

Uniwersytet Mikołaja Kopernika w Toruniu
Katedra Ekonometrii i Statystyki

Mariola Pilatowska

SKUMULOWANY BŁĄD PROGNOZ JAKO METODA WYBORU MODELU

Zarys treści. Celem artykułu jest prezentacja i wykorzystanie skumulowanego błędu prognoz na jeden okres naprzód (APE) nie tylko jako metody (strategii) wyboru modelu, ale również jako narzędzia do wyboru samej strategii (meta-wyбір). Na przykładach empirycznych metoda APE jest porównywana z metodami wykorzystującymi kryteria informacyjne (AIC i BIC). Otrzymane wyniki wskazują na dużą praktyczną przydatność metody APE.

Słowa kluczowe: wybór modelu, meta-wyбір, kryteria informacyjne, skumulowany błąd prognoz.

WSTĘP

W literaturze można znaleźć różne metody (strategie) wyboru modelu ekonometrycznego, m.in. podejście bazujące na sekwencji testów statystycznych, podejście informacyjne bazujące na kryteriach informacyjnych typu Akaike'a, podejście bazujące na kryteriach predykcyjnych, które można zaliczyć do dominujących podejść w wyborze modelu ekonometrycznego. Ze względu na to, że prawdziwy model generujący dane jest w praktyce nieznanym, uwaga w metodach wyboru modelu wyraźnie przesuwana się z kwestii odkrywania jednego, prawdziwego modelu na zagadnienie wyboru na podstawie danych najlepszego modelu spośród danego zestawu modeli czy też wyboru kilku wiarygodnych modeli, wśród których model najlepszy ma relatywnie niewielką przewagę nad pozostałymi (Burnham, Anderson, 2002). Wybór modelu najlepszego czy wnioskowanie

na podstawie wielu modeli (*multi-model inference*) zakłada oczywiście, że modele zostały starannie dobrane do zestawu modeli, ponieważ w każdym zestawie modeli można wybrać model relatywnie najlepszy, który jednak w sensie absolutnym będzie słaby.

W ramach każdej strategii określa się pewien algorytm, który dla danego zbioru danych statystycznych pozwala wybrać najlepszy (pod pewnym względem) model spośród określonego zestawu modeli (ogólnie mogą to być modele zagnieżdżone lub niezagnieżdżone, mogą bazować na różnych teoriach czy założeniach modelowych). Problem wyboru modelu obejmuje jednak nie tylko wybór modelu w ramach danej strategii, ale również wybór samej strategii. Większość opracowań skupia się na wyborze modelu w ramach danej strategii, ewentualnie na porównaniu kilku strategii pod względem wyboru najlepszego modelu, jednak bez dotykania zagadnienia wyboru strategii.

Wybór strategii może zależeć od celu, jakiemu ma służyć wybrany model (estymacja, predykcja), od wielkości próby (strategie te zachowują się różnie w małych i dużych próbach), od znajomości modelu generującego dane¹. W praktyce zatem pojawia się potrzeba stworzenia adaptacyjnej procedury bazującej na danych statystycznych, która służyłaby do identyfikacji właściwej strategii wyboru modelu. Identyfikacja taka nazywa się meta-wyborem (ang. *meta-selection*, De Luna, Skouras, 2003). Podejście to nie zakłada znajomości modelu generującego dane oraz spełnia sekwencyjnie predykcyjną zasadę (*predictive sequential – ‘prequential’ principle*; Dawid, 1984), która porzuca cel wyboru prawdziwego modelu na rzecz poszukiwania najmniejszego błędu prognozy poprzez porównanie prognoz z realizacjami zmiennej niezależnie od tego, jaki model został użyty do prognozowania (Clarke, 2001). Punktem wyjścia tego podejścia jest założenie, że adekwatność modelu musi być odzwierciedlona w przydatności do prognozowania, tj. dobry model pod względem estymacji, identyfikacji czy testowania hipotez statystycznych powinien dać dobre prognozy.

Celem artykułu jest prezentacja i wykorzystanie skumulowanego błędu prognoz nie tylko jako metody (strategii) wyboru modelu, ale również jako narzędzia do wyboru samej strategii.

¹ Pewne strategie są optymalne w zależności od włączenia (lub nie) modelu generującego dane do początkowego zestawu modeli-kandydatów (Shao, 1997).

1. SKUMULOWANY BŁĄD PROGNOZ NA JEDEN OKRES NAPRZÓD

Wybór modelu z wykorzystaniem skumulowanego błędu prognoz na jeden okres naprzód (ang. *accumulative prediction error*, APE) bazuje na ocenie, w jakim stopniu modele są w stanie prognozować przyszłą obserwację x_{t+1} . Według metody APE najbardziej użyteczny jest model, któremu odpowiada najmniejszy skumulowany błąd prognoz *ex post* na jeden okres naprzód. Jednak ze względu na niedostępność obserwacji x_{n+1} błąd ten nie może być wyznaczony. Można natomiast wyliczyć błąd prognoz *ex post*, bazując na poprzednich obserwacjach x^i ($0 < i < n$), jako sumę poprzednich błędów prognoz *ex post* na jeden okres naprzód.

Załóżmy, że mamy szereg czasowy o n obserwacjach, $x^n = (x_1, x_2, \dots, x_n)$. Wyznaczenie skumulowanego błędu prognoz *ex post*, APE, odbywa się przez obliczanie w sposób sekwencyjny prognoz na jeden okres naprzód, bazując na stopniowo powiększanej liczebności szeregu. Dla modelu M_j odbywa się to następująco (Wagenmaker, Grünwald, Steyvers, 2006):

1. Ustala się najmniejszą liczbę s obserwacji (długości szeregu), dla której model jest możliwy do estymacji. Ustalmy $i = s + 1$, stąd $i - 1 = s$.
2. Bazując na pierwszych $i - 1$ obserwacjach, oblicza się prognozę \hat{p}_i dla następnej obserwacji i .
3. Oblicza się błąd prognozy dla obserwacji i , np. kwadrat różnicy między zaobserwowaną wartością x_i a prognozą \hat{p}_i .
4. Zwiększa się i o jeden i powtarza się kroki 2) i 3) aż do $i = n$.
5. Sumuje się błędy prognoz *ex post* na jeden okres naprzód – wynik daje APE.

Dla modelu M_j skumulowany błąd prognoz *ex post* jest równy:

$$APE(M_j) = \sum_{i=s+1}^n d[x_i, (\hat{p}_i | x^{i-1})],$$

gdzie d oznacza specyficzną funkcję straty, która mierzy rozbieżność między obserwacjami a prognozami.

Korzystając z metody APE, należy rozważyć problem wyboru, jaką formę powinny przyjąć prognozy, tj. formę punktową (Skouras, Dawid, 1998) czy rozkładu prawdopodobieństwa (Aitchison, Dunsmore, 1975). W pierwszym przypadku prognozy \hat{p}_i są prognozami wartości średniej dla i -tego wyniku x_i . W drugim przypadku prognozy \hat{p}_i tworzą rozkład na zbiorze możliwych wyników x_i .

Kolejnym problemem jest wybór funkcji straty, której zadaniem jest pomiar rozbieżności między obserwowanymi i prognozowanymi wartościami. Można to mierzyć w różny sposób. W przypadku prognoz punktowych najczęściej używa

się błędu kwadratowego $(x_i - \hat{p}_i)^2$, ale może to również być inna funkcja straty, np. absolutny błąd $|x_i - \hat{p}_i|$, czy też ogólnie funkcja straty typu $\alpha, |x_i - \hat{p}_i|^\alpha$, gdzie $\alpha \in [1, 2]$ (Rissanen, 2003). W przypadku rozkładu prognoz najczęściej używa się logarytmicznej funkcji straty $-\ln \hat{p}_i(x_i)$, stąd strata zależy od masy prawdopodobieństwa (lub gęstości) przypisanej \hat{p}_i jako prognozie obserwowanej wartości x_i . Im większe jest prawdopodobieństwo, tym mniejsza jest strata².

Metodę APE można również zastosować do wyboru strategii (De Luna, Skouras, 2003). Załóżmy, że w ramach każdej z q potencjalnych strategii wyboru modelu, $S_1, S_2, \dots, S_q, k = 1, 2, \dots, q$, dysponujemy pewnym zestawem M -modeli $P_p(\theta_p), p = 1, 2, \dots, M$, które stanowią aproksymację modelu generującego dane. Z każdym modelem związane są parametry θ_p , które muszą być oszacowane. Jeżeli każda ze strategii prowadzi do wyboru tego samego modelu, to wtedy nie ma potrzeby wyboru strategii. W przypadku gdy pojawia się różnica w wyborze modelu przez różne strategie, wybiera się tę strategię $S_k, k = 1, 2, \dots, q$, dla której skumulowany błąd prognoz

$$APE(S_k) = \sum_{i=m}^n L(x_i, \hat{x}^{i-1}(S_k))$$

jest minimalny, gdzie $\hat{x}^{i-1}(S_k)$ jest prognozą $\hat{x}^{i-1}(p)$ otrzymaną z modelu p wybranego za pomocą strategii S_k na podstawie podpróby x_1, x_2, \dots, x_{i-1} .

Zatem $APE(S_k)$ pokazuje zachowanie w prognozowaniu, gdy strategia S_k została użyta do obliczenia sekwencyjnych prognoz, przy czym w każdym kroku, dla $i = m, \dots, n$, rekalkulacji podlega estymacja parametrów, jak i wybór modelu. Meta-wyбір sprowadza się nie tylko do minimalizacji $APE(S_k)$, ale też do śledzenia zmian $APE(S_k)$ w miarę zwiększającej się wielkości próby.

2. PRZYKŁAD EMPIRYCZNY

W celu prezentacji zachowania się skumulowanego błędu prognoz *ex post* na jeden okres naprzód (APE) w odniesieniu do wyboru modelu, jak i wyboru strategii wyboru modelu wykorzystano dane z bazy Maddisona (2001), zawierającej roczne szeregi PKB dla 36 krajów. W badaniu wzięto pod uwagę PKB dla dwóch przykładowych krajów (Francja dla okresu³ 1947–2003 i Polska dla okresu 1952–2003). Dane zostały wyrażone w mln USD w cenach stałych z 1990 roku z uwzględnieniem parytetu siły nabywczej.

² Przyjęcie logarytmicznej funkcji straty sprawia, że metoda APE staje się zgodna z maksymalizacją funkcji wiarygodności oraz bayesowskim wnioskowaniem (m.in. Wagenmaker, Grünwald, Steyvers, 2006).

³ W artykule skorzystano ze zaktualizowanej bazy Maddisona ze strony www.ggdc.net.

Kluczowym zagadnieniem przy wyborze modelu jest kwestia określenia początkowego zestawu modeli (kandydatów). W prezentowanym przykładzie zestaw ten obejmował dwa modele: ARIMA(1,1,0) oraz model trendu liniowego i autoregresji rzędu drugiego (T+AR(2)). Uzasadnieniem dla takiego zestawu jest pewna tradycja w podejściu do badań empirycznych fluktuacji PKB. Mianowicie badania te w ciągu ostatnich trzydziestu lat koncentrowały się na weryfikacji hipotezy o pierwiastku jednostkowym (co oznacza, że PKB jest niestacjonarny w wariancji lub ma trend stochastyczny) czy też hipotezy o stacjonarnych odchyleniach wokół trendu deterministycznego (co oznacza, że PKB jest niestacjonarny w średniej). Mimo bardzo licznych badań poświęconych temu zagadnieniu⁴ nie udało się uzyskać jednoznacznej odpowiedzi na to pytanie.

Model ARIMA(1,1,0) został wybrany spośród różnych specyfikacji modeli ARIMA(p, d, q), dla $p, q = 0, 1, 2, d = 0, 1$, z wykorzystaniem różnicy kryteriów Akaike'a⁵, tj. $\Delta_i = AIC_i - AIC_{\min}$, gdzie AIC_i – kryterium AIC dla i -tego modelu, a AIC_{\min} – wartość kryterium AIC dla najlepszego modelu. Modele te były szacowane dla tej samej liczebności próby, tj. 1947–2003. Im większa różnica Δ_i , tym mniej prawdopodobne jest, że dany model jest dobrym modelem w sensie kryterium Kullbacka-Leiblera⁶. W praktyce akceptuje się modele (Burnham, Anderson, 2002), dla których $\Delta_i < 4$. Na podstawie Δ_i można też uzyskać wagi AIC, które są przydatne do oceny relatywnej przewagi modelu najlepszego (o największej wadze) nad pozostałymi modelami ze zbioru R -modeli-kandydatów. Wagi te oblicza się następująco (Burnham, Anderson,

⁴ Do prac poruszających kwestię wyboru między trendem stochastycznym (niestacjonarność w wariancji) a trendem deterministycznym (niestacjonarność w średniej) dla szeregu PKB należą m.in.: Nelson, Plosser, 1982; Stock, Watson, 1986; Quah, 1987; Perron, Phillips, 1987; Christiano, Eichenbaum, 1990; Rudebusch, 1993; Diebold, Senhadji, 1996; Murray, Nelson, 1998. Wskazuje się (Haubrich, Lo, 2001), że powodem tego jest błędne założenie o prawdziwości jednej w powyższych hipotez. W ten sposób uwzględnia się jedynie możliwość wystąpienia w badanych szeregach zależności o charakterze trwałego trendu (niestacjonarność w wariancji) i przejściowych fluktuacji (niestacjonarność w średniej), a pomija się zależności o charakterze pośrednim (tzw. długa pamięć), które można by opisać przez ułamkowy proces ARFIMA.

⁵ W pracy zastosowano zmodyfikowane kryterium AIC, tj. $AIC_c = AIC + \frac{2K(K+1)}{n-K-1}$, gdzie $AIC = -2 \ln L + 2K$, K oznacza liczbę szacowanych parametrów, n – liczbę obserwacji. Standardowe kryterium AIC może zachowywać się słabo (tj. wybierać mało oszczędne w parametry modele), jeżeli występuje zbyt dużo parametrów w stosunku do liczby obserwacji. Stosowanie AIC_c zaleca się, gdy stosunek n / K jest mniejszy niż 40 (Sugiura, 1978). Dla przejrzystości zapisu zachowano zapis AIC.

⁶ Odległość (informacja) Kullbacka-Leiblera (K-L) jest miarą rozbieżności między prawdziwym i aproksymacyjnym modelem. Akaike (1973) pokazał, że wybór modelu z najniższą oczekiwaną stratą informacji (tj. modelu, który minimalizuje oczekiwaną miarę K-L) jest asymptotycznie równoważny wyborowi modelu, który ma najniższą wartość kryterium AIC.

2002; Piłatowska, 2009, 2010): $w_i = \exp(-0,5\Delta_i) / \sum_{r=1}^R \exp(-0,5\Delta_r)$, przy czym $\sum_{i=1}^R w_i = 1$. Dla modelu ARIMA(1,1,0) różnica Δ_i wynosiła zero, czyli był to model najlepszy, a dla pozostałych modeli $\Delta_i < 3$, zatem były też wiarygodne modele w sensie kryterium K-L. Jednak przewaga modelu ARIMA(1,1,0) nad pozostałymi modelami była znacząca, gdyż model ARIMA(1,1,0) miał wyraźnie dominującą wagę, tj. ok. 0,55.

W podobny sposób ustalono specyfikację alternatywnego modelu, tj. modelu trendu liniowego i autoregresji rzędu drugiego, T+AR(2), przy czym maksymalny testowany rząd autoregresji wynosił 3.

W celu dokonania wyboru między modelem ARIMA(1,1,0) a modelem T+AR(2) zastosowano trzy strategie wyboru modelu, tj. kryterium informacyjne AIC i BIC oraz skumulowany błąd prognoz na jeden okres naprzód (APE), przy czym jako funkcję straty przyjęto błąd kwadratowy (oznaczenie APE_SE) oraz błąd absolutny (oznaczenie APE_AE)⁷. Minimalna liczebność próby, dla której rozpoczynała się estymacja⁸ modeli, wynosiła 11 obserwacji (1947–1957), następnie liczebność ta była powiększana o jeden aż do n (do roku 2000). Na każdym etapie były wyznaczane prognozy z poszczególnych modeli oraz skumulowane błędy prognoz *ex post* na jeden okres naprzód (APE_SE i APE_AE). Wyniki, w formie różnic między AIC, BIC i APE dla poszczególnych modeli w zależności od liczby obserwacji, prezentują wykresy 1 i 2.

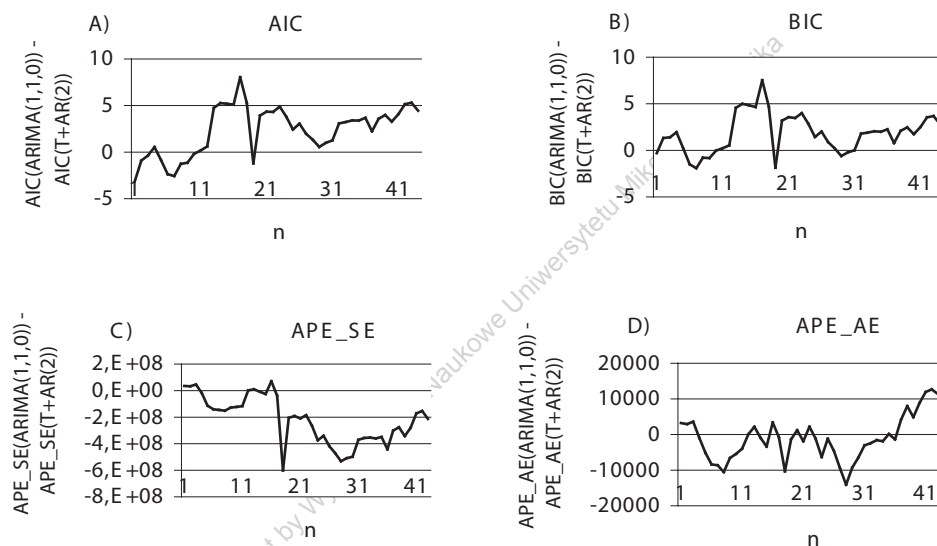
Wykres 1 (panele A i B) pokazuje, że w miarę wzrostu liczby obserwacji kryteria AIC i BIC wskazują wyraźnie na przewagę modelu T+AR(2), ponieważ różnica kryteriów $AIC(\text{ARIMA}) - AIC(\text{T+AR}(2))$ oraz $BIC(\text{ARIMA}) - BIC(\text{T+AR}(2))$ jest dodatnia (oznacza to niższe kryterium AIC i BIC dla modelu T+AR(2)); z wyjątkiem, że prognozując PKB (Francja) na 18. okres (rok 1975) – por. panel A – oraz na okres 28. i 29. (rok 1985 i 1986) – por. panel B – różnica kryteriów przyjmuje wartość ujemną, co wskazywałoby na przewagę modelu ARIMA w tych okresach prognozowania.

Śledząc przebieg różnic skumulowanego błędu prognoz na jeden okres naprzód (APE_SE oraz APE_AE) obliczanych dla obu modeli, nie otrzymuje się tak jednoznacznej przewagi modelu T+AR(2) – por. panele C i D. Na podstawie przebiegu różnic między skumulowanym błędem APE_SE dla obu modeli można zauważyć, że niemal w całym okresie prognozowania różnica ta jest ujemna, co wskazywałoby na zdecydowaną przewagę modelu ARIMA przy prognozowaniu PKB – por. panel C (z wyjątkiem okresu od 1. do 3. (lata 1958–1960), 12. do 13.

⁷ Skumulowany błąd prognoz (APE) został wyznaczony na podstawie własnej procedury zapisanej w języku skryptowym programu Gretl.

⁸ Model ARIMA(1,1,0) szacowany był według metody największej wiarygodności, a model T+AR(2) – według metody najmniejszych kwadratów.

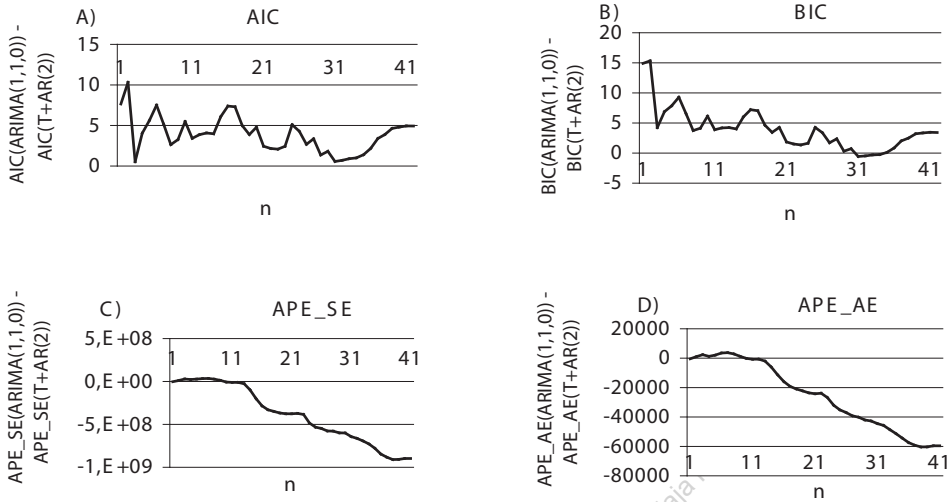
(lata 1969–1970)). Natomiast przebieg różnic między skumulowanym błędem APE_AE dla obu modeli wskazuje na przewagę modelu ARIMA w okresie prognozowania od 4. do 35. (dla lat 1961–1993) – bo różnica kryteriów APE_SE dla modeli niemalże w całym okresie prognozowania jest ujemna (czyli model ARIMA daje niższe błędy prognoz na jeden okres naprzód), a od okresu 36. do 43. (dla lat 1994–2000) – na przewagę modelu T+AR(2), bo różnica kryteriów APE_SE dla dwóch modeli jest dodatnia (czyli model T+AR(2) daje niższe błędy prognoz na jeden okres naprzód). Wynika z tego, że przyjęcie różnych postaci funkcji straty będzie miało wpływ na decyzję o wyborze modelu.



