

**Michał MACKIEWICZ, Piotr KRAJEWSKI, Zbigniew KUCHTA,
Małgorzata MAZUREK, Maciej NOWAKOWSKI**

Długookresowe prognozowanie zmiennych ekonomicznych na podstawie bardzo krótkich szeregów czasowych

Analiza szeregów czasowych jest jednym z podstawowych narzędzi ekonometrycznych znajdujących zastosowanie w prognozowaniu. Powszechnie znane metody analizy szeregów czasowych, takie jak wygładzanie wykładnicze, metoda średniej ruchomej, modele trendu (Cieślak, 2005) czy model autoregresyjny (często stosowany do prognozowania inflacji) (Stock, Watson, 1999) mają ograniczone zdolności predykcyjne. Metody te należy uznać za efektywne w przypadku krótkiego okresu prognozy *ex ante*. Im okres tej prognozy się wydłuża, tym większy błąd jest wprowadzany do otrzymywanych predykcji zmiennej. Przyczyny te stanowią podstawowy czynnik ograniczający wykorzystanie tych metod w praktycznej analizie danych statystycznych.

Pośród podstawowych narzędzi ekonometrycznych znajdują się także modele strukturalne (Welfe, 2003). Mają one przewagę nad analizą szeregów czasowych, ponieważ mogą odwzorowywać teorię ekonomiczną. Jednak narzędzie to ma istotne ograniczenia. Są nimi konieczność dogłębnej analizy badanego procesu, nałożenie odpowiednich restrykcji wynikających z teorii ekonomii oraz wykorzystanie podejścia *from general to specific* (Hendry, 1995). Przy formułowaniu prognoz *out of sample* istotnym problemem jest również określenie przyszłych wartości zmiennych egzogenicznych. W praktycznym zastosowaniu ograniczeniem jest ponadto liczba dostępnych obserwacji, która wraz z liczbą parametrów do oszacowania determinuje liczbę stopni swobody. To wymaganie właściwie wyklucza wykorzystanie modelu strukturalnego jako narzędzia do prognozowania na podstawie bardzo krótkich szeregów czasowych.

Celem artykułu jest przedstawienie metody prognozowania długookresowego zjawisk opartej na szeregach czasowych. Narzędzie to stanowi modyfikację wcześniej wymienionych metod. Ma ono szczególne zastosowanie przy prognozowaniu na podstawie ultrakrótkich szeregów czasowych. Nadrzędnym celem omawianej metody jest jej praktyczne wykorzystanie do prognozowania długookresowego, dlatego głównym kryterium porównań powinna być jedna z miar błędów prognozy *ex post* (np. MSE — *Mean Square Error*) (Clements, Hendry, 1988). Właściwości predykcyjne narzędzia zostały zweryfikowane przy wykorzystaniu techniki *rolling window* na podstawie prognoz poza próbą.

Proponowana metoda może być szczególnie efektywna (z punktu widzenia własności prognostycznych) przy predykcji wskaźników struktury. Zgodnie

z intuicją ekonomiczną poziom wskaźników struktury powinien się stabilizować w długim okresie, nie wykraczając poza przedział (0;1). Jednak analizując wykresy szeregów czasowych (przedstawiających wskaźniki struktury) można bardzo często zauważyć tendencję rozwojową. Jest to wyraźnie widoczne szczególnie przy krótkich szeregach czasowych. Wykorzystując do prognoz model ARIMA (p, d, q) (Harris, Sollis, 2003) konieczny jest wybór wartości parametrów p, d, q (gdzie parametry p, q oznaczają odpowiednio rząd autoregresji AR i średniej ruchomej MA, a parametr d — stopień zróżnicowania szeregu).

