

Uniwersytet Mikołaja Kopernika w Toruniu
Katedra Ekonometrii i Statystyki

Jarosław Krajewski

ZASTOSOWANIE DYNAMICZNEGO MODELU CZYNNIKOWEGO DO MODELOWANIA I PROGNOZOWANIA PKB W POLSCE[†]

Zarys treści. Referat traktuje o podstawach konstrukcji dynamicznych modeli czynnikowych i ich zastosowaniu empirycznym. DFM stosuje się do prognozowania, konstruowania głównych wskaźników koniunktury, analiz polityki monetarnej i badania międzynarodowych cykli koniunkturalnych. W referacie oszacowano DFM PKB w Polsce w latach 1997–2008, a także oceniono trafność uzyskanych na jego podstawie prognoz w porównaniu do modelu AR i modelu symptomatycznego. Zbiór danych wykorzystanych do badania zawiera 41 zmiennych makroekonomicznych. Najlepszym ze statystycznego punktu widzenia okazał się model z 3 czynnikami.

Słowa kluczowe: dynamiczny model czynnikowy, metoda głównych składowych, PKB.

1. WSTĘP

Dynamiczne modele czynnikowe (*Dynamic Factor Models – DFM*) w ostatnim czasie stały się bardzo popularne w empirycznych analizach makroekonomicznych. Zauważyć można także znaczny rozwój tego narzędzia pod względem metodologicznym. Za pionierów w zakresie DFM uważa się Geweke'a (1977) oraz Simsa i Sargenta (1977), którzy zastosowali ten typ modeli do małych zbiorów danych. DFM stosuje się do prognozowania, konstruowania głównych wskaźników koniunktury, analiz polityki monetarnej i badania mię-

[†] Praca naukowa współfinansowana ze środków Europejskiego Funduszu Społecznego i Budżetu Państwa w ramach Zintegrowanego Programu Operacyjnego Rozwoju Regionalnego, Działania 2.6 „Regionalne Strategie Innowacyjne i transfer wiedzy” projektu własnego Województwa Kujawsko-Pomorskiego „Stypendia dla doktorantów 2008/2009 – ZPORR”



ZPORR
Zintegrowany Program
Operacyjny
Rozwoju Regionalnego



WOJEWÓDZTWO
KUJAWSKO-POMORSKIE



dzynarodowych cykli koniunkturalnych. Dynamic Factor Models prezentują ateoretyczne podejście do modelowania ekonometrycznego (por. Sims, 1980).

Celem prezentowanego artykułu jest skonstruowanie dynamicznego modelu czynnikowego PKB w Polsce w latach 1997–2008. Ponadto w pracy przedstawiona zostanie istota DFM i jedna z metod ich estymacji. Kolejny punkt zawierał będzie opis danych wykorzystanych w obliczeniach, a także empiryczny model PKB.

2. DYNAMICZNY MODEL CZYNNIKOWY

Koncepcja modeli czynnikowych opiera się na założeniu, że zachowanie się większości zmiennych makroekonomicznych może być dobrze opisane za pomocą małej liczby nieobserwowalnych wspólnych czynników. Czynniki te często są interpretowane jako wiodące siły w ekonomii. Poszczególne zmienne mogą wtedy zostać wyrażone jako liniowa kombinacja mniej niż 20 czynników, które wyjaśniają znaczącą część ich zmienności (por. Kotłowski, 2008).

Niech y_t oznacza pewien szereg czasowy i \mathbf{X}_t wyraża wektor N zmiennych w postaci szeregów czasowych zawierających informacje użyteczne w modelowaniu, a także prognozowaniu wartości y_t . W dynamicznym modelu czynnikowym zakładamy, że wszystkie zmienne x_{it} zawarte w wektorze \mathbf{X}_t mogą zostać wyrażone jako liniowa kombinacja bieżących i opóźnionych nieobserwowalnych czynników f_{it}

$$x_{it} = \lambda_i(L)\mathbf{f}_t + e_{it}, \text{ dla } i = 1, \dots, N, \quad (1)$$

gdzie $\mathbf{f}_t = [f_{1t}, f_{2t}, \dots, f_{\bar{r}t}]'$ jest wektorem \bar{r} nieobserwowalnych wspólnych czynników w momencie t , $\lambda_i(L) = \lambda_{i0} + \lambda_{i1}L + \lambda_{i2}L^2 + \dots + \lambda_{iq}L^q$ jest operatorem opóźnień, natomiast e_{it} wyraża swoisty błąd zmiennej x_{it} (por. Stock, Watson, 1998).

