

Iwona Kasprzyk

Akademia Ekonomiczna w Katowicach

WYKORZYSTANIE UKRYTYCH MODELI MARKOWA DO KLASYFIKACJI SPÓŁEK GIEŁDOWYCH

1. Wstęp

Modele klas ukrytych wprowadzone przez Lazarsfelda [Lazarsfeld, Henry 1968] pozwalają na analizę danych mierzonych na najniższych skalach pomiaru. W przypadku analizy tego typu modeli wymagane jest, aby dane zostały przedstawione za pomocą tablicy kontyngencji.

Ukryte modele Markowa (*latent Markov model*) zostały zaproponowane przez Wigginsa [1973] i znalazły swoje zastosowanie przede wszystkim w psychologii oraz socjologii. Modele Markowa zostały opisane przede wszystkim przez Langeheine'a [1994; 2000] oraz Vermunta [Vermunt, Langeheine, Böckenholt 1999].

Celem prezentowanego artykułu jest zastosowanie ukrytych modeli Markowa do analizy wybranych danych finansowych spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie. Tego typu modele pozwolą na klasyfikację badanych spółek oraz na analizę zmian danego zjawiska na przestrzeni kilku okresów.

2. Modele klas ukrytych

Założmy, że mamy daną tablicę kontyngencji z trzema zmiennymi obserwowalnymi: A ($i = 1, 2, 3, \dots, I$), B ($j = 1, 2, \dots, J$) oraz C ($k = 1, 2, \dots, K$). Zmienna ukryta X przyjmuje wartości $w = 1, 2, \dots, W$, gdzie W oznacza liczbę klas. Model z jedną zmienną ukrytą X można przedstawić za pomocą poniższego równania:

$$\pi_{ijk} = \sum_{w=1}^W \pi_{ijkw}^{ABCX}, \quad (1)$$

gdzie π_{ijkw}^{ABCX} oznacza prawdopodobieństwo warunkowe tego, że i -ta, j -ta, k -ta kategoria zmiennej A, B oraz C znajdzie się w klasie ukrytej w .

Wykorzystanie wzoru (1) wymaga spełnienia założenia o lokalnej niezależności zmiennych:

$$\pi_{ijkw}^{ABCX} = \pi_w^X \pi_{iw}^{A \setminus X} \pi_{jw}^{B \setminus X} \pi_{kw}^{C \setminus X}, \quad (2)$$

gdzie π_w^X oznacza prawdopodobieństwo tego, że dla danej obserwacji zmienna ukryta X przyjmie wartość t , zaś $\pi_{iw}^{A \setminus X}$ – prawdopodobieństwo warunkowe tego, że i -ta kategoria zmiennej A znajdzie się w klasie ukrytej t .

Prawdopodobieństwa po prawej stronie równania (2) wymagają spełnienia odpowiedniego założenia:

$$\sum_{w=1}^W \pi_w^X = \sum_{i=1}^I \pi_{iw}^{A \setminus X} = \sum_{j=1}^J \pi_{jw}^{B \setminus X} = \sum_{k=1}^K \pi_{kw}^{C \setminus X} = 1. \quad (3)$$

3. Prosty model Markowa

Prosty model Markowa (*simple Markov model*) [Bishop, Fienberg, Holland 1975], gdy pod uwagę brane są trzy okresy, można zapisać zgodnie ze wzorem:

$$P_{ijk} = \delta_i^1 \tau_{ji}^{2 \setminus 1} \tau_{kj}^{3 \setminus 2}, \quad (4)$$

gdzie δ_i^1 – obserwowany początkowy rozkład prawdopodobieństwa w pierwszym analizowanym okresie, $\tau_{ji}^{2 \setminus 1}, \tau_{kj}^{3 \setminus 2}$ – prawdopodobieństwa przejścia ze stanu w punkcie 1 do stanu w punkcie 2 oraz ze stanu w punkcie 2 do stanu w punkcie 3.

4. Ukryty model Markowa

W wyniku połączenia modelu oznaczonego równaniem (2) oraz prostego modelu Markowa danego równaniem (4) otrzymujemy ukryty model Markowa (*latent Markov model*), który można zapisać za pomocą wzoru:

$$P_{ijk} = \sum_{a=1}^A \sum_{b=1}^B \sum_{c=1}^C \delta_{w_1}^1 \rho_{i \setminus w_1}^1 \tau_{w_2 \setminus w_1}^{2 \setminus 1} \rho_{j \setminus w_2}^2 \tau_{w_3 \setminus w_2}^{3 \setminus 2} \rho_{k \setminus w_3}^3, \quad (5)$$

gdzie: $\delta_{w_1}^1$ – początkowe prawdopodobieństwo ukrytej klasy w_1 dla pierwszego okresu, $\tau_{w_2 \setminus w_1}^{2 \setminus 1}, \tau_{w_3 \setminus w_2}^{3 \setminus 2}$ – ukryte prawdopodobieństwa przejścia, tj. prawdopodobieństwo przejścia z ukrytej klasy w_1 w pierwszym okresie do klasy w_2 w drugim oraz z ukrytej klasy w_2 w drugim okresie do w_3 w trzecim okresie, $\rho_{i \setminus w_1}^1, \rho_{j \setminus w_2}^2$ – prawdopodobieństwo warunkowe znalezienia się kategorii zmiennej i, j w klasie ukrytej w_1, w_2 .

Podobnie jak modele klas ukrytych, tak i ukryty model Markowa (5) wymaga spełnienia następującego ograniczenia:

$$\sum_{w_t=1}^W \delta_{w_t}^t = 1; \sum_i \rho_{i/w_t}^t = 1; \sum_{w_{t+1}} \tau_{w_{t+1}/w_t}^{t+1} = 1. \quad (6)$$

5. Założenia ukrytych modeli Markowa

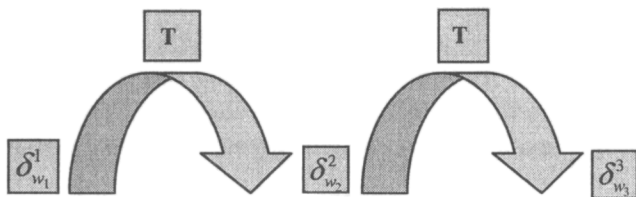
Niech $\mathbf{R}$ oznacza macierz zawierającą prawdopodobieństwa warunkowe danych kategorii zmiennych w klasie ukrytej w_t oraz $\mathbf{T}$ oznacza macierz prawdopodobieństw przejścia. Gdy macierze $\mathbf{R}$ oraz $\mathbf{T}$ są jednorodne w czasie, tzn. macierze te dla poszczególnych okresów są sobie równe:

$$\begin{aligned} \mathbf{R}^1 &= \mathbf{R}^2 = \mathbf{R}^3 = \mathbf{R} \\ \mathbf{T}^{21} &= \mathbf{T}^{32} = \mathbf{T}, \end{aligned} \quad (7)$$

mówimy o tzw. jednorodnym ukrytym modelu Markowa. Gdy założenie oznaczone równaniem (7) nie jest spełnione, to taki model nazywamy niejednorodnym ukrytym modelem Markowa.

W przypadku jednorodnego modelu Markowa rozkłady prawdopodobieństw klas ukrytych dla poszczególnych okresów obliczamy poprzez przemnożenie prawdopodobieństwa klas ukrytych z poprzedniego okresu przez macierz prawdopodobieństwa przejścia, tj.:

$$\delta_{(w_t)}^{(t)} = \delta_{(w_{t-1})}^{(t-1)} \cdot \mathbf{T}. \quad (8)$$



Rys. 1. Rozkład prawdopodobieństw klas ukrytych dla trzech okresów

Źródło: opracowanie własne.

Na rysunku 1 pokazano dodatkowo schemat rozkładu klas ukrytych.

6. Miary jakości dopasowania modelu

W analizie modeli dla tablic kontyngencji można posłużyć się różnymi statystykami pozwalającymi na wybór modelu najlepiej dopasowanego do danych. Do najbardziej popularnych można zaliczyć statystykę chi-kwadrat (χ^2), statystykę największej wiarygodności (L^2) oraz Cresie-Reada (CR):

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^l \frac{(n_{ijk} - \hat{n}_{ijk})^2}{\hat{n}_{ijk}}, \quad (9)$$

$$L^2 = 2 \left(\sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^l n_{ijk} \log \left(\frac{n_{ijk}}{\hat{n}_{ijk}} \right) \right), \quad (10)$$

$$CR = \frac{2}{\lambda(\lambda+1)} \sum_{ijk} n_{ijk} \left[\left(\frac{n_{ijk}}{\hat{n}_{ijk}} \right)^\lambda - 1 \right]. \quad (11)$$

W przypadku statystyki Cressie-Reada przyjmuje się, że $\lambda = \frac{2}{3}$. Gdy $\lambda = 1$, to statystyka CR sprowadza się do statystyki χ^2 , a gdy $\lambda \rightarrow 0$, to statystyka CR równoważna jest statystyce L^2 .