Z punktu widzenia prognozy długookresowej kluczowa jest wartość parametru d . Wykorzystanie procedury testowej przy określaniu wartości parametru d dla zmiennej charakteryzującej się tendencją rozwojową może zasugerować wartość parametru równą 1, czyli fakt, że proces stochastyczny, generujący dany szereg statystyczny, jest zintegrowany w stopniu pierwszym I (1) (Granger, Newbold, 1974). W praktyce oznacza to, że prognoza długookresowa będzie charakteryzowała się tendencją rozwojową zaobserwowaną w próbie. Może to prowadzić do rezultatów nieakceptowalnych z punktu widzenia teorii ekonomii, w tym również do sytuacji, w której prognozowana wartość wskaźnika struktury przyjmie wartość poniżej zera lub powyżej jedności. W przeciwieństwie do modelu ARIMA, problem ten nie występuje w proponowanej metodzie. Metoda ta uwzględnia zarówno fakt, że obserwowane w próbie tendencje rozwojowe mogą być wynikiem zmian strukturalnych w gospodarce, jak i — oczekiwaną często w przypadku wskaźników struktury — stabilizację w długim okresie. Dlatego w opisywanej metodzie, po okresie dostosowawczym, w którym tendencja zaobserwowana w próbie jest kontynuowana, prognozowane wartości stabilizują się.

METODA KONSTRUKCJI PROGNOZY

Opisywana metoda jest przeznaczona w zasadzie do prognozowania danych rocznych, w analizie pominięto zatem występowanie zjawiska sezonowości. Podstawową trudnością przy formułowaniu prognozy na podstawie danych o rocznej częstotliwości, szczególnie w warunkach polskich, jest zarówno brak odpowiednio długich szeregów czasowych, jak i porównywalnych danych statystycznych. W dużym stopniu ogranicza to stosowanie większości narzędzi ekonometrycznych. Dlatego podjęto próbę konstrukcji alternatywnej metody, dzięki której można otrzymać zgodne z intuicją ekonomiczną długookresowe prognozy na podstawie dostępnych, stosunkowo krótkich, szeregów czasowych.

Konstrukcja prognozy przy użyciu proponowanej metody polega na połączeniu wniosków płynących z dwóch, często wykorzystywanych, metod prognozowania: średniej ważonej oraz trendu liniowego. Zastosowanie metody średniej ważonej polega na obliczeniu prognozy dla badanej zmiennej y zgodnie ze wzorem:

$$\hat{y}_{T+1} = \sum_{t=1}^T y_t w_t \quad (1)$$

gdzie:

T — długość szeregu czasowego,

$\hat{y}_{T+1}$ — prognozowana wartość zmiennej y dla okresu $T+1$,

y_t — znane, wcześniej uzyskane bądź pochodzące z próby wartości obserwacji zmiennej y ,

w_t — waga dla okresu t , przy czym $w_t \in \langle 0,1 \rangle$ i $\sum_{t=1}^T w_t = 1$.

Metoda ta może być w różny sposób modyfikowana, m.in. poprzez wykorzystanie przyrostów zmiennych bądź też przez nadanie poszczególnym składnikom równych wag, co sprowadza ją do średniej arytmetycznej.

Zastosowanie metody trendu liniowego polega na oszacowaniu parametrów równania regresji¹ postaci:

$$y_t = \alpha + \beta t + \xi_t \quad (2)$$

gdzie:

t — zmienna czasowa,

α, β — parametry funkcji trendu liniowego,

ξ_t — składnik losowy spełniający założenia procesu Gaussa-Markowa.

Parametry równania (2) są szacowane metodą najmniejszych kwadratów. Na tej podstawie dokonuje się prognozy, podstawiając kolejne wartości zmiennej czasowej.

Obydwie niestrukturalne metody prognozowania mają wiele ograniczeń. Jednak ich odpowiednie połączenie wydaje się w satysfakcjonujący sposób realizować założenia stawiane przed metodą:

- prognozowanie na podstawie krótkich szeregów czasowych,
- minimalizacja błędów prognozy długookresowej *ex ante*,
- możliwość wykorzystania wiedzy *a priori*.

Konstrukcja prognoz odbywa się etapami. Na początku dokonuje się oszacowania parametrów modelu trendu liniowego na podstawie n -elementowej próby metodą najmniejszych kwadratów. Przyjmujemy, że model ma postać $y_t = \alpha + \beta t + \xi_t$, analogiczną do (2), gdzie t jest zmienną czasową ($t = T - (n - 1), T - (n - 2), \dots, T$), zaś T jest to długość dostępnego szeregu czasowego.

¹ Często wykorzystuje się również trendy nieliniowe, które charakteryzują się bardziej złożonymi formami funkcyjnymi. Jednak ich wykorzystanie pociąga za sobą zwiększoną liczbę parametrów poddawanych estymacji, co znacznie utrudnia ich zastosowanie w przypadku bardzo krótkich szeregów czasowych.

Przy prognozowaniu na podstawie bardzo krótkich szeregów czasowych wartość parametru n najczęściej jest determinowana przez dostępność porównywalnych danych statystycznych ($n = T$). W empirycznej części opracowania przedstawione zostaną rezultaty badania wpływu długości próby wykorzystywanej do estymacji parametrów modelu na jakość prognoz, rozważane będą także próby krótsze niż maksymalna dostępna próba ($n < T$).

Drugi etap konstrukcji prognoz obejmuje ustalenie długookresowego poziomu prognozowanej zmiennej:

$$\hat{y}_{T+m} = \hat{\alpha} + \hat{\beta}(T+m) \quad (3)$$

m — liczba okresów poza próbą, determinująca poziom długookresowy prognozowanej zmiennej.

W ostatnim etapie konstrukcji prognozy oblicza się prognozy wartości analizowanej zmiennej. W tym celu wykorzystuje się metody średniej ważonej oraz trendu liniowego.

Przyszłe wartości $\hat{y}_t$ otrzymuje się zgodnie z formułą rekurencyjną:

$$\hat{y}_{T+s} = (1-k)\hat{y}_{T+s-1} + k\hat{y}_{T+m} \quad (4)$$

$$k \in (0,1)$$

gdzie:

k — współczynnik wagi,

s — numer iteracji w procedurze rekurencyjnej, $s = 1, 2, \dots, S$,

S — horyzont prognozy.

Wartość współczynnika wagi ma kluczowy wpływ na prognozę. W przypadku gdy ustalony współczynnik będzie bardzo bliski 0, to prognozę na przyszły okres można sprowadzić do wartości poprzedniej prognozowanej zmiennej. Natomiast w przypadku, gdy ustalony współczynnik będzie bardzo bliski jedności, to prognoza będzie bardzo bliska asymptotycznemu poziomowi długookresowemu. Będzie więc w największym stopniu zależała od parametrów strukturalnych otrzymanej linii trendu.

Korzystając z formuły (3) oraz (4), podstawiając rekurencyjnie, można otrzymać wzór pozwalający wyznaczyć wartość prognozowanej zmiennej $\hat{y}_{T+s}$ bez konieczności stosowania procedury rekurencyjnej:

$$\hat{y}_{T+s} = (1-k)^s y_T + (1-(1-k)^s)[\hat{\alpha} + \hat{\beta}(T+m)] \quad (5)$$

Opisywane narzędzie ma zastosowanie w prognozowaniu długookresowym, zatem dopuszcza się sytuację, w której horyzont prognozy znacznie przekracza długość próby, na podstawie której estymowane były parametry trendu ($S > n$). Należy jednak wziąć pod uwagę, że wraz ze wzrostem horyzontu prognozy wzrasta niepewność co do przyszłości.

Przedstawiona metoda pozwala uwzględnić specyfikę wskaźników struktury, charakteryzujących się tendencją rozwojową w okresie zmian strukturalnych w gospodarce oraz relatywnie stałym poziomem w długim okresie. Wzór (4) pokazuje, że prognoza na okres $T + s$ jest średnią ważoną z prognozy wartości zmiennej na okres $T + s - 1$ i długookresowej prognozy wartości zmiennej (3). Długookresowa prognoza wartości zmiennej jest otrzymywana poprzez ekstrapolację trendu liniowego (m okresów poza próbą), oszacowanego na podstawie próby. Taka konstrukcja prognozy pozwoli uwzględnić tendencję rozwojową, o ile występuje ona w próbie. W długim okresie własnością metody jest prognozowane ustabilizowanie wartości wskaźnika:

$$\lim_{s \rightarrow \infty} \hat{y}_{T+s} = \hat{\alpha} + \hat{\beta}(T + m) \quad (6)$$

Otwarta pozostaje praktyczna kwestia wyboru wartości parametrów n , m oraz k . Wartość parametru n powinna być wystarczająco duża, by możliwe było oszacowanie parametrów trendu liniowego. Jednak należy uwzględnić fakt, że wydłużenie próby wykorzystanej do oszacowania parametrów strukturalnych linii trendu może wprowadzić do prognozy informacje, które pogorszą jej własności. Przykładowo, nadmierne zwiększenie próby wstecz zwiększa prawdopodobieństwo uwzględnienia w próbie przeszłych zmian strukturalnych. Parametr m determinuje asymptotyczny długookresowy poziom badanej zmiennej (6), może być on skalibrowany na podstawie wiedzy *a priori*.

EMPIRYCZNY TEST WŁASNOŚCI PROGNOSTYCZNYCH

Własności prognostyczne przedstawionej metody prognozowania zostaną ocenione na podstawie danych panelowych wskaźników zatrudnienia w 34 krajach OECD w ujęciu rocznym w następujących dziedzinach:

- rolnictwo,
 - budownictwo,
 - przemysł,
 - przemysł wytwórczy
- oraz w sektorach:
- prywatnym
 - i usług.

Dane pochodzą z bazy danych AMECO Komisji Europejskiej i obejmują okres 1960—2007 (nie dla wszystkich wskaźników dostępna jest pełna próba).

Wskaźnik zatrudnienia został zdefiniowany jako udział liczby zatrudnionych w danym sektorze gospodarki w całkowitej populacji danego kraju.

Analizowane szeregi zróżnicowane są pod względem średniej, rozproszenia oraz występowania i kształtu trendu. Dlatego są one odpowiednie do oceny własności prognostycznych opisywanej metody. Jako podstawę do porównań tych własności zastosowano prognozy otrzymane na podstawie modelu ARIMA (p, d, q). Wykorzystano tu model ARIMA (1, 0, 0) ze względu na bardzo krótkie szeregi czasowe. Modele ARIMA z parametrem d różnym od 0 nie zostały wzięte pod uwagę ze względu na fakt, że długookresowe prognozy wskaźników struktury generowane przez takie modele przyjmują często nieakceptowalne wartości.

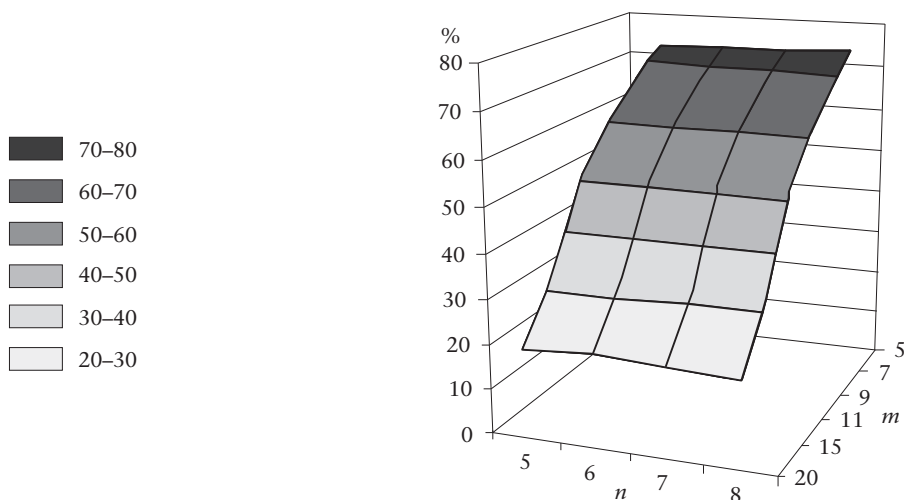
Przedmiotem badania jest porównanie trafności 10-letnich prognoz na podstawie błędu średniokwadratowego (MSE). Podjęto również próbę określenia optymalnej wartości parametrów n (ilość obserwacji wykorzystana do estymacji parametrów modelu trendu liniowego) oraz m (parametr determinujący asymptotyczny poziom prognozowanej zmiennej). W wielu przypadkach dokładna analiza badanego zjawiska oraz wiedza *a priori* umożliwiają wybór wartości tych parametrów. Jednak w przypadku prognozowania dużej ilości zmiennych analiza danych może okazać się czasochłonna, dlatego uzasadniona wydaje się próba jej zautomatyzowania.

