EKONOMETRIA ECONOMETRICS 4(42) • 2013
ISSN 1507-3866

Marcin Łupiński
Narodowy Bank Polski

PORÓWNANIE JAKOŚCI PROGNOZOWANIA POLSKIEGO PKB DYNAMICZNYMI MODELAMI CZYNNIKOWYMI ORAZ CZYNNIKOWYMI MODELAMI MIDAS

Streszczenie: Artykuł jest kontynuacją cyklu opracowań autora [2007; 2009; 2012] dotyczących optymalnych metod prognozowania polskich zmiennych makroekonomicznych na przykładzie dynamiki produktu krajowego brutto. W ramach wykonanego na potrzeby artykułu badania porównano jakość nowcastów („prognoz" teraźniejszości) i właściwych prognoz określonych na podstawie proponowanego przez Mariano i Murasawę [2003] dynamicznego modelu czynnikowego z obsługą mieszanych częstotliwości danych wejściowych i braków danych (MFDG-DFM) oraz rozszerzonego o strukturę czynnikową modelu regresji MIDAS (DFM-MIDAS) opracowanego pierwotnie przez Marcellino i Schumachera [2008]. Przedstawiono również zaplecze matematyczne obu modeli, wskazując na kombinowane podejście filtru Kalmana oraz metody największej wiarygodności jako metodę estymacji obu struktur. Uzyskane wyniki wskazują na przewagę modelu Mariano i Murasawy (o ok. 15% bardziej trafne prognozy niż w przypadku konkurenta), choć w zakresie nowcastów i prognoz na kwartał do przodu model ten musi uznać wyższość czynnikowego wariantu podejścia MIDAS.

Słowa kluczowe: dynamiczne modele czynnikowe, MIDAS, prognozy polskiego PKB.

1. Wstęp

Dostęp do trafnej oceny stanu bieżącego oraz prognoz sytuacji makroekonomicznej danego kraju stanowi jeden z fundamentalnych warunków prowadzenia polityki społecznej, fiskalnej, monetarnej oraz makrostabilnościowej. Wśród grupy dostarczonych przez wspomniane środowiska prognoz zmiennych szczególna uwaga decydentów koncentruje się na dynamice produktu krajowego brutto odzwierciedlającej bieżące i przyszłe trendy rozwoju gospodarczego przedmiotowego kraju oraz stanowiącej zmienną wejściową wielu analiz spodziewanych reakcji gospodarki na prowadzone polityki gospodarcze.

Na potrzeby decydentów odpowiedzialnych za poszczególne rodzaje polityk budowane są rozbudowane zaplecza analityczne, w ramach których prognozy otoczenia makroekonomicznego formułowane są za pomocą złożonych środowisk analitycznych, takich jak modele strukturalne czy też dynamiczne stochastyczne modele równowagi ogólnej (DSGE). Mając na uwadze zgłaszany przez decydentów popyt na modele prognoz kluczowych zmiennych makroekonomicznych, wśród badaczy akademickich, począwszy od przełomu lat 80. i 90. ubiegłego wieku, dużą popularność w zakresie prognozowania szeregów czasowych zdobyły dynamiczne modele czynnikowe (*dynamic factor models*), które na podstawie niewielkiej grupy zmiennych makroekonomicznych (z reguły liczebność zbioru nie przekraczała 10 szeregów czasowych) pozwalają na konstruowanie prognoz (szczególnie prognoz krótkookresowych) o jakości porównywalnej z uzyskiwanymi za pomocą modeli strukturalnych lub DSGE lub wyższej od nich. W klasie wspomnianych modeli szczególne uznanie zdobyły dynamiczne modele czynnikowe pozwalające na łączenie różnych częstotliwości danych wejściowych oraz zapewniające obsługę braków obserwacji. Modele te pozwoliły na uchwycenie dynamiki oraz podstawowych zależności pomiędzy kluczowymi zmiennymi makroekonomicznymi, przez co zaczęły być stosowane jako modele pomocnicze lub nawet jako modele referencyjne przez instytucje międzynarodowe, banki centralne oraz urzędy państwowe odpowiedzialne za przygotowanie analiz na potrzeby kluczowych polityk prowadzonych w skali makro.

Na początku bieżącego stulecia akademicy opracowali alternatywne podejście do modelowania w ramach jednej struktury szeregów czasowych zbieranych z różniącymi się częstotliwościami. Mowa tu o modelach regresyjnych z mieszanymi częstotliwościami próbkowania (MIDAS – *MIxed DAta Sampling regressions*). W pierwszym etapie modele te występowały jako proste struktury jednorównaniowe, po kilku latach rozwoju ewoluowały do postaci struktur wielowymiarowych, zapewniających możliwość modelowania nieobserwowalnych komponentów (czynników).

Celem niniejszej pracy jest dokonanie porównania obu podejść (dynamicznych modeli czynnikowych z mieszanymi częstotliwościami i obsługą braków danych oraz modeli MIDAS ze strukturami czynnikowymi), w zakresie zarówno teoretycznym, jak i praktycznym, do prognozowania teraźniejszości (nowcasty) i przyszłego przebiegu zmiennych makroekonomicznych na przykładzie dynamiki polskiego produktu krajowego brutto. W ramach komparatystyki omówione zostaną zarówno podstawy teoretyczne obu metod, jak i sposoby ich estymacji. Kluczowym elementem wykonanego badania jest porównanie wielkości błędów nowcastów i prognoz uzyskanych za pośrednictwem obu metod pozwalające na wskazanie metody dostarczającej wyższej jakości przewidywań dotyczących aktualnego i przyszłego rozwoju sytuacji.

2. Przegląd literatury

Dynamiczne modele czynnikowe zostały wprowadzone do analizy ekonomicznej oraz finansowej przez Geweke [1977], a następnie spopularyzowane przez serię artykułów Stocka i Watsona (m.in. [1989; 1998]). Modele obu wymienionych badaczy miały charakter niewielkich struktur analitycznych obejmujących swoim zakresem maksymalnie kilkanaście zmiennych. Próbę generalizacji modeli czynnikowych do obliczeń wykorzystujących duże zbiory zmiennych podjęli Forni, Hallin, Lippi i Reichlin [2001] w swoim frameworku, zwanym uogólnionymi modelami czynnikowym (*generalized dynamic factor models*). Praktyczne zastosowanie modeli czynnikowych do prognozowania zmiennych makroekonomicznych wskazało jednak na inne niż rozmiar próby problemy, co do których stawiano hipotezy, że ich rozwiązanie może przynieść istotną poprawę trafności oczekiwanych scenariuszy rozwoju sytuacji. Wzmiankowanymi problemami okazały się różnorodność częstotliwości danych wejściowych oraz brak części obserwacji dla wybranych zmiennych. Wariant dynamicznych modeli czynnikowych rozwiązujących oba przedstawione problemy zaproponowali Mariano i Murasawa [2003]. Praktyczne zastosowania przedmiotowej struktury analitycznej przedstawili niemal równolegle Camacho i Perez-Quirios [2009] oraz Arouba, Diebold i Scotti [2009].

Podstawy modeli MIDAS przedstawione zostały w serii artykułów [Ghysels, Santa-Clara i Valkanov 2004; Ghysels, Sinko i Valkanov 2006]. Połączenie obu podejść (regresji MIDAS oraz modeli dynamicznych czynnikowych) dokonane zostało po raz pierwszy w pracy Marcellino i Schumachera [2008]. We wspomnianej pracy autorzy dokonali oceny jakości prognozowania niemieckiego PKB za pomocą kombinowanej metody czynnikowego MIDAS. Praktyczne zastosowanie czynnikowej odmiany modeli MIDAS można również znaleźć w pracy Frale i Montefortego [2010].

Autor niniejszego artykułu konsekwentnie rozwija metody analityczne służące do prognozowania kluczowych krajowych zmiennych makroekonomicznych. Opis podejścia czynnikowego można znaleźć w jego pracach z 2007 oraz 2009 r. Z kolei w 2012 r. opublikowany został artykuł, w którym porównana została jakość prognoz krajowego PKB uzyskanych za pomocą podstawowego dynamicznego modelu czynnikowego oraz modelu czynnikowego obejmującego mieszane częstotliwości i obsługę braku danych. Przedmiotowy artykuł jest więc częścią większej serii opracowań mających na celu dobór optymalnych metod prognozowania polskich zmiennych makroekonomicznych.

3. Dynamiczny model czynnikowy z mieszanymi częstotliwościami i niezbilansowaną próbą

Jako model referencyjny przedstawiony zostanie model zaproponowany przez Mariano i Murasawę [2003], który następnie rozwinięty został przez Camacho i Pereza-Quiriosa [2009; 2010]. Mariano, Murasawa oraz Camacho i Perez-Quirios w prosty koncepcyjnie sposób rozszerzyli możliwości standardowego środowiska Stocka i Watsona [1989; 1998], opartego na dynamicznym podejściu czynnikowym, opisanym w przestrzeni stanów przez obsługę mieszanych częstotliwości i niezbilansowanej próby (*ragged edges*). W pierwszej części bieżącej sekcji przedstawiony zostanie szkic bazowego modelu dynamicznych czynników, a następnie omówione zostaną elementy uzupełniające jego strukturę dodane przez wymienionych badaczy.

Przyjmijmy, że dana jest sekwencja (długości T) wektorów obserwacji N makroekonomicznych szeregów czasowych:

$$Y = [Y_1, Y_2, ..., Y_T] \tag{1}$$

Wspomniane szeregi opisują bieżącą sytuację gospodarczą danego kraju. Wektor obserwacji tychże szeregów zebrany dla danego okresu t może być postrzegany jako realizacja wielowymiarowego procesu stochastycznego Y, który będzie modelowany w ramach zaproponowanej struktury ekonometrycznej. Przyjmowane jest również dodatkowe założenie, iż siłą wiodącą wspomnianego procesu stochastycznego jest grupa nieobserwowalnych komponentów (czynników) reprezentujących bieżący stan, w którym znajduje się analizowana gospodarka.

Wygodnym sposobem zapisu modelu zawierającego nieobserwowalne czynniki jest przestrzeń stanów (*state-space*) składająca się z dwóch bloków równań:
- pomiarowego, określającego relację pomiędzy obserwowanymi zmiennymi i nieobserwowalnymi czynnikami,
- przejścia, określającego dynamikę w czasie nieobserwowalnych komponentów:

$$Y_t = H\beta_t + Az_t + e_t \tag{2}$$

$$\beta_t = \mu + F\beta_{t-1} + \nu_t \tag{3}$$

$$\begin{pmatrix} e_t \\ \nu_t \end{pmatrix} \sim N\left(0, \begin{pmatrix} R & 0 \\ 0 & Q^* \end{pmatrix}\right) \tag{4}$$

gdzie: H, A, F są macierzami o wymiarach $(n \times r)$, $(n \times k)$ i $(r \times r)$, a e_t i v_t reprezentują błędy typu i.i.d pochodzące z rozkładu gaussowskiego z odpowiadającymi im macierzami R i Q.

Przy tak zdefiniowanym modelu celem ekonometryka jest estymacja nieobserwowalnych czynników zgrupowanych w wektorze $\beta_{t\text{-}1}$ dla wszystkich okresów próby ($t = 1, 2, …, T$). Estymacja nieobserwowalnych czynników może zostać przeprowadzona pod warunkiem znajomości parametrów strukturalnych modelu (elementów macierzy H, A, F, R, Q). Jako że parametry te nie są znane, etap właściwego obliczania czynników musi zostać poprzedzony estymacją tychże parametrów. Z reguły w tym celu stosowana jest metoda największej wiarygodności (MNW), chociaż w literaturze relatywnie często używane jest również podejście bayessowskie. Poniżej przedstawiona zostanie iteracyjna estymacja nieobserwowalnych czynników za pomocą filtru Kalmana; sposób użycia MNW do określenia elementów macierzy opisany zostanie w kolejnej części sekcji. W każdej iteracji filtru Kalmana (indeksowanej subskryptem t) na podstawie zbioru informacji dostępnego w okresie t–1 $\psi_{t\text{-}1}$,/$\psi_{t\text{-}1} = \{Y_{t\text{-}1},Y_{t\text{-}2}, ..., Y_1,z_{t\text{-}1},z_{t\text{-}2}, ..., z_1\}$/formułowana jest warunkowa wartość oczekiwana wektora nieobserwowalnych czynników przypisana do referencyjnego okresu t:

$$\beta_{t|t-1} = E[h_t \mid \psi_{t-1}] \tag{5}$$

Ze sformułowaną warunkową wartością oczekiwaną związana jest macierz wariancji kowariancji błędu, określona równaniem:

$$P_{t|t-1} = E[(\beta_t - \beta_{t|t-1})(\beta_t - \beta_{t|t-1})'] \tag{6}$$

W praktyce w każdej z iteracji obliczana jest prognoza wektora nieobserwowalnych czynników na jeden okres do przodu (dla okresu t na bazie okresu $t-1$)

$$\beta_{t|t-1} = \mu + F\beta_{t-1|t-1} \tag{7}$$

oraz powiązana z nią macierz błędu średniokwadratowego:

$$P_{t|t-1} = F\beta_{t-1|t-1}F' + Q_j^* \tag{8}$$

Następnie w celu zapewnienia zgodności z używanymi w modelu obserwowanymi danymi na bazie obliczonych nieobserwowalnych komponentów formułowana jest ich jednookresowa prognoza:

$$\eta_{t|t-1} = Y_t - H\beta_{t|t-1} - Az_t \tag{9}$$

i obliczana pochodna dla niej macierz średniokwadratowego błędu prognozy:

$$f_{t|t-1} = HP_{t|t-1}H' + R \tag{10}$$

Wielkość błędu stanowi podstawę do korekty jednookresowej prognozy wektora nieobserwowalnych komponentów:

$$\beta_{t|t} = \beta_{t|t-1} + P_{t|t-1}H'[f^{(}_{t|t-1}]^{-1}\eta_{t|t-1} \tag{11}$$

Korekta prognozy pociąga za sobą również zmianę elementów macierzy błędów

$$P_{t|t} = (I - P_{t|t-1}H'[f^{(}_{t|t-1}]^{-1}H)P_{t|t-1} \tag{12}$$

Powyższe równanie kompletuje pojedynczą iterację filtru Kalmana.

Jak zostało to opisane wcześniej, estymacja nieobserwowalnych komponentów wymaga zastosowania MNW do estymacji strukturalnych parametrów modelu. Dla modelu zapisywana jest łączna funkcja wiarygodności oparta na łącznej funkcji gęstości (zakładana jest niezależność obserwacji):

$$l(\theta \mid Y) = \prod_{t-1}^{T} f(Y_t|\psi_{t-1}, \theta] \tag{13}$$

gdzie θ jest zbiorem jego strukturalnych parametrów. W zastosowaniach praktycznych obliczana jest logarytmiczna postać funkcji wiarygodności:

$$ll(\theta \mid Y) = \sum_{t-1}^{T} f(Y_t|\psi_{t-1}, \theta] \qquad (14)$$

Przy (stosunkowo restrykcyjnym) założeniu, że obserwacje szeregów czasowych używanych do estymacji pochodzą z rozkładu normalnego, poprzednie równanie uzyskuje skonkretyzowaną postać:

$$ll(\theta \mid \tilde{y}) = \sum_{t=1}^{T} (2\pi)^{-\frac{N}{2}} |f_{t|t-1}|^{-\frac{1}{2}} exp[-\frac{1}{2}\eta'_{t|t-1} f_{t|t-1} n_{t|t-1}] \qquad (15)$$

Jednocześnie przy zastosowaniu nierówności Cramera-Rao możliwe jest obliczenie macierzy kowariancji estymatora wektora parametrów $\hat{\theta}$. Jest to odwrotność ujemnego Hessianu funkcji logarytmicznej wiarygodności:

$$\left[-\frac{\partial^2 ll(\theta \mid Y)}{\partial\theta\partial\theta'}\right]^{-1} \qquad (16)$$

dla $\theta = \hat{\theta}$.

Zaproponowane przez Mariano i Murasawę oraz przez Pereza-Quiriosa rozszerzenia standardowego dynamicznego modelu czynnikowego zapisanego w przestrzeni stanów pozwalają na prowadzenie estymacji, w przypadku gdy dostępne dane mają różne częstotliwości oraz mają niezbilansowany koniec próby. Pierwsze z rozszerzeń pozwala na zapisanie relacji pomiędzy danymi o niskiej częstotliwości i wysokiej częstotliwości na podstawie przybliżenia geometryczną średnią średniej arytmetycznej. Powinno ono być stosowane wyłącznie do danych mających charakter przepływów. W poniższym przykładzie relacją związane są dane kwartalne Y z danymi miesięcznymi X:

$$\begin{aligned}
&Y_t = 3\frac{X_t + X_{t-1} + X_{t-2}}{3} \backsimeq 3(X_t X_{t-1} X_{t-2})^{1/3} \\
&ln(Y_t) = ln(3) + \tfrac{1}{3}ln(X)_t + \tfrac{1}{3}ln(X_{t-1}) + \tfrac{1}{3}ln(X_{t-2}) \\
&ln(Y_t) - ln(Y_{t-3}) = \tfrac{1}{3}(lnX_t - \tfrac{1}{3}lnX_{t-3}) + \tfrac{1}{3}(lnX_{t-1} - \tfrac{1}{3}lnX_{t-4}) + .. \\
&Y_t^{\triangle q} = \tfrac{1}{3}X_t^{\triangle m} + \tfrac{2}{3}X_{t-1}^{\triangle m} + X_{t-2}^{\triangle m} + \tfrac{2}{3}X_{t-3}^{\triangle m} + \tfrac{1}{3}X_{t-4}^{\triangle m}
\end{aligned} \qquad (17)$$

gdzie: $Y_t^{\triangle q} = Y_t - Y_{t\text{-}3}$, $X_t^{\triangle m} = X_t - X_{t\text{-}1}$.

W przypadku gdy wszystkie dane są dostępne, powyższa aproksymacja pozwala zapisać model jako:

$$
\begin{aligned}
&\begin{bmatrix} Y_{1,t} \\ Y_{j,t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_j \end{bmatrix} + \begin{bmatrix} \gamma_1(\frac{1}{3}h_t + \frac{2}{3}h_{t-1} + h_{t-2} + \frac{2}{3}h_{t-3} + \frac{1}{3}h_{t-4}) \\ \gamma_j h_t \end{bmatrix} \\
&+ \begin{bmatrix} \frac{1}{3}i_{1,t} + \frac{2}{3}i_{1,t-1} + i_{1,t-2} + \frac{2}{3}i_{1,t-3} + \frac{1}{3}i_{1,t-4} \\ i_{j,t} \end{bmatrix} \\
&\Phi(L)h_t = v_{h,t} \\
&\Theta(L)i_{j,t} = v_{i,j,t} \\
&\begin{bmatrix} v_{h,t} \\ v_{i,j,t} \end{bmatrix} \sim N\left(0 \; , \begin{bmatrix} \sigma_h & 0 \\ 0 & \Sigma_i \end{bmatrix}\right)
\end{aligned} \tag{18}
$$

gdzie wektor β zostaje rozbity na czynniki wspólne (h) i idiosynkratyczne (i)

Drugim z wprowadzonych rozszerzeń jest zagnieżdżony mechanizm obsługi braków obserwacji wewnątrz oraz na końcu próby (niezbilansowany koniec próby). Rozwiązanie przyjęte dla obu wariantów jest takie samo w przypadku obu wariantów. Brakujące obserwacje są zastępowane wartościami wylosowanymi z rozkładu gaussowskiego o parametrach równych ich estymatorom obliczonym dla wejściowych szeregów czasowych.

$$
\bar{Y}_{k,t} = \begin{cases} Y_{k,t}\, jeśli\, Y_{k,t}\, jest\, dostępne \\ w_{k,t}\, jeśli\, Y_{k,t}\, jest\, niedostępne \end{cases} \tag{19}
$$

gdzie $w_{j,t} \backsim N(\mu,\sigma)$, μ,σ są estymatorami dwóch pierwszych momentów. Ponadto obsługa braków danych pociąga za sobą konieczność modyfikacji struktury modelu oraz czyni proces estymacji dwuwariantowym (przykład dla obsługi braków danych pierwszej zmiennej):

$$
\begin{aligned}
&\begin{bmatrix} Y_{1,t} \\ Y_{j,t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_j \end{bmatrix} + \begin{bmatrix} \gamma_1(\frac{1}{3}h_t + \frac{2}{3}h_{t-1} + h_{t-2} + \frac{2}{3}h_{t-3} + \frac{1}{3}h_{t-4}) \\ \gamma_j h_t \end{bmatrix} \\
&+ \begin{bmatrix} \frac{1}{3}i_{1,t} + \frac{2}{3}i_{1,t-1} + i_{1,t-2} + \frac{2}{3}i_{1,t-3} + \frac{1}{3}i_{1,t-4} \\ i_{j,t} \end{bmatrix} \\
&jeśli\; Y_{1,t} jest\; dostępne \\
&\begin{bmatrix} Y_{1,t} \\ Y_{j,t} \end{bmatrix} = \begin{bmatrix} 0 \\ \mu_j \end{bmatrix} + \begin{bmatrix} 0 \\ \gamma_j h_t \end{bmatrix} + \begin{bmatrix} 0 \\ i_{j,t} \end{bmatrix} + \begin{bmatrix} w_{1,t} \\ 0 \end{bmatrix} \; jeśli\; \tilde{y}_{1,t} nie\; jest\; dostępne \\
&w_{1,t} = N(\delta_1, \sigma_1) \\
&\ldots
\end{aligned} \tag{20}
$$

Jak widać z powyższego opisu, rozszerzenia zaproponowane przez Mariano i Murasawę oraz Pereza-Quiriosa pozwalają na relatywnie proste rozwiązanie niezwykle często spotykanego w przypadku estymacji modeli makroekonometrycznych problemu dostępności danych z mieszanymi częstotliwościami oraz niezbilansowanej próby.

4. Modele regresyjne z mieszanymi częstotliwościami próbkowania (MIDAS)

Jak zostało to opisane w sekcji poświęconej analizie literatury przedmiotu, alternatywna koncepcja modelowania w ramach jednej struktury analitycznej danych o różniących się częstotliwościach została przedstawiona w pracach [Ghysels, Santa-Clara, Valkanov 2004; Ghysels, Sinko, Valkanov 2006]. Opracowane przez wypomnianych ekonometryków modele regresyjne z mieszanymi częstotliwościami próbkowania MIDAS (*MIxed DAta Sampling regressions*) wykorzystywały schematy agregacji obserwacji umożliwiające zapisanie w oszczędny (*parsimonious*) sposób relacji pomiędzy dynamiką zmiennych obserwowanych z małą i dużą częstotliwością. W swoim podstawowym zakresie miały one charakter jednorównaniowych struktur z opóźnieniami (*distributed lags* MIDAS, DL-MIDAS), uwzględniających ewentualne komponenty autoregresyjne (*autoregressive distributed lags* MIDAS*,* ADL-MIDAS). Porównując te modele z przedstawionymi w poprzedniej sekcji dynamicznymi modelami czynnikowymi z mieszanymi częstotliwościami wykorzystującymi filtr Kalmana, można stwierdzić, iż mają one uproszczoną strukturę, która pozwala na uniknięcie modelowania dynamik danych o wysokich częstotliwościach używanych do estymacji zmiennych wyjaśnianych. Biorąc pod uwagę popularność dynamicznych modeli czynnikowych, Marcellino i Schumacher [2008], a następnie Ghysels, Andreou i Kourtellos [2010] zaprezentowali strukturę MIDAS rozszerzoną o modelowanie dynamicznych czynników (*dynamic factor models* MIDAS, DFM-MIDAS). W dalszej części niniejszej sekcji przedstawione zostaną modele MIDAS zarówno w wersji podstawowej, jak i w wariancie rozszerzonym o komponent czynnikowy.

Załóżmy, że chcemy uzależnić w ramach modelu regresyjnego zmienną o niskiej częstotliwości (NC) od zmiennej o wysokiej częstotliwości (WC). W najprostszym wariancie moglibyśmy taką zależność zapisać jako:

$$y_t^{NC} = \mu + \sum_{i=1}^{N_{WC}} \beta_i x_{N_{WC-i},t}^{WC} + u_t \tag{21}$$

gdzie N_{WC} oznacza liczbę obserwacji zmiennej zbieranej z wysoką częstotliwością zawartych w okresie pomiędzy poszczególnymi odczytami zmiennej o niskiej częstotliwości. Jak widać z powyższego równania, wariant ten prowadzi do problemu nadmiernej parametryzacji (np. w przypadku gdy dane o niskiej częstotliwości zbierane są kwartalnie, a dane o wysokiej częstotliwości dziennie, estymacja modelu wiązałaby się z koniecznością obliczenia 66 parametrów β_j. Aby uniknąć tego problemu w ramach modeli MIDAS relacja pomiędzy danymi o wysokich i niskich częstotliwościach determinowana jest za pomocą wielomianu opóźnień wykorzystującego zdefiniowany schemat ważenia obserwacji:

$$W(L^{N_{WC}}; \theta_X^{WC}) x_t^{WC} = \sum_{i=1}^{N_{WC}} w_i(\theta_X^{WC}) x_{N_{WC-i},t}^{WC} \tag{22}$$

Jednym z najczęściej używanych wielomianów opóźnień ze schematem ważenia jest wykładniczy wielomian Almona:

$$w_i(\theta_1, \theta_2) = \frac{exp(\theta_1 i + \theta_2 i^2)}{\sum_{i=1}^{m} exp(\theta_1 i + \theta_2 i^2)} \tag{23}$$

Wykorzystanie wielomianu opóźnień umożliwia zapisanie zależności pomiędzy zmiennymi niskich i wysokich częstotliwości za pomocą równania

$$y_t^{NC} = \mu + \beta \sum_{j=1}^{q} \sum_{i=1}^{N_{WC}} w_{i+j*N_{NC}}(\theta_X^{WC}) x_{N_{WC}-i,t-j}^{WC} + u_t \tag{24}$$

zwanego filtrem DL-MIDAS(q), przy czym q definiuje liczbę okresów niskiej częstotliwości rozważanych wstecz w modelu. Aby umożliwić identyfikację modelu, przyjmowane jest założenie, że $w_{i+j*NNC}(\theta_X^{WC}) \in (0,1)$ oraz że wagi sumują się do 1 dla wszystkich wartości q $\left(\sum_{j=1}^{q} \sum_{i=1}^{N_{WC}} w_{i+j*N_{NC}}(\theta_X^{WC}) = 1\right)$. Estymacja parametrów modelu DL-MIDAS(q) prowadzona jest zazwyczaj za pomocą metody nieliniowych najmniejszych kwadratów (NLK).

Ze względu na fakt, że makroekonomiczne szeregi czasowe charakteryzują się z reguły istotną statystycznie autokorelacją składowych obserwacji, podstawowy model MIDAS (DL-MIDAS) uzupełniany jest o składnik autoregresyjny rzędu p:

$$y_t^{NC} = \mu + \sum_{j=1}^{p} \alpha_j y_{t-j-1}^{NC} + \beta \sum_{j=1}^{q} \sum_{i=1}^{N_{WC}} w_{i+j*N_{NC}}(\theta_X^{WC}) x_{N_{WC}-i,t-j}^{WC} + u_t \tag{25}$$

Powyższa specyfikacja oznaczana jest jako ADL-MIDAS(p,q). Model ADL-MIDAS(1,q):

$$y_t^{NC} = \mu + \alpha y_{t-1}^{NC} + \beta \sum_{j=1}^{q} \sum_{i=1}^{N_{WC}} w_{i+j*N_{NC}}(\theta_X^{WC}) x_{N_{WC}-i,t-j}^{WC} + u_t \tag{26}$$

można równoważnie zapisać za pomocą struktury DL-MIDAS(q) z autokorelacją błędów prognozy:

$$y_t^{NC} = \mu(1-\alpha)^{-1} + \beta \sum_{j=1}^{q} \sum_{i=1}^{N_{WC}} w_{i+j*N_{NC}}(\theta_X^{WC})(1 - L^{NC})^{-1} x_{N_{WC}-i,t-j}^{WC} + \bar{u}_t \tag{27}$$

gdzie $u_t = (1 - \alpha L^{NC})^{-1} u_t$ jest operatorem opóźnień danych o niskiej częstotliwości. Wadą takiego rozwiązania jest występowanie niemającej odzwierciedlenia w rzeczywistych danych kumulacji wag w odstępach co N_{WC} obserwacji ze względu na wzajemne iteracje wielomianów opóźnień niskich i wysokich częstotliwości. Złożona struktura błędów modelu pociąga za sobą konieczność używania bardziej złożonej metody estymacji, np. metody uogólnionych najmniejszych kwadratów.

Możliwym rozwiązaniem problemów związanych z podstawowym wariantem autoregresyjnej wersji modelu MIDAS jest użycie zmiennej niskiej częstotliwości agregującej obserwacje zmiennej wysokiej częstotliwości:

$$x_t^{NC}(\theta_X^{WC}) = \sum_{i=1}^{N_{WC}} w_i(\theta_X^{WC}) x_{N_{WC}-i,t}^{WC} \tag{28}$$

do określenia specyfikacji modelu z multiplikatywnym schematem ważenia obserwacji w wielomianie opóźnień (ADL-MIDAS-M(p_y,p_x)):

$$y_t^{NC} = \mu + \sum_{j=1}^{p_y} \alpha_j y_{t-j-1}^{NC} + \sum_{k=1}^{p_x} \beta_k x_{t-k}^{NC}(\theta_X^{WC}) + u_t \tag{29}$$

W celu uproszczenia struktury modelu opisanego powyższym równaniem wprowadzana jest jednolita struktura wielomianu opóźnień dla części autoregresyjnej i modelującej wpływ zmiennej wysokich częstotliwości:

$$y_t^{NC} = \mu + \alpha \sum_{k=1}^{p[r]_y} w_k(\theta_y^{NC}) y_{t-k-1}^{NC} + \beta \sum_{k=1}^{p[r]_x} w_k(\theta_y^{NC}) x_{t-k}^{NC}(\theta_X^{WC}) + u_t \tag{30}$$

Struktura ta określana jest w literaturze mianem ADL-MIDAS-M($p[r]_y$,$p[r]_x$)

W cytowanej już pracy Ghysels, Andreou i Kourtellos [2010] opisali możliwość połączenia modeli klasy MIDAS ze strukturą dynamicznych czynników. Przedstawili oni zarówno model ADL-MIDAS(p,q) rozszerzony o czynniki niskiej częstotliwości (FADL-MIDAS(p_y,p_f,q)):

$$\begin{gathered} y_t^{NC} = \mu + \textstyle\sum_{j=1}^{p_y} \alpha_j y_{t-j-1}^{NC} \\ + \textstyle\sum_{k=1}^{p_f} \beta_k f_{t-k-1}^{NC} + \gamma \sum_{j=1}^{q} \sum_{i=1}^{N_{WC}} w_{i+j*N_{NC}}(\theta_X^{WC}) x_{N_{WC}-i,t-j}^{WC} + u_t \end{gathered} \tag{31}$$

jak i rozbudowaną wersję modelu ADL-MIDAS-M(p_y,p_x) – FADL-MIDAS--M(p_y,p_f,p_x):

$$y_t^{NC} = \mu + \sum_{j=1}^{p_y} \alpha_j y_{t-j-1}^{NC} + \sum_{k=1}^{p_f} \beta_k f_{t-k-1}^{NC} + \sum_{k=1}^{p_x} \gamma_k x_{t-k}^{NC}(\theta_X^{WC}) + u_t \tag{32}$$

Modele te, wykorzystujące jednocześnie zalety modeli MIDAS i DFM, estymowane są za pomocą filtru Kalmana z wykorzystaniem metody największej wiarygodności opartej na rozkładzie gaussowskim.

5. Dane wykorzystane w badaniu

Do estymacji modelu dynamicznych czynników z mieszanymi częstotliwościami i obsługą braków danych (MFDG-DFM) oraz modelu MIDAS uzupełnionego o dynamiczne czynniki (DFM-MIDAS), a następnie prognozowania polskiego PKB wykorzystano zestaw danych używanych wcześniej przez autora [Łupiński 2012] do porównania jakości modeli MFDG-DFM oraz standardowego modelu czynnikowego opracowanego przez Stocka i Watsona [1989]. Wśród wybranych wówczas danych znalazły się (zmienne w ujęciu dynamiki w odniesieniu do analogicznego okresu roku poprzedniego):

- PKB Polski w cenach stałych,
- indeks produkcji przemysłowej,
- indeks obrotów handlu detalicznego,
- eksport,
- import.

Pierwsza ze zmiennych miała częstotliwość kwartalną, cztery pozostałe częstotliwość miesięczną. Dobór powyższych wskaźników został uzasadniony zarówno badaniem własności statystycznych (za pomocą korelacji krzyżowej oraz koherencji zbadana została siła zależności pomiędzy grupą 20 kandydujących zmiennych wejściowych), jak i ich istotnością w opisie zjawisk gospodarczych. I tak indeks produkcji przemysłowej oraz wskaźnik importu pozwala na obserwowanie procesu tworzenia krajowego PKB, wskaźnik eksportu daje podstawę do śledzenia procesów dystrybucji PKB, natomiast indeks obrotów handlu detalicznego reprezentuje stronę popytową krajowej gospodarki. W tabeli 1 zebrano podstawowe informacje na temat zmiennych użytych w estymacji.

Tabela 1. Zmienne makroekonomiczne użyte do estymacji modeli MFDG-DFM oraz DFM-MIDAS

Zmienna	Częstotliwość	Źródło
PKB	kwartalna	GUS
Indeks produkcji przem.	miesięczna	GUS
Indeks obrotów handlu det.	miesięczna	GUS
Import	miesięczna	GUS
Eksport	miesięczna	GUS

Źródło: opracowanie własne.

Do estymacji modeli wykorzystano dane z lat 1997-2006, natomiast prognozy zmiennych PKB uzyskano na podstawie danych z lat 2007-2010. Zastosowanie wzmiankowanych szeregów czasowych do estymacji obu modeli wymagało ich wcześniejszej transformacji. Wszystkie zmienne poddano procedurze usunięcia efektów sezonowych oraz dni roboczych za pomocą metody TRAMO/SEATS dostępnej w udostępnianym na zasadzie *open source* pakiecie Demetra+. Dane pozbawione obu rodzajów efektów zbadano na obecność pierwiastka jednostkowego. Podobnie jak we wcześniejszym badaniu autora [Łupiński 2012] w przypadku wszystkich pięciu zmiennych na podstawie testu ADF (*Augmented Dickey-Fuller*) stwierdzono brak stacjonarności. Ostatecznie do estymacji użyto więc pierwszych różnic logarytmów zmiennych wejściowych.

