ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 225, 2009

*Andrzej Czajkowski**

FORECASTING INCOME DISTRIBUTIONS OF HOUSEHOLDS IN POLAND ON THE BASIS OF MARKOV CHAINS

Abstract

In order to forecast income distributions of population, we can make use of, among others, stochastic processes. These processes can be used to determine probabilities of transition of households from one income class to another.

The paper attempts to present an application of homogenous Markov chains in the process of forecasting the income structure of six socio-economic groups of population in Poland for the years 2004, 2006 and 2008. Forecasts are based on results of individual household budgets surveys.

Key words: income distributions, Markov chains, forecast.

1. Introduction

In case of forecasting income distributions of households, especially for shorter forecast horizon, stochastic processes used to determine the volume of probabilities of households' transition from one particular income class to another, can be applicable. Here, we mean Markov chains (G a j e k, 1998).

We say that a sequence of experiments is bonded in the Markov chain if a series of distributions of probabilities of transition to a random stage depends on the result of the previous stage but does not depend on results of earlier stages.

In practical applications a particularly important case of the Markov chain is the situation when the transition distribution does not depend on the number of the stage, i.e. if at k-th and l-th random stages, the result ω_i is satisfied, then at (k+1)-th and (l+1)-th stages respectively, the result ω_j will be fulfilled with the

* Ph.D., Chair of Statistical Methods, University of Łódź.

same probability. The Markov chain of this property is called homogenous Markov chain (K u b o l e, K r u p o w i c z, 1982).

Let **P** denote a stochastic matrix of transition including probabilities (p_{ij}) of transition from the state "*i*" to the state "*j*". For the random sequence $(\omega_{i_0}, \omega_{i_1}, \ldots \omega_{i_n})$, we can determine the following probability distribution:

$$\mathbf{P}^{(n)}(\omega_{i_0}, \omega_{i_1}, \ldots, \omega_{i_n}) = P_0(\omega_{i_0}) p_{i_0} p_{i_1} p_{i_1} p_{i_2} \cdots p_{i_{n-1} i_n} \tag{1}$$

where:

P_0 is the initial random distribution defined in a set $\Omega = \{\omega_{i_0}, \omega_{i_1}, \ldots, \omega_{i_n}\}$,

$p_{i_k i_{k+1}}$ – elements of matrix of transition **P**.

From the point of view of forecasting income distributions of population, the distribution P_n which is a marginal distribution $\mathbf{P}^{(n)}$, is of great importance. The distribution determines probabilities of stages "*i*" at *n*-th moment or *n*-th stage as $p_i(n)$ or $P_n(\omega_i)$. Hence,

$$p_i^{(n)} = \sum_{j_0, j_1, \ldots, j_{n-1}} p_{j0}(0) \cdot p_{j_0 j_1} \cdot \ldots \cdot p_{j_{n-1} j_n} \tag{2}$$

Determining an element in *i*-th row and *j*-th column of matrix $\mathbf{P}^{(n)}$, we obtain a formula for the vector of probabilities at *n*-th stage

$$\mathbf{p}^{(n)} = (\mathbf{P}^T)^n \mathbf{p}^{(0)} \tag{3}$$

where:

$\mathbf{p}^{(n)}$ denotes the matrix of transition after "*n*" steps,

$\mathbf{p}^{(0)}$ is a vector of initial probabilities.

2. Application of Markov chains to forecasting income distribution of households

Let us denote thresholds of income drawn by the household over the particular time by $d_1, d_2, \ldots, d_t$ and reached income class also in time *t* by y_t. Then, the probability that the household will be situated in *i*-th income class in time *t* can be defined by means of the following formula:

$$P(y_t = d_i) = p_i(t) \tag{4}$$

Above all, the probability depends on what income class the household was situated earlier.

Let us denote by $p_{ij}(t)$ the conditional probability that the household which, in period $t-1$ was situated in i-th income class, in period t will be situated in j-th income class. And if, at the same time, the probability does not depend on the period t, then this process is called the Markov chain, i.e.

$$P(y_t = d_j | y_{t-1} = d_i) = p_{ij}(t) = p_{ij} \tag{5}$$

The probability of transition of households from the specific income class in the period $t-1$ to specific income classes in time t can be defined as follows:

$$P(y_{t-1} = d_i, y_t = d_j) = P(y_{t-1} = d_i) P(y_t = d_j | y_{t-1} = d_i) = p_i(t-1) \tag{6}$$

The total probability that the household will reach j-th income class in period t can be defined as follows:

$$P(y_t = d_j) = \sum_{i=1}^{k} p_i(t-1) p_{ij}; \qquad j = 1, 2, \ldots k \tag{7}$$

Assuming that the number of income classes that the household can reach in period from $t-1$ to t equals k, then probabilities $p_i(t-1)$ and $p_i(t)$ are column vectors of k components and probabilities p_{ij} create the square matrix of probabilities $\mathbf{P}$ which has the following form:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{bmatrix} \tag{8}$$

while: $p_{ij} \geq 0$ $(i, j = 1, 2, \ldots, k)$,

$$\sum_{j=1}^{k} p_{ij} = 1, (i = 1, 2, \ldots, k).$$

Elements of matrix $\mathbf{P}$ have a practical interpretation. On the main diagonal there are fractions (indices of structure) of households whose income did not undergo significant changes in time unit t and which remained in the same income class. Whereas, elements outside the main diagonal $p_{ij}(i \neq j)$ denote the fraction of households which, in time unit t, will transit from i-th to j-th income class.

The probability of reaching specific income thresholds in period t can be presented in the form of the following matrix formula:

$$\mathbf{p}(t) = \mathbf{P}^T \mathbf{p}(t-1) \tag{9}$$

where:

$$\mathbf{p}(t) = \begin{bmatrix} p_1(t) \\ p_2(t) \\ \vdots \\ p_k(t) \end{bmatrix}, \quad \mathbf{p}(t-1) = \begin{bmatrix} p_1(t-1) \\ p_2(t-1) \\ \vdots \\ p_k(t-1) \end{bmatrix}$$

$\mathbf{P}$ is defined as above.

The vector $\mathbf{p}(t)$ defines structures (distribution) of income in period t, $\mathbf{p}(t-1)$ defines structures of income in the initial period $t-1$, while $\mathbf{P}^T$ is a transposed matrix of transition.

Probabilities of transition p_{ij} by means of which P_n can be determined at a random moment, can be estimated from the following formula:

$$y_j(t) = \sum_{i=1}^{k} p_{ij} y_i(t-1) + \varepsilon_j(t), i, j = 1, ..., k \tag{10}$$

where:

n_{jt} denotes the share of households in the total number of households which were situated in i-th income class at the moment t.

3. Forecast of income distributions of main socio-economic groups of population in Poland

In the described survey, the estimate of elements of transition matrix, i.e. elements p_{ij}. was based on aggregated data on observed unconditional distributions of the form:

$$y_{it} = \frac{n_{it}}{\sum_i n_{it}} \tag{11}$$

where:

n_{it} denotes the number of households which, in period t, were situated in i-th income class.

Putting the observed distributions y_{it} into the formula (10) in place of unknown unconditional distributions, we obtain

$$y_{jt} = \sum_{jt}^{k} p_{ij} y_{it-1} + \varepsilon_{jt}, \quad j = 1, 2, ..., r \tag{12}$$

In the formula, transition probabilities p_{ij} are parameters of regression function y_{jt} in relation to variables $y_{1t-1}, y_{2t-1}, ..., y_{rt-1}$, while p_{ij} fulfils conditions $p_{ij} \geq 0\,(i, j = 1, ..., r)$ and $\sum_{j=1}^{r} p_{ij} = 1 (j = 1, 2, ..., r); \varepsilon_{jt}$ – random component.

On the basis of the information from individual household budgets surveys[1], five income classes were formed. They were characterized by the particular percentage of median, namely: up to 60% Me (1), 60–100% Me (2), 100–150% Me (3), 150–200% Me (4) and above 200% Me (5). This classification is the same as in researches which were done earlier by K u d r y c k a (1998). Then, for each of the surveyed years (1999–2003), for each of six socio-economic groups, empirical unconditional distributions y_{it} were calculated as quotients of the number of households in the particular income classes by the total size corresponding with the particular income class. Finally, for the total income, per person in the household, matrices of probabilities of transition for six types of households were determined. Further on, we present estimated matrices of transition.

1. Households of employees

$$\mathbf{P}_1 = \begin{bmatrix} 0.7533 & 0.2352 & 0 & 0.0115 & 0 \\ 0.1040 & 0.7659 & 0.1095 & 0.0206 & 0 \\ 0.0371 & 0.1183 & 0.7409 & 0.0871 & 0.0166 \\ 0 & 0 & 0.2144 & 0.5870 & 0.1986 \\ 0.0300 & 0 & 0.0003 & 0.2033 & 0.7664 \end{bmatrix}$$

As it is known, elements on the main diagonal of matrix (p_{ii}) define the probability that the household will remain in the same income class for a year. From conducted calculations, it results that in the surveyed group of households these probabilities, except for the fourth income class, are very strong and at approximate level (from 0.7409 to 0.7674).

Analysing probabilities outside the main diagonal $p_{ij} (i \neq j)$, we note that their values are relatively big for households transiting from the 1st to the 2nd income class $(p_{12} = 0.2352)$, from the 4th to the 3rd income class $(p_{43} = 0.2144)$,

[1] Household budgets surveys are conducted by means of representative method which enables to generalize (with a specific error size) obtained results for the particular socio-economic groups of households. The size of the sample of households sampled annually amounts to thirty odd thousand which account for from 0.2 to 0.3% of households in Poland in total.

from the 5th to the 4th income class ($p_{54} = 0.2033$) and from the 4th to the 5th income class ($p_{45} = 0.1986$). Whereas, transition of households from the 1st to the 3rd, from the 1st to the 5th, from the 2nd to the 5th, from the 4th to the 1st, from the 4th to the 2nd and from the 5th to the 2nd income class ($p_{13} = p_{15} = p_{25} = p_{41} = p_{42} = p_{52} = 0$) over the year is impossible.

2. Households of employees – farmers

The matrix of transition for this socio-economic group has the following form:

$$\mathbf{P}_2 = \begin{bmatrix} 0.7187 & 0.0978 & 0.0870 & 0.0557 & 0.0408 \\ 0.0788 & 0.8189 & 0.0422 & 0.0599 & 0.0002 \\ 0.0658 & 0.1014 & 0.8046 & 0.0250 & 0.0032 \\ 0 & 0.0980 & 0.1275 & 0.6100 & 0.1646 \\ 0 & 0.0430 & 0.0885 & 0.1703 & 0.6982 \end{bmatrix}$$

In this group of households, the strongest probability of the household's remaining in the particular income class is noticeable in income classes 2 and 3 ($p_{33} = 0.8189$, $p_{44} = 0.8046$). Probabilities of transition from the 5th to the 4th income class ($p_{54} = 0.1703$) and from the 4th to the 5th income class $p_{45} = (0.1646)$ are relatively strong. In the light of conducted calculations, however, transition of the household from the 4th to the 1st and from the 5th to the 1st income class is impossible. What is more, it is worth drawing our attention to the fact that the remaining probabilities of transition from i-th to j-th income class ($i \neq j$) are, in comparison with the particular probabilities of the previously analysed households group (households of employees), generally stronger.

3. Households of farmers

The matrix of transition is as follows:

$$\mathbf{P}_3 = \begin{bmatrix} 0.8227 & 0.0978 & 0.0795 & 0 & 0 \\ 0.1000 & 0.7455 & 0.0516 & 0.0900 & 0.0129 \\ 0.0900 & 0.0649 & 0.8451 & 0 & 0 \\ 0.0560 & 0.0980 & 0.660 & 0.7800 & 0 \\ 0.0034 & 0.0430 & 0 & 0.0117 & 0.9419 \end{bmatrix}$$

The household's remaining in the highest (the fifth) income class ($p_{55} = 0.9419$) is the most probable. However, probabilities of staying in the remaining income classes over one year are also strong. Whereas, household's transition from the 1st do the 4th, from the 1st to the 5th, from the 3rd to the 4th, from the 3rd to the 5th, from the 4th to the 6th and from the 5th to the 3rd income class, taking into consideration the probability, is impossible. The remaining probabilities of transition from i-th to j-th income class ($i \neq j$) are not very strong which, taking into account the above remark, can prove that households of farmers have quite time-stable income level.

4. Households of self-employed

The matrix of transition for this group of households is as follows:

$$\mathbf{P}_4 = \begin{bmatrix} 0.9032 & 0.0870 & 0.0098 & 0 & 0 \\ 0.0190 & 0.8910 & 0.0900 & 0 & 0 \\ 0 & 0.0900 & 0.7509 & 0.0829 & 0.0762 \\ 0.0541 & 0.0980 & 0.1700 & 0.6100 & 0.0679 \\ 0.0034 & 0.0430 & 0 & 0.1210 & 0.8326 \end{bmatrix}$$

In the light of conducted calculations, also this group of households has the time-stable income. Probabilities of remaining in income classes 1 and 2 which equal 0.9032 and 0.8910, respectively, reach the highest level. Similarly to the socio-economic groups analysed above, probabilities of staying in the remaining income classes are also strong. On the other hand, the number of income classes whose probabilities of transition to another income class equal zero, is relatively big. And, probabilities of transition from the 1st to the 4th and the 5th income class, from the 2nd to the 4th and the 5th income class, from the 3rd to the 1st and from the 5th to the 3rd income class equal zero. The remaining elements of the matrix of transition are, generally, not too big.

5. Households of pensioners and retirees

Estimated matrix of transition is as follows:

$$\mathbf{P}_5 = \begin{bmatrix} 0.8150 & 0.0980 & 0.0870 & 0 & 0 \\ 0.7970 & 0.7169 & 0.1939 & 0.0036 & 0.0060 \\ 0 & 0.1998 & 0.6838 & 0.0695 & 0.0469 \\ 0.0560 & 0.0753 & 0.1700 & 0.6100 & 0.0887 \\ 0.0034 & 0.0430 & 0.0120 & 0.3387 & 0.6029 \end{bmatrix}$$

Similarly to the groups of households analysed above, the strongest probabilities concern remaining of the household in the same income class over the year. And, at the same time, as a rule, the probability of remaining in the income class 1 ($p_{11} = 0.8150$) is the strongest. It is also worth noticing that probabilities of staying in the remaining income classes are far weaker. Probabilities of transition from the 5th to the 4th income class are relatively strong. While, probabilities of transition from the 1st to the 4th and the 5th income class and from the 3rd to the 1st income class equal zero.

6. Households maintained from non-earned sources

For this group of households, estimated matrix of transition is as follows:

$$\mathbf{P}_6 = \begin{bmatrix} 0.7950 & 0.0978 & 0.0870 & 0.0202 & 0 \\ 0.1000 & 0.7321 & 0.0820 & 0.0409 & 0.0450 \\ 0.0900 & 0.1100 & 0.7200 & 0.0800 & 0 \\ 0.0560 & 0.0715 & 0.0980 & 0.6100 & 0.1645 \\ 0 & 0.0430 & 0.1210 & 0.1116 & 0.7244 \end{bmatrix}$$

Still, probabilities of remaining in the same income class run at the highest level, while the probability of households' remaining in the income class 1 ($p_{11} = 0.7950$) is the strongest. Probabilities of changing the income class over the year are not too strong – the strongest is the probability of transition from the 4th to the 6th income class ($p_{46} = 0.1645$). In the light of data, however, transition from the 1st and the 3rd income class to the 5th income class and from the 5th to the 1st income class is impossible.

Income distributions from the year 2003 and matrices of probabilities of transitions $\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_6$ are the basis of the forecast of income distributions. On this basis, forecasts of income distributions of 6 types of households for the years 2004, 2006 and 2008 were made.

In the forecast, it was assumed that in the following (future) points of time, analysed income distributions would be homogeneous Markov chains. If the condition is fulfilled, it is possible to forecast the income distribution by means of the matrix $\mathbf{P}$ containing the estimate of probabilities of transition obtained from the previous period t. Results of the forecast are presented in Tables 1, 2 and 3.

Assuming that income of households of employees will develop in the future, according to assumptions made above, we should expect, on the one

hand, the increase in the number of households in the last two (4 and 5) income classes, and on the other hand, the decrease in this number in income classes 2 and 3.

In case of households of employees- farmers we forecast that in the future, we will have to do with the increase in the number of households of the income from income classes 1 and 5 with the simultaneous significant drop in their number in income classes 2 and 4.

In case of households of farmers, we can forecast that in the future, the number of households in the lowest and in the medium (3) income class will increase. What is more, the number of households in the highest income class may decrease.

The situation in households of self-employed shall be different. In this group of households the decrease in the number of households of the income from income classes 5, 1 and 3 is forecasted. It seems that the number of households of the income from the income class 2 will increase significantly.

The structure of the income of households of pensioners and retirees shall be fairly stable in the future. We forecast that the number of households from the income class 1 will increase a little. At the same time, the number of households from income classes 2 and 3 shall decrease a little.

In case of households maintained from non-earned sources, we can expect the significant decrease in their number in the highest income class and, to a smaller extent, in the income class 2. Whereas, the number of households of the income from the remaining income classes shall increase a little.

Table 1

Forecasts of distributions of the total income per capita in households of employees and employees-farmers

Year	Income classes (in % of the median)				
	up to 60%	60–100%	100–150%	150–200%	above 200%
Households of employees					
1999	0.1598	0.3430	0.2703	0.1174	0.1095
2000	0.1707	0.3294	0.2662	0.1121	0.1216
2001	0.1763	0.3238	0.2648	0.1120	0.1231
2002	0.1788	0.3212	0.2489	0.1170	0.1341
2003	0.1828	0.3172	0.2443	0.1161	0.1399
forecast					
2004	0.1840	0.3148	0.2405	0.1178	0.1429
2006	0.1849	0.3111	0.2361	0.1202	0.1477
2008	0.1847	0.3084	0.2339	0.1220	0.1510

Table 1 (contd.)

Year	Income classes (in % of the median)				
	up to 60%	60–100%	100–150%	150–200%	above 200%
Households of employees-farmers					
1999	0.1577	0.3423	0.2872	0.1183	0.0945
2000	0.1588	0.3412	0.2945	0.1247	0.0808
2001	0.1620	0.3380	0.2935	0.1116	0.0949
2002	0.1709	0.3291	0.2872	0.1201	0.0928
2003	0.1578	0.3422	0.2778	0.1214	0.1008
forecast					
2004	0.1839	0.3148	0.2405	0.1179	0.1429
2006	0.1848	0.3112	0.2361	0.1202	0.1477
2008	0.1846	0.3084	0.2339	0.1221	0.1510

Source: own calculations on the basis of household budgets surveys.

