

Lech Kujawski*

Zastosowanie danych o różnej częstotliwości w prognozowaniu makroekonomicznym na podstawie modeli dynamicznych

Wstęp

Rozwój technik elektronicznego magazynowania i przesyłania danych umożliwił dostęp on-line do baz danych ekonomicznych zawierających wiele (często setki i tysiące) szeregów czasowych zmiennych odnoszących się do funkcjonowania realnej gospodarki i rynków finansowych. Dostęp do tak licznych zasobów wywołuje naturalną chęć i potrzebę wykorzystania informacji w tych danych zawartych. Jednakże, jak wiadomo, typowe modele szeregów czasowych (wykorzystywane w prognozowaniu krótkookresowym), takie jak ARIMA czy VAR, wykorzystują jeden (ARIMA) lub co najwyżej kilka (VAR) szeregów czasowych. Modelem, w którym jednocześnie można wykorzystać wiele (kilkadziesiąt, a nawet kilkaset) szeregów czasowych, jest model zaproponowany przez J. Stocka, M. Watsona [2002], zwany w literaturze dynamicznym modelem czynnikowym (DFM). Niewątpliwie atrakcyjna możliwość uwzględnienia w jednym modelu zmienności bardzo wielu zmiennych oraz nieskrępowany dostęp do baz danych przyczyniły się do spopularyzowania modelu DFM zarówno w odniesieniu do modelowania rynków finansowych, jak i modelowania makroekonomicznego [por. Artis i inni, 2003; Schneider, Spitzer, 2004; Forni i inni, 2005, Boivin, Ng, 2006]. DFM jest jednak narzędziem konstruowanym i estymowanym na podstawie danych o jednolitej częstotliwości, co oznacza, że dla badaczy zajmujących się aspektami makroekonomicznymi dostępne są w zasadzie dane o częstotliwości kwartalnej lub miesięcznej. Przykładowo, makroekonomiczny problem modelowania i prognozowania produktu krajowego brutto o częstotliwościach szeregów wyższych niż roczne lub kwartalne podejmowany jest bardzo rzadko, do wyjątków w tym zakresie należą prace G. Chow, A. Lin [1971], w których wykorzystuje się techniki interpolacji. Częściej spotyka się próby modelowania kwartalnego PKB przy zastosowaniu szeregów czasowych o wyż-

* Dr, Katedra Ekonometrii, Wydział Zarządzania, Uniwersytet Gdański, ul. Armii Krajowej 101, 81-824 Sopot, lech.kujawski@ug.edu.pl

szych (miesięcznych) częstotliwościach [Miller, Chin, 1996; Marcellino i inni, 2006]. Cechą wspólną ostatnich wymienionych publikacji jest stosowanie regresji, w których w roli zmiennej objaśniającej występują obserwacje wybranego miesiąca danego kwartału, nie jest więc wykorzystywana pełna wysokoczęstotliwościowa zmienność wszystkich danych miesięcznych wyróżnionego kwartału.

W niniejszym artykule proponuje się zastosowanie regresji, w której zmienna zależna wyrażona jest w postaci szeregu czasowego o niskiej częstotliwości (w tym badaniu o częstotliwości kwartalnej), natomiast zmienne niezależne wprowadzane są do modelu w dwóch częstotliwościach: równych częstotliwości zmiennej zależnej oraz wyższych niż częstotliwość zmiennej zależnej (w tym badaniu będą to zmienne miesięczne), przy czym *ex post* wykorzystywana jest cała wiedza zawarta w szeregach miesięcznych wyróżnionego kwartału.

Techniką ekonometryczną pozwalającą połączyć szeregi o zróżnicowanej częstotliwości jest model oznaczony skrótem MIDAS (MIXed DATA Sampling) [Ghysels i inni, 2004a, 2004b]. Model MIDAS wywodzi się z rynków finansowych [Chen, Ghysels, 2009; Andreou i inni, 2010], gdzie szczególnie atrakcyjnym jest fakt potencjalnie możliwej śródkresowej korekty prognozy wraz z napływem informacji pochodzących z danych o wysokiej częstotliwości. W modelowaniu makroekonomicznym owa śródkresowa korekta prognoz jest oczywiście nie mniej atrakcyjna, dołączając do niej brak utraty informacji na skutek agregacji zmiennych, otrzymujemy dwie główne przesłanki przemawiające za stosowaniem modeli MIDAS. Przytoczone cechy regresji MIDAS wydają się na tyle ciekawe i atrakcyjne, że właściwe wydaje się empiryczne zweryfikowanie dopasowania prognoz makroekonomicznych uzyskiwanych na podstawie tej klasy modeli i porównanie tych prognoz z innymi prognozami uzyskanymi z „typowych” modeli szeregów czasowych (ARIMA, VAR, DFM) stosowanych w krótkookresowym prognozowaniu gospodarki.

Celem opracowania jest weryfikacja tezy postawionej następująco: uwzględnienie w modelu dynamicznym zmiennych o częstotliwości wyższej niż częstotliwość zmiennej objaśnianej statystycznie istotnie poprawia jakość (dopasowanie) prognoz makroekonomicznych.

Weryfikacja tezy nastąpi poprzez oszacowanie modeli MIDAS, ARIMA, DFM i VAR, na ich podstawie wyznaczone zostaną prognozy krótkookresowe produktu krajowego brutto w Stanach Zjednoczonych,

a następnie porównana zostanie precyzja tych prognoz. Do oszacowania modeli wykorzystana zostanie baza makroekonomicznych danych czasu rzeczywistego. Zastosowanie danych czasu rzeczywistego ma kluczowe znaczenie (co uzasadnione zostanie w dalszej części artykułu) dla zaplanowanych badań. Ponieważ autor nie ma dostępu do bazy danych czasu rzeczywistego polskiej gospodarki, stąd badanie dotyczy PKB w USA. Dodatkowo w artykule autor zaproponował niewielką modyfikację oryginalnej regresji MIDAS, tak by na wzór modelu DFM możliwe było jednocześnie wykorzystanie informacji płynącej z dużej liczby szeregów czasowych.

Tekst podzielony został na trzy główne części. Część pierwsza poświęcona została opisowi modeli MIDAS i DFM, ponadto wyjaśniona została istota zastosowanej modyfikacji oryginalnego modelu MIDAS. W części drugiej scharakteryzowano dane czasu rzeczywistego użyte w badaniu oraz uzasadniono konieczność posłużenia się danymi czasu rzeczywistego. Część trzecia jest opisem wyników oszacowań samych modeli, a także prognoz wyliczonych na ich podstawie. W tej części dokonano również porównania jakości prognoz, co ostatecznie pozwoliło sformułować wnioski odnoszące się to tezy postawionej w niniejszym artykule.

1. Metoda badawcza

Ponieważ modele szeregów czasowych zastosowane w niniejszym badaniu, takie jak ARIMA czy VAR, są powszechnie znane, a ich podręcznikowy opis jest powszechnie dostępny, szczególna uwaga skupiona zostanie na modelach MIDAS i DFM. Jednorównaniowy model klasy MIDAS o horyzoncie prognozy $h=1$ można zapisać w sposób następujący [Clements, Galvao, 2006]:

$$y_t = \beta_0 + \beta_1 B(L^{1/m}; \theta) x_{t-1}^m + \xi_t^m \quad (1)$$

$$L^{s/m} x_{t-1}^m = x_{t-1-s/m}^m \quad (2)$$

$$B(L^{1/m}; \theta) = \sum_{k=1}^K b(k; \theta) L^{(k-1)/m} \quad (3)$$

Formuła (1) opisuje jednorównaniowy model MIDAS, wzór (2) definiuje operator opóźnień zmiennych wysokiej częstotliwości, formuła (3) wyjaśnia, w jaki sposób rozwijany jest wielomian skalarny względem operatora opóźnień zmiennych wysokiej częstotliwości. Indeks zmiennej

niskiej częstotliwości oznaczony jest jako t , litera m oznacza stałą liczbę obserwacji wysokiej częstotliwości w jednostce czasu niskiej częstotliwości, natomiast K określa maksymalny rząd rozłożonych opóźnień w szeregu czasowym wysokiej częstotliwości. Formuły (1)–(3) nie definiują jednakże sposobu liczenia $b(k; \theta)$, czyli wag przypisywanych poszczególnym opóźnionym obserwacjom z szeregu wysokiej częstotliwości. W literaturze przedmiotu [Ghysels i inni, 2005; Foroni, Marcelliano, 2013] wymienia się pięć podstawowych sposobów parametryzacji rozłożonych opóźnień stosowanych w modelach MIDAS: liniową, hiperboliczną, gamma, beta, wykładniczą. Do badacza należy rozstrzygnięcie, jakiego rodzaju parametryzację wybrać, ponadto po wyborze rodzaju parametryzacji należy zdecydować o samej liczbie parametrów wewnątrz wybranej parametryzacji.

Wskazane powyżej źródła literatury sugerują, że dostatecznie elastyczne i zarazem oszczędne parametryzacje uzyskuje się dla opóźnień zgodnych ze schematem beta i wykładniczym.

Schemat zmienności wag występujących przy parametrach strukturalnych jest zgodny ze schematem beta, jeśli wyraża się go wzorem:

$$b(k; \theta) = \frac{f(k/K, \theta_1, \theta_2)}{\sum_{k=1}^K f(k/K, \theta_1, \theta_2)} \quad (4)$$

gdzie $f(k/K, \theta_1, \theta_2)$ jest gęstością rozkładu beta z parametrami θ_1, θ_2 .

Schemat zmienności wag występujących przy parametrach strukturalnych opisany formułą wykładniczą zapisuje się:

$$b(k; \theta) = \frac{\exp(\theta_1 k + \dots + \theta_Q k^Q)}{\sum_{k=1}^K \exp(\theta_1 k + \dots + \theta_Q k^Q)} \quad (5)$$

Badania prowadzone przez Foroni i Marcelliano [2013] sugerują ponadto, że dostateczną zmienność wag wykładniczych uzyskuje się przy dwóch parametrach θ , a zwiększanie ich liczby nie prowadzi do polepszenia własności modeli, powodując jedynie nadmierną parametryzację. Dlatego ostatecznie na potrzeby dalszego badania przyjęto schemat zmienności wag zgodny z formułą:

$$b(k; \theta) = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{k=1}^K \exp(\theta_1 k + \theta_2 k^2)} \quad (6)$$

Równania (1)–(3) i (4) oraz (6) tłumaczą zatem podstawowy model MIDAS. Przyjmując przykładowo $m=3$ i $K=12$, samo równanie regresji można zapisać:

$$y_t = \beta_0 + \beta_1 [b(1; \theta)x_{t-1}^3 + b(2; \theta)x_{t-1-1/3}^3 + \dots + b(12; \theta)x_{t-4-2/3}^3] + \xi_t^3 \quad (7)$$

Jeśli indeks t oznacza obserwację z pierwszego kwartału roku 2015 (2015q1), wówczas x_{t-1}^3 jest obserwacją na zmiennej o częstotliwości miesięcznej z okresu 2014m12, $x_{t-1-1/3}^3$ oznacza obserwację z 2014m11, natomiast $x_{t-4-2/3}^3$ obserwację z okresu 2014m01.

Model sformułowany zgodnie z równaniem (1) umożliwia wyznaczanie prognoz o horyzoncie $h=1$ tylko i wyłącznie na podstawie zmiennych wysokiej częstotliwości odnoszących się do okresów wcześniejszych niż bieżąca obserwacja dokonywana na zmiennej niskiej częstotliwości. W takim przypadku nie byłaby możliwa śródokresowa korekta prognoz w oparciu o najnowszą dostępną informację. Z uwagi na niniejsze badanie, tj. kwartalnego PKB przy zastosowaniu miesięcznych zmiennych objaśniających, niewątpliwie najciekawszą wersją modelu MIDAS jest ta, w której możliwe jest liczenie prognoz kwartalnych na podstawie danych miesięcznych pochodzących z miesiąca/miesiący wyróżnionego prognozowanego kwartału. Model taki można przedstawić następująco:

$$y_t = \beta_0 + \beta_1 B(L^{1/m}; \theta)x_{t-2/3}^m + \xi_t^m \quad (8)$$

Horyzont prognozy ustalony jest wówczas na $h=2/3$, czyli wskazuje na dostępność danych pochodzących z pierwszego miesiąca danego kwartału¹, model ten uwzględnia zatem możliwość oddziaływania danych miesięcznych wyszczególnionego kwartału na kwartalne prognozy zmiennej objaśnianej.

Z punktu widzenia poprawności klasyfikacji równanie (8) jest modelem dynamicznym *sensu largo*. Nie występują w nim zmienne endogeniczne opóźnione w czasie w roli zmiennych objaśniających. Dlatego nie spełnia on wymogów modelu dynamicznego *sensu stricto*. Dołączenie części autoregresyjnej do modelu (7) prowadzi do postaci AR-MIDAS sformułowanej następująco:

$$y_t = \beta_0 + \lambda y_{t-1} + \beta_1 B(L^{1/m}; \theta)x_{t-2/3}^m + \xi_t^m \quad (9)$$

¹ Horyzont $h=1/3$ wskazuje na dostępność danych pochodzących z dwóch pierwszych miesięcy danego kwartału.