Stąd też y_t może być zapisane jako funkcja bieżących i opóźnionych wspólnych czynników zawartych w wektorze \mathbf{f}_t oraz opóźnionych wartości y_t w następujący sposób:

$$y_t = \beta(L)\mathbf{f}_t + \gamma(L)y_t + e_t. \quad (2)$$

Zatem można powiedzieć, że dynamiczny model czynnikowy składa się z równań (1) i (2).

3. ESTYMACJA PARAMETRÓW DFM I SPECYFIKACJA LICZBY CZYNNIKÓW

Jedną z najczęściej używanych metod estymacji parametrów i czynników w modelach czynnikowych jest metoda głównych składowych. W metodzie tej obie macierze, czynników i parametrów, są nieznane. Model przedstawiony jako równanie (1) może zostać zapisany w następującej formie macierzowej:

$$\mathbf{X} = \mathbf{F}\mathbf{H}\mathbf{H}^{-1}\mathbf{\Lambda}' + \mathbf{e}, \quad (3)$$

gdzie \mathbf{H} jest niejednostkową macierzą o wymiarach $r \times r$. Niezbędne jest wykonanie odpowiedniej normalizacji macierzy \mathbf{H} . Stock i Watson (1998) zaproponowali warunek $(\mathbf{\Lambda}'\mathbf{\Lambda}/N) = \mathbf{I}_r$, który może zostać nałożony na parametry modelu i sprawi, że macierz \mathbf{H} będzie ortonormalna.

Estymacja macierzy \mathbf{F} i $\mathbf{\Lambda}$ przy użyciu metody głównych składowych polega na znalezieniu takich estymatorów macierzy $\hat{\mathbf{F}}$ i $\hat{\mathbf{\Lambda}}$, które będą minimalizować sumę kwadratów reszt równania (3) wyrażoną w następujący sposób:

$$V(\mathbf{F}, \mathbf{\Lambda}) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \mathbf{\Lambda}_i' \mathbf{F}_t)^2. \quad (4)$$

W pierwszym kroku należy dokonać minimalizacji funkcji (4) w odniesieniu do macierzy czynników \mathbf{F} przy założeniu, że macierz $\mathbf{\Lambda}$ jest znana i stała. W wyniku tego otrzymany zostanie estymator $\hat{\mathbf{F}}$, jako funkcja $\mathbf{\Lambda}$, który następnie zastępuje w powyższym równaniu prawdziwe wartości \mathbf{F} . W drugim kroku minimalizowana jest funkcja (4) w odniesieniu do macierzy $\mathbf{\Lambda}$ z warunkiem normalizacji $(\mathbf{\Lambda}'\mathbf{\Lambda}/N) = \mathbf{I}_r$, w ten sposób otrzymany zostaje bezpośrednio estymator $\hat{\mathbf{\Lambda}}$. Warto zauważyć, że jest to równoznaczne z maksymalizacją wyrażenia $tr[\mathbf{\Lambda}'(\mathbf{X}'\mathbf{X})\mathbf{\Lambda}]$.

Kolejne kolumny macierzy $\hat{\mathbf{\Lambda}}$ są wektorami własnymi, macierzy $\mathbf{X}'\mathbf{X}$ pomnożonej przez \sqrt{N} , odpowiadającymi największym wartościom własnym tej macierzy. Z kolei estymator macierzy \mathbf{F} jest wyrażony jako

$$\hat{\mathbf{F}} = (\mathbf{X}\hat{\mathbf{\Lambda}}) / N. \quad (5)$$

Stock i Watson podkreślają, że jeżeli liczba zmiennych jest wyższa od liczby obserwacji, tzn. $N > T$, wtedy z obliczeniowego punktu widzenia łatwiejsza do zastosowania jest procedura polegająca na oszacowaniu $\tilde{\mathbf{F}}$ przez minimalizację (4) z uwzględnieniem dla \mathbf{F} warunku $\mathbf{F}'\mathbf{F}/T = \mathbf{I}_r$. Macierz $\tilde{\mathbf{F}}$ zawiera wówczas wektory własne z macierzy $\mathbf{X}'\mathbf{X}$ odnoszące się do r największych wartości własnych z tej macierzy pomnożonej przez \sqrt{T} . Z kolei estymator macierzy $\tilde{\mathbf{\Lambda}}$ przyjmie następującą formę:

$$\tilde{\Lambda}' = (\tilde{\mathbf{F}}' \mathbf{X}) / T. \quad (6)$$

Oba estymatory $\hat{\mathbf{F}}$ i $\tilde{\mathbf{F}}$ są równoważne.

W praktyce liczba czynników, niezbędna do pokazania związków pomiędzy zmiennymi, jest zazwyczaj nieznana. Istnieją jednak kryteria, które mogą zostać użyte do wyznaczenia liczby czynników. Bai i Ng (2002) zaproponowali w tym celu następujące kryteria informacyjne:

$$IC_1(k) = \ln(\hat{V}(k)) + k \left(\frac{N+T}{NT} \right) \ln \left(\frac{NT}{N+T} \right) \quad (7)$$

$$IC_2(k) = \ln(\hat{V}(k)) + k \left(\frac{N+T}{NT} \right) \ln C_{NT}^2 \quad (8)$$

$$IC_2(k) = \ln(\hat{V}(k)) + k \left(\frac{\ln C_{NT}^2}{C_{NT}^2} \right) \quad (9)$$

W powyższych wzorach $\hat{V}(k)$ oznacza sumę kwadratów reszt z k – czynnikowego modelu, a $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$.

4. DANE I ANALIZA WYNIKÓW EMPIRYCZNYCH

W badaniu zastosowanie znalazły kwartalne dane makroekonomiczne charakteryzujące polską gospodarkę. Zbiór danych zawiera 41 zmiennych w postaci szeregów czasowych o częstotliwości kwartalnej. Dane dotyczą okresu od 1 kwartału 1997 do 3 kwartału 2008 roku, a więc każdy szereg składa się z 47 obserwacji. Jako zmienna objaśniana posłużyło PKB. Wszystkie dane pochodzą z internetowej strony Głównego Urzędu Statystycznego¹ i poddane zostały odpowiednim transformacjom. Sprowadzono je do cen stałych z 1 kwartału 1997 roku i oczyszczono z wahań sezonowych. W kolejnym kroku dane zlogarytmowano i zróżnicowano odpowiednio do szeregów, w celu sprowadzenia ich do stacjonarności (por. Greene, 2003). Na koniec dane poddane zostały standaryzacji. W obliczeniach wykorzystano dane dotyczące wielkości sprzedaży produkcji przemysłowej ogółem, jak i jej części składowych, budownictwa w różnych aspektach, handlu krajowego i zagranicznego, inflacji i rynku pracy w różnych ujęciach, sfery budżetowej, a także charakterystyki szeroko rozumianej sfery polityki pieniężnej.

Po wstępnym przygotowaniu danych zastosowana została metoda głównych składowych w celu wyznaczenia czynników. Następnie wyznaczone zostały wartości kryteriów informacyjnych Bai'a i Ng w celu specyfikacji ich liczby. Tabela 1 pokazuje wartości tych kryteriów informacyjnych dla różnej liczby

¹ www.stat.gov.pl.

czynników w modelu, a także wartości własne i ich udział w całkowitej zmienności. Ostatecznie w modelu zostały uwzględnione trzy czynniki, ponieważ wskazały na to pierwsze i drugie kryterium. Trzecie kryterium wskazało na konieczność uwzględnienia 10 czynników, jednakże ma ono tendencję do zawyżania ich liczby. Trzy czynniki wyjaśniają niemal 82% całkowitej zmienności, co jest wartością wysoce zadowalającą.

Część badaczy stoi na stanowisku, że czynniki wyznaczone metodą głównych składowych nie mają interpretacji ekonomicznej. Jednakże w niniejszym artykule dokonana została próba ich przybliżonej interpretacji. W tym celu przeprowadzona została analiza współczynników R – kwadrat z regresji pomiędzy poszczególnymi czynnikami a zmiennymi pierwotnymi. Założono, że w skład czynnika wchodzi ta zmienna, której R – kwadrat jest najwyższy. Z powyższego wynika, że pierwszy czynnik reprezentuje głównie rynek pracy i handel zagraniczny. Na drugi czynnik wpływ mają ceny, wynagrodzenia i przychody z działalności gospodarczej. Trzeci czynnik zawiera szeroko pojętą sprzedaż.

Następnie za pomocą kryterium BIC wybrano opóźnienia zarówno dla zmiennej zależnej, jak i dla czynników. Okazało się, że najlepszy jest model zawierający jedynie czynniki bez opóźnień. Wyniki estymacji tego modelu prezentuje tabela 2. Wszystkie współczynniki tego modelu okazują się być istotne na poziomie nieprzekraczającym 5%. Tak oszacowany model opisuje kształtowanie się PKB w ponad 70% i nie występuje w nim autokorelacja. Rzeczywiste wartości PKB i wartości obliczone na podstawie modelu czynnikowego prezentuje wykres 1.

Tabela 1. Wybór liczby czynników do modelu

Liczba czynników	Wartości własne	Udział w zmienności	Skumulowany udział w zmienności	IC1	IC2	IC3
1	76,949	0,721	0,721	-3,694	-3,491	-3,833
2	6,249	0,059	0,779	-3,584	-3,177	-3,861
3	4,278	0,040	0,819	-4,555	-3,945	-4,971
4	2,921	0,027	0,846	-4,490	-3,677	-5,044
5	2,460	0,023	0,870	-4,467	-3,451	-5,160
6	1,976	0,019	0,888	-4,335	-3,115	-5,166
7	1,551	0,015	0,903	-4,192	-2,769	-5,162
8	1,429	0,013	0,916	-4,109	-2,482	-5,216
9	1,282	0,012	0,928	-4,246	-2,417	-5,493
10	1,066	0,010	0,938	-4,126	-2,093	-5,511

Źródło: obliczenia własne.

W tym miejscu sprawdzone zostało również czy kolejność rozważania czynników w procedurze badawczej ma wpływ na wynik końcowy modelowania. Zbadany został wpływ zmiany kolejności czynników na: wartości kryteriów informacyjnych Bai'a i Ng oraz wartości kryterium BIC, jak również na

wartości parametrów modelu i podstawowe miary dobroci modelu. W każdym przypadku okazało się, że zmiana kolejności rozważanych w badaniu czynników nie ma wpływu na wynik końcowy modelowania.

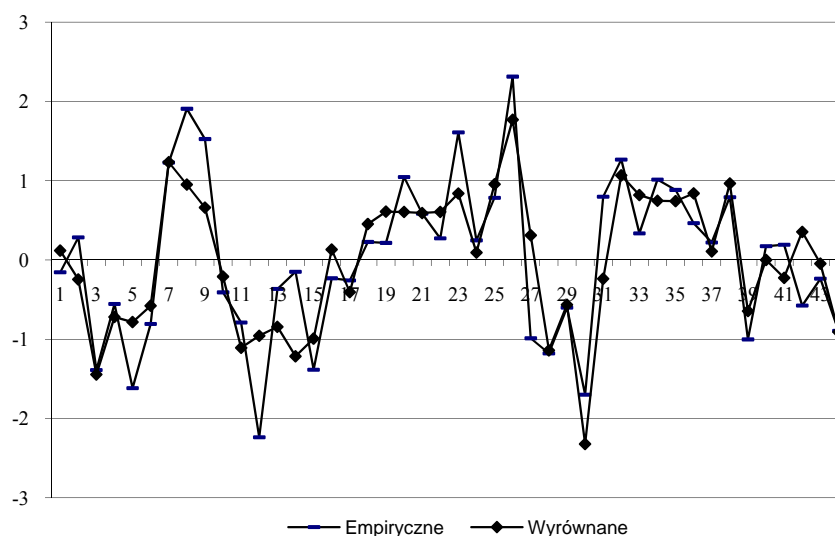
Kolejnym krokiem analizy było sprawdzenie czy opóźnienie niektórych zmiennych pierwotnych ma wpływ na ostateczną postać modelu.

Działanie takie spowodowało zmianę ostatecznej postaci modelu, a dokładniej zwiększenie liczby uwzględnionych w nim czynników do 4. Wynik estymacji modelu czteroczynnikowego prezentuje tabela nr 3.

Tabela 2. Dynamiczny model czynnikowy PKB Polski w latach 1997–2008

Zmienna zależna: PKB				
	Współczynnik	Błąd std.	Statystyka t	P
F1	-0,0362	0,0095	-3,7985	0,0005
F2	0,0683	0,0335	2,0404	0,0478
F3	0,3707	0,0405	9,161	0,0000
R-kwadrat	0,7142	Kryterium Akaika		1,7264
Dopasowany R-kwadrat	0,7003	Kryterium Schwarza		1,8481
Durbin-Watson	2,1454	Kryt. Hannana-Quinna		1,7715

Źródło: obliczenia własne.



Wykres 1. Rzeczywiste i dopasowane na podstawie DFM wartości PKB w Polsce w latach 1997–2008

Źródło: opracowanie własne.

Jak nietrudno zauważyć, zmiana pierwotnego zbioru danych spowodowała również polepszenie się statystycznych właściwości modelu. Współczynnik R –

kwadrat wzrósł o około 7 punktów procentowych. Spadły także empiryczne poziomy istotności parametrów znajdujących się przy poszczególnych czynnikach.

Ostatnim etapem analizy było wyznaczenie prognozy i jej błędów na podstawie dynamicznych modeli czynnikowych. Błędy te następnie poddane zostały porównaniu z analogicznymi otrzymanymi z modelu autoregresyjnego. Modele klasy AR stanowią bowiem najczęstszy, choć nie jedyny, punkt odniesienia w literaturze tematu.

Autorzy najczęściej przeciwstawiają prognozy z modeli z wieloma zmiennymi modelom z jedną zmienną (por. Marcellino, Stock, Watson, 2001). Do porównań przyjęty został, na podstawie kryterium BIC, stacjonarny model AR(1). Ponadto do wyznaczenia prognozy posłużył również klasyczny model symptomatyczny PKB w Polsce. W modelu symptomatycznym za zmienne objaśniające posłużyły wielkości reprezentujące produkcję sprzedaną przemysłu i przeciętne zatrudnienie.

Modele prognostyczne oszacowane zostały na próbie skróconej do końca roku 2007, co nie wywarło istotnego wpływu na ich jakość. Prognoza została wyznaczona na pierwszy kwartał roku 2008. Najbardziej trafną prognozą okazała się ta otrzymana na podstawie pierwszego DFM, czego potwierdzenie znajduje się w tabeli 4.

Tabela 3. Dynamiczny model czynnikowy PKB Polski w latach 1997–2008 – po modyfikacji zbioru danych

Zmienna zależna: PKB				
	Współczynnik	Błąd. Std.	Statystyka t	P
F21	0,1722	0,0273	6,316	0,0000
F22	0,1686	0,0329	5,132	0,0000
F23	-0,2548	0,0361	-7,0669	0,0000
F24	0,2241	0,04	5,609	0,0000
R-kwadrat	0,7909	Kryterium Akaike'a		1,4858
Dopasowany R - kwadrat	0,7748	Kryterium Schwarz		1,6497
Durbin-Watson	2,2511	Kryt. Hannana-Quinna		1,5463

Oznaczenie w postaci F21 oznacza, że jest to pierwszy czynnik ze zmodyfikowanego o odpowiednie opóźnienia zbioru danych.

Źródło: obliczenia własne.

Tabela 4. Błędy prognoz

	MAPE	RMSE	R-kwadrat
AR	90,8002	0,8116	0,1246
DFM	1,7898	0,016	0,7003
DFM2	12,869	0,115	0,7748
Model symptomatyczny	4,9267	0,044	0,8694

Źródło: obliczenia własne.

5. PODSUMOWANIE

Przeprowadzona analiza doprowadziła do zredukowania liczby pierwotnych zmiennych objaśniających z 41 do 3 czynników, co otrzymane zostało przez zastosowanie metody głównych składowych. Dzięki temu otrzymano dynamiczny model czynnikowy opisujący gospodarkę Polski w zakresie PKB w sposób zadowalający ze statystycznego punktu widzenia.

Modyfikacja zbioru danych pierwotnych w postaci opóźnienia poszczególnych zmiennych o jeden okres w tył ma wpływ na ostateczną postać modelu. W powyższej analizie spowodowała zwiększenie się liczby czynników w modelu i polepszenie poziomu jego dopasowania do danych empirycznych. Niestety, nie przyniosła oczekiwanej poprawy wyników prognozy.

Z dynamicznego modelu czynnikowego wzrostu PKB w Polsce w latach 1997–2008 opartego na niezmodyfikowanym zbiorze zmiennych pierwotnych otrzymaliśmy prognozę charakteryzującą się najwyższą trafnością spośród narzędzi rozważanych w powyższym artykule.

LITERATURA

- Bai J., Ng S. (2002), *Determining the Number of Factors in Approximate Factor Models*, "Econometrica", 70, 191–221.
- Geweke J. (1977), *The Dynamic Factor Analysis of Economic Time Series*, [w:] Aigne D. J., Goldberger A. S. (red.), *Latent Variables in Socio – Economic Models*, Amsterdam, North Holland.
- Greene W. H. (2003), *Econometric Analysis*, Pearson Education, New Jersey.
- Marcellino M., Stock J. H., Watson M. W. (2001), *Macroeconomic Forecasting in the Euro Area: Country Specific versus Area – Wide Information*, Working Paper, 201, Innocenzo Gasparini Institute for Economic Research.
- Sargent T., Sims C. (1977), *Business Cycle Modelling without Pretending to have too much a-priori Economic Theory*, [w:] Sims C. (red.), *New Methods in Business Cycle Research*, Minneapolis, Federal Reserve Bank of Minneapolis.
- Sims C. A. (1980), *Macroeconomics and Reality*, "Econometrica", 48, 1–48.
- Kotłowski J. (2008), *Forecasting Inflation with Dynamic Factor Model – the Case of Poland*, Working Papers, 2-08, SGH, Warszawa.
- Stock J., Watson M. W. (1998), *Diffusion Indexes*, Working Paper, 6702, National Bureau of Economic Research.

ESTIMATING AND FORECASTING GDP IN POLAND
WITH DYNAMIC FACTOR MODEL

A b s t r a c t. Presented paper concerns the dynamic factors models theory and application in the econometric GDP in Poland analyses. DFMs are used for construction of the economic indicators and in forecasting. They are applied in macroeconomics analyses, mainly in regard to the monetary policy and international business cycles. In the article we compare forecast accuracy of dynamic factor models with the forecast accuracy of two competitive models: univariate autoregressive model and symptomatic model. We have used 41 quarterly time series from the Polish economy. The results are encouraging. The dynamic factor model outperforms other models. The best fitted to empirical data was model with 3 factors.

K e y w o r d s: Dynamic factor models, principal components analysis, GDP.