Podstawą analizy jest obliczenie błędu prognozy *out of sample* MSE dla obydwu porównywanych metod. Parametry modelu trendu liniowego oraz parametry modelu ARIMA są estymowane na podstawie n -elementowej próby. Prognoza jest wyznaczana na 10 lat. Obliczenia przeprowadzono metodą *rolling window*, czyli w pierwszym kroku parametry modeli są szacowane na podstawie obserwacji $x_1, x_2, \dots, x_n$, w drugim kroku wyznaczane są prognozy na 10 okresów $\hat{x}_{n+1}, \hat{x}_{n+2}, \dots, \hat{x}_{n+10}$. Następnie obliczony oraz porównany jest MSE dla obu metod, przy czym do sformalizowanego porównania wykorzystano test Diebolda-Mariano² (Diebold, Mariano, 1995). W kolejnym kroku cała procedura jest powtarzana, z tym że parametry modeli są szacowane na podstawie próby $x_2, x_3, \dots, x_{n+1}$, zaś prognozy zostaną wyznaczone na okres $\hat{x}_{n+2}, \hat{x}_{n+3}, \dots, \hat{x}_{n+11}$. Kolejne etapy analizy są powtarzane w sposób analogiczny. Taka procedura została przeprowadzona dla każdego z 34 państw i dla każdej z 6 zmiennych. Obliczony został udział przypadków, w których opracowana metoda okazała się statystycznie istotnie lepsza od modelu ARIMA (MSE był statystycznie istotnie niższy) w ogólnej liczbie przeprowadzonych estymacji.

Ponadto, w celu ustalenia optymalnych wartości parametrów m i n , analiza została powtórzona dla różnych wartości parametrów n ($n = 5, 6, 7, 8$) i m ($m = 5, 7, 9, 11, 15, 20$). Wyniki badania empirycznego zostały przedstawione na poniższym wykresie i w tablicy.

² Test ten wykorzystywany jest do porównywania trafności dwóch konkurencyjnych prognoz tego samego szeregu czasowego. Hipoteza zerowa zakłada, że wartość oczekiwana z różnicy funkcji strat związanych z poszczególnymi prognozami jest równa 0.

PORÓWNANIE PROGNOZ GENEROWANYCH OPISYWANĄ METODĄ ORAZ PRZY UŻYCIU MODELU ARIMA



U w a g a. Na wykresie przedstawiony został procentowy udział przypadków, w których MSE uzyskany na podstawie opracowanej metody okazał się istotnie niższy od MSE uzyskanego na podstawie modelu ARIMA w ogólnej liczbie przeprowadzonych estymacji w zależności od wartości parametrów m i n .

