TOMASZ SKĄPSKI

METODY ESTYMACJI ROZKŁADÓW OPÓŹNIEŃ CZASOWYCH

Z odmienności form równań regresji z opóźnionymi zmiennymi objaśniającymi wynikają różnice w ich zastosowaniu i interpretacji. Równania te mogą zawierać po kilka kolejnych opóźnień i wówczas obrazują procesy stopniowego wygasania wpływu pewnego zjawiska na zjawisko objaśniane w równaniu. Często wprowadza się do równania zmienną z jednym tylko opóźnieniem. Wpływ tej zmiennej na zjawisko objaśniane wynika z przeciętnego upływu czasu od powzięcia pewnej decyzji do jej zrealizowania lub też upływu czasu od powzięcia decyzji do pojawienia się jej skutku. Te średnie opóźnienia czasowe są najczęściej spotykane w analizach ekonomicznych.

Nasuwa się naturalne zagadnienie wyboru jednej z form stosowania opóźnionych zmiennych. Jak zobaczymy niżej nie zawsze włączenie do równania pełnego rozkładu opóźnień prowadzi do najlepszych rezultatów. Najwygodniejsze i najbardziej efektywne, szczególnie przy wielu zmiennych opóźnionych, jest użycie opóźnień średnich.

W zależności od faktu, czy opóźnionymi zmiennymi są zmienne endo- czy egzogeniczne, mamy do czynienia z różnymi rodzajami równań ekonometrycznych. Gdy w równaniu zmienna endogeniczna Y_t jest funkcją wartości tej samej zmiennej z okresów wcześniejszych, tzn. jest funkcją o postaci

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-L})$$

równanie takie zwane jest w literaturze równaniem autoregresyjnym. W przypadku, gdy funkcja przybiera inną postać

$$y_t = f(x_t, x_{t-1}, \dots, x_{t-K})$$

mamy do czynienia z równaniem o rozłożonych opóźnieniach zmiennej objaśniającej.

Są to dwie podstawowe postacie funkcji dynamicznych w badaniach ekonomicznych, choć w praktyce często spotykamy się z mieszanymi formami modeli, zawierającymi opóźnienia kilku zmiennych endo- i egzogenicznych

$$y_t = f(y_{t-1}, \dots, y_{t-L}, x_{it}, x_{it-1}, \dots, x_{it-K})$$

Jakkolwiek równania ze zmiennymi opóźnionymi posiadają najczęściej określony sens ekonomiczny i znajdują szerokie zastosowanie w analizach układów pewnych zjawisk jak również dla celów predykcji, to jednak przy ich adaptacji wyłania się wiele problemów natury metodologicznej. Ze względu na to, że problemy te w większości nie są rozstrzygnięte jednoznacznie w literaturze, konieczne jest ich wyraźne zasygnalizowanie w celu uniknięcia nieporozumień w interpretacji uzyskanych rezultatów.

Pierwszym z problemów jest zagadnienie agregacji czasowej w modelach ze zmiennymi opóźnionymi. Należy zwrócić uwagę na fakt, że struktura wzajemnych związków między elementami badanego układu w ujęciu dynamicznym staje się coraz bardziej skomplikowana w miarę jak układ w swoim zachowaniu jest uzależniony od coraz większej liczby opóźnień czasowych. Zależy to w dużej mierze od rozpiętości przedziałów czasowych przyjętych za podstawę analizy. Z drugiej strony, oczywiście, o ile układ w przedziałach rocznych mógł nie wykazywać wcale reakcji opóźnionych, o tyle przy przechodzeniu do mniejszych jednostek czasowych może on wykazywać liczne opóźnienia.

Stosowanie opóźnień czasowych w modelach prostych, jednorównaniowych daje najczęściej jednoznaczną interpretację ekonomicznych zależności. Modele rekurencyjne przedstawiają pod tym względem układy o znacznie bardziej złożonej strukturze wewnętrznej. Oczywiście rekurencyjny model wielorównaniowy nie nastręcza nowych problemów typu statystycznego w porównaniu z modelami jednorównaniowymi. Jeżeli równania zawierają tylko pojedyncze zmienne o rozłożonych opóźnieniach wypływają stąd jedynie trudności w interpretacji ekonomicznej takiego modelu. Jeżeli, mianowicie, na daną zmienną objaśnianą działa zmienna objaśniająca z określonym opóźnieniem, to efekt tego oddziaływania rozprzestrzeni się od razu na dalsze zmienne objaśniane w równaniu, lecz tylko te, które stoją w kolejności jednoczesnych oddziaływań w powstałym, wskutek uszeregowania równań, łańcuchu sprzężeń w stosunku do wyróżnionej zmiennej. Ma to istotne znaczenie z punktu widzenia cybernetycznych modeli sprzężeń analizujących układ ekonomiczny pod kątem możliwości jego regulacji i sterowania.

Inny problem występuje w sytuacji, gdy zmienna objaśniana jest funkcją kombinacji więcej niż jednej zmiennej objaśniającej lub gdy rozpatrujemy zagadnienie jednoczesnego szacunku parametrów przy opóźnionych zmiennych w modelach równań współzależnych. Proponowane metody odnośnie do pierwszego przypadku angażują bardzo zawiły aparat matematyczny i opierając się na wielu założeniach nie dają w pełni efektywnych rezultatów[1]. Przypadek szacowania rozkładów opóźnień czasowych w modelach równań jednoczesnych nie znalazł dotąd zadowalają-

[1] Metody takie podaje R. J. Hannan w pracy pt. *Time Series Analysis,* Londoni 1960, R. 3 pn. Regression for Time Series.

cego rozwiązania w literaturze. Przypuszczać należy, że podstawową rolę w tym wypadku odgrywają zagadnienia identyfikacji.

Kolejnym problemem, który należy zasygnalizować w związku z wprowadzeniem zmiennych opóźnionych do równania regresji, to problem wyboru metod estymacji parametrów przy tych zmiennych. Przy małych próbach stosowanie metody najmniejszych kwadratów w postaci klasycznej prowadzi do wzrostu błędów szacunku parametrów ze względu na osłabienie założeń o nielosowości zmiennych objaśniających, lecz nawet, jak wykazał Koyck [2], w dużych próbach metoda ta daje obciążone szacunki parametrów. Błąd ten spowodowany jest tym, że równanie ze zmienną opóźnioną zawiera składnik losowy z okresu (t-k), który może nie być niezależny od opóźnionej o ten sam okres zmiennej objaśniającej. W dodatku składnik losowy tego równania wykazywać może autokorelację w czasie. Poza tym występuje często współliniowość opóźnionych zmiennych objaśniających, która daje duży stopień niepewności co do zgodności i nieobciążoności estymatorów.

Na kanwie tych właśnie zastrzeżeń rozwinęło się szereg metod szacunku parametrów przy zmiennych opóźnionych przy czym chronologicznie pierwszą była metoda dwustopniowego szacowania parametrów zaproponowana przez L. Koycka w cytowanej już pracy. Kolejne próby szły w kierunku uproszczenia metody Koycka lub też przyjmowania a priori niewystępowania wymienionych zastrzeżeń i szacowania ex post autokorelacji składnika losowego i stopnia skorelowania jego ze zmiennymi objaśniającymi. Jedną z tych metod jest metoda L. R. Kleina, w której autor przyjmując niezależność składników losowych kilku równań modelu proponuje estymację parametrów równań ze zmiennymi opóźnionymi przy pomocy metody największej wiarygodności z układu równań współzależnych[8]. Oryginalną metodę przy zastosowaniu twierdzenia Beyesa dla kolejnych wariantów odnośnie do założeń co do kształtowania się składnika losowego podali A. Zellner i M. S. Geisel[4].

Wszystkie teoretyczne rozważania dążące do statystycznego określenia opóźnień czasowych ujmują ten problem jako badanie wpływu ważonej sumy wszystkich opóźnionych wartości zmiennej niezależnej ná zmienną objaśnianą. Rozważa się więc model o postaci:

$$y_t = \sum_{i=0} w_i x_{t-i}$$

lub bardziej dokładnie

$$y_t = a \sum_{i=0} w_i x_{t-i} \, .$$

[2] L. Kóyck, *Distributed Lags and Investment Analysis*$_y$ Amsterdam 1954.

[3] L. R. Klein, *The Estimation of Distributed Lags,* Econometrica, 1958, ss. 553 - - 565.

[4] A. Zellner, M. S. Geisel, *Analysis of Distributed Lag Models with Application to Consumption Function Estimations,* Econometrica, 1970, nr 4.

Jest oczywiste, że bezpośrednia estymacja parametrów a i w_i przy użyciu danych z szeregów y i x jest niemożliwa bez uprzedniego przyjęcia określonych założeń co do rozkładu wag w_i w czasie. Pierwsze próby określenia wzoru zachowania się w_i nastąpiły w pracach I. Fishera [5]. Jako naturalne ograniczenie przyjął on fakt, że wagi w_i maleją przy przechodzeniu do coraz to dalszych okresów czasu. W konsekwencji tego założenia przyjmował on malejący postęp arytmetyczny za rozkład współczynników przy kolejnych opóźnieniach zmiennej niezależnej. Jego teorię przypomniał i rozwinął L. Koyck [6], zakładając że

$$w_i = \lambda^k, \quad \text{gdzie} \quad 0 < \lambda < 1$$

tzn., że wagi w_i przy kolejnych opóźnieniach zmiennej objaśniającej tworzą malejący w czasie postęp geometryczny.

Tego rodzaju założenia przyjmują a priori, że wpływ zmiennej niezależnej jest największy w okresie bieżącym *(t)*, a wraz z uwzględnieniem coraz wcześniejszych okresów czasu wpływ ten jest systematycznie malejący. Formułowanie takiego założenia opiera się więc na postępowaniu, które preferuje informacje pochodzące z późniejszych okresów czasu względem informacji wcześniejszych [7]. Postępowanie takie jest uzasadnione w okolicznościach, gdy zmienna endogeniczna wykazuje natychmiastową reakcję na zmiany zmiennej egzogenicznej, tzn. gdy istnieje rzeczywiste sprzężenie dwóch zjawisk w okresie bieżącym *(t)*. Nie zawsze jednak zjawiska ekonomiczne wpływają na siebie według powyższego modelu. Częste są opóźnione reakcje jednych zjawisk ekonomicznych na zmiany drugich. W tych okolicznościach proponowany monotoniczny rozkład wag przy zmiennych opóźnionych nie znajduje realnej interpretacji. Stąd powstały kolejne modyfikacje modelu L. Koycka, stąd też próby uogólnienia i uelastycznienia teoretycznych konstrukcji rozkładów współczynników przy zmiennych opóźnionych, stąd wreszcie całkowicie nowe próby empirycznego ich ustalenia.

Najprostszym, lecz bardzo kłopotliwym rachunkowo rozwiązaniem jest założenie, że rozkład opóźnień (a raczej rozkład współczynników przy zmiennych opóźnionych) kształtuje się hiperbolicznie przyjmując postać

[5] I. Fisher, *Note on Short-Cut Method for Calculating Distributed Lags,* Bulletin de l'Institut International de Statistique, 1937, nr 3.

[6] L. Koyck, op. cit.

[7] Przypisanie większych wag danym, których dostarczają informacje bardziej „aktualne" jest charakterystyczne dla badań szeregów czasowych. W literaturze polskiej system wag harmonicznych dla kolejnych obserwacji w szeregach czasowych podał Z. Hellwig w artykule pt. *Schemat budowy prognozy statystycznej metodą wag harmonicznych,* Przegląd Statystyczny, 1967, nr 2. Por. też M. Nerlove, S. Wage, *On the Optimality of Adaptive Forecasting,* Management Science 1964. System wag proponowany przez tych autorów wykorzystany został dla określenia rozkładu współczynników opóźnień, w: R. Ch. Sims, *Discrete Approximation of Continuous Time Distributed Lags,* Econometrica, 1971, ss. 646 - 668.

wielomianu określonego stopnia. W ten sposób współczynniki opóźnień zależą od kilku parametrów wielomianu, które szacuje się empirycznie przy użyciu różnych metod. Parametrami tymi są: stopień wielomianu, ilość składników wielomianu oraz współczynniki przy kolejnych składnikach wielomianu.

Innym sposobem jest uogólnienie założenia Koycka o geometrycznym uszeregowaniu współczynników opóźnień. Według tej teorii współczynniki w_i przyjmują rozkład Pascala z dwoma parametrami. Rozkład ten w zależności od wartości tych parametrów prowadzi do różnych modeli opóźnień. Wagi w_i determinowane są więc następującą formułą:

$$w_i = \binom{r+i-1}{i}(1-\lambda)\, r\lambda^i$$

gdzie λ i r są parametrami rozkładu, $0<\lambda<1$, a r jest liczbą całkowitą dodatnią. Średnia rozkładu Pascala jest równa $r\lambda/ (1—\lambda)$ z wariancją $r\lambda/ (1—\lambda)^2$. Wartość modalna jest równa $(r\lambda—1)/(1—\lambda)$. Moda jest zwykle mniejsza niż średnia, więc rozkład jest prawostronnie asymetryczny. Wraz z większym λ i mniejszym r asymetria jest większa. Gdy $\lambda \to 0$ i $r \to m$, rozkład Pascala dąży do rozkładu Poissona ze średnią m. Gdy m jest duże, granicę stanowi rozkład Gaussa. Rozkład geometryczny, należy zauważyć, jest jedną z form rozkładu Pascala dla $r=1$.

Przy braku jakichkolwiek informacji a priori o rozkładzie wag w_i widzimy, że wybór teoretycznej postaci tego rozkładu ma charakter w dużej mierze arbitralny. W literaturze brak jest wiarygodnych wskazówek co do wyboru analitycznej postaci rozkładu. Rozkład Pascala ze względu na posiadanie dwóch parametrów wydaje się być najbardziej elastyczny dla celów estymacji rozkładów opóźnień czasowych w równaniach regresji. Problemem staje się estymacja parametrów l i r. Szczególnie skomplikowana staje się estymacja, gdy nie znamy wcześniej rzędu wielkości r. [8]

Jednakże istnieją metody pozwalające z dostateczną dokładnością (mierzoną m. in. wariancją średniego opóźnienia, konstrukcją jego przedziału ufności lub podaniem określonego prawdopodobieństwa) na ustalenie rozkładu współczynników przy zmiennych opóźnionych. Dużą rolę odgrywają nadal przy stosowaniu tych metod sposoby iteracyjne.