Różnica w specyfikacji modelu MIDAS i modelu AR-MIDAS prowadzi do różnic w sposobie estymacji parametrów strukturalnych. Jak wykazano w literaturze [Ghysels i inni, 2004a] przy typowych założeniach odnoszących się do struktury stochastycznej modelu MIDAS², estymator nieliniowej MNK jest co najmniej zgodny, taki zatem estymator zastosowano na potrzeby niniejszej publikacji. Jednakże, jak wykazano w innym miejscu w literaturze [Ghysels i inni, 2004c], proste dołączenie do modelu MIDAS części AR, tak jak zrobiono to w równaniu (9), zawsze prowadzi do powstania sezonowej zmienności zmiennych o częstotliwości miesięcznej, niezależnie od tego, czy owa zmienność sezonowa w danych miesięcznych występowała, czy zmienne zostały poddane procedurze odsezonowania, wynika to bowiem z faktu multiplikatywnego nałożenia opóźnień kwartalnych i miesięcznych.

Rozwiązaniem problemu jest dwustopniowa procedura estymacji wywodząca się z metody czynników wspólnych [Hendry, Mizon, 1978; Clements, Galvao, 2006]. Procedura rozpoczyna się oszacowaniem „standardowego” modelu MIDAS postaci (8), następnie na podstawie reszt modelu oszacowaniu podlega parametr λ części AR. Kolejny etap polega na λ różnicowym przekształceniu zarówno zmiennych objaśniających (miesięcznych), jak i zmiennej objaśnianej (kwartalnej) i powtórnym szacowaniu jedynie parametrów β i θ relacji (9). Całość wymaga dwukrotnego szacowania nieliniową MNK parametrów β i θ oraz „doszacowania” parametru λ . Opisaną procedurę zastosowano do szacowania dynamicznych modeli MIDAS na potrzeby niniejszej publikacji.

Dysponując bogatą bazą danych, do modelu MIDAS postaci (9) dołączyć można zmienne o częstotliwości miesięcznej odnoszące się do okresów wcześniejszych niż badany kwartał, oraz dowolne zmienne o częstotliwości kwartalnej. Uwzględnienie dodatkowych zmiennych miesięcznych i kwartalnych prowadzi do modelu użytego w niniejszym badaniu³:

$$y_t = \beta_0 + \lambda y_{t-1} + \sum_{i=0}^4 \beta_i B(L^{1/m}; \theta_i) x_{i,t-2/3}^m + \sum_{j=0}^4 \alpha_j B(L^{1/m}; \theta_j) z_{j,t-1}^m +$$

² Odnosnie do struktury stochastycznej modeli klasy MIDAS czyni się założenia takie jak w klasycznym schemacie Gaussa-Markowa. Dołączenie części AR do modeli postaci (1) lub (8) motywowane jest typowo, czyli: polepszeniem własności prognostycznych, eliminacją autokorelacji.

³ Jest to zatem połączenie modelu AR-MIDAS i DL.

$$+ \sum_{l=0}^4 \sum_{g=1}^4 \gamma_{lg} W_{lg,t-g} + \xi_t^m \quad (10)$$

Specyfikacja AR-MIDAS (9) pozwala na:

- uwzględnienie do czterech zmiennych ($i = 0, 1, \dots, 4$) o częstotliwości miesięcznej pochodzących z pierwszego miesiąca wyróżnionego kwartału;
- uwzględnienie do czterech zmiennych o częstotliwości miesięcznej ($j = 0, 1, \dots, 4$) pochodzących z okresów poprzedzających wyróżniony kwartał; ponadto dla wszystkich zmiennych miesięcznych przewidziano maksymalny rząd opóźnień równy $K = 12$;
- do czterech opóźnionych zmiennych ($l = 0, 1, \dots, 4$) kwartalnych, o maksymalnym stopniu rozłożonych opóźnień równym 4 ($g = 1, 2, 3, 4$).

Warto podkreślić, że używając dynamicznego modelu (10) możliwe staje się wykorzystanie najnowszej wiedzy/informacji płynącej z bieżących miesięcznych danych makroekonomicznych. Ograniczeniem modelu (10) są arbitralnie przyjęte maksymalne liczby zmiennych o częstotliwości miesięcznej, kwartalnej oraz arbitralnie przyjęte maksymalne rzędy opóźnień obu rodzajów zmiennych. Są to zarazem ograniczenia programu samodzielnie napisanego przez autora w środowisku R. Niemniej jednak należy zwrócić uwagę, że maksymalnie wykorzystana specyfikacja modelu (10) przy parametryzacji beta lub wykładniczej powoduje konieczność oszacowania 32 parametrów strukturalnych i 20 parametrów określających schemat zmienności wag. Nawet w dobie dzisiejszych szybkich komputerów rodzi to poważne problemy numeryczne, stąd przyjęte ograniczenia.

Dynamicznym modelem czynnikowym DFM wprowadzonym i spopularyzowanym przez J. Stocka i M. Watsona [2002] jest:

$$y_t = \alpha_h + \beta_h(L)F_{t-h} + \gamma_h(L)y_{t-h} + \xi_t \quad (11)$$

gdzie: y_t jest zmienną objaśnianą/prognozowaną; F_t jest macierzą czynników (w praktyce macierzą głównych składowych) oszacowaną na podstawie dużego zbioru zmiennych makroekonomicznych mających potencjalny wpływ na kształtowanie się zmiennej prognozowanej, h oznacza horyzont prognozy. Pozostałe składowe to parametry strukturalne i składnik zakłócający, odnośnie którego zakłada się spełnienie typowych założeń.

Przy założeniu skończonego rzędu rozkładu opóźnień wersją modelu używaną do prognozowania jest:

$$y_t = \hat{\alpha}_h + \hat{\beta}_h(L)F_{t-h} + \hat{\gamma}_h(L)y_{t-h} \quad (12)$$

Wybór rzędu autoregresji w procesie $\hat{\gamma}_h(L)y_{t-h}$ i rozłożonych opóźnień w procesie $\hat{\beta}_h(L)F_{t-h}$, czyli de facto wybór „optymalnego modelu DFM” dokonywany jest na podstawie kryteriów pojemności informacyjnej (Stock i Watson stosowali bayesowskie kryterium informacyjne BIC, w niniejszej pracy wykorzystano również to kryterium). Po ustaleniu rzędów procesów AR i DL parametry modelu (11) szacowane⁴ są osobno dla każdego spośród założonych horyzontów prognozy h , a następnie przy użyciu formuły (12) liczone są prognozy.

Kończąc opis zastosowanego modelu DFM, warto dodać, że na podstawie DFM technicznie możliwe jest prognozowanie bezpośrednie⁵ i iteracyjne⁶, w badaniach zastosowano obie techniki.

2. Dane

W badaniu zastosowano dane czasu rzeczywistego pochodzące z bazy danych Banku Rezerwy Federalnej w Filadelfii (RTDS-*Real Time Data Set*) [<http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-ata/data-files>]. Użyto 38 zmiennych kwartalnych z okresu 1995q1–20014q4, dołączono ponadto zmienną objaśnianą/prognozowaną, czyli annualizowane PKB w Stanach Zjednoczonych. Zmienne miesięczne dostępne w bazie podzielono na dwa rozłączne zbiory. Do pierwszego zbioru zaliczono 20 zmiennych z okresu 1995m1–2014m10. Dobór zmiennych do omawianego zbioru podyktowany był momentem publikacji informacji o zmiennej. Założono, że do pierwszego zbioru mogą należeć jedynie zmienne miesięczne, które publikowane są w trakcie wyróżnionego (prognozowanego) kwartału. Drugi zbiór objął 7 zmiennych miesięcznych pochodzących z okresu 1995m1–2014m9. Dla tych zmiennych nie są znane obserwacje miesięczne w prognozowanym (najnowszym) kwartale⁷.

⁴ Wykazano [por. Stock, Watson, 1999], że przy spełnieniu typowych założeń odnoszących się do struktury stochastycznej (tj. założeń schematu Gaussa-Markowa), estymator MNK modelu DFM jest nieobciążony i najefektywniejszy, dlatego ten estymator zastosowano w niniejszej pracy. Stock i Watson wykazali, że prognozy uzyskane z modelu DFM są asymptotycznie efektywne w tym sensie, że błąd MSE zmierza do optymalnego MSE, o ile $N, T \rightarrow \infty$.

⁵ Ang. *multi-step, direct*, tj. prognozowanie z okresu t na $t+h$.

⁶ Ang. *iterated forecast*, tj. krokowe prognozowanie za każdym razem na okres $t+1$ aż do $t+h$.

⁷ Dane zawarte w bazie RTDS pozbawione są sezonowości. Ponieważ modele MIDAS, DFM, ARMA, VAR są modelami stacjonarnych szeregów czasowych, zmienne zostały

Danymi typu *real-time* określa się w literaturze zbiór danych (o charakterze ekonomicznym) zawierających szeregi czasowe obserwacji o zróżnicowanej w czasie wiarygodności informacji. Z uwagi na fakt, że dane ekonomiczne podlegają rewizjom, w których uwzględnia się nowe, uprzednio niedostępne informacje, jak również doskonalą się metody gromadzenia, przetwarzania danych, oraz doskonalą same techniki statystyczne, rewizje zmieniają wiarygodność informacji o kategoriach ekonomicznych opisywanych przez wyróżnione zmienne. W zamyśle rewizje prowadzą do polepszenia jakości danych.

Dane czasu rzeczywistego pojedynczej zmiennej mają formę macierzy (górnotrójkątnej), przy czym kolumnom odpowiadają momenty publikacji informacji o kształtowaniu się zmiennej, wierszom odpowiadają daty określające okres, którego dotyczy informacja. Dzięki takiemu układowi zbioru w dowolnym wierszu można sprawdzić, jak zmieniały się wartości zmiennej (jej oszacowania) będącej przedmiotem badania, wraz z kolejnymi publikacjami dostępnymi w kolejnych kolumnach. Ponadto szereg czasowy dostępny w każdej kolumnie rozumiany jest jako najnowsza dostępna informacja o kształtowaniu się zmiennej w danym momencie publikacji. Elementy głównej przekątnej macierzy zawierają informacje o pierwszych publikacjach zmiennej (czyli bez rewizji), takich jakie były dostępne w momencie publikacji ówczesnie dla ówczesnych badaczy [<http://www.philadelphiafed.org/research-and-data/real-time-center>].

W literaturze [Clements, Galvao, 2010; Koenig i inni, 2003] wyróżnia się dwie ważne przyczyny rewizji danych:

- hipotezę błędu pomiaru (*noise hypothesis*),
- hipotezę prognoz efektywnych (*news hypothesis*).

Jeżeli symbolem y_t^s oznaczmy ocenę zmiennej y dokonaną na okres t w momencie s ($t=1, \dots, T$; $s \geq t$), wówczas ocenę y_t^s można dekomponować na: prawdziwą wartość zmiennej, oznaczoną \tilde{y}_t , zakłócenia ε_t^s i innowacji v_t^s , stąd:

$$y_t^s = \tilde{y}_t + \varepsilon_t^s + v_t^s \quad (13)$$

Rewizje są zgodne z hipotezą prognoz efektywnych, jeśli innowacje nie są skorelowane z bieżącymi ocenami zmiennej, czyli jeśli

doprowadzone do stacjonarności. Przekształcenia zmiennych sprowadzające je do stacjonarności takiej jak różnicowanie, dwukrotne różnicowanie, różnicowanie logarytmów itp., zaczerpnięto z pracy [Stock, Watson, 2002].

$\text{cov}(v_t^s, y_t^s) = 0$. Mówi się wówczas, że pierwotnie dokonane oceny są optymalnymi prognozami bieżących ocen. Natomiast hipoteza błędu pomiaru dla rewizji jest spełniona, o ile zakłócenia nie są skorelowane z prawdziwą wartością zmiennej, czyli o ile zachodzi $\text{cov}(\varepsilon_t^s, \tilde{y}_t) = 0$.

Hipoteza błędu pomiaru i hipoteza prognoz efektywnych w „czystej” postaci przyjęłyby formy odpowiednio: $y_t^s - \tilde{y}_t = \varepsilon_t^s$ i $y_t^s - \tilde{y}_t = v_t^s$, formuła (13) wyraża tzw. rewizje mieszane. Weryfikacja empiryczna obu hipotez nie dostarcza jednoznacznych wskazań na korzyść którejkolwiek z nich. C. Richardson [2003], J. Faust, J. Rogers, J. Wright [2005], wykazali, że rachunki narodowe Wielkiej Brytanii zachowują się zgodnie z hipotezą prognoz efektywnych. N. G. Mankiw, M. D. Shapiro [1986], G. Kapetanios, T. Yates [2004] dowodzą, że rachunki narodowe amerykańskiego i brytyjskiego PNB zachowują się zgodnie z hipotezą błędu pomiaru.

Empiryczne ustalenie lub przynajmniej aprioryczne założenie, czy rewizje zachowują się zgodnie z którąś z przytoczonych hipotez ma decydujące znaczenie dla konstrukcji modelu stosowanego następnie do formułowania prognoz. Przyjmując za właściwą hipotezę prognoz efektywnych, w konsekwencji uznaje się, że rewizje są nieprognozowalne. Przyjmując natomiast za obowiązującą hipotezę błędu pomiaru, pojawiają się dwie możliwe strategie: podjęcie próby estymacji błędu pomiaru na podstawie oceny obciążenia y_t^s , podjęcie próby oceny błędu pomiaru na podstawie dodatkowych danych [Chamberlin, 2007, 2010].

W końcu, jeśli przyjmie się za obowiązujące obie hipotezy (wersja mieszana), czyli uwzględni się wpływ na rewizję zarówno błędów pomiaru, jak i innowacji, można [Jacobs, van Norden, 2011] zapisać model (13) uwzględniający całą historię rewizji w przestrzeni stanów i stosując filtr Kalmana próbować szacować czynniki nieobserwowalne.

Prognozowanie rewizji ma oczywiście na celu podniesienie jakości danych, jednak efektywność takich prognoz jest w świetle literatury co najmniej wątpliwa [Patterson, 2002; Brown i inni 2010]. Wspomniani autorzy, korzystając z szeregów czasowych o bardzo dużej długości, wykazali, że rewizje są niestacjonarne. Potwierdzono istnienie wielu trendów stochastycznych, zależnych od momentu przeprowadzania rewizji i daty publikacji szeregu czasowego. Autorzy skłaniali się ku konkluzji, iż skoro rewizje nie mają wspólnego wzorca zmienności, próby ich prognozowania (które jak każda prognoza jest próbą ekstrapolacji przeszłości w przyszłość), mogą pogorszyć zamiast polepszyć

jakość danych, a tym samym pogorszyć jakość prognoz konstruowanych na podstawie danych z uwzględnionymi prognozami rewizji.

Niezależnie od przyjętej lub faktycznie obowiązującej hipotezy odnoszącej się do rewizji (błędu pomiaru, prognoz efektywnych, mieszanej), faktem jest, że rewizje mogą wywierać istotny wpływ na jakość prognoz. W literaturze przedmiotu [Stark, Croushore, 2002; Croushore, 2010] wymienia się trzy sposoby oddziaływania rewizji na prognozy ekonomiczne:

- bezpośrednio, poprzez zmianę wartości zmiennej prognozowanej, wówczas ten sam model w różnych okresach dostarcza prognoz różniących się precyzją,
- pośrednio, prowadząc do zmiany oszacowań parametrów strukturalnych modelu,
- pośrednio, prowadząc do zmiany specyfikacji modelu polegającej na zmianie wyboru zmiennych objaśniających lub ich struktury opóźnień.

Wymienieni autorzy jednoznacznie stwierdzają, że konstruując model do prognozowania ekonomicznego, błędem jest nieposługiwanie się danymi czasu rzeczywistego. Niezależnie od tego, kiedy szacujemy dany model, powinniśmy posługiwać się tymi samymi zbiorami danych i o takiej samej jakości, jakie dostępne były ówczesnym badaczom. Każda później dokonana rewizja, podnosząc jakość danych (zmniejszając błąd pomiaru, czy uwzględniając innowacje) pozwala oszacować potencjalnie lepszy model (lepiej dopasowany do danych empirycznych) nieosiągalny w przeszłości.

Dostęp do bazy danych czasu rzeczywistego oferowany przez Bank Rezerwy Federalnej w Filadelfii jest nieodpłatny. Nieskrępowany dostęp do bazy ograniczył pole badania do amerykańskiego (a nie krajowego) PKB. Zdaniem autora w świetle argumentów przytoczonych w niniejszym paragrafie zastosowanie danych czasu rzeczywistego miało kluczowe znaczenie dla rzetelnego porównania jakości/precyzji prognoz generowanych przez różne modele, w tym dynamiczne modele o zmiennych o zróżnicowanej częstotliwości.

3. Wyniki

Badanie zostało zaplanowane jako symulacja ośmiu sesji prognozytycznych (stąd zastosowanie danych czasu rzeczywistego pozwalające symulować sesje prognoz *ex ante*). W każdej z sesji wyznaczano prognozy krótkookresowe o horyzoncie $h=1$ ($h=2/3$ dla regresji MIDAS).

W pierwszej sesji prognozowano PKB na okres 2013q1, w kolejnych sesjach uzupełniano zbiory zmiennych i wyliczono prognozy na kolejne kwartały roku 2013, a następnie roku 2014, kończąc prognozą na 2014q4.

Procedura estymacji i prognozowania z użyciem modelu MIDAS obejmowała:

1. Oszacowanie wszystkich możliwych modeli dla założonych rzędów $i, j, l = 0, 1, \dots, 4, g = 1, 2, 3, 4, m = 3, K = 12$.
2. Testowanie autokorelacji składników zakłócających modeli i wybór do dalszego badania tych spośród nich, dla których nie znaleziono podstaw do odrzucenia hipotezy o braku autokorelacji (testowano autokorelację do rzędu 4 łącznie).
3. Na podstawie kryterium BIC wybór „najlepszego” modelu [por. procedura wyboru najlepszego modelu – Stock, Watson, 2006].
4. Oszacowanie prognozy na podstawie modelu „najlepszego”.

Identycznie przebiegały procedury sesji prognostycznych zastosowane dla pozostałych klas modeli. W modelu DFM założono maksymalnie cztery procesy DL, każdy z nich maksymalnie czwartego rzędu, ponadto maksymalnie cztery procesy AR, każdy z nich maksymalnie czwartego rzędu. W modelu ARIMA(p, d, q) założono maksymalne rzędy $p=6$ i $q=6$, podobnie jak dla modeli wcześniejszych, szacowano każdą możliwą kombinację p i q .

W modelu wektorowo-autoregresyjnym wykorzystano zmienne z wielorównaniowego modelu gospodarki amerykańskiej FAIRMODEL [Fair, 2013]. Wektor zmiennych tworzyły: PKB, wartość produkcji sektora cywilnego, wartość produkcji sektora militarnego, liczba przepracowanych roboczogodzin w obu sektorach (uwzględniania w postaci dwóch osobnych zmiennych), zyski przed opodatkowaniem. Dla modelu VAR założono maksymalny rząd procesu wynoszący 4, ale ostateczną decyzję odnoszącą się do rzędu procesu VAR podejmowano na podstawie porównania miar pojemności informacyjnej (BIC) oraz wyników testów autokorelacji dla poszczególnych równań modelu VAR. Wybrano jedynie spośród tych rzędów procesu, dla których wynik testów autokorelacji wskazywał na brak podstaw do odrzucenia hipotezy zerowej. Ponadto nie przeprowadzono strukturalizacji modelu z uwagi na jego pomocniczy charakter, prognozy modeli VAR w niniejszym badaniu użyte zostały jedynie jako pewnego rodzaju benchmark dla prognoz uzyskiwanych z modeli o zmiennych o zróżnicowanej częstotliwości.

Po oszacowaniu prognoz wyznaczone zostały ich błędy *ex post* oraz miary MSE, RMSE i V. Minimalny błąd RMSE wskazywał na metodę dostarczającą najbardziej precyzyjnych prognoz (tablica 1).

Tablica 1. Prognozy tempa wzrostu annualizowanego PKB w USA oraz miary błędów prognoz *ex post* z ośmiu symulowanych sesji prognostycznych

Kwar- tał	PKB [%]	AR-MIDAS		MIDAS		DFMD	DFMI	ARI- MA	VAR	NAIW- NA
		exp	beta	exp	beta					
2013q1	2,714	2,917	2,925	2,952	3,154	3,225	3,125	2,256	2,853	0.062
2013q2	1,756	2,071	2,182	2,553	2,68	2,811	2,914	2,991	2,258	2.714
2013q3	4,442	3,255	3,185	2,081	2,014	2,004	2,057	1,995	3,201	1.756
2013q4	3,455	2,896	2,781	2,891	2,708	2,612	2,599	2,55	3,387	4.442
2014q1	-2,12	0,859	0,995	1,792	1,855	1,987	1,972	1,994	2,55	3.455
2014q2	4,515	2,412	2,384	2,057	1,993	1,985	1,944	1,921	2,481	-2.124
2014q3	4,877	3,188	2,952	3,111	3,019	2,99	3,014	2,955	2,849	4.515
2014q4	2,616	2,257	2,384	2,149	2,114	2,057	2,02	1,959	2,336	4.877
MSE		2,271	2,536	3,912	4,175	4,400	4,393	4,544	3,999	12,073
RMSE		1,507	1,593	1,978	2,043	2,098	2,096	2,132	2,000	3,475
V		60,7%	64,4%	80,8%	83,7%	85,3%	85,4%	91,6%	73,0%	141,1%

W modelach AR-MIDAS i MIDAS skrót „exp” oznacza, że stosowano wykładniczy, zgodny z formułą (6), sposób szacowania wag; analogicznie skrót „beta” oznacza zastosowanie formuły (4) zgodnej z rozkładem beta. DFMD, oraz DFMI oznaczają prognozy z modeli DFM uzyskane bezpośrednio (D), iteracyjnie (I).

Źródło: Obliczenia własne.

Z zestawienia zawartego w tablicy 1 wynika, że najniższy błąd RMSE zanotowano dla modelu AR-MIDAS z wagami rozłożonych opóźnień przy zmiennych o częstotliwości miesięcznej liczonymi według formuły wykładniczej (6). Zgodnie z przyjętymi kryteriami model ten uznano za narzędzie dostarczające najbardziej precyzyjnych prognoz, a same prognozy zostały uznane za „najlepsze”.

W kroku kolejnym parami testowano identyczność prognoz pozostałych z prognozami „najlepszymi”, tj. testowano, czy precyzja prognoz uzyskanych z modelu AR-MIDAS(exp) jest statystycznie istotnie lepsza od precyzji prognoz z dowolnego innego modelu. W tym celu posłużono się testem Diebolda-Mariano z poprawką małopróbkową [Diebold, 2012]. W tablicy 2 zebrano statystyki testu Diebolda-Mariano (DM), a w nawiasach kwadratowych empiryczne prawdopodobieństwa odrzucenia hipotezy zerowej o identycznej dokładności/precyzji prognoz.

Tablica 2. Wyniki testu Diebolda-Mariano (z poprawką małopróbkową)

AR-MIDAS		MIDAS		DFMD	DFMI	ARIMA	VAR	NAIWNA
exp	beta	exp	beta					
DM	-2,063	-1,997	-2,148	-2,169	-2,217	-2,337	-1,071	-1,950
[prob]	[0,038]	[0,046]	[0,034]	[0,033]	[0,031]	[0,026]	[0,159]	[0,046]

Źródło: Obliczenia własne.

Wnioski z testowania identyczności precyzji prognoz przedstawiają się następująco:

1. Wyniki testu Diebolda-Mariano wskazują, że stosując typowe poziomy istotności, nie można odrzucić hipotezy (zerowej) o identycznej precyzji prognoz „najlepszych”, tj. uzyskanych z modelu AR-MIDAS (exp) i prognoz pochodzących z modelu VAR.
2. Wyłączony model VAR i stosując 5% lub 10% poziom istotności można stwierdzić, że model AR-MIDAS(exp) dostarcza bardziej precyzyjnych prognoz niż pozostałe modele użyte w niniejszym badaniu.
3. Stosując 1% poziom istotności, z wyników testu DM wynika, że precyzja dowolnych prognoz (w tym prognozy naiwnej) nie jest gorsza niż precyzja prognoz uznanych za „najlepsze”.

Należy stwierdzić, że wnioski różnią się zasadniczo w zależności od przyjętego we wnioskowaniu poziomu istotności. W skrajnych przypadkach (por. podpunkt 1) otrzymujemy dwa najlepsze modele (i ich prognozy), lub wręcz przeciwnie, modele i ich prognozy nierozróżnialne od prognozy naiwnej (por. podpunkt 3).

Postawiono hipotezę, że otrzymany, nie do końca konkluzyjny wynik związany jest ze zbyt krótkim szeregiem porównywanych prognoz, pomimo zastosowanej poprawki małopróbkowej polepszającej moc testu Diebolda-Mariano. W celu sprawdzenia tej hipotezy metodą bootstrapową wydłużono próbkę o kolejne osiem prognoz, tak by w każdej iteracji bootstrapowej porównywanych było szesnaście par prognoz. Iteracje bootstrapowe (tj. losowanie z powtórzeniami par prognoz) powtarzano 10 000 razy. Bootstrapowe statystyki testu Diebolda-Mariano, wraz z prawdopodobieństwami empirycznymi odrzucenia hipotezy zerowej zawarto w tablicy 3.

Tablica 3. Wyniki bootstradowego testu Diebolda-Mariano (z poprawką małopróbkową)

AR-MIDAS		MIDAS		DFMD	DFMI	ARIMA	VAR	NAIWNA
exp	beta	exp	beta					
DM-	-2,830	-3,059	-3,406	-3,539	-3,370	-3,866	-1,092	-2,997
[prob-	[0,006]	[0,006]	[0,002]	[0,002]	[0,001]	[0,001]	[0,146]	[0,006]

Źródło: Obliczenia własne.

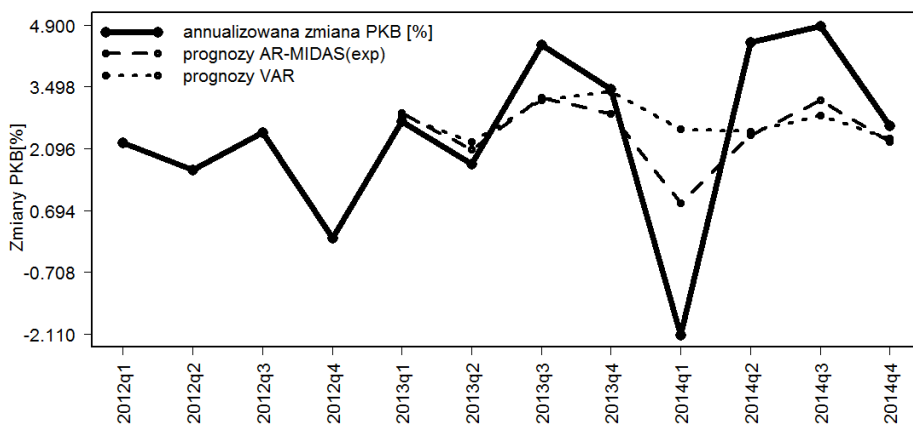
Zastosowana w teście DM technika repróbki, prowadząca *de facto* do wydłużenia próby, a przez to prowadząca do zwiększenia mocy testu, pozwala na dwa zasadnicze wnioski:

1. Precyzja prognoz PKB w USA pochodzących z modelu AR-MIDAS(exp) i modelu VAR jest statystycznie nieodróżnialna, pomimo że (próbki) błąd RMSE prognoz modelu MIDAS jest niższy niż analogiczny błąd modelu VAR.
2. Stosując dowolny z powszechnie używanych poziomów istotności, test DM wskazuje na wyższą precyzję prognoz modelu AR-MIDAS (exp) w porównaniu do prognoz uzyskanych z pozostałych modeli (wyłączywszy prognozy model VAR).

Wnioski wynikające z testów DM, zarówno w wersji zawartej w tabelicy 2, jak i tabelicy 3 nie pozwalają na jednoznaczne poparcie tezy pracy, która brzmiała: uwzględnienie w modelu dynamicznym zmiennych o częstotliwości wyższej niż częstotliwość zmiennej objaśnianej statystycznie istotnie poprawia jakość (dopasowanie) prognoz makroekonomicznych. Nie stwierdzono statystycznie lepszej precyzji/dopasowania prognoz w modelu AR-MIDAS w stosunku do precyzji prognoz modelu VAR.

W ostatnim etapie badania wykonano wykres zmiennej prognozowanej, prognoz dla $h=2/3$ uzyskanych z modelu AR-MIDAS(exp) oraz prognoz dla $h=1$ uzyskanych z modelu VAR (rysunek 1). Z rysunku 1 można odczytać, że zarówno prognozy uzyskane z modelu AR-MIDAS(exp) (oznaczone linią przerywaną), jak i prognozy uzyskane z modelu VAR (oznaczone linią kropkowaną) charakteryzują się zbyt małą zmiennością w porównaniu do zmienności PKB obserwowanego (oznaczonego pogrubioną ciągłą linią). Wykresy obu rodzajów prognoz są spłaszczone w stosunku do wykresu zmiennej prognozowanej, wariancja prognoz jest zbyt niska.

Rysunek 1. Zmienna prognozowana i najlepiej dopasowane prognozy krótkookresowe pochodzące z modeli AR-MIDAS(exp) i VAR



Źródło: Opracowanie własne.

Ponadto z wykresu można odczytać, że w okresach 2013q2, 2013q3, 2014q1, 2014q3 wystąpiły tzw. punkty zwrotne w kształtowaniu się zmiennej objaśnianej. W każdym przypadku zostały one prawidłowo prognozowane na podstawie modelu AR-MIDAS(exp), prawdopodobnie na skutek uwzględnienia najnowszej pochodzącej z danego kwartału informacji miesięcznej. Jednocześnie prognozy z modelu VAR nie wskazały poprawnie wszystkich punktów zwrotnych. W okresie 2013q4 prognoza PKB rosła, podczas gdy okazało się, że obserwowane PKB w istocie zmalało, w okresie 2014q2 prognoza wskazuje na spadek PKB, podczas gdy faktyczne PKB wzrosło. Poprawność prognozowania punktów zwrotnych uzyskana na podstawie modelu AR-MIDAS jest o tyle cenna w prognozowaniu krótkookresowym, że badacz często zainteresowany jest informacją o tym, czy trend wzrostowy (spadkowy) zostanie w przyszłości utrzymany, czy raczej należy spodziewać się odwrócenia tendencji. Wydaje się, że skuteczność prognozowania punktów zwrotnych przez regresję MIDAS wynika z unikalnej zdolności tego typu modeli do adaptacji informacji pochodzących ze zmiennych o wyższej częstotliwości a odnoszących się bezpośrednio do prognozowanego okresu. Pomimo empirycznie stwierdzonej jednakowej precyzji prognoz AR-MIDAS i VAR, te pierwsze jawią się jako bardzo ciekawa alternatywa dla „tradycyjnego” modelowania i prognozowania wektorowo-autoregresyjnego.

Zakończenie

Przeprowadzone badania wskazują na najlepsze dopasowanie prognoz krótkookresowych wyznaczanych na podstawie modeli dynamicznych łączących zmienne o zróżnicowanej częstotliwości, czyli modeli klasy MIDAS, oraz statystycznie taką samą precyzję prognoz uzyskiwanych na podstawie modeli VAR. Spośród tych dwóch klas modeli jedynie regresja MIDAS dostarczała prognoz prawidłowo reagujących na punkty zwrotne, przypuszcza się, że cecha ta wynika ze zdolności modeli MIDAS do uwzględnienia w prognozie najnowszej informacji pochodzącej ze zmiennych wysokiej częstotliwości odnoszących się do prognozowanego okresu zmiennej o niskiej częstotliwości. Z badań wynika, iż można rekomendować wykorzystanie modeli AR-MIDAS do prognoz typu *now-casting*.

Literatura

1. Andreou E., Ghysels E., Kourtellos A. (2010), *Forecasting with mixed-frequency data*, w: *Oxford Handbook on Economy Forecasting*, Clements M. P., Hendry D. F. (ed.).
2. Artis M., Banerjee A., Marcelino M. (2003), *Factor forecast for the UK*, „Bacconi University Working Paper”, No. 203.
3. Boivin J., Ng S. (2006), „Are more data always better for factor analysis?”, *Journal of Econometrics*, Vol. 132, No. 1.
4. Brown G., Buccellato T., Chamberlin G., Dey-Chowdhury D., Youl R. (2010), *Understanding the quality of early estimates of Gross Domestic Product*, „Economic & Labour Market Review”, Vol. 4, No. 6.
5. Chamberlin G. (2007), *Forecasting GDP using external data sources*, „Economic and Labour Market Review”, Vol. 1, No. 8.
6. Chamberlin G. (2010), *Real time data*, „Economic and Labour Market Review”, Vol. 4, No. 12.
7. Chen X., Ghysels E. (2009), *News – good or bad – and its impact on predicting future volatility*, „Review of Financial Studies”, Vol. 24, No. 1.
8. Chow G., Lin A. (1971), *Best linear unbiased interpolation, distribution and extrapolation of time series by related time series*, „Review of Economics and Statistics”, Vol. 53.
9. Clements M. P., Galvao A. B. (2010), *Real-time Forecasting of Inflation and Output growth in the Presence of Data Revisions*, „Warwick Economic Research Papers”, No. 953.

10. Clements M. P., Galvao A.B. (2006), *Macroeconomic Forecasting with Mixed Frequency Data: Forecast of US output growth and inflation*, „Warwick Economic Research Papers”, No. 773.
11. Croushore D. (2005), *Forecasting with Real-Time Data Vintages*, „University of Richmond Working Paper”.
12. Diebold F. (2012), *Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Test*, „University of Pennsylvania Working Paper”, No. 7.
13. Fair R. (2013), *Macroeconometric Modeling*, <http://fairmodel.econ.yale.edu/mmm/mm.pdf>.
14. Faust J., Rogers J., Wright J. (2005), *News and noise in G7 announcements*, „Centre for Economic Policy Research”, Vol. 12.
15. Forni M., Hallin M., Lippi M., Reichlin L. (2005), *The Generalized Dynamic Factor Model*, „Journal of the American Statistical Association”, Vol. 100, No. 471.
16. Foroni C., Marcellino M. (2013), *A survey of econometric methods for mixed-frequency data*, „Working Paper”, No. 6 Norge Bank Research.
17. Ghysels E., Santa-Clara P., Valkanov R. (2004a), *The MIDAS touch: Mixed Data Sampling regression models*, Chapel Hill, N.C.
18. Ghysels E., Santa-Clara P., Valkanov R. (2004b), *Predicting volatility: Getting the most out of return data sampled at different frequencies*, NBER Working Paper No. 10914.
19. Ghysels E., Sinko A., Valkanov R. (2004c), *MIDAS regression: Further results and new directions*, „Econometric Reviews”, Vol. 26.
20. Ghysels E., Santa-Clara P., Valkanov R. (2005), *There is a risk-return trade-off after all*, „Journal of Financial Economics”, Vol. 76, No. 3.
21. Hendry D. F., Mizon G.E. (1978), *Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England*, „Economic Journal”, Vol. 88.
22. Jacobs J., van Norden S. (2011), *Modeling data revisions: Measurement error and dynamics of „true” values*, „Journal of Econometrics”, Vol. 161.
23. Kapetanios G., Yates T. (2004), *Estimating time-variation in measurement error from data revision; an application to forecasting in dynamic models*, „Bank of England Working Papers”, No. 238.
24. Koenig E. F., Doms S., Piger J. (2003), *The use and abuse of real-time data on economic forecasting*, „The Review of Economic and Statistics”, Vol. 85, No. 3.

25. Mankiw N. G., Shapiro M. D. (1986), *News of noise. An analysis of GNP revision*, „Survey of Current Business”.
26. Marcellino M., Stock J., Watson M. (2006), *A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series*, „Journal of Econometrics”, Vol. 135, No. 1–2.
27. Miller P. J., Chin D. M. (1996), *Using monthly data to improve quarterly model forecasts*, „Federal Reserve Bank Minneapolis Quarterly Review”, Vol. 20.
28. Patterson K. (2002), *The data measurement process for UK GNP: stochastic trends, long memory and unit roots*, „Journal of Forecasting”, Vol. 21.
29. Richardson C. (2003), *Revision analysis: a time series approach*, „Economic Trends”, Vol. 12.
30. Schneider M., Spitzer M. (2004), *Forecasting Austrian GDP using the generalized dynamic factor model*, „Oesterreichische Nationalbank Working Paper”, Vol. 89.
31. Stark T., Croushore D. (2002), *Forecasting with a real time data set for macroeconomists*, „Journal of Macroeconomics”, Vol. 24.
32. Stock J., Watson M. (1999), *Forecasting Inflation*, „Journal of Monetary Economics”, Vol. 44.
33. Stock J., Watson M. (2002), *Macroeconomic Forecasting Using Diffusion Indexes*, „Journal of Business and Economic Statistics”, Vol. 20, No. 2.
34. Stock J., Watson M. (2005), *Implications of Dynamic Factor Models for VAR Analysis*, „NBER Working Papers”, No. 11467.
35. Stock J., Watson M. (2006), *Forecasting with Many Predictors*, „Handbook of Economic Forecasting”, Vol. 1.

Streszczenie

Celem badania było sprawdzenie, czy uwzględnienie w modelu dynamicznym zmiennych o częstotliwości wyższej niż częstotliwość zmiennej objaśnianej statystycznie istotnie poprawia jakość (dopasowanie) prognoz makroekonomicznych. Prognozy makroekonomiczne wyznaczone były z modeli dynamicznych łączących zmienne o zróżnicowanej częstotliwości znanych jako regresja MIDAS, następnie porównywane były z prognozami tych samych kategorii ekonomicznych uzyskanych z modeli DFM, ARIMA, VAR. Do oszacowania modeli, prognozowania krótkookresowego (horyzont prognozy $h=1$) i symulacji ośmiu sesji prognostycznych wykorzystano dane czasu rzeczywistego odnoszące się do gospodarki USA. Z przeprowadzonych badań wynika, że dynamiczny model MIDAS (AR-MIDAS) i model VAR dostarczyły najbardziej precyzyjnych prognoz, prognozy z modeli DFM i ARIMA okazały się statystycznie mniej precyzyjne. Spośród dwóch klas modeli dostarczających

najbardziej precyzyjnych prognoz (AR-MIDAS, VAR) jedynie AR-MIDAS wygenerował prognozy, prawidłowo przewidując wszystkie punkty zwrotne, cech tej nie miały prognozy VAR. Z tego powodu można rekomendować wykorzystanie modeli AR-MIDAS do prognoz typu *now-casting*.

Słowa kluczowe

prognoza, zmienne zróżnicowanej częstotliwości, MIDAS, DFM

Mixed Frequency Data Dynamic Models in Macroeconomic Forecasting (Summary)

The aim of the study was to find out whether the use of mixed frequency data dynamic models can improve the accuracy of quarterly forecasts of selected macroeconomic variables. The research tools were the MIDAS, AR-MIDAS, DFM, ARIMA and VAR models with quarterly forecasts as reference points. The study used a simulation of eight (*ex ante*) forecasting sessions. To that end, it was necessary to use variables from the real-time database. Yet, due to the limited access to such databases, the study focused on U.S. GDP. The results indicate that the AR-MIDAS class model, which directly incorporates the latest available monthly information as well as „classic” VAR model provided significantly more accurate forecasts of GDP than DFM and ARIMA models. Comparing AR-MIDAS and VAR forecasts (which were equally accurate in terms of Diebold-Mariano test) only the previous model proved to be able to provide accurate forecasts of all turning points which occurred within the sample. This unique feature enables to recommend AR-MIDAS as a now-casting or short term forecasting tool.

Keywords

mixed frequency data, real-time forecasting, MIDAS, AR-MIDAS, DFM