Z punktu widzenia finalnego zastosowania estymowanych modeli (nowcasty i prognozowanie w pseudoczasie rzeczywistym krajowego PKB, *pseudo real-time nowcasts and forecasts*) niezbędne było odpowiednie przygotowanie zbioru danych odzwierciedlającego napływ w czasie poszczególnych obserwacji znajdujących się na końcu próby. W celu symulacji rosnącego zakresu próby z lat 2007-2010 wykorzystano polski Kalendarz Udostępniania Danych (*Data Release Calendar*) udostępniany na stronie internetowej hurtowni danych EBC (*EBC Statistical Data Warehouse*, http://sdw.ecb.int). Konstrukcja „rosnącego" panelu danych na potrzeby nowcastów i progoz zmiennych wejściowych użytych modeli była jednym z najbardziej wymagających czasowo elementów badania.

6. Estymacja

Na podstawie opisanego w poprzedniej sekcji zbioru danych dokonano estymacji obu modeli: referencyjnej struktury Mariano-Murasawy (MFDG-DFM) oraz modelu czynnikowego modelu MIDAS (DFM-MIDAS). W obu przypadkach parametry modelu obliczone zostały na podstawie metody największej wiarygodności (MNW) połączonej z filtrem Kalmana. Identyfikacja odpowiedniej struktury modeli (stosownej liczby opóźnień w równaniach determinujących dynamikę zmiennych obserwowalnych i nieobserwowalnych czynników) odbyła się za pomocą kryteriów informacyjnych Akaikego i Bayesa. Do prognozowania PKB wybrane zostały modele z najmniejszą wartością kryteriów informacyjnych.

Wykorzystana w praktyce struktura modelu MFDG-DFM przedstawiała się następująco:

$$
\begin{aligned}
&\begin{bmatrix} Y_{1,t} \\ Y_{j,t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_j \end{bmatrix} + \begin{bmatrix} \gamma_1(\frac{1}{3}h_t + \frac{2}{3}h_{t-1} + h_{t-2} + \frac{2}{3}h_{t-3} + \frac{1}{3}h_{t-4}) \\ \gamma_j h_t \end{bmatrix} \\
&+ \begin{bmatrix} \frac{1}{3}i_{1,t} + \frac{2}{3}i_{1,t-1} + i_{1,t-2} + \frac{2}{3}i_{1,t-3} + \frac{1}{3}i_{1,t-4} \\ i_{j,t} \end{bmatrix} \; jeśli\, Y_{1,t} jest\ dostępne \\
&\begin{bmatrix} Y_{1,t} \\ Y_{j,t} \end{bmatrix} = \begin{bmatrix} 0 \\ \mu_j \end{bmatrix} + \begin{bmatrix} 0 \\ \gamma_j h_t \end{bmatrix} + \begin{bmatrix} 0 \\ i_{j,t} \end{bmatrix} + \begin{bmatrix} w_{1,t} \\ 0 \end{bmatrix} \; jeśli\, \tilde{y}_{1,t} nie\ jest\ dostępne \\
&w_{1,t} = N(\delta_1, \sigma_1) \\
&\Theta(L) i_{j,t} = v_{i,j,t} \\
&\begin{bmatrix} v_{h,t} \\ v_{i,j,t} \end{bmatrix} \sim N\left(0\ , \begin{bmatrix} \sigma_h & 0 \\ 0 & \Sigma_i \end{bmatrix}\right)
\end{aligned}
\tag{33}
$$

(gdzie j = 2,3,4,5). Zmienna $Y_{1,t}$ zarezerwowana została dla kwartalnego PKB, zmienne $Y_{2,t}, Y_{3,t}$ przypisano indeksowi produkcji przemysłowej oraz indeksowi obrotów handlu detalicznego, z kolei zmienne $Y_{4,t}, Y_{5,t}$ odpowiadały importowi i eksportowi. Na podstawie kryteriów informacyjnych AIC i BIC określono, że błędy idiosynkratyczne zmiennych makroekonomicznych oraz dynamika nieobserwowalnego czynnika powinny być modelowane za pomocą procesów typu AR(1). Tabela 2 przedstawia wyniki estymacji modelu MFDG-DFM zrealizowane za pomocą metody największej wiarygodności.

Tabela 2. Parametry modelu MFDG-DFM wyestymowane za pomocą MNW

Parametr	Wartość	Błąd standardowy
γ_1	0,62	0,11
γ_2	0,24	0,08
γ_3	0,36	0,09
γ_4	0,67	0,10
γ_5	0,74	0,08
ϕ	–0,33	0,07
θ_1	–0,89	0,09
θ_2	–0,21	0,06
θ_3	–0,35	0,09
θ_4	–0,56	0,12
θ_5	–0,42	0,10
σ_1	0,73	0,14
σ_2	0,83	0,12
σ_3	0,69	0,09
σ_4	0,39	0,13
σ_5	0,40	0,08
Wartość funkcji wiarygodności: 351,85		
AIC: –3,37, BIC: –3,32		

Źródło: obliczenia własne.

Drugą klasą modeli użytych do formułowania nowcastów i prognoz polskiego PKB były modele MIDAS uzupełnione o dynamikę wspólnego czynnika (DFM-MIDAS). Podobnie jak w przypadku dynamicznego modelu czynnikowego o mieszanych częstotliwościach i obsłudze braków danych wejściowy wektor zmiennych wejściowych podzielony został na dwie części: zmiennych obserwowanych z niską częstotliwością (NC) i zmiennych charakteryzujących się wysoką częstotliwością: (WC). W konkretnym modelu użytym na potrzeby niniejszej pracy zbiór zmiennych o niskiej częstotliwości (kwartalnej) jest jednoelementowy (PKB, $Y_{1,t}^{NC}$), natomiast zmienne o wyższych częstotliwościach (miesięcznej) obejmują bezrobocie, wskaźnik produkcji przemysłowej, indeks obrotów w handlu detalicznym, import oraz eksport $X_{j,t}^{WC}$, $j = 1,2, ..., 4$. Relacja pomiędzy zmiennymi o wysokich i niskich częstotliwościach określana jest jako:

$$Y_{j,t}^{NC} = b(L, \theta_X^{WC})X_{j,t}^{WC} \tag{34}$$

W celu realizacji praktycznej estymacji zależność ta opisywana jest wielomianem Almona (rząd wielomianu równy jest 2) .

Struktura analityczna DFM-MIDAS pozwalająca na jednoczesne wykorzystanie zmiennych niskiej i wysokiej częstotliwości ma postać:

$$\begin{aligned} &\begin{bmatrix} Y_{1,t}^{NC} \\ b(L,\theta_X^{WC})X_{j,t}^{WC} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_j \end{bmatrix} + \begin{bmatrix} \alpha_1(L)Y_{1,t}^{NC} \\ 0 \end{bmatrix} + \begin{bmatrix} \gamma_1 h_t^{NC} \\ \gamma_j h_t^{NC} \end{bmatrix} + \begin{bmatrix} i_{1,t} \\ i_{j,t} \end{bmatrix} \\ &\Phi(L)h_t^{WC} = v_{h,t} \\ &\Theta(L)i_{j,t} = v_{i,j,t};\ j = 1,2,3,4,5 \\ &\begin{bmatrix} v_{h,t} \\ v_{i,j,t} \end{bmatrix} \sim N\left(0\ ,\begin{bmatrix} \sigma_h & 0 \\ 0 & \Sigma_i \end{bmatrix}\right) \end{aligned} \tag{35}$$

Podobnie jak w przypadku modelu MFDG-DFM wartości kryteriów informacyjnych wskazały, że dynamika nieobserwowalnego czynnika oraz błędów idiosynkratycznych zmiennych obserwowalnych może być modelowana za pomocą struktury AR(1).