Jeżeli szeregi czasowe zmiennych w modelu charakteryzują się istnieniem bardziej lub mniej prawidłowych oscylacji o charakterze sezonowym bądź cyklicznym, to rozkłady opóźnień oszacowane na podstawie tych szeregów będą rozkładami wielomodalnymi. Pomijając zjawiska sezonowe i ich wpływ na rozkład opóźnień czasowych[9], zwrócimy uwagę

[8] Por. Z. Z. Gritliches, *Distributed Lags. A Survey.* Econometrica 1967.

[9] Z badanego szeregu czasowego należy wyeliminować sezonowość przechodząc do rocznych obserwacji lub uwzględniając w modelu liczbę opóźnianych okresów równą np. cztery przy danych kwartalnych lub 12 przy danych miesięcznych.

na postać rozkładu w przypadku występowania prawidłowości o charakterze cykli gospodarczych.

Wielomodalny rozkład opóźnień otrzymujemy wówczas, gdy analizujemy model

$$y_t = \sum_{i=0}^{L} w_i x_{t-i}$$

gdzie *L=n—l,* a *l* to średnia długość cyklu, przy $n=2$, 3, 4, ..., *N.* Należy zauważyć, że przy cykliczności zjawisk gospodarczych postępowanie takie nosi w sobie błąd wobec zmiennej długości cyklu gospodarczego. Innymi słowy jakkolwiek dobrana długość *L* zawierać w sobie może elementy cykliczności szeregu czasowego badanej zmiennej. Problem doboru „optymalnego" *L* doczekał się opracowania we wcześniejszej literaturze estymacji opóźnień czasowych i w obecnej chwili istnieje szereg metod pozwalających z określonym błędem lub prawdopodobieństwem na estymację „najlepszej" długości *L,* minimalizującej ryzyko włączenia do analizy elementów cyklu gospodarczego[10]. W najnowszej literaturze obserwuje się odchodzenie od bezpośredniej estymacji *L* drogą uogólniania modelu poprzedniego równania na:

$$y_t = \sum_{i=0}^{\infty} w_i x_{t-i}$$

W niektórych przypadkach celowe okazuje się włączenie do badania opóźnionych o *L>21* zmiennych egzogenicznych, a to ze względu na niewygaśnięcie wpływu wcześniejszych wahań cyklicznych (szczególnie przy cyklach krótkich 3 - 4-letnich). Aproksymantą rozkładu opóźnień może być wówczas jedna z funkcji drgań tłumionych lub suma funkcji okresowych zestawionych w szereg Founiera przy pomocy analizy harmonicznej [11].

Istnieje w obecnej chwili stosunkowo bogata bibliografia dotycząca metod estymacji rozkładów opóźnień czasowych. Większość z nich nie była jednak dotąd weryfikowana na materiale empirycznym. Ponieważ wiele z tych metod wymaga poważnego przygotowania matematycznego, to w świetle niepewnych wartości uzyskiwanych rezultatów nie będziemy w tym miejscu zajmowali się niektórymi z nich.

Część z tych propozycji powstałych w stosunkowo niedawnym czasie, prowadzi do zastosowań analizy harmonicznej. Uzyskanie rezultaty (rozkłady) generowane widmową funkcją gęstości interpretowane są jako estymatory empirycznych rozkładów opóźnień zmiennej endogenicznej w czasie[12].

Inną grupę stanowią metody zakładające, że opóźnienia tworzą male-

[10] Por. np. V. K. Chetty, *Estimation of Solow's Distributed Lag Models,* Econometrica 1971, nr 1.

[11] Por. np. J, Hanan, op. cit.

[12] Por. w szczególności R, Ch. Sims, op. cit.

jący postęp geometryczny (a więc opierają się na założeniu Koycka). W grupie tej posługującej się metodami Monte Carlo szczególny nacisk kładzie się na własności składników losowych modelu, tzn. na ich autokorelacji w czasie, stałości wariancji i stopnia skorelowania z kalejnymi opóźnionymi zmiennymi egzogenicznymi[13].

W niniejszym artykule[14] zainteresujemy się bliżej metodami korelacji z opóźnieniem oraz estymacją parametrów rozkładu Pascala. Ponieważ proponowane metody nie były stosowane dotąd w polskiej literaturze, przedstawimy pokrótce ogólne zasady, na których się one opierają i ich praktyczny algorytm.

Metoda wielomianu interpolacyjnego zaproponowana przez S. Almon[15] jest sposobem empirycznego ustalenia rozkładu współczynników przy zmiennych opóźnionych przybierającego postać wielomianu stopnia n. Punktem wyjścia jest skończony szereg kolejno opóźnionych zmiennych objaśniających przy czym współczynniki korelacji całkowitej między zmienną objaśnianą a kolejno opóźnionymi zmiennymi objaśniającymi są użyte do określenia odcinka czasu (t_l), w ciągu którego kolejne opóźnione zmienne objaśniające istotnie wpływają na zmienną objaśnianą. Moment czasu, w którym opóźniona zmienna objaśnia zmienną endogeniczną „prawie tak samo dobrze" jak zmienna nieopóźniona, autorka przyjmuje jako ocenę długości t_L. Dla pewności szacunku t_L przyjmuje się okres $t_{L\pm K}$ przy czym K dobierane jest arbitralnie. Celem tego postępowania jest oszacowanie okresu, w ciągu którego zaznacza się wpływ zmiennych opóźnionych na zmienną objaśnianą, przy czym jako kryterium optymalnego jej doboru przyjmuje się kwadrat współczynnika korelacji wielorakiej między wybranymi zmiennymi (R^2). Dla tak dobranej kombinacji opóźnień zmiennej objaśniającej, zmienna endogeniczną wyrażona będzie funkcją:

$$y_t = \sum_{i=0}^{L} w_i x_{t-i}$$

W przypadku gdy liczba x_{t-i} jest mała i nie występuje między nimi współliniowość można, przy założeniu liniowej zależności, szacować w_i bezpośrednio przy użyciu metody najmniejszych kwadratów. W przeciwnym wypadku należy określić rozkład w_i w czasie. Według S. Almon przyjmuje on postać wielomianu stopnia n. Wstępną wartość n określa się na podstawie dostępnej liczby informacji z analizy rozkładu współczynników korelacji całkowitej. Do „najlepszej" oceny n dochodzi się drogą iteracji. Rozkład w_i otrzymuje się z funkcji:

[13] L. R. Klein, op. cit. i E. Malinvaud, *The Estimation of Distributed Lags, A Comment,* Econometrica 1961, ss. 430 - 433.

[14] Artykuł jest częścią szerszego opracowania.

[15] S. Almon, *The Distributed Lag between Capital Appropriations and Expenditures,* Econometrica 1965, nr 1.

$$w_i = \sum_{j=1}^{n} \Phi_j(i) b_j \qquad j = 1, 2 \ldots, n$$

gdzie b_j to parametry funkcji

$$y_t = \sum_{j=1}^{n} b_j Z_{tj}$$

szacowane metodą najmniejszych kwadratów, a

$$Z_{tj} = \sum_{i=1}^{L-1} \Phi_j(i) x_{t-i}$$

a Φ_j *(i)* to współczynniki interpolacyjne Lagrange'a o wzorze ogólnym:

$$\Phi_{n+1}(i) = \frac{(i-i_0)(i-i_1)\ldots(i-i_n)}{(i_{n+1}-i_0)(i_{n+1}-i_1)\ldots(i_{n+1}-i_n)}$$

Wartości poszczególnych w_i w określonym rozkładzie interpretuje się jako udział kolejnych opóźnionych zmiennych objaśniających w wyjaśnianiu zmienności zmiennej objaśnianej. W kolejnym etapie następuje wybór jednej lub kilku zmiennych o najwyższych wartościach w_i. Wybrane zmienne włącza się do modeli i estymuje się parametry równania dla *n—k* zmiennych opóźnionych przy pomocy klasycznej metody najmniejszych kwadratów. Autorka uważa jednocześnie, iż to samo równanie może zawierać inne zmienne o różnych rozkładach opóźnień lub/i zmienne nie opóźnione.

Autorka nie podaje żadnych założeń co do możliwości stosowania takiej właśnie metody i jej efektywności. Trudno jest więc ocenić poszczególne etapy z punktu widzenia ich statystycznej prawidłowości. Na marginesie przedstawionej metody nasuwają się jednak pewne uwagi.

Otóż wiadomo, że R^2, choć najbardziej naturalna miara gdy posługujemy się współczynnikami korelacji całkowitej dla ustalenia siły związku między zmiennymi, preferuje w tym przypadku maksymalną wartość *L*, szczególnie gdy liczba uwzględnionych zmiennych zbliża się do liczby obserwacji w szeregu. Wskazane byłoby więc zastosowanie również i innych miar doboru zmiennych opóźnionych do równania. Interesujące z punktu widzenia efektywności wyników byłoby użycie pojemności integralnych nośników informacji Z. Hellwiga. Korzystanie z tej metody byłoby tym bardziej naturalne, gdyż dysponujemy już wektorem współczynników korelacji całkowitej między zmienną objaśnianą a zmiennymi objaśniającymi.

Wartość stosowanej metody obniża fakt trudności w oszacowaniu rzędu n. Iteracja staje się bardzo czasochłonna przy rosnącej liczbie opóźnionych zmiennych przy konieczności obliczania kolejnych współczynników interpolacyjnych Lagrange'a. Częste stosowanie metody najmniejszych kwadratów prowadzi również do niepewności w wiarygodności uzyskanych szacunków.

Jak wspomnieliśmy najbardziej istotną przeszkodą użycia klasycznej metody najmniejszych kwadratów dla estymacji parametrów funkcji regresji z opóźnionymi zmiennymi jest istnienie autokorelacji reszt. Istnienie autokorelacji prowadzi do uzyskiwania obciążonych ocen parametrów a i w_i, a przede wszystkim niezgodnych ocen realnego opóźnienia *(t—i),* Wspomniane wyżej metody (L. R. Kleina i innych) prowadzące do uzyskania zgodnych i nieobciążonych ocen a, w_i i L odnoszą się tylko do przypadku monotonicznego rozkładu wag w_i. Zainteresowania badaczy skupiają się ostatnio wokół wspomnianego już rozkładu Pascala jako rozkładu współczynników przy zmiennych opóźnionych, przy czym równanie:

$$y_t = a \sum_{i=0}^{L} \binom{r+i-1}{i} (1-\lambda)^r \lambda^i x_{t-i}$$

zwane jest, od nazwiska jego współtwórcy, modelem Solowa [16]. Zauważmy, że na zasadzie rozwinięcia prawej strony tego równania w szereg dwumianowy o wykładniku ujemnym, otrzymamy:

$$\sum_{i=0}^{L} \binom{r+i-1}{i} \lambda^i = (1-\lambda)^{-r}$$

Wprowadzając l jako operator opóźnienia o własności:

$$l x_t = x_{t-1}$$

otrzymujemy

$$y_t = a(1-\lambda)^r (1-\lambda l)^{-r} x_t$$

Funkcję tę możemy przekształcić w

$$y_t = \binom{r}{1} y_{t-1} + \binom{r}{2} \lambda^2 y_{t-2} - \ldots + (-1)^r \lambda^r y_{t-r} + (1-\lambda)^r x_t$$

Zakładając a priori brak autokorelacji składnika losowego szukane oceny parametrów a i Λ dla danego r otrzymujemy minimalizując różnicę kwadratów między empirycznymi wartościami y_t a wartością przekształconej powyżej funkcji. Rozwiązując przy pomocy rachunku różniczkowego otrzymujemy dla danego r układ równań z niewiadomymi a i λ. Pierwiastkami równania dla λ są wszystkie wartości z zamkniętego przedziału od 0 do 1. Najlepszą wartość r wyznacza się iteracyjnie, przy czym kryterium wyboru r jest wariancja lub błąd standardowy szacunku [17]. Obliczenia znacznie się komplikują wraz z wydłużeniem okresu opóźnienia

[16] R. M. Solow, *On the Family of Lag Distribution,* Ecanometrdca 1960, por. też A. S. Goldberger, *Teoria ekonometrii,* Warszawa 1972, s. 353.

[17] Ze względu na brak miejsca nie przytaczamy w tym miejscu wszystkich formuł matematycznych, które mają dość skomplikowaną postać. Zainteresowanych odsyłamy do cytowanego artykułu R. M. Solowa.

i wraz ze wzrostem r. W tym świetle wybór możliwych wartości *r* staje się arbitralny.

Wad takich nie ma metoda estymacji parametrów modelu Solowa przy pomocy twierdzeń rachunku prawdopodobieństwa. Metoda ta[18] wykorzystuje twierdzenie Bayesa dla ustalenia rozkładów prawdopodobieństwa a priori dla parametrów r, a, i λ Traktuje się więc parametry te jako zmienne losowe, przy czym λ jest zmienną losową ciągłą w przedziale 0,1, natomiast r jest zmienną losową skokową, przy czym $r=1, 2, \ldots, m$. W myśl tej metody funkcja wiarygodności $l(r, \lambda, a, \sigma, x, y)$ przybiera postać

$$\frac{1}{\sigma^T} \exp - \left\{ \frac{1}{2\sigma^2} Q \right\}$$

gdzie $$Q = \sum_{t=1}^{T} (w_t - a z_t)^2,$$

$$w_t = \sum (1 - \lambda l)^r y_t$$

gdzie l jest operatorem opóźnienia

$$z_t = \sum (1 - \lambda)^r x_t$$

Zakładając rozkłady prawdopodobieństwa a priori

$$p(\lambda) = d\lambda \qquad 0 < \lambda < 1$$

$$p(r) = \frac{1}{m} \qquad r = 1, 2, \ldots, m$$

po przeprowadzeniu całkowania funkcji wiarygodności ze względu na a, x, y, σ, otrzymujemy ostatecznie

$$p(\lambda, r) = \left(\sum_{t=1}^{T} z_t^2 \right)^{-\frac{1}{2}} . \qquad s^{-(T-1)}$$

gdzie

$$s = \sum_t (w_t - \bar{a} z_t)^2 : (T-1)$$

natomiast

$$\bar{a} = \sum_t z_t w_t : \sum_t z_t^2$$

$p(\lambda, r)$ jest funkcją gęstości prawdopodobieństwa zmiennej losowej dwuwymiarowej i stąd szukane $p(r)$ i $p(\lambda)$ jako funkcje prawdopodobieństwa brzegowego otrzymujemy z

$$p(r) = \int_{-\infty}^{+\infty} p(r, \lambda) \, d\lambda$$

$$p(\lambda) = \sum_r p(r, \lambda)$$

[18] V. K. Chetty, op. cit

Z otrzymanych rozkładów wybieramy tę wartość r dla której $p(r)$ ma wartość modalną. Oceną X jest wartość oczekiwana funkcji $p(\lambda)$. Uzyskane w ten sposób oceny parametrów określają rozkład współczynników opóźnień w_i poprzez kolejne podstawianie i=0, 1, . . . , L do rozkładu Pascala.