Wykres 1. Różnice między kryteriami wyboru dla modelu ARIMA(1,1,0) i modelu T+AR(2) wyznaczającego prognozy PKB (Francja). Panel A – kryterium AIC, panel B – kryterium BIC, panel C – kryterium APE_SE, panel D – kryterium APE_AE

Źródło: opracowanie własne.

Dla szeregu PKB (Polska) – por. wykres 2 (panele A i B) – kryteria AIC i BIC dają przewagę modelowi T+AR(2), ponieważ różnica kryteriów $AIC(ARIMA) - AIC(T+AR(2))$ oraz $BIC(ARIMA) - BIC(T+AR(2))$ jest dodatnia. Jednakże w przypadku kryterium BIC przewaga modelu T+AR(2) zmniejsza się w miarę wzrostu liczby obserwacji, ponieważ różnica kryteriów maleje (wykres 2, panel B). Natomiast przebieg skumulowanego błędu prognoz na jeden okres naprzód (APE_SE, APE_AE – panele C i D) wyraźnie wskazuje na przewagę modelu ARIMA, ponieważ różnice między skumulowanymi błędami APE_SE

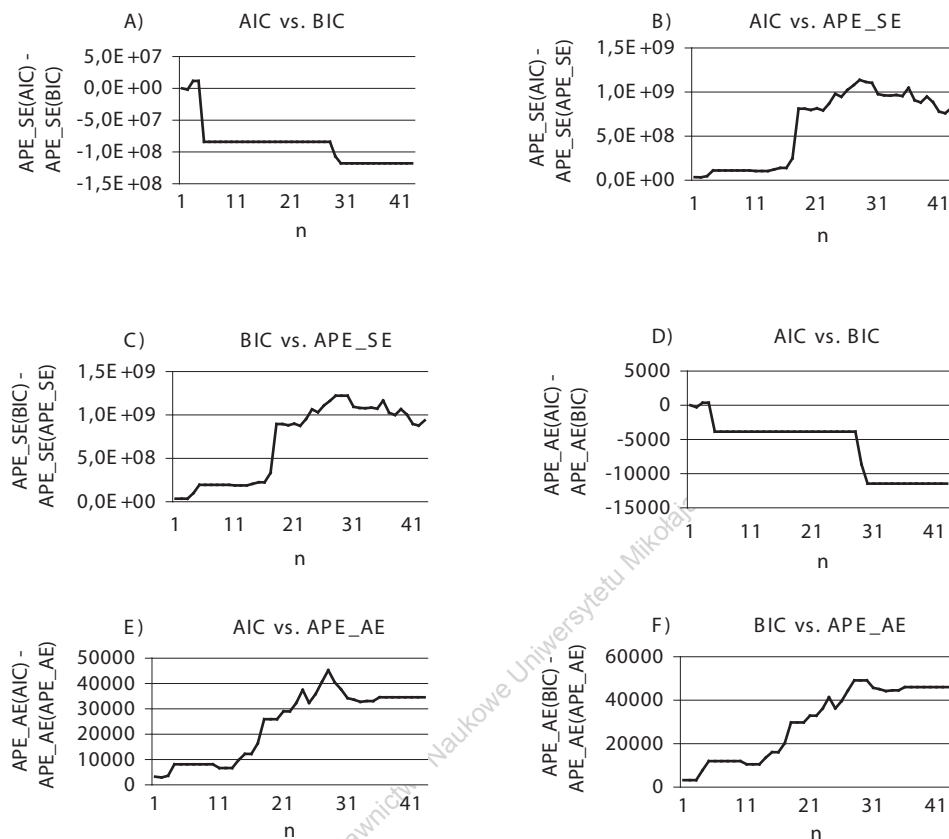


Wykres 2. Różnice między kryteriami wyboru dla modelu ARIMA(1,1,0) i modelu T+AR(2) wyznaczającego prognozy PKB (Polska). Panel A – kryterium AIC, panel B – kryterium BIC, panel C – kryterium APE_SE, panel D – kryterium APE_AE

Źródło: opracowanie własne.

(a także APE_AE) dla obu modeli są ujemne w niemal całym okresie prognozowania (z wyjątkiem okresu od 2. do 9., tj. dla okresu 1960 do 1968), co oznacza mniejsze błędy prognoz z modelu ARIMA.

Skumulowany błąd prognoz na jeden okres naprzód (APE) może być również zastosowany do oceny strategii wyboru modelu w kontekście zachowania w prognozowaniu (meta-wybór). W takim przypadku celem jest ocena prognozytnych walorów poszczególnych strategii (AIC, BIC, APE), a nie poszczególnych modeli (ARIMA, T+AR(2)). Podobnie jak obliczenie APE we wcześniejszym podejściu, w procedurze meta-wyboru konieczne jest również oszacowanie modeli ARIMA i T+AR(2) na podstawie kolejnych prób powiększających się o jedną obserwację. Wartość predycyjna np. strategii AIC jest określana z wykorzystaniem skumulowanego błędu prognoz (APE) obliczanego dla modeli wybranych za pomocą kryterium AIC. Przykładowo, założmy, że dla zwiększającej się liczby obserwacji aż do $n = 20$ kryterium AIC preferuje model ARIMA, a następnie dla $n > 20$ kryterium AIC zaczyna preferować model T+AR(2). Wtedy skumulowany błąd prognoz (APE) jest sumą błędów prognoz *ex post* otrzymanych z modeli ARIMA i T+AR(2) (odpowiednio dla pierwszej i drugiej części próby) wybranych według minimum kryterium AIC. Obliczając różnice APE dla różnych strategii wyboru modelu, uzyskuje się relatywną ocenę przydatności



Wykres 3. Meta-wybór pomiędzy strategiami: AIC, BIC, APE_SE i APE_AE dla szeregu PKB (Francja)

Źródło: opracowanie własne.

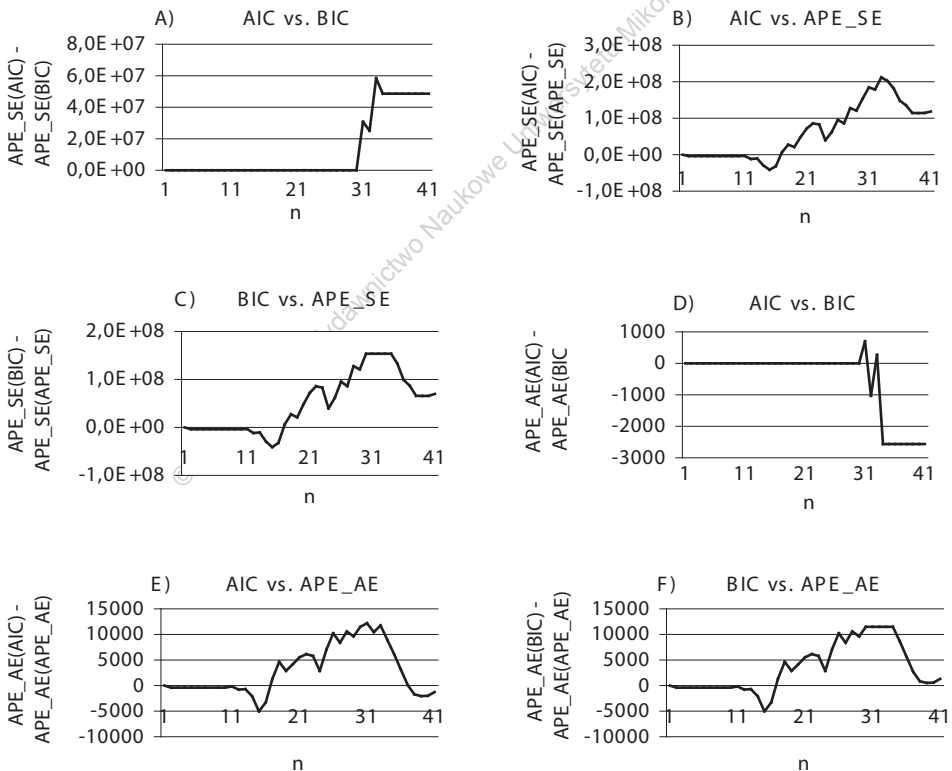
poszczególnych strategii wyboru modelu. Na wykresie 3 przedstawione są różnice między skumulowanymi błędami prognoz (APE) otrzymanymi w wyniku zastosowania poszczególnych strategii wyboru modelu (tj. AIC, BIC, APE_SE i APE_AE).

Dla badanego szeregu PKB (Francja) zastosowanie strategii AIC do wyboru modelu daje mniejsze błędy prognoz niż w przypadku strategii BIC – por. panel A, na którym różnica między skumulowanym błędem APE_SE obliczonym przy zastosowaniu strategii AIC i BIC ($APE_SE(AIC) - APE_SE(BIC)$)⁹

⁹ Skrót np. $APE_SE(AIC)$ oznacza skumulowany błąd prognoz (jako funkcję straty przyjęto błąd kwadratowy) obliczony w przypadku, gdy do wyboru spośród dwóch modeli, ARIMA i T+AR(2), zastosowano strategię AIC.

jest ujemna. Należy zauważyć, że poziome odcinki na wykresie 3 oznaczają, że różnice w błędach prognoz dla dwóch strategii wyboru modelu (np. AIC i BIC, panel A) nie zmieniają się. Występuje to wtedy, gdy dwie strategie wyboru modelu preferują ten sam model. Podobne wnioski otrzymuje się w przypadku zastosowania absolutnego błędu (AE) jako funkcji straty (por. panel D).

Porównując zachowanie strategii AIC i APE_SE oraz BIC i APE_SE (panele B i C, wykres 3), wyraźnie widać, że stosując strategię APE_SE do wyboru modelu, otrzymuje się mniejsze skumulowane błędy prognoz niż przy zastosowaniu strategii AIC czy BIC (ponieważ różnice $APE_SE(AIC) - APE_SE(APE_SE)$ są dodatnie) w całym okresie prognozowania. Podobne wyniki uzyskano, gdy jako funkcję straty przyjęto błąd absolutny (panele E i F), przy czym dla ostatnich dziesięciu okresów różnica błędów APE_AE dla poszczególnych strategii jest stała, co sugeruje podobne zachowanie się tych strategii przy wyborze modelu.



Wykres 4. Meta-wybór pomiędzy strategiami: AIC, BIC, APE_SE i APE_AE dla szeregu PKB (Polska)

Źródło: opracowanie własne.

Ogólnie jednak zastosowanie strategii APE_SE (czy APE_AE) pozwala otrzymać mniejsze skumulowane błędy prognoz na jeden okres naprzód niż w przypadku strategii AIC czy BIC.

Dla szeregu PKB (Polska) – wykres 4 – strategię AIC i BIC dają ten sam wybór (różnica $APE_SE(AIC) - APE_SE(BIC)$ jest równa zero) dla pierwszych 30 okresów prognozowania (tj. dla okresu (1960–1990), natomiast dla pozostałego okresu zastosowanie strategii BIC do wyboru modelu daje mniejsze błędy prognoz niż w przypadku strategii AIC (wykres 4, panel A), ponieważ różnica błędów APE_SE dla strategii AIC i BIC jest dodatnia. Jeżeli jednak jako funkcję straty wybierze się błąd absolutny (AE), to wniosek jest odwrotny, tj. przewagę ma strategia AIC, ponieważ różnica błędów APE_AE dla strategii AIC i BIC jest ujemna (por. panel D). Potwierdza to wcześniejszy wniosek o wpływie wyboru funkcji straty na wybór modelu i na wybór samej strategii.

Z porównania zachowania strategii AIC i APE_SE oraz BIC i APE_SE (panele B i C, wykres 4) wynika, że stosując strategię APE_SE do wyboru modelu, otrzymuje się mniejsze skumulowane błędy prognoz niż przy zastosowaniu strategii AIC czy BIC (ponieważ różnice $APE_SE(AIC) - APE_SE(APE_SE)$ są dodatnie) w prawie całym okresie prognozowania (z wyjątkiem pierwszych 15 okresów – lata 1959–1965, gdy różnice błędów APE_SE dla strategii AIC i APE_SE oraz BIC i APE_SE są ujemne, co wskazuje na przewagę strategii odpowiednio AIC i BIC). Podobne wyniki otrzymuje się w wyniku porównania strategii AIC i APE_AE oraz BIC i APE_AE (panele E i F, wykres 4), przy czym pod koniec badanego okresu zauważyć można relatywny spadek tej przewagi, ponieważ różnica błędów APE_AE dla odpowiednich par strategii znacznie spadła, choć nadal jest dodatnia.

W celu sprawdzenia poprawności wyboru modelu ARIMA czy T+AR(2) według skumulowanego błędu prognoz *ex post* (APE_SE i APE_AE) wyznaczono prognozy PKB (Francja) i PKB (Polska) na jeden okres naprzód dla okresu 2001–2003 oraz błędy prognoz *ex post* (bezwzględne δ_T i względne δ'_T) na podstawie obu modeli. Wyniki przedstawiają tabele 1 i 2.

Tabela 1. Prognozy PKB (Francja) na jeden okres naprzód i błędy prognoz *ex post* z modeli ARIMA(1,1,0) i T+AR(2)

Lata	Realizacja	Model: ARIMA(1,1,0)			Model: T+AR(2)		
		prognoza	δ_T	δ'_T	prognoza	δ_T	δ'_T
2001	1289387	1297071	-7684,3	-0,60%	1292864	-3477,0	-0,27%
2002	1305136	1312186	-7050,8	-0,54%	1309083	-3947,3	-0,30%
2003	1315601	1323622	-8021,1	-0,61%	1321281	-5680,3	-0,43%

Źródło: obliczenia własne.

Tabela 2. Prognozy PKB (Polska) na jeden okres naprzód i błędy prognoz *ex post* z modeli ARIMA(1,1,0) i T+AR(2)

Lata	Realizacja	Model: ARIMA(1,1,0)			Model: T+AR(2)		
		prognoza	δ_T	δ'_T	prognoza	δ_T	δ'_T
2001	281508	286913	-5406	-1,92%	286307	-4798,8	-1,70%
2002	285365	284901	464	0,16%	283789	1575,8	0,55%
2003	296237	289382	6856	2,31%	288394	7843,2	2,65%

Źródło: obliczenia własne.

Na podstawie tabeli 1 można zauważyć, że błędy prognoz *ex post* na jeden okres naprzód są mniejsze w przypadku modelu T+AR(2), co potwierdzałyby wskazanie skumulowanego błędu prognoz APE_AE (por. wykres 1, panel D). Jednakże błędy prognoz *ex post* w przypadku modelu ARIMA(1,1,0) są również niewielkie, co wskazywałoby także na przydatność tego modelu w prognozowaniu. Oznacza to, że w przypadku szeregu PKB (Francja) oba modele dobrze zachowują się w prognozowaniu, choć przewagę ma model T+AR(2). Podobne wyniki uzyskuje się przy prognozowaniu PKB (Polska), z tym że w tym przypadku mniejsze błędy prognoz otrzymuje się z modelu ARIMA, na który wskazywała metoda APE (por. wykres 2, panele C i D).

PODSUMOWANIE

Zaprezentowany przykład empiryczny wskazuje na przydatność skumulowanego błędu prognoz na jeden okres naprzód jako metody wyboru modelu. Dodatkowo śledzenie zmian skumulowanego błędu prognoz w miarę zwiększającej się liczby obserwacji pokazuje wyraźnie, że pojęcie najlepszego modelu należy odnosić do liczebności próby, tzn. model najlepszy przy danej wielkości próby może być, w przypadku większej liczby obserwacji, zastąpiony przez inny model, który daje prognozy o mniejszych błędach.

Metoda ta może być stosowana w odniesieniu zarówno do modeli zagnieżdżonych, jak i niezagnieżdżonych oraz w przypadku, gdy w zestawie modeli nie występuje model generujący dane. Niewątpliwą zaletą skumulowanego błędu prognoz jest to, że może być stosowany nie tylko do wyboru modelu w ramach danej strategii wyboru, ale również do wyboru samej strategii, a przez to do porównywania różnych strategii wyboru modelu. Rozszerza to zagadnienie wyboru modelu i sprawia, że metoda ta może mieć duże znaczenie praktyczne.

LITERATURA

- Aitchison J., Dunsmore I. R. (1975), *Statistical Prediction Analysis*, Cambridge University Press, Cambridge.
- Akaike H. (1973), *Information Theory and an Extension of the Maximum Likelihood Principle*, [w:] Petrov B. N., Csaki F., *Second International Symposium on Information Theory*, Akademia Kiado, Budapest.
- Burnham K. P., Anderson D. R. (2002), *Model Selection and Multimodel Inference*, Springer.
- Christiano L. J., Eichenbaum M. (1990), *Unit Roots in Real GNP: Do We Know and Do We Care?*, „Carnegie-Rochester Conference Series on Public Policy”, nr 32, 7–61.
- Clarke B. (2001), *Combining Model Selection Procedures for Online Prediction*, „Sahkhyā: The Indian Journal of Statistics”, 63, seria A, 229–249.
- Dawid A. P. (1984), *Statistical Theory: the Prequential Approach*, „Journal of Royal Statistical Society Series B”, 147, 278–292.
- De Luna X., Skouras K. (2003), *Choosing a Model Selection Strategy*, „Scandinavian Journal of Statistics”, 30, 113–128.
- Diebold F. X., Senhadji A. (1996), *Deterministic vs. Stochastic Trend in U.S. GNP. Yet again*, NBER Working Papers, nr 5481.
- Haubrich J. G., Lo A. W. (2001), *The source and nature of long-term memory in aggregate output*, Federal Reserve Bank of Cleveland, „Economic Review”, QII, 15–30.
- Maddison A. (2001), *The World Economy – a millennial perspective*, OECD Development Centre, Paris.
- Murray C., Nelson C. (1998), *The Uncertain Trend in U.S. GNP*, Discussion Papers in Economics at the University of Washington, nr 0074.
- Nelson C. R., Plosser C. I. (1982), *Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications*, „Journal of Monetary Economics”, 10(2), 139–162.
- Perron P., Phillips P. C. B. (1987), *Does GNP Have a Unit Root?*, „Economics Letters”, 23, 129–145.
- Piłatowska M. (2009), *Prognozy kombinowane z wykorzystaniem wag Akaike’a*, „Acta Universitatis Nicolai Copernici. Ekonomia”, 39, 51–62.
- Piłatowska M. (2010), *Kryteria informacyjne w wyborze modelu ekonometrycznego*, „Studia i Prace Uniwersytetu Ekonomicznego w Krakowie”, 25–37.
- Quah D. (1987), *What do we Learn from Unit Roots in Macroeconomic Series?*, NBER Working Papers, nr 2450.
- Rissanen J. (2003), *Complexity of Simple Nonlogarithmic Loss Function*, „IEEE Transactions on Information Theory”, 49, 476–484.
- Rudebusch G. D. (1993), *The Uncertain Unit Root in Real GNP*, „American Economic Review”, 83(1), 264–272.
- Shao J. (1997), *An Asymptotic Theory for Linear Model Selection*, „Statistica Sinica”, 7, 221–264.

- Skouras K., Dawid A. P. (1998), *On Efficient Point Prediction Systems*, „Journal of Royal Statistical Society” B, 60, 765–780.
- Sugiura N. (1978), *Further Analysis of the Data by Akaike’s Information Criterion and the Finite Corrections*, „Communications in Statistics, Theory and Methods”, A7, 13–26.
- Stock J., Watson M. (1986), *Does GNP Have a Unit Root?*, „Economics Letters”, 22(2/3), 147–151.
- Wagenmaker E.-J., Grünwald P., Steyvers M. (2006), *Accumulative Prediction Error and the Selection of Time Series Models*, „Journal of Mathematical Psychology”, 50, 149–166.

ACCUMULATIVE PREDICTION ERROR AS A METHOD OF MODEL SELECTION

Abstract. The purpose of the paper is to present and apply the accumulative one-step-ahead prediction error (APE) not only as a method (strategy) of model selection, but also as a tool of model selection strategy (meta-selection). The APE method is compared with the information approach to model selection (AIC and BIC information criteria), supported by empirical examples. Obtained results indicated that the APE method may be of considerable practical importance.

Key words: model selection, meta-selection, information criteria, accumulative prediction error.