Wyniki estymacji zestawiono w tab. 3.

Tabela 3. Parametry modelu DFM-MIDAS wyestymowane za pomocą MNW

Parametr	Wartość	Błąd standardowy
α_1	0,79	0,18
γ_1	0,53	0,09
γ_2	0,21	0,04
γ_3	0,32	0,06
γ_4	0,65	0,13
γ_5	0,67	0,09
ϕ	–0,29	0,05
θ_1	–0,76	0,11
θ_2	–0,31	0,08
θ_3	–0,35	0,09
θ_4	–0,52	0,1
θ_5	–0,37	0,09
σ_1	0,84	0,16
σ_2	0,81	0,09
σ_3	0,63	0,12
σ_4	0,43	0,09
σ_5	0,36	0,12
Wartość funkcji wiarygodności: 383.62		
AIC: –3,83, BIC: –3,79		

Źródło: obliczenia własne.

W ramach procedury estymacji modeli MFDG-DFM i DFM-MIDAS obliczone zostały wspólne komponenty (odpowiednio h_t i h_t^{LF}). Po przetransformowaniu wraz z szeregiem czasowym PKB do postaci rocznych wzrostów (analogiczny kwartał roku poprzedniego równy 100) zmienne te zostały użyte do nowcastów i prognozowania polskiego produktu krajowego brutto na maksymalnie 4 okresy do przodu (h = 0,1,2,3,4), zgodnie z formułą:

$$y_{t+h} = \alpha_0 + \sum_{i=1}^{k} \alpha_i y_{t-i} + \sum_{j=0}^{l} \beta_j . x_{t-j} \tag{36}$$

W powyższym równaniu y jest prognozowaną zmienną (PKB), a x egzogeniczną zmienną wspomagająca prognozowanie (w tym przypadku odpowiednio nieobserwowalnym wspólnym komponentem h_t lub h_t^{NC}), k jest rzędem opóźnień procesu prognozowanej zmiennej, natomiast l rzędem opóźnień procesu zmiennych pomocniczych. Nowcasty (h = 0) oraz prognozy referencyjnej zmiennej uzyskane za pomocą obu użytych metod zostały porównane za pomocą standardowych miar: średniokwadratowego błędu prognozy (*Root Mean Square Error*, RMSE), średniego bezwzględnego błędu prognozy (*Mean Absolute Error*, MAE) oraz średniego błędu prognozy (*Mean Error*, ME).

7. Porównanie prognoz PKB uzyskanych referencyjnymi modelami

Niniejsza część artykułu przeznaczona została na prezentację oceny jakości nowcastów oraz prognoz polskiego PKB uzyskanych za pomocą ćwiczenia wykonanego w pseudoczasie rzeczywistym. Jak zostało to opisane wcześniej, ćwiczenie to zostało wykonane za pomocą kroczącego okna o długości maksymalnie 4 kwartałów z zakresu czasowego obejmującego lata 2007-2010, czyli okresu ogólnoświatowego kryzysu gospodarczego oraz finansowego, mającego swój początek na przełomie lat 2007 i 2008. Jakość uzyskanych wyników porównano za pomocą trzech kryteriów: RMSE, MAE i ME. Tabela 4 przedstawia porównanie jakości nowcastów (h = 0) uzyskanych za pomocą obu modeli, MFDG-DFM oraz DFM-MIDAS.

Tabela 4. Porównanie jakości nowcastów uzyskanych za pomocą alternatywnych modeli

Model	RMSE	MAE	ME
MFDG-DFM	0,43	0,31	0,11
DFM-MIDAS	0,46	0,29	0,09

Źródło: obliczenia własne.

Uzyskane wyniki nie pozwalają jednoznacznie stwierdzić, który z dwóch modeli pozwala na uzyskanie bardziej wiarygodnych nowcastów. Chociaż według kryterium błędu średniokwadratowego mniejszym błędem obciążone są „prognozy” teraźniejszości uzyskane za pomocą dynamicznego modelu czynnikowego, to jednak w zakresie średniego błędu absolutnego i błędu średniego przewagę zyskują nowcasty otrzymane za pomocą wariantu modelu MIDAS. Należy jednocześnie zwrócić uwagę, że dzięki możliwości włączania do obliczeń najbardziej aktualnych danych o miesięcznych częstotliwościach oba modele pozwalają na uzyskanie dobrej jakości „prognoz" wartości bieżących.

W drugim etapie badania jakości dwóch zaprezentowanych modeli porównano jakość prognoz uzyskiwanych w 4 horyzontach, od 1 do 4 kwartałów naprzód (h = 1, 2, 3, 4). Dla każdego z horyzontów obliczono osobne statystyki i zastawiono je w tab. 5.

Tabela 5. Porównanie jakości *h*-okresowych prognoz uzyskanych za pomocą alternatywnych modeli

Model	RMSE	MAE	ME
h = 1			
MFDG-DFM	0,91	0,72	**0,11**
DFM-MIDAS	**0,88**	**0,72**	0,13
h = 2			
MFDG-DFM	**1,02**	**0,83**	**0,15**
DFM-MIDAS	1,12	0,98	0,25
h = 3			
MFDG-DFM	**1,59**	**1,28**	**0,52**
DFM-MIDAS	1,73	1,57	0,87
h = 4			
MFDG-DFM	**2,25**	**1,91**	**0,94**
DFM-MIDAS	2,61	2,13	1,36

Źródło: obliczenia własne.

Na podstawie przedstawionych wyników można wyciągnąć wniosek, że czynnikowe modele MIDAS pozwalają na sformułowanie prognoz, które lepiej sprawdzają się w bardzo krótkim horyzoncie czasu (1 kwartał). W dłuższej perspektywie czasu przewagę zaczynają uzyskiwać prognozy obliczone z wykorzystaniem modelu MFDG-DFM. Posiłkując się kryterium RMSE, prognozy na dwa okresy uzyskane za pomocą modelu MFDG-DFM są o ok. 10% bardziej trafne od prognoz otrzymanych za pomocą modelu DFM-MIDAS. W perspektywie jednego roku (h = 4) prognozy uzyskane za pomocą modelu czynnikowego z mieszanymi częstotliwościami i obsługą braków danych są już o ponad 15% lepsze niż w przypadku konkurencyjnej metody. Jednocześnie niezależnie od użytej metody jakość prognoz systematycznie maleje wraz z długością przewidywanego horyzontu czasowego. Jak wynika z porównania tab. 4 i 5, wielkość błędu prognozy (obliczonego za pomocą RMSE) dla dokładniejszej metody jest w przypadku prognozy na 1 rok do przodu niemal sześciokrotnie wyższa niż dla błędu charakteryzującego nowcast.

8. Podsumowanie

Głównym celem artykułu było porównanie podstaw matematycznych oraz jakości nowcastów i prognoz PKB Polski uzyskanych za pomocą dwóch alternatywnych podejść: dynamicznych modeli czynnikowych z mieszanymi częstotliwościami i obsługą braków danych (MFDG-DFM) oraz dynamicznych modeli czynnikowych MIDAS (DFM-MIDAS). W ramach teoretycznej części pracy porównano podstawy matematyczne obu metod oraz metody estymacji zbudowanych struktur analitycznych pozwalające na włączenie do procesu estymacji zmiennych o różniących się częstotliwościach. Na podstawie próby obejmującej swoim zakresem lata 2007-2010 przeprowadzono właściwą ewaluację trafności nowcastów i prognoz polskiego PKB na podstawie kroczącej próby uwzględniającej historię napływu obserwacji na końcu próby zmiennych używanych w badaniu (ćwiczenie pseudoczasu rzeczywistego).

Uzyskane wyniki nie pozwoliły na udzielenie jednoznacznej odpowiedzi na pytanie, która z badanych metod pozwala na uzyskanie bardziej adekwatnych wyników. W przypadku „prognozowania teraźniejszości” oraz krótkiego horyzontu czasowego (jednego kwartału) oba podejścia dawały podstawę do formułowania porównywalnych prognoz (z lekkim wskazaniem na czynnikowy model MIDAS). W dłuższej perspektywie swoją wyższość (mierzoną trafnością prognoz) zaczęło uwidaczniać podejście MFDG-DFM. W perspektywie jednego roku sformułowane za pomocą tej metody prognozy były o ok. 15% bardziej trafne (według kryterium RMSE) niż prognozy modelu konkurencyjnego.

Trudno jest stawiać jednoznaczne tezy dotyczące przyczyn tego stanu rzeczy. Wiadomo, że prognozy obliczone za pomocą zarówno modeli regresyjnych, jak i dynamicznych modeli czynnikowych w długim okresie zbiegają do średniej. W krótkiej perspektywie oba modele zapewniają pewną przewagę nad konkurencją przez wykorzystanie do prognozowania zmiennych kwartalnych danych zebranych z wyższą częstotliwością (w tym przypadku miesięcznych). Jak się wydaje, w dłuższej perspektywie zapewniana przez model MFDG-DFM struktura dynamiki zmiennych obserwowalnych i zależności pomiędzy zmiennymi niskich i wysokich częstotliwości jest w warunkach polskich bardziej adekwatna do rzeczywistości niż ta zaoferowana przez konkurenta. Mając na uwadze wcześniejsze badania autora, niezbędne jest prowadzenie dalszych prac analitycznych mających na celu monitoring alternatywnych podejść pozwalających na optymalne modelowanie kluczowych polskich zmiennych makroekonomicznych.

Literatura

Arouba S.B., Diebold F. X., Scotti C., *Real-time measurement of business conditions,* “Journal of Business and Economic Statistics” 2009, nr 27 (4), s. 417-27.

Camacho M., Perez-Quiros G., *Ń-STING: España Short Term INdicator of Growth*, “Banco de Espana Working Papers” nr 0912, Banco de España, 2009.

Camacho M., Perez-Quiros G., *Introducing the euro-sting: Short-term indicator of euro area growth,* “Journal of Applied Econometrics” 2010, nr 25(4), s. 663-694.

Diebold F.X., Mariano R.S., *Comparing predictive accuracy*, “Journal of Business & Economic Statistics, American Statistical Association” 1995, nr 13(3), s. 253-63.
Forni M. i in., *Coincident and leading indicators for the euro area*, “Economic Journal, Royal Economic Society” 2001, nr 111(471), s. 62-85.
Frale C., Monteforte L., *FaMIDAS: A Mixed Frequency Factor Model with MIDAS structure*, „Italian Ministry of Economy and Finance Working Paper“ nr 3, 07/2010.
Geweke J., *The Dynamic Factor Analysis of Economic Time-Series Models*, [w:] *Latent Variables in Socio-Economic Models,* D. Aigner, A. Goldberger (ed.), North-Holland, New York 1977.
Ghysels E., Santa-Clara P., Valkanov R., *The MIDAS Touch: Mixed Data Sampling Regression Models,* “CIRANO Working Paper” 2004, CIRANO.
Ghysels E., Sinko A., Valkanov R., *MIDAS regressions: further results and new directions*, “Econometric Reviews” 2006, nr 26, s. 53-90.
Ghysels E., Andreou E., Kourtellos A., *Regression models with mixed sampling frequencies,* “Journal of Econometrics” 2010, nr 158, s. 246-261.
Hamilton J.D., *Time Series Analysis*, Princeton University PressPrinceton, New York 1994.
Łupiński M., *Konstrukcja wskaźnika wyprzedzającego aktywności ekonomicznej w Polsce*, praca doktorska, Uniwersytet Warszawski, Warszawa 2007.
Łupiński M., *Four years after expansion. are Czech Republic, Hungary and Poland closer to core or periphery of EMU?*, “Ekonomia” 2009, nr 22, s. 75-100.
Łupiński M., *Short-term forecasting and composite indicators construction with help of dynamic factor models handling mixed frequencies data with ragged edges*, „Przegląd Statystyczny” 2012, nr 59(1), s. 48-73.
Marcellino M., Schumacher C., *Factor-MIDAS for now- and forecasting with ragged-edge data: a model comparison for German GDP*, CEPR Discussion Paper nr 6708 Mariano R.S., 2008.
Mariano R., Murasawa Y., *A new coincident index of business cycles based on monthly and quarterly series*, “Journal of Applied Econometrics” 2003, nr 18(4), s. 427-443.
Stock J.H., Watson M.W., *New Indexes of Coincident and Leading Economic Indicators* NBER Chapters, “NBER Macroeconomics Annual” 1989, nr 4, s. 351-409, NBER.
Stock J.H., Watson M.W., *Business Cycle Fluctuations in U.S. Macroeconomic Time Series*, ”NBER Working Papers” nr 6528, NBER 1998.

COMPARISON OF POLISH GDP FORECASTING QUALITY WITH DYNAMIC FACTOR MODELS AND MIDAS EMBEDDED WITH FACTOR STRUCTURE

Summary: The article is a continuation of the previous author’s papers (2007, 2009, 2012) devoted to the optimal methods of forecasting Polish macroeconomic variables, with the sample of GDP. The research was aimed at a comparison of the quality of nowcasts (”forecasts” of the present time) and forecasts prepared with a dynamic factor model with mixed frequency and data gaps handling (MFDG-DFM) proposed by Mariano and Murasawa [2003] and MIDAS model augmented with factor structure (DFM-MIDAS), described for the first time in the paper of Marcellino and Schumacher [2008]. Mathematical backgrounds of both models were presented and a combination of Kalman filter and Maximum likelihood estimation was hinted as the estimation framework for both of them. The gained results show an advantage of Mariano and Murasawa approach in the field of the forecasts (approximately 15% more adequate forecasts for 2 and more quarters ahead) but this model is less adequate than the competitor for one quarter ahead forecasts and nowcasts.

Keywords: dynamic factor model, MIDAS, Polish GDP forecasts.