Table 2

Forecasts of distributions of the total income per capita in households of farmers and self-employed

Year	Income classes (in % of the median)				
	up to 60%	60–100%	100–150%	150–200%	above 200%
Households of farmers					
1999	0.2903	0.2100	0.2138	0.1189	0.1670
2000	0.2743	0.2257	0.1854	0.1152	0.1994
2001	0.2671	0.2329	0.1920	0.1085	0.1995
2002	0.2859	0.2145	0.1972	0.1155	0.1870
2003	0.2691	0.2313	0.2192	0.1072	0.1732
forecast					
2004	0.2708	0.2309	0.2257	0.1065	0.1661
2006	0.2748	0.2410	0.2359	0.1051	0.1532
2008	0.2788	0.2317	0.2437	0.1041	0.1417
Households of self-employed					
1999	0.1776	0.3224	0.2513	0.1160	0.1327
2000	0.1776	0.3226	0.2501	0.1098	0.1399
2001	0.1968	0.3032	0.2303	0.1259	0.1438
2002	0.2035	0.2970	0.2345	0.1249	0.1401
2003	0.2246	0.2756	0.2355	0.1108	0.1534
forecast					
2004	0.2146	0.3038	0.2413	0.0871	0.1532
2006	0.1957	0.3441	0.2454	0.0648	0.1500
2008	0.1797	0.3702	0.2471	0.0570	0.1460

Source: own calculations on the basis of household budgets surveys.

Table 3

Forecasts of distributions of the total income per capita in households of pensioners and retirees and maintained from non-earned sources

Year	Income classes (in % of the median)				
	up to 60%	60–100%	100–150%	150–200%	above 200%
Households of pensioners and retirees					
1999	0.1433	0.3569	0.3469	0.1027	0.0502
2000	0.1551	0.3449	0.3359	0.1094	0.0548
2001	0.1585	0.3416	0.3233	0.1160	0.0607
2002	0.1652	0.3348	0.3557	0.1135	0.0609
2003	0.1675	0.3326	0.3240	0.1114	0.0646
forecast					
2004	0.1694	0.3308	0.3203	0.1135	0.0660
2006	0.1723	0.3278	0.3165	0.1160	0.0647
2008	0.1740	0.3260	0.3148	0.1173	0.0679
Households of maintained from non-earned sources					
1999	0.2141	0.2863	0.2315	0.1201	0.1480
2000	0.2137	0.2866	0.2319	0.1251	0.1427
2001	0.2391	0.2612	0.2177	0.1147	0.1674
2002	0.2231	0.2769	0.2285	0.1081	0.1634
2003	0.2391	0.2612	0.2165	0.1087	0.1746
forecast					
2004	0.2418	0.2537	0.2299	0.1186	0.1561
2006	0.2480	0.2480	0.2426	0.1255	0.1359
2008	0.2533	0.2472	0.2469	0.1263	0.1263

Source: own calculations on the basis of household budgets surveys.

4. Final remarks

The results of forecasts made for the years 2004, 2006 and 2008 support the situation that took place earlier, Kudrycka, Radziunkiewicz (1998). However, we should pay attention to the significant increase in the percentage of households of the poverty sphere (the income up to 60% of the median) within the group of households maintained from non-earned sources.

References

Gajek L. (1998), *Wnioskowanie statystyczne*, Wyd. Naukowo-Techniczne, Warszawa.

Kubole L., Krupowicz A. (1982), *Wprowadzenie do rachunku prawdopodobieństwa i jego zastosowanie*, PWN, Warszawa.

Kudrycka L., Radziunkiewicz M. (1998), *Możliwości pozyskiwania dochodów przez poszczególne grupy społeczne – zmiany w rozkładach dochodów i prognozy*, Warszawa.

Andrzej Czajkowski

Prognozowanie rozkładów dochodów gospodarstw domowych w Polsce w oparciu o łańcuchy Markowa

Do prognozowania rozkładów dochodów ludności wykorzystywane być mogą m. in. procesy stochastyczne służące do określania wielkości prawdopodobieństw przejścia gospodarstw domowych z określonej grupy dochodowej do innej.

W opracowaniu podjęto próbę wykorzystania jednorodnych łańcuchów Markowa do sporządzania prognoz struktury dochodów sześciu grup społeczno-ekonomicznych ludności w Polsce dla lat 2004, 2006 i 2008. Podstawę szacunków stanowiły wyniki badań budżetów indywidualnych gospodarstw domowych.