$ form a matrix:

$$\mathbf{P}(\lambda) = \{p_{i,j}(\lambda)\}_{i,j=1,2,\dots,S}$$

- Transition rules are reflected by positions of non-zero elements of $\mathbf{P}(\lambda)$ that correspond to events of zero claims, 1 claim, 2 claims,...
- Non-zero elements of $\mathbf{P}(\lambda)$ equal probabilities of these events for a driver characterised by the value λ of the parameter Λ
- Transition rules ensure that $\mathbf{P}(\lambda)$ is ergodic for all $\lambda > 0$.

About the population of drivers we assume:

- The population is closed.
- Population consists of N drivers characterised by positive values $\lambda_1, \lambda_2, ..., \lambda_N$ of risk parameter.
- Migration processes of individual drivers are independent.

In the paper the **continuous variant** of the model is also considered, where the discrete distribution $\lambda_1, \lambda_2, ..., \lambda_N$ is replaced by a continuous density $f(\lambda)$.
For brevity, the continuous variant **is not presented here.**

3. MODEL WITH DISCRETE DISTRIBUTION OF RISK PARAMETER Λ

We consider a **two stage experiment**:

- At the first stage a driver is randomly drawn from the population. Denoting the result (number of driver being drawn) by *K*, we have:

  $$\Pr(K = k) = N^{-1}, \quad k = 1,2,,,,N$$

- At the second stage the process of migration of the driver on the space of BM classes takes place. Conditionally (given $K = k$) the process satisfy assumptions of the homogeneous Markov Chain.

Essence:

- The process concerns changes in time of the joint distribution over members of the population and BM classes. **Marginal distribution** over BM classes and **conditional distributions** change in a different way. An exception is a homogeneous population case, when marginal and all conditional distributions are identical.

Basic conditional characteristics of the process (provided the event $K = k$ has occurred at the first stage) are:

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

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

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

which under assumed ergodicity converges to the limit:

$$\mathbf{h}_k := \lim_{t\to\infty} \mathbf{h}_k^{(t)} \text{ for } k = 1,2,\ldots,N.$$

Of course the limit satisfies the system of equations:

$$\mathbf{h}_k = \mathbf{h}_k \mathbf{P}_k .$$

Basic marginal characteristics of the process are:

- Sequence $\left\{\mathbf{h}^{(t)}\right\}_{t=1}^{\infty}$ of vectors $\mathbf{h}^{(t)} := \left[h_1^{(t)} \quad h_2^{(t)} \quad \ldots \quad h_S^{(t)}\right]$ of probabilities of staying in year t in classes $i = 1,2,\ldots,S$, where:

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

- And sequence of matrices $\mathbf{P}^{(t)} = \left\{p_{i,j}^{(t)}\right\}_{i,j=1,\ldots,S}$ with elements:

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

representing probability of transition from class i in year t to class j in the next year.

The result confirm that matrices $\mathbf{P}^{(t)}$ are in general **time-dependent**, except when population is homogeneous, i.e. when $\lambda_1 = \lambda_2 = \ldots = \lambda_N$.

4. Where the confusion comes from

For the marginal process the following formula holds:

$$\mathbf{h}^{(t+1)} = \mathbf{h}^{(t)}\mathbf{P}^{(t)} .$$

Despite matrices $\mathbf{P}^{(t)}$ are time-dependent, convergence of conditional probability vectors $\mathbf{h}_k^{(t)} \to \mathbf{h}_k$ for each $k = 1,2,...,N$ ensures that:

- Marginal probability vectors converge to their limiting value: $\mathbf{h}^{(t)} \to \mathbf{h}$,
- As a result $\mathbf{P}^{(t)} \to \mathbf{P}$ as well
- And so the limiting values satisfy the system of equations: $\mathbf{h} = \mathbf{hP}$.

The above results may suggest that errors made when treating the true migration process as a homogeneous Marcov Chain:

- are significant for small t (**true**)
- but for large t could be neglected (**false**)

In the paper, distortions due to heterogeneity of population on properties/characteristics of the homogeneous Marcov Chain are studied. The study focus on the following properties/characteristics:

- The ability to measure the rate of convergence of vectors $\mathbf{h}^{(t)}$ to their limit $\mathbf{h}$ by the second-largest eigenvalue of $\mathbf{P}$
- Basic characteristics of the distribution of vector of frequencies of drivers staying in classes $1,2,...,S$ in a given year t
- Basic characteristics of the conditional distribution of that vector given its value from year $t-1$
- Bias and variance of estimators of transition probabilities based on empirical data on number of transitions between classes at the end of year $t-1$

Some of results are illustrated by examples, others are presented in a more general form.

