DYNAMIC ECONOMETRIC MODELS
Vol. 8 – Nicolaus Copernicus University – Toruń – 2008

Katarzyna Osiecka
Warsaw University of Technology
Józef Stawicki
Nicholas Copernicus University in Toruń

Markov Set-Chains as a Tool for an Analysis of Household Expenditure Structure in Poland Between 1993–2005

Markov Chain Model describes a stochastic process on a finite or countable state space and in discrete time t = 1, 2, 3, … . Basic concepts are explained by the formula:

$$P\left\{X^{(t+1)} = j \middle| X^{(t)} = i, X^{(t-1)} = i_{t-1}, \ldots\right\} = P\left\{X^{(t+1)} = j \middle| X^{(t)} = i\right\} = p_{ij}(t)$$

The probability distribution of $X^{(t)}$ depends only on the state occupied at the end of the previous experiment. Probabilities $p_{ij}(t)$ are called transition probabilities and the matrix $P_t = \lfloor p_{ij}(t) \rfloor$ is a one-step transition matrix. According to the random value of the initial state as vector $\bar{x}_t$ and using the definition of conditional probability in the matrix form P_t we have the formula

$$\bar{x}_{t+1} = \bar{x}_t \cdot P_t .$$

The Markov Chain is called a homogeneous chain, if $P_t = P$ for all t.

If $P_t = P(\bar{z}_t)$, where $\bar{z}_t$ is a vector of observed factors, we have particular non-homogeneous Markov Chains

$$\bar{x}_{t+1} = \bar{x}_t \cdot P(\bar{z}_t) .$$

In this paper we use Markov Chain for the purpose of describing changes in the structure of the household expenditures. The structure is understood as a vector

$$\bar{x}_t = \left(x_{t1}, x_{t2}, \cdots, x_{tr}\right),$$

where x_{ti} denotes the share of expenditures spent in time t for the i-group of commodities and services in total expenditures.

The vector $\bar{x}_t$ satisfies the conditions

$$\underset{t}{\exists}\underset{i}{\exists}\, x_{ti} \geq 0 \quad \text{and} \quad \underset{t}{\exists} \sum_{i=1}^{r} x_{ti} = 1 .$$

In this case, the Markov Chain Model describes a stochastic process of spending a money unit for commodities and services by typical household. There is no clear interpretation of money spending, however this model is very useful in predicting the structure of expenditures.

The Markov Set-Chains are a new proposal consisting in looking for the non-homogeneous transition matrix with the use of an interval of matrices. Hartfiel (1999) introduced the models based on the Markov-Set Chains concept. The above mentioned theory is grounded on an assumption that the transition matrix in each step comes from the set of matrices defined by the intervals giving the bounds in which elements of matrices can be changed. The initial structure is defined as the vector interval.

The set

$$[\bar{\mathbf{p}}, \bar{\mathbf{q}}] = \{\bar{\mathbf{x}} \in R^n ; \bar{\mathbf{p}} \leq \bar{\mathbf{x}} \leq \bar{\mathbf{q}}\},$$

where

$$\bar{\mathbf{q}} \geq \bar{\mathbf{p}} \geq \bar{\mathbf{0}} \text{ and}$$

$$\sum_{i} x_i = 1 ,$$

is called the interval of stochastic vectors.

The interval of stochastic vector is a convex polytope. In particular, if $\bar{\mathbf{q}} = \bar{\mathbf{p}}$, the vector interval comes down to the point. In the case of vectors from the space R^2, the interval is a segment of line, and in the R^3 space it is a convex polygon with maximum six vertices.

The tight intervals are important in this concept. The interval $[\bar{\mathbf{p}}, \bar{\mathbf{q}}]$ is called tight if all coordinates p_i and q_i are tight. The coordinates of these vectors satisfy following conditions:

$$p_i = \min_{\bar{x} \in [\bar{p}, \bar{q}]} x_i ,$$

$$q_j = \max_{\bar{x} \in [\bar{p}, \bar{q}]} x_j .$$

Similarly the matrix intervals are defined.