Dodatkowymi miarami mającymi na celu ocenę dopasowania danego modelu są kryteria informacji, jak np. kryterium Akaikego (AIC) oraz kryterium BIC:

$$AIC = L^2 - 2 \cdot (\text{liczba stopni swobody}), \quad (12)$$

$$BIC = L^2 - \ln(n) \cdot (\text{liczba stopni swobody}). \quad (13)$$

Za pomocą tych kryteriów wybierany jest ten model, dla którego danego kryterium przyjmuje wartość najmniejszą.

7. Zastosowanie

Wykorzystanie modeli Markowa pokazano na przykładzie analizy spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie od I kwartału 2005 r. do I kwartału 2006 r. Postanowiono poddać badaniu wskaźnik zysku przypadającego na jedną akcję (EPS). Z analizy wyeliminowano te spółki, dla których przynajmniej w jednym kwartale nie było danych o tym wskaźniku ekonomiczno-finansowym. W efekcie analiza została przeprowadzona w odniesieniu do 213 spółek.

Wskaźnik ten jest jednym z najważniejszych wskaźników dotyczących potencjalnego inwestora inwestującego w akcje danej spółki na giełdzie. Aby możliwe było przedstawienie modeli Markowa, dla każdego okresu dokonano dychotomizacji wskaźnika EPS, przyjmując za „0” – wartość wskaźnika na jedną akcję mniejszą bądź równą 0 oraz „1” – wartość wskaźnika EPS większą od zera.

W ocenie dopasowania modelu pod uwagę wzięto prosty model Markowa, jednorodny ukryty model Markowa oraz niejednorodny model Markowa. Jak pokazują wartości statystyki χ^2 , L^2 czy też CR, najlepiej dopasowanym modelem jest jednorodny model Markowa. Dodatkowo kryteria informacyjne AIC oraz BIC wskazują, że ten model jest dobrze dopasowany.

Tabela 1. Ocena jakości dopasowania modelu

Model	df	χ^2	p	L^2	p	CR	p	AIC	BIC
Prosty model Markowa	28	110,748	0,0000	83,202	0,0000	96,801	0,0000	27,202	-66,782
Jednorodny ukryty model Markowa	26	37,800	0,0632	38,918	0,0496	36,907	0,0762	-13,082	-100,353
Niejednorodny ukryty model Markowa	20	32,441	0,0388	33,081	0,0331	31,630	0,0474	-6,919	-74,051

Źródło: opracowanie własne.

W tabeli 2 przedstawiono szczegółową analizę niejednorodnego modelu Markowa. Można zauważyć, że dla $t = 1$ (I kwartał 2005 r.) większość spółek (61,48%) notowanych na giełdzie warszawskiej osiągnęło wskaźnik zysku na jedną akcję większy od zera i znalazło się w klasie 1. Klasę tę charakteryzuje bardzo duża liczba spółek (93,72%) osiągających dodatnią wartość wskaźnika EPS, podczas gdy w klasie 2 można zaobserwować, że wśród 64,93% spółek notowanych na giełdzie w Warszawie wartość wskaźnika EPS była mniejsza od zera bądź równa zero.

Macierz przejścia z t do $t+1$ pokazuje, że prawdopodobieństwo przejścia z klasy 1 do klasy 2 wynosi 3,85%. W tabeli 2 znajdują się jeszcze inne macierze przejścia, które w sposób bardziej wyraźny obrazują to, co wydarzy się w wyniku przejścia z pierwszego analizowanego okresu do np. ostatniego: można zauważyć, że w wyniku przejścia z I kwartału 2005 r. do I kwartału 2006 r. wśród 86,75% spółek można było odnotować to, że wartość wskaźnika EPS była większa od zera, podczas gdy blisko 21% spółek miało problem z osiągnięciem dodatniego wskaźnika zysku na jedną akcję w ostatnim okresie.

Tabela 2. Estymowane parametry dla jednorodnego modelu Markowa

Rozkłady prawdopodobieństw klas ukrytych (δ)							
Ukryta kategoria 1		0,6148					
Ukryta kategoria 2							
0,3852							
Macierze prawdopodobieństw przejścia (T)							
Macierz prawdop. przejścia z t do $t+1$ klas		Z $t=1$ do $t=3$		Z $t=1$ do $t=4$		Z $t=1$ do $t=5$	
0,9615	0,0385	0,9268	0,0730	0,8956	0,1044	0,8675	0,1325
0,0609	0,9391	0,1157	0,8842	0,1651	0,8349	0,2096	0,7904
Macierz prawdopodobieństw warunkowych (R)							
Kategoria 1	Kategoria 2						
0,9372	0,0628						
0,3507	0,6493						

Źródło: opracowanie własne z wykorzystaniem programu LEM.

Jeśli stosuje się równanie (8), to istnieje możliwość obliczenia poszczególnych rozkładów prawdopodobieństw klas ukrytych dla wszystkich analizowanych okresów, tj. od I kwartału 2005 r. do I kwartału 2006 r. Dla $t = 2, \dots, 5$ wynoszą one: $\delta_{w_2}^2 = (0,6060 \ 0,3392)$, $\delta_{w_3}^3 = (0,5908 \ 0,4118)$, $\delta_{w_4}^4 = (0,5908 \ 0,4231)$, $\delta_{w_5}^5 = (0,5843 \ 0,4333)$. Można zauważyć, że prawdopodobieństwo dla drugiej kategorii wzrasta z

0,3852 (dla $t = 1$) do 0,4333 (dla $t = 5$), czyli w ostatnim okresie 43,33% spółek notowanych na giełdzie osiągało wartość wskaźnika na jedną akcję mniejszą od zera lub równą zeru.

8. Podsumowanie

Modele Markowa niewątpliwie są bardzo przydatnym narzędziem stosowanym do badania zmian zachodzących w danej zmiennej jakościowej na przestrzeni pewnego czasu.

Modele te stały się przede wszystkim bardzo popularne w naukach społecznych. W artykule została podjęta próba zastosowania ukrytych modeli Markowa do analizy finansowej, w szczególności wskaźnika zysku na jedną akcję spółek notowanych na Giełdzie Papierów Wartościowych w czasie pięciu kwartałów, tj. od I kwartału 2005 r. do I kwartału 2006 r.

Za pomocą obliczenia tzw. macierzy prawdopodobieństw przejścia można było zauważyć, że z kwartału na kwartał zmniejszał się odsetek spółek osiągających dodatnią wartość wskaźnika EPS. Zastosowanie ukrytych modeli Markowa pozwoliło na zobrazowanie ogólnej sytuacji finansowej spółek notowanych na giełdzie.

Literatura

- Bijleveld C.C.J.H. (2003), *Latent Markov Modelling of Recidivism Data*, „Statistica Neerlandica” vol. 57, nr 3, s. 305-320.
- Hagenaars J.A. (1990), *Categorical Longitudinal Data: Log-Linear Panel, Trend, and Cohort Analysis*, Newbury Park, Ca: Sage.
- Langeheine R. (1994), *Latent Variables Markov Models*, [w:] A. von Eye, C.C. Clogg (red.), *Latent Variables Analysis: Applications for Developmental Research*, Sage Publications.
- Langeheine R., van de Pol F. (2000), *Latent Markov Chains*, [w:] J.A. Hagenaars, A.L. McCutcheon (red.), *Applied Latent Class Analysis*, Cambridge University Press, Cambridge.
- Lazarsfeld P.F., Henry N.W. (1968), *Latent Structure Analysis*, Houghton Mill, Boston.
- Read T.R.C., Cressie, N.A.C. (1988), *Goodness-of-Fit Statistics for Discrete Multivariate Data*, Springer, New York.
- Reinecke J. (1997), *Testing the Theory of Planned Behavior with Latent Markov Models*, [w:] J. Rost, R. Langeheine (red.), *Applications of Latent Trait and Latent Class Models in the Social Sciences*, Waxmann Verlag.
- Vermunt J.K. (1997), *LEM: A General Program for the Analysis of Categorical Data*, Tilburg University, Tilburg.

- Vermunt J.K., Langeheine R., Böckenholt U. (1999), *Discrete-Time Discrete-State Latent Markov Models with Time-Constant and Time-Varying Covariates*, „Journal of Educational and Behavioral Statistics” nr 24, s. 178-205.
- Wiggins L.M. (1973), *Panel Analysis: Latent Probabilistic Models for Attitude and Behavior Processes*, Elsevier Scientific Publishing Company, Amsterdam.

APPLICATION OF LATENT MARKOV MODEL TO COMPANIES STOCK CLASSIFICATION

Summary

The main purpose of this paper is to show the latent Markov models, which we can use to financial analysis. This article shows the earning per share ratio analysis as one of the most important ratios for the potential investment on stock exchange.

This kind of models has been proposed by Wiggins [1973] and they are combining of the latent class models and simple Markov model. These models can be used to describe individual change in categorical variables.