5. THE SECOND-LARGEST (IN ABSOLUTE TERMS) EIGENVALUE OF $\mathbf{P}$

Let us consider a simple 3-class no claim discount system with a transition probability matrix:

$$\mathbf{P} = \begin{bmatrix} q & p & 0 \\ q & 0 & p \\ q & 0 & p \end{bmatrix}$$

Where p is a no-claim probability for a given driver, and $q = 1 - p$. The stationary distribution on the space of classes is simply given by:

$$\mathbf{h} = \begin{bmatrix} q & qp & p^2 \end{bmatrix}$$

as the current position of a driver depends only on two last years. Eigenvalues of matrix $\mathbf{P}$ equal $\{\rho_1, \rho_2, \rho_3\} \approx \{1, 0, 0\}$, so the second largest eigenvalue is zero, which confirms the fact that the sequence $\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \mathbf{h}^{(3)}, \ldots$ converge to $\mathbf{h}$ in a finite number of steps.

Let us take now a population of drivers that differ by no-claim probability p. The limit of the sequence $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \mathbf{P}^{(3)}, \ldots$ of transition probability matrices equals now:

$$\mathbf{P} = \begin{bmatrix} q_{(1)} & p_{(1)} & 0 \\ q_{(2)} & 0 & p_{(2)} \\ q_{(3)} & 0 & p_{(3)} \end{bmatrix}.$$

Where $p_{(i)}$ denote probability of no-claim for a driver being in class number i in the stationary state, and $q_{(i)} = 1 - p_{(i)}$.

Eigenvalues of this matrix are the following:

$\rho_1 = 1$,

$$\rho_{2,3} = \frac{1}{2}\left\{ (p_{(3)} - p_{(1)}) \pm \sqrt{(p_{(3)} - p_{(1)})^2 + 4p_{(1)}(p_{(3)} - p_{(2)})} \right\}.$$

Obviously the latter two equal zero only when $p_{(1)} = p_{(2)} = p_{(3)}$.

However, it can be formally shown that if there is any diversity of parameter q in the population, then no-claim probabilities satisfy:

$$p_{(1)} < p_{(2)} < p_{(3)}.$$

The interpretation is straightforward:

- structure of class 1 is biased towards higher share of bad drivers,
- structure of class 3 is biased towards higher share of good drivers,
- class 2 is in this respect a medium one.

On the other hand, we know that all drivers reach their stationary state in two years, so all the population attain the stationary state in two years, too. Thus the conclusion reads:

> The second largest (in absolute terms) eigenvalue of matrix **P looses the property of reflecting the rate of convergence** once we depart from the homogeneity assumption.

6. NOTATIONS FOR ANALYSIS OF FREQUENCIES

Let us denote by:

- $N_{i,j}^{(t)}$ - the number of drivers migrating from class i in year t to class j in the next year
- $N_i^{(t)} := \sum_{j=1}^{S} N_{i,j}^{(t)}$ - the number of drivers in class i in year t,
- $\mathbf{N}^{(t)} := \begin{bmatrix} N_1^{(t)} & N_2^{(t)} & \dots & N_s^{(t)} \end{bmatrix}$ - the vector of the above numbers

Further on we focus on 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 conventional, as $\mathbf{N}^{(t)}$ depends also on starting conditions. The convention is justified by focussing on the case of large t, when that dependence disappears.

7. Moments of the unconditional distribution of $\mathbf{N}^{(t)}$

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

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

Due to assumed independency of migration processes of individual drivers the covariance matrix is just a simple sum:

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

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

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

Where:

- The first term appearing on the RHS is a covariance matrix adequate for the case of homogeneous population where $\mathbf{h}_1^{(t)}=\mathbf{h}_2^{(t)}=\ldots=\mathbf{h}_N^{(t)}=\mathbf{h}^{(t)}$.
- Second term measures heterogeneity within the set of vectors $\mathbf{h}_1^{(t)},\mathbf{h}_2^{(t)},\ldots,\mathbf{h}_N^{(t)}$. It is in fact an *N*-multiple of the covariance matrix of the random vector $\mathbf{h}_K^{(t)}$ being a function of random variable *K*.
- For $t\to\infty$ the second matrix stabilises. It represents the tendency (realised over a number of years) towards concentration of good drivers in bonus classes and bad drivers in malus classes.