Set of matrix

$$[\mathbf{P},\mathbf{Q}] = \{\mathbf{A}; \mathbf{P} \le \mathbf{A} \le \mathbf{Q}\},$$

where $\mathbf{A} = \lfloor a_{ij} \rfloor$ is a stochastic matrix, $\mathbf{P} = \lfloor p_{ij} \rfloor$ and $\mathbf{Q} = \lfloor q_{ij} \rfloor$ are nonnegative matrices and $\mathbf{Q} \ge \mathbf{P} \ge \mathbf{0}$, is called the matrix interval.

Let $\mathbf{M}$ be the interval of matrices; let $\mathbf{M}^k$ be a set of matrices defined as

$$\mathbf{M}^k = \{\mathbf{A};\ \mathbf{A} = \mathbf{A}_1 \cdot \mathbf{A}_2 \cdot \ldots \cdot \mathbf{A}_k\},\ \text{where}\ \underset{i}{\forall}\ \mathbf{A}_i \in \mathbf{M}.$$

The sequence

$$\mathbf{M}, \mathbf{M}^2, \mathbf{M}^3, \ldots, \mathbf{M}^k, \ldots$$

is called the Markov Set-Chain.

The non-homogeneous Markov Chain can be described by the set $\mathbf{M}$ and the initial interval of stochastic vectors $\mathbf{S}_0$.

We call $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \ldots$ a distribution set and

$$\mathbf{S}_k = \mathbf{S}_0 \cdot \mathbf{M}^k = \{\overline{x} : \overline{x} = \overline{y} \cdot \mathbf{A}, \quad \text{where} \quad \overline{y} \in \mathbf{S}_0 \quad \text{and} \quad \mathbf{A} \in \mathbf{M}^k\}.$$

The vertices of the convex polytopes $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \ldots$ can be computed be the use of[1]:

- the vertices method,
- the Hi-Lo method,
- the Monte Carlo method.

A practical point of the discussed issue consists in defining the matrix set $\mathbf{M}$ in a form of a matrix interval $[\mathbf{P}, \mathbf{Q}]$ and a set of stochastic vectors $\mathbf{S}_0$. A compact set $\mathbf{S}_0$ can be determined as a vector of expenditure structure characterizing a particular household. This will enable to predict this structure in successive years as a vector interval.

In the analysis of the household expenditure structure four following spending groups have been taken into account:

1) Food expenditures.
2) Non-food expenditures (incl.: clothes and shoes, accommodation, hygiene articles).
3) Service expenditures (incl.: transport and communication, recreation and entertainment, education, health).
4) Other expenditures.

[1] The vertices method and the Hi-Lo method are proposed by Hartfiel (1999), the Monte-Carlo method is proposed by Samuels (2001).

The structure of household expenditures depends on many different factors. Basing on the household data the structure of household expenditures is analysed with regard to social-economic groups, income groups and the number of family members.

In the nineties of 20th century significant changes in the structure of household expenditures took place. Table 1 contains data concerning the general structure of household expenditures divided into 4 groups. In the analysed period 1993–2005 the share of food spending was the biggest, however, systematically decreasing. The reason of such behaviour lays in increasing prices of basic products and the commercialisation of health service and education related services accompanied by a slow-down of food prices increase in recent years. As a result of increasing usual living costs related to accommodation including rent, heating and electricity, a constant slow increase of non-food expenditures share in total household spending was notified. The share of service expenditures (such as health, transport and communication) also increased.

Table 1. The expenditure structure of households in Poland in 1993–2005 (in %)

Year	Food expenditures	Non-food expenditures	Service expenditures	Other expenditures
1993	41.50	29.90	19.24	9.36
1994	39.86	30.50	19.85	9.80
1995	39.71	30.75	19.60	9.95
1996	37.80	31.39	20.68	10.13
1997	36.04	32.13	21.09	10.75
1998	31.53	32.11	22.06	14.30
1999	29.02	33.20	24.46	13.31
2000	28.75	31.77	26.01	13.46
2001	28.89	31.57	25.60	13.94
2002	27.49	32.73	25.63	14.15
2003	26.24	33.36	26.32	14.09
2004	26.11	32.67	27.08	14.14
2005	27.80	33.00	27.20	12.00

In the following part of the analysis, the observations on households divided into workers' households, workers' households exploiting farms, farmers' households, pensioners' households, households of self-employed have been used. The observed structures concern the above mentioned groups in total and also with division into the size of household. The analysis considers households starting with one person until to six and more persons.

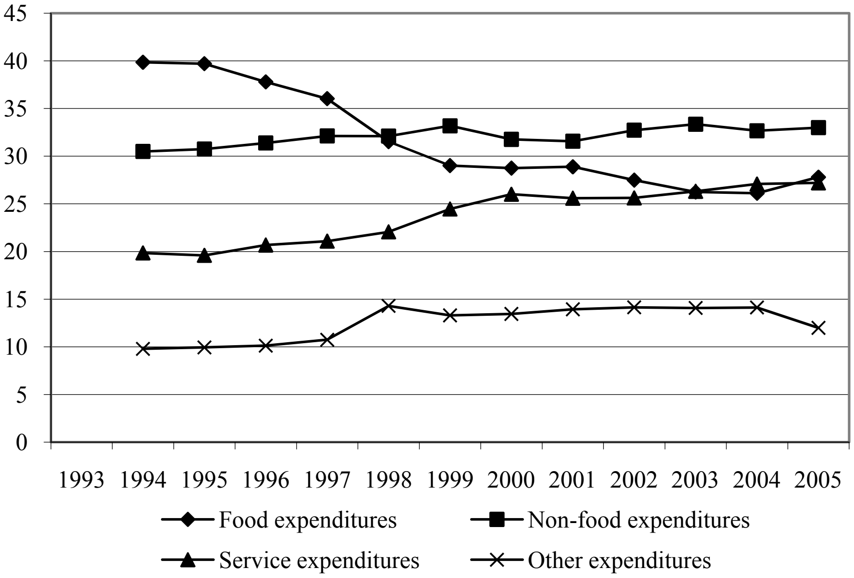

Figure 1. The expenditure structure of households in Poland in 1993-2005 (expenditure share in %)

The stochastic data are available in a form of a matrix $X_g = \left[x_{ti}^g\right]_{13\times 4}$, which presents the expenditure structure of individual social-economic groups. Using the method of transition matrix estimation with a criterion of absolute deviation[2] for aggregated data the transition matrices for individual groups have been given in the form of $\mathbf{P}_{\mathrm{g}} = \left[p_{ij}^g\right]_{4\times 4}$.

Matrices $\mathbf{P}$ and $\mathbf{Q}$ determine an interval $[\mathbf{P}, \mathbf{Q}]$ and can be defined as follows:

$$p_{ij} = \min_g \left\{p_{ij}^g\right\},$$

$$q_{ij} = \max_g \left\{p_{ij}^g\right\}.$$

In a similar way a vector interval $\mathbf{S}_0$ for the last observation period has been defined.

Both the vector interval $\mathbf{S}_0$ and the matrix interval $[\mathbf{P}, \mathbf{Q}]$ have been subjected to a flexing procedure. Forecasts for two successive years have been set by the Hi-Lo method. Besides, the vector interval as the matrix interval and as the matrix intervals being the forecasts of the expenditure structure have been pre-

[2] See: Lee, Judge and Zelner (1970).

sented below. The forecasts have to be referred to any household which can change the number of persons or income source.

The bottom matrix for the analysed structure changes has the following form:

$$P = \begin{bmatrix} 0.7924 & 0.0036 & 0.0130 & 0.0000 \\ 0.0000 & 0.7166 & 0.0037 & 0.0000 \\ 0.0000 & 0.0000 & 0{,}6363 & 0.0000 \\ 0.0211 & 0.0059 & 0.0038 & 0.5807 \end{bmatrix}$$

And the upper matrix has a form of

$$Q = \begin{bmatrix} 0.9562 & 0.1163 & 0.0913 & 0.0000 \\ 0.0302 & 0.9837 & 0.2034 & 0.1309 \\ 0.1127 & 0.2578 & 0.9753 & 0.1058 \\ 0.2079 & 0.2233 & 0.1676 & 0.9579 \end{bmatrix}$$

The vector interval set for the last observation year as minimum and maximum shares of expenditures in the analysed product groups, calculated for all social-economic groups, has the following form:

$$S_T = [\overline{\mathbf{p}}, \overline{\mathbf{q}}] = [(18.91 \quad 27.38 \quad 22.38 \quad 6.26) \quad (40.68 \quad 38.06 \quad 35.87 \quad 11.29)]$$

This interval is presented in Figure 2.