Wyznaczymy obecnie rozkład realizacji złożonych zamówień na statki towarowe do przewozu ładunków suchych, czyli napływ nowo zbudowanego tonażu do eksploatacji zamówionego L lat wcześniej w stoczniach światowych.

Oznaczając y_t jako tonaż w eksploatacji w okresie t, a przez x_t portfel zamówień tonażowych w tym samym okresie, określimy rozkład opóźnień zmiennej x_t względem y_t przy pomocy metod omówionych na stronach poprzednich.

Zakładamy, że cykl realizacji zamówień tonażowych nie przekracza czterech lat oraz że średnio cykl ten trwa dłużej niż jeden rok. Rozumiemy więc, że opóźnienia y_t względem x_t nie rozkładają się w szeregu geometrycznym w czasie.

Szacujemy więc średni czas realizacji inwestycji tonażowych przy czym dokonamy tego przy pomocy modelu Solowa dla $r=2$ i dla $r=3$. Dokonamy następnie porównania otrzymanego rozkładu opóźnień z rozkładami otrzymanymi przy pomocy wielomianu interpolacyjnego. Otrzymane wartości $w\backslash$ interpretować będziemy jako udziały zrealizowanych w kolejnych latach zamówień złożonych w stoczniach $t—1$, $t—2$, $t—3$ i t — 4 lat wcześniej we wzroście tonażu będącego w eksploatacji. Możemy tak postąpić, bowiem

$$\sum_i w_i = 1$$

W modelu Solowa jako kryterium przyjęcia odpowiednich r i λ przyjmiemy obliczone według odpowiednich formuł wartości wariancji s_r^2 oraz wartość sumy w_i. Za najlepszą kombinację r i λ traktujemy tę, dla której s_r^2 jest najmniejsza i Σw_i jest najbliższa jedności. Warto w tym miejscu zwrócić uwagę, iż fakt, że suma w_i jest bliska jedności oznacza pełną realizację złożonych zamówień w badanym okresie. Tym samym więc suma w_i jest jednocześnie kryterium „dobroci" parametru L.

Dla $r=2$ otrzymaliśmy następujące równanie ze względu na λ

$1250216064\ \lambda^8—1973896777\ \lambda^7+1385050249\ \lambda^6+$
$—364379810\ \lambda^5+2500539731\ \lambda^4—4007453912\ \lambda^3+$
$+2878299741\ \lambda^2—756941890\ \lambda—1874770=0$

Odpowiednio dla $r=3$

$20532000\text{-}27135972\ \lambda^{11}+1703075411\ \lambda^{10}—6093279275\ \lambda^9+$
$+16474471959\ \lambda_8—29875857387\ \lambda^7+447119885989\ \lambda^6+$
$—27380419678\ \lambda^5+31310887065\ \lambda^4—9223166087\ \lambda^3+$
$+6434604267\ \lambda^2—1125012533\ \lambda=0$

Szukając pierwiastków dla obu równań w przedziale od 0 do 1 otrzymujemy dla pierwszego równania $\lambda = 0{,}58141$ oraz dla drugiego równania $\lambda = 0{,}25368$. Wariancje s_r^2 wynosiły odpowiednio $s_{r=2}^2 = 630$ i $s_{r=3}^3 = 1067$. Otrzymujemy następujące rozkłady współczynników w_i: dla $r = 2$

0,17522 0,22375 0,17767 0,13775 0,10011

dla $r=3$

0,28618 0,31166 0,13319 0,04745 0,01524

przy czym dla pierwszego równania $\Sigma w_i = 0{,}81452$, natomiast dla drugiego równania $\Sigma w_i = 0{,}79372$

Oceny parametrów r i λ otrzymane przy pomocy metody zaproponowanej przez Chetty'go przedstawiają się następująco

r	1	2	3	4	5
$p(r)$	0,01	0,23	0,65	0,09	0,02

$E(\lambda) = 0{,}27$

Oceną parametru r jest jego wartość, dla której prawdopodobieństwo brzegowe jest najważniejsze, czyli dla $r=3$. Ocena parametru λ nie odbiega znacznie od wartości uzyskanej przy pomocy modelu Solowa dla $r=3$.