8. The conditional distribution of $\mathbf{N}^{(t)}$ given $\mathbf{N}^{(t-1)}$

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

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

where in the case of homogeneous population we have:

- $\mathrm{E}(\mathbf{N}^{(t)} \big| \mathbf{N}^{(t-1)}) = \mathbf{N}^{(t-1)}\mathbf{P}$, so that:
- $\mathrm{E}(\boldsymbol{\varepsilon}^{(t)} \big| \mathbf{N}^{(t-1)}) = 0$,
- and $\mathrm{cov}(\boldsymbol{\varepsilon}^{(t)}, \boldsymbol{\varepsilon}^{(t)} \big| \mathbf{N}^{(t-1)})$ dependent on $\mathbf{N}^{(t-1)}$.

Despite the last complication it is a system of linear regressions, and so covariance matrix of random terms can be easily calculated:

$$\mathrm{cov}(\boldsymbol{\varepsilon}^{(t)}, \boldsymbol{\varepsilon}^{(t)}) = \mathrm{cov}(\mathbf{N}^{(t)}, \mathbf{N}^{(t)}) - \mathbf{P}'\,\mathrm{cov}(\mathbf{N}^{(t-1)}, \mathbf{N}^{(t-1)})\mathbf{P} .$$

The decomposition takes an especially nice form for large *t*:

$$\lim_{t\to\infty} \mathrm{cov}(\boldsymbol{\varepsilon}^{(t)}, \boldsymbol{\varepsilon}^{(t)}) = N\left\{diag(\mathbf{h}) - \mathbf{P}'diag(\mathbf{h})\mathbf{P}\right\}.$$

All the above properties are destroyed once we admit heterogeneity in the population, as $\mathbf{P}^{(t)}$ differs then from its 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 a driver can enter class j only from class i. We can focus then on the conditional expectation $\mathrm{E}(N_j^{(t)} \middle| N_i^{(t-1)})$. Direct calculations for extreme cases $N_i^{(t-1)} = N$ and $N_i^{(t-1)} = 1$ render following results:

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

which generally differs from $Np_{i,j}^{(t-1)}$, and:

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

which differs from $p_{i,j}^{(t-1)}$ as well.

It is fairly difficult to derive more general results. However, examples studied strongly support the following interpretation:

- For large t expected structure of the bonus (malus) class is biased towards larger share of good (bad) drivers
- However, the information that in year t there are more drivers than expected in the class implies that the bias is weaker (the structure of the class is more alike the structure of the population)
- To the contrary, unexpectedly small number of drivers in the class implies stronger selection, and so stronger bias
- Hence, if for some large t the probability of no-claim in class i is greater (smaller) than average in the population, then this probability conditional on number of drivers in this class $N_i^{(t)}$ is a decreasing (increasing) function of $N_i^{(t)}$.

The vector of random terms of the system of equations:

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

can be decomposed into two effects:

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

- 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.
The conclusion is not derived in full generality, but rather based on examples studied.

9. ESTIMATING TRANSITION PROBABILITIES WHEN DATA ON TRANSITIONS ARE AVAILABLE

Transition probabilities could be estimated on the basis of empirical data on number of transitions $N_{i,j}^{(t)}$. A natural estimator of the transition probability $p_{i,j}^{(t)}$ can be defined then as:

$$\hat{p}_{i,j}^{(t)} := \frac{N_{i,j}^{(t)}}{N_i^{(t)}},$$

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 the following approximations:

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

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

Both formulas are exact in respect of terms of order $1/N$ or larger, whereas terms of order N^{-2} and smaller are neglected.

- In both cases the first component represents the classical estimator of the conditional probability of some event when the condition is satisfied for an individual observation with probability $h_i^{(t)}$.
- Heterogeneity results in bias of order $1/N$.
- Heterogeneity results also in efficiency gain that is of order $1/N$ as well as the whole variance. Thus the relative efficiency gain does not disappear when population size N increases.

Approximations for covariances are as follows:

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

when $i \neq m$, i.e. transitions are from two different classes, and:

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

when transitions to different classes ($j \neq n$) are from the same class i.

- A separate effect independent of heterogeneity concerns only transitions from the same class to two different classes.
- Under heterogeneity all transitions are in a way dependent events. Larger $N_{i,j}^{(t)}$ than expected (given $N_i^{(t)}$) imply larger fraction of bad (good) drivers than expected among those moving from class i to j, and so smaller fraction of them in the rest of population.

10. SUMMARY

Major implications of the heterogeneity of the population:

- Transition probabilities vary in time for small t (*not really new*)
- Different properties of the process $\mathbf{N}^{(t)}$ even for large t:
    - Probability of loss depending on class (*well known, but 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 when estimating transition probabilities (*new result*)

Lack of awareness of the implications may lead to misinterpretations of various anomalies that might be observed in empirical research. These anomalies are quite common and may come from:

- the hunger for bonus phenomenon
- various factors that change parameters of the migration process in subsequent calendar years.

REFERENCES

[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*