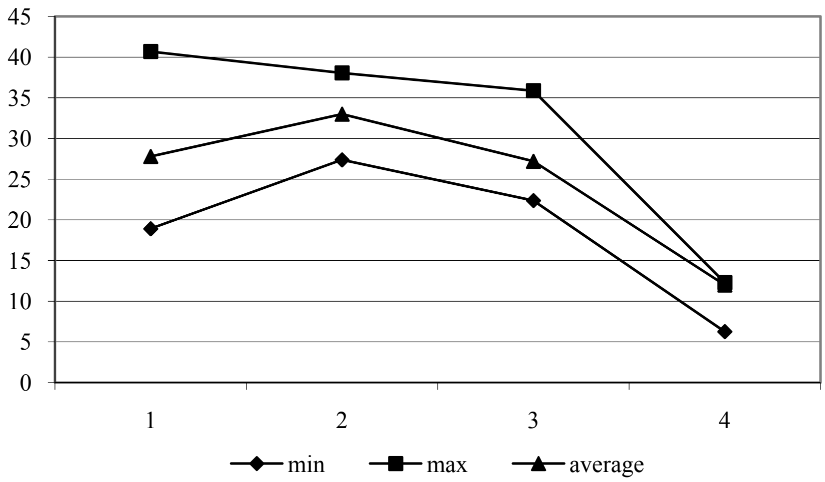

Figure 2. The vector interval of household expenditures in Poland in 2005

An average vector being a structure for all households has been taken as a basis for the forecast. The forecast for 2006 was set with the use of the simulation method. At first a stochastic matrix P_s from matrix interval $[\mathbf{P}, \mathbf{Q}]$ was generated, and then the forecast from the following formula has been calculated:

$$\bar{x}^p{}_{2006} = \bar{x}_{2005} \cdot P_s$$

From the set of forecasts the vector interval was set using minimal and maximal elements:

$$S_{T+1} = [\bar{\mathbf{p}}, \bar{\mathbf{q}}]^p = [(25.05 \quad 32.89 \quad 26.39 \quad 11.56) \quad (28.65 \quad 33.58 \quad 28.71 \quad 12.96)]$$

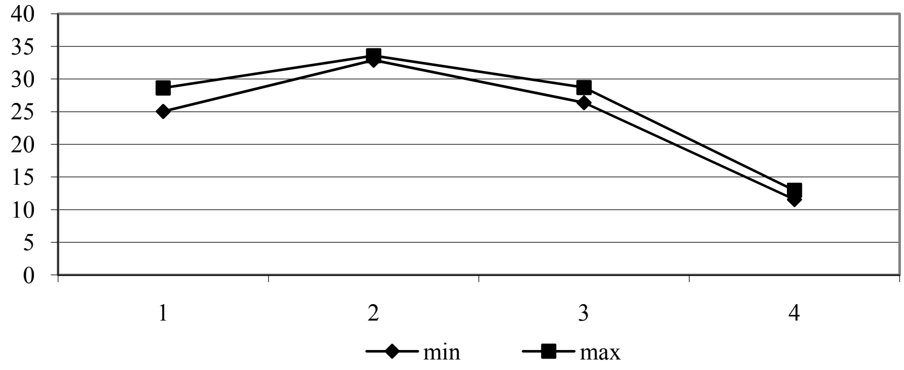

Figure 3. The forecast interval for household expenditure structure in 2006

The attention should be paid to the significant share of food spending, which is equal to 3.6, and to the fact that the share of non-food expenditures is very stabile. Large volatility characterizes service expenditures (2.32). The obtained forecasts also confirm the general trend in changes of expenditure structure, such as declining the food share and the increase of the share of non-food products and services in total expenditure structure.

References

Hartfiel, D.J. (1999), *Markov Set-Chains*, Springer, New York.

Lee, T. C., Judge, G. G., Zelner, A. (1970), *Estimating the Parameters of Markov Probability Model from Aggregate Time Series Date*, Nort-Holland Publ. Co., Amsterdam.

Samuels, C. L. (2001), *Markov Set-Chains as Models of Plant Succession*, University of Tennessee, Knoxville.

Stawicki, J. (2002), *Łańcuchy Markowa w analizie rynku kapitałowego*, Toruń, Wydawnictwo UMK.