Przy zastosowaniu metody wielomianu interpolacyjnego przyjęliśmy do badania zmienną x_{t-L} dla $L=0, 1, 2, 3, 4$. Wydłużenie szeregu analizy nie znajduje pokrycia w praktyce ze względu na fakt, że więcej niż 4-letnia realizacja kontraktu pod budowę statku jest zjawiskiem niesłychanie rzadkim nawet w okresach pełnego wykorzystania zdolności produkcyjnych stoczni światowych.

Jako postać rozkładu współczynników w_i założyliśmy wielomian drugiego stopnia. Po obliczeniu wartości współczynników interpolacyjnych przy przyjętych na poprzednich stronach sposobach postępowania rozkład współczynników opóźnień przedstawia się następująco:

0,2034 0,2550 0,2274 0,0901 0,0033

$$\sum w_i = 0{,}7792$$

Interpretując uzyskane rezultaty możemy stwierdzić, że jakkolwiek uzyskaliśmy trzy różne rozkłady opóźnień, to jednak we wszystkich przypadkach posiadają one wartość modalną dla $i = 1$ rok. Widzimy więc, że średni okres realizacji zamówionego tonażu wynosił na przestrzeni analizowanych lat (1950 - 1971) jeden rok. We wszystkich trzech przypadkach obserwujemy jednak duże spłaszczenie rozkładu współczynników, gdyż wartości sąsiednie w_i w stosunku do wartości modalnych niezbyt istotnie się od nich różnią. Szczególnie dużą dyspersję wykazuje rozkład w_i dla $r=2$ oszacowany przy pomocy modelu Solowa. Sugeruje on ok. 10-procentowy udział w nowo zbudowanym tonażu inwestycji sprzed czterech lat.

Największą asymetrią charakteryzuje się drugi rozkład współczynników opóźnień (dla $r=3$). We wszystkich trzech przypadkach obserwujemy jednak wyraźną asymetrię prawostronną, przy czym rozkład współczynników uzyskany metodą wielomianu interpolacyjnego wykazuje bardzo łagodną asymetrię. Może się to wiązać z przyjęciem jako podstawy interpolacji wielomianu stopnia drugiego.

Zaobserwowana asymetria dodatnia znajduje swoje wytłumaczenie w tendencjach współczesnego budownictwa okrętowego. Biorąc pod uwagę postęp techniczny i organizacyjny w przemyśle stoczniowym wyrażający się między innymi w mechanizacji prac stoczniowych, szerokim stosowaniu metody połówkowej przy budowie statków, stosowaniu nowych rodzajów stali i tworzyw sztucznych w tym przemyśle[19], najbardziej realną kombinacją wydaje się być rozkład opóźnień oszacowany przy pomocy modelu Solowa dla $r=3$ i $\lambda=0,25$. Prawie 30-procentowy udział w nowo zbudowanym tonażu zamówień złożonych w tym samym roku nadając rozkładowi charakter skrajnie asymetryczny wyraża tendencje panujące na rynku budownictwa okrętowego a dążące do dalszego skracania cyklu budowy statku.

Wytłumaczenie zaobserwowanej dyspersji leży oczywiście w znacznym zróżnicowaniu średniego cyklu budowy statków w różnych stoczniach W poszczególnych krajach. Należy zwrócić uwagę, że dyspersja ta może być spowodowana zarówno rzeczową (w skład portfelu zamówień na tonaż „suchy" wchodzi co najmniej kilkanaście typów statków o zróżnicowanych parametrach, a tym samym o zróżnicowanej pracochłonności ich budowy), jak i czasową agregacją danych. Wydaje się że badanie opóźnień imiennej x_t w krótszych jednostkach czasu pozwoliłaby na bardziej jednoznaczne wyodrębnienie średniej długości cyklu realizacji zamówień tonażowych.

Zaobserwowana dyspersja utrudnia znacznie wybór konkretnego opóźnienia dla zmiennej x w przypadku traktowania tej zmiennej jako objaśniającej w modelu ekonométrycznym. Najlepszym rozwiązaniem w tym przypadku jest wprowadzenie do równania modelu zmiennej będącej ważoną sumą wartości tej zmiennej (x_{t-i}), przy czym wagami tej sumy są obliczone współczynniki opóźnień — w_i.

Konstrukcja takiej zmiennej pozwala na uchwycenie pełnego wpływu wszystkich opóźnionych wartości zmiennej x na wybraną zmienną objaśnianą bez konieczności uwzględniania w modelu wszystkich opóźnionych wartości x_{t-i} i jednocześnie bez daleko idących uproszczeń przy uwzględnianiu tylko średniego opóźnienia tej zmiennej.

[19] Por. K. Wandelt, *Studia nad postępem technicznym i organizacyjnym,* Prace Komisji Nauk Ekonomicznych PTPN, Poznań 1972.

ESTIMATION METHODS OF THE LAG DISTRIBUTION

Summary

The author deals with the principle estimation methods of the lag distribution. The lag estimation with help of the interpolation polynomial and with help of the Pascal distribution has been broad discussed.