Ź r ó d ł o: dane AMECO, obliczenia własne

PORÓWNANIE PROGNOZ GENEROWANYCH PRZEZ OPISYWANE NARZĘDZIE I MODEL ARIMA

$n \backslash m$	20	15	11	9	7	5
5	20,6	32,9	51,1	60,4	67,2	72,7
6	21,6	33,2	50,9	61,1	67,7	73,2
7	20,5	32,2	51,4	61,4	67,9	73,4
8	19,8	32,2	51,3	61,1	68,4	74,2

U w a g a. W tabelicy przedstawiony został procentowy udział przypadków, w których MSE uzyskany na podstawie opracowanej metody okazał się istotnie niższy od MSE uzyskanego na podstawie modelu ARIMA w ogólnej liczbie przeprowadzonych estymacji w zależności od wartości parametrów m i n .

Ź r ó d ł o: dane AMECO, opracowanie własne.

Można zaobserwować znaczny wzrost procentowego udziału przypadków, w których proponowana metoda okazała się statystycznie istotnie lepsza od modelu ARIMA w ogólnej liczbie przeprowadzanych estymacji, wraz ze spadkiem wartości parametru m . Rekomendowana wartość parametru m przy prognozach długookresowych (10-letnich) wskaźników struktury wynosi $m = 5$. Należy jednak zaznaczyć, że optymalna wartość parametru m w dużej mierze zależy od specyfiki prognozowanego szeregu statystycznego, dlatego trzeba podchodzić z dużą ostrożnością do uogólniania uzyskanych wyników. Parametr m może być również skalibrowany na podstawie wiedzy *a priori* (np. o długości trwania

zmian strukturalnych w innych krajach). W badaniu przyjęto założenie o braku wiedzy *a priori*, a zatem nie podlegały testowaniu wartości parametru $m < 5$, gdyż takie wartości świadczyłyby o posiadaniu wiedzy o zakończeniu zmian strukturalnych w krótkim okresie.

Analizując wpływ parametru n na rezultat badania można dojść do wniosku, że wartość tego parametru nie zmienia w istotny sposób badanego udziału procentowego. Biorąc pod uwagę konieczność prognozowania na podstawie krótkich szeregów czasowych, wydaje się rozsądne przyjęcie parametru n na poziomie $n = 5$. W badaniu nie były brane pod uwagę różne wartości parametru k , został on przyjęty na stałym poziomie równym $k = 0,3$.

Podsumowanie

W artykule przedstawiono metodę prognozy stworzoną na potrzeby prognozowania długookresowego zmiennych rocznych na podstawie krótkich szeregów czasowych. W tym celu skonstruowana została metoda prognozy wykorzystująca własności średniej ważonej oraz trendu liniowego. W części empirycznej dokonano prognozy wskaźników zatrudnienia przy użyciu proponowanej metody oraz porównania własności prognostycznych z modelem ARIMA przy użyciu testu Diebolda-Mariano. Dokonana predykcja długookresowa (horyzont prognozy równy 10 okresom), przekraczająca znacznie długość szeregu czasowego, na podstawie którego estymowane były parametry, pozwoliła ocenić efektywność skonstruowanego narzędzia. W konkluzji można stwierdzić, że metoda ta pozwala z powodzeniem uzyskiwać wartości przyszłe przy prognozach długookresowych.

dr Michał Mackiewicz, dr Piotr Krajewski, mgr Zbigniew Kuchta, Małgorzata Mazurek, Maciej Nowakowski — Uniwersytet Łódzki

LITERATURA

- Cieślak M. (2005), *Prognozowanie praktyczne*, PWN
- Clements M. P., Hendry D. F. (1988), *Forecasting Economic Time Series*, Cambridge University Press
- Diebold F. X., Mariano R. S. (1995), *Comparing Predictive Accuracy*, „Journal of Business & Economic Statistics”
- Granger C. W. J., Newbold P. (1974), *Spurious Regression in Econometrics*, „Journal of Econometrics”
- Harris R., Sollis R. (2003), *Applied Time Series Modelling and Forecasting*, John Wiley & Sons
- Hendry D. F. (1995), *Dynamic Econometrics*, Oxford University Press
- Stock J., Watson M. (1999), *Forecasting Inflation*, MA: National Bureau of Economic Research, Cambridge
- Welfe A. (2003), *Ekonometria*, PWE