



Anna Marynowska

Uniwersytet Łódzki
Wydział Ekonomiczno-Socjologiczny
Katedra Ekonometrii
amarynowska6@gmail.com

PROGNOZOWANIE WYBRANEGO SUBINDEKSU CPI PRZY UŻYCIU DANYCH GOOGLE TRENDS¹

Streszczenie: Prognozy inflacji mają bezpośredni wpływ na prowadzenie polityki monetarnej państwa oraz odgrywają dużą rolę w uświadamianiu społeczeństwa o potrzebie wprowadzenia ewentualnych zmian w sposobie jej prowadzenia. W celu zapewnienia trafności konstruowanych prognoz stale poszukuje się zmiennych, które istotnie wpływają na inflację. Celem artykułu jest sprawdzenie, czy dane udostępniane przez serwis Google Trends mogą poprawić dokładność prognoz komponentów CPI związanych z sektorem transportowym. Dla każdego z subindeksów zbudowano modele prognozytyczne uwzględniające zmienne wpływające na poziom wybranych cen, modele wzbogacone o hasła z wyszukiwarki Google i modele zawierające w swojej specyfikacji wspólne czynniki opisujące zmienność 32 wybranych haseł Google.

Słowa kluczowe: prognozowanie, inflacja, CPI, analiza czynnikowa, Google Trends.

JEL Classification: E31, E37.

Wprowadzenie

Jednym z głównych zadań Narodowego Banku Polskiego jest utrzymanie inflacji na poziomie zbliżonym do ustalonego celu inflacyjnego. Wymaga to ciągłego badania czynników, które mają wpływ na kształtowanie się sytuacji gospodarczej państwa. Jednak w przypadku utrzymywania stabilnej inflacji, ważna jest nie tylko bieżąca analiza kondycji gospodarki, ale również przewidywanie jej przyszłego poziomu.

¹ Autorka pragnie złożyć serdeczne podziękowania dr. hab. Pawłowi Baranowskiemu za naukowe wsparcie przy pisaniu artykułu.

Analiza oraz prognozy inflacji są dla banku centralnego punktem wyjścia do prowadzenia polityki monetarnej państwa. Stanowią one bowiem najbardziej prawdopodobny scenariusz kształtowania się sytuacji gospodarczej, a co za tym idzie, zawczasu informują o potrzebie wprowadzenia ewentualnych zmian w sposobie prowadzenia polityki pieniężnej [Sławiński, 2011]. Należy więc formułować prognozy obciążone jak najmniejszymi błędami, m.in. dlatego ciągle poszukuje się nowych determinant, które mogłyby pomóc w prognozowaniu inflacji.

Celem artykułu jest sprawdzenie, czy za pomocą danych udostępnianych przez serwis Google Trends możliwe jest obniżenie błędów prognozy wybranych komponentów CPI. Co więcej, badanie stanowi próbę odpowiedzi na pytanie, czy trafność tak skonstruowanych prognoz może poprawić zastosowanie analizy czynnikowej.

Zgodnie z wiedzą autorki niniejsze badanie stanowi pierwszą próbę wykorzystania danych Google Trends do prognozowania indeksów cen w Polsce. Proponowane zastosowanie danych firmy Google wynika z faktu, iż w państwie, w którym aż 72,4% populacji stanowią internauci², naturalnym krokiem wydaje się szukanie informacji na temat czynników, które mogłyby rozszerzyć specyfikację prognoz inflacji, w Internecie.

1. Prognozowanie zjawiska inflacji

Słowo „inflacja” pochodzi z języka łacińskiego (*inflatio*) i w dosłownym tłumaczeniu oznacza nadęcie, rozdęcie. Najczęściej spotykana definicja określa inflację jako „proces wzrostu ogólnego poziomu cen” [np. Kwiatkowski, Milewski, 2005, s. 409].

Banki centralne wielu państw definiują optymalny poziom inflacji, tzw. cel inflacyjny. Polski cel inflacyjny sugeruje, że pożądane jest występowanie stopy inflacji na niskim poziomie (2,5% +/-1 p.p.). Występowanie inflacji pełzającej wpływa korzystnie na przebieg procesów gospodarczych danego państwa [Kwiatkowski, Milewski, 2005, s. 414].

Przy świadomości negatywnych skutków wywołanych przez wysoką oraz nieoczekiwaną przez społeczeństwo inflację bardzo ważną kwestią staje się pomiar tego zjawiska i przewidywanie jego dalszego poziomu.

Prognozując zjawisko inflacji, należy zadbać o zidentyfikowanie czynników, które mają największy wpływ na jego kształtowanie. Wymaga się, aby wybrane determinanty miały potwierdzenie w teorii ekonomii i zarazem były jak

² Statystyki podane przez Internet Live Stats [www 1], stan rzeczy na 2016 r.

najbardziej aktualne. Zdaniem autorki dane udostępniane przez serwis Google Trends spełniają powyższe kryteria oraz stanowią cenne źródło informacji na temat reakcji społeczeństwa na zachodzące zjawiska gospodarcze.

W niniejszym artykule analiza błędów *ex post* daje odpowiedź na pytanie, czy dane udostępniane przez serwis Google Trends mogą poprawić dokładność prognoz inflacji. W praktyce oznacza to sprawdzenie, czy uwzględnienie w modelu prognostycznym tych danych zmniejsza błędy prognoz.

2. Google Trends oraz przegląd dotychczasowych badań związanych z usługą

Google Trends jest usługą firmy Google, która przechowuje informacje na temat zapytań wprowadzanych przez internautów do wyszukiwarki internetowej tej firmy. Niniejsze dane są udostępniane użytkownikom w postaci statystyk podsumowujących częstotliwość wyszukiwania danego hasła. Wprowadzając konkretne hasło do wyszukiwarki w serwisie Google Trends, użytkownik otrzymuje informacje na temat częstości i pochodzenia interesującego go hasła, a także ma możliwość pobrania wygenerowanego szeregu czasowego.

Firma Google dostosowuje dane w celu zapewnienia porównywalności pomiędzy różnymi hasłami. Pobrany szereg czasowy ma formę indeksu jednopodstawowego, przyjmującego wartość 100 w okresie największego zainteresowania internautów danym hasłem.

Wszystkie powyższe cechy serwisu sprawiły, że Google Trends stał się ważnym narzędziem wykorzystywanym przy prowadzeniu działań marketingowych, m.in. do badania korelacji między częstotliwością wyszukiwania konkretnych haseł w wyszukiwarce Google a fluktuacjami na rynku finansowym [por. Preis, Reith, Stanley, 2010].

W dzisiejszych czasach Internet stał się powszechnie używanym źródłem wiedzy, a ludzie coraz częściej pierwsze kroki przy szukaniu odpowiedzi na nurtujące ich pytania kierują właśnie w stronę przeglądarki internetowej. Wyszukiwarka firmy Google jest według portalu ranking.pl najpopularniejszą wyszukiwarką według domeny (96,31%) oraz według silnika (97,53%)³. Zdaniem autorki to właśnie ona może stać się dla prognostyków źródłem cennych danych na temat trendów i tematów zainteresowania internautów. Te z kolei mogą stanowić cenne informacje na temat bieżących sytuacji politycznych, społecznych i ekonomicznych.

³ Dane według ranking.pl [www 2], stan rzeczy na luty 2017 r.

Temat wzbogacania prognoz o dane udostępniane przez firmę Google zyskuje coraz większą popularność wśród badaczy. Pierwszą publikacją, w której poszukiwano odpowiedzi na pytanie, czy hasła wprowadzane przez użytkowników Google pomagają w prognozowaniu aktywności ekonomicznej danego kraju, był artykuł Choi i Variana [2009]. Publikacja stanowiła bezpośrednią inspirację do przeprowadzenia niniejszego badania.

Autorzy badania postawili hipotezę, iż ilość wyszukiwań pewnego hasła wprowadzanego przez internautów do wyszukiwarki Google jest skorelowana z bieżącym stopniem aktywności ekonomicznej w danej gałęzi sektora [Choi, Varian, 2009]. Dane Google Trends odpowiadają liczbie wyszukiwań na dany miesiąc (stanowią bieżące statystyki) i pozwala to przewidywać, co się stanie w najbliższej przyszłości (*nowcasting*). Wyniki analizy błędów MAE sugerują, iż zastosowanie modeli autoregresyjnych uwzględniających istotne zmienne reprezentujące wyszukiwania Google poprawiają trafność prognoz (w porównaniu do modelu bez tych zmiennych) o 5-20% (w zależności od prognozowanej zmiennej).

Zbliżoną tematykę zawiera praca Guzmana [2011]. Autor opisał badanie oczekiwań inflacyjnych przy użyciu danych Google Trends, wykorzystując w tym celu nowy wskaźnik Google Inflation Search Index (GISI). Wskaźnik GISI mierzy zmiany w częstotliwości wpisywania do wyszukiwarki Google hasła „inflacja”. Autor badania założył, że wzrost liczby zapytań dotyczących tego hasła jest odpowiedzią na zwiększone obawy społeczeństwa przed rosnącymi cenami. Celem tej pracy [Guzman, 2011] stało się zbadanie, czy dane Google (dane o wyższej częstotliwości) mogą poprawić prognozy inflacji w porównaniu do klasycznego podejścia do prognozowania opartego na danych o niższej częstotliwości. Analiza wyników wskazuje, że prognoza oparta na wskaźniku GISI pozwala przewidzieć poziom inflacji z najmniejszym błędem predykcji (najmniejsze błędy prognozy otrzymano dla modelu opartego na wskaźniku GISI, błąd RMSE = 0,29 p.p.).

Dane Google Trends były dotychczas wykorzystywane również do prognozowania bezrobocia [por. Askitas, Zimmermann, 2009] czy konsumpcji indywidualnej [Schmidt, Vosen, 2011] – w tym przypadku porównywano prognozę wzbogaconą o wyszukiwania Google z powszechnie stosowanymi indeksami konsumpcji.

Wszystkie wymienione badania prowadzą do satysfakcjonującej konkluzji, iż wykorzystanie danych oferowanych przez serwis Google Trends wpływa korzystnie na trafność przeprowadzanych prognoz.

3. Wykorzystane dane

W badaniu zastosowano szeregi reprezentujące trzy komponenty CPI: transport (CPI_tran), transport pasażerski powietrzny (CPI_lot) oraz paliwa i smary do prywatnych środków transportu (CPI_pal). Wszystkie komponenty są związane z sektorem transportowym, który stanowi prężnie rozwijającą się gałąź polskiej gospodarki. Cechuje się on dużą zmiennością cen, a ich poziom jest determinowany przez wiele czynników (m.in. nastroje społeczne, ceny transportu zagranicznego, duże uzależnienia od dostawców paliw).

Badanie zostało przeprowadzone na podstawie danych miesięcznych obejmujących okres od stycznia 2003 r. do grudnia 2016 r. (168 obserwacji). Analizowany okres charakteryzował się stabilnym poziomem inflacji, która wynosiła średnio 1,97%. Autorka posługiwała się szeregami reprezentującymi łańcuchowe indeksy cen (w ujęciu miesiąc do miesiąca).

Poniższe zestawienie przedstawia użyte w badaniu zmienne objaśniające, wpływające na poziom cen, wraz z ich opisem (tabela 1).

Tabela 1. Zestawienie zmiennych objaśniających, wykorzystanych w badaniu

Zmienna	Opis
Odpowiednie opóźnienia zmiennych objaśnianych (autoregresja)	Wykorzystanie opóźnień zmiennej objaśnianej pozwala otrzymać prognozę autoregresyjną. Taki typ prognoz zakłada, że prognozowaną wartość badanej zmiennej można otrzymać, opierając się na wartościach niniejszej zmiennej z poprzednich okresów
Dynamika cen ropy	Cena ropy naftowej ma kluczowe znaczenie w kształtowaniu się cen na rynku transportowym, gdyż otrzymuje się z niej m.in. benzynę, naftę czy smary – substancje, bez których gałąź transportu nie mogłaby funkcjonować. Wzrost cen tego surowca, wraz ze wszystkimi jego pochodnymi, przekłada się bezpośrednio na zwiększone koszty produkcji (nie tylko w przemyśle transportowym, ale w całej gospodarce), co zmusza przedsiębiorstwa do podniesienia cen wytwarzanych dóbr
Dynamika kursów USD/PLN oraz EUR/PLN	Kurs walutowy jest bardzo ważną zmienną przy wyznaczaniu ścieżek polityki pieniężnej państwa. Zwiększenie kosztu związanego z wymianą złotego (słaba waluta krajowa) powoduje zmniejszenie importu, ponieważ importowane produkty są relatywnie drogie. Droższe półprodukty i surowce zagraniczne skutkują wyższymi kosztami produkcji krajowej, co z kolei przekłada się na wzrost jej cen (zwiększenie poziomu inflacji). W badaniu wykorzystano notowania kursów średnich euro oraz dolara amerykańskiego w polskim złotym (waluty najmocniej oddziałujące na cenę kursu PLN). Euro jest walutą większości państw członkowskich Unii Europejskiej, natomiast dolar amerykański to waluta, w której dokonuje się większości rozliczeń na rynku surowców petrochemicznych
Zmiana stawek akcyzy na benzynę oraz olej napędowy	Podatek akcyzowy (akcyza) jest podatkiem pośrednim, którego stawka powiększa wartość kupowanego dobra lub nabywanej usługi. Wpływy z podatku akcyzowego zwiększają w znaczny sposób budżet państwa [Podstawka, 2010, s. 247, 542]. Akcyza, obok opłaty paliwowej oraz podatku VAT, jest jednym z czynników wpływających na poziom cen benzyny w Polsce [por. Gradzewicz i in., 2013]. Im wyższa stawka akcyzy, tym wyższy poziom cen w przemyśle transportowym i zwiększone koszty produkcji w całej gospodarce (przyczyniające się do wzrostu poziomu inflacji)

Kolejną grupą zmiennych, które wykorzystano w badaniu, są hasła z serwisu Google Trends (tabela 2). Należy podkreślić, że ich dobór był całkowicie subiektywnym wyborem autorki (celem było uwzględnienie tylko zapytań ściśle kojarzonych z tematyką zmian cen w gałęzi transportu).

Tabela 2. Zestawienie wybranych do badania haseł z serwisu Google Trends

Nazwa zmiennej	Zapytanie	Nazwa zmiennej	Zapytanie
google_1	bezrobocie w Polsce	google_17	paliwo lotnicze
google_2	inflacja	google_18	promocja bilety lotnicze
google_3	inflacja w Polsce	google_19	Ryanair tanie loty
google_4	kurs walutowy	google_20	samoloty pasażerskie
google_5	podwyżka	google_21	tanie bilety
google_6	cena ropy	google_22	tanie bilety lotnicze
google_7	cena	google_23	weekend majowy
google_8	akcyza na paliwo	google_24	wyjazdy
google_9	cena baryłki ropy	google_25	ceny paliw
google_10	ceny paliwa	google_26	cena dolara
google_11	marża	google_27	cena euro
google_12	opłata paliwowa	google_28	cena benzyny
google_13	paliwo	google_29	bilety pkp
google_14	tanie paliwo	google_30	bilety autobusowe
google_15	bilety lotnicze	google_31	bilety mpk
google_16	długi weekend	google_32	cena biletów pkp

4. Przygotowanie danych do modelu prognostycznego

Kierując się logiką, dobór zmiennych do modelu musi być uzasadniony merytorycznie (powinien być poparty teorią ekonomii). Jednak przed przystąpieniem do budowy modelu, należy zadbać również o jak najlepsze własności statystyczne wykorzystywanych danych.

W tym celu autorka zbadała, czy analizowane wartości charakteryzują się sezonowością. Wszystkie szeregi reprezentujące badane komponenty CPI zostały jednorazowo oczyszczone z sezonowości przy pomocy procedury TRAMO/SEATS (metoda szeroko wykorzystywana w dotychczasowych badaniach, m.in. w artykule [Baranowski, Leszczyńska, Szafranski, 2010]). W badaniu uwzględniono sezonowość multiplikatywną w przypadku szeregu dotyczącego transportu pasażerskiego powietrznego, a także addytywną w przypadku transportu oraz paliw i smarów do prywatnych środków transportu.

Kolejny etap polegał na zbadaniu stacjonarności zmiennych przy użyciu testu ADF i sprowadzeniu zmiennych niestacjonarnych do postaci stacjonarnej (zastosowano pierwsze przyrosty).

5. Analiza czynnikowa

Badając zjawiska występujące w rzeczywistości (np. wskaźniki makroekonomiczne, zależności gospodarcze), należy pamiętać o oddaniu występujących między nimi skomplikowanych zależności. W celu ich określenia wykorzystuje się m.in. analizę wielowymiarową, która prowadzi do uproszczenia struktury danych poprzez redukcję dużego zbioru i w konsekwencji uzyskania klarownej interpretacji otrzymanych wyników.

W przypadku niniejszego opracowania analiza wielowymiarowa została użyta do redukcji wymiaru szeregów odzwierciedlających liczbę wyszukiwań konkretnego hasła Google. Można podejrzewać, że są to obiekty charakteryzujące się podobną strukturą, pozwalające na wyciągnięcie ogólnych wniosków opisujących zbiór wszystkich trzydziestu dwóch zapytań. Posłużono się analizą czynnikową (*factor analysis*), która opiera się na założeniu, iż każda zmienna stanowi kombinację liniową nieskorelowanych czynników wspólnych dla całego zbioru oraz czynnika specyficznego opisującego jedynie konkretną zmienną [Aczel, 2005].

Jedną z najpopularniejszych metod wykorzystywanych przy analizie czynnikowej jest analiza głównych składowych (*principal component analysis* – PCA), polegająca na dekompozycji zmienności charakteryzującej analizowany zbiór danych na zbiór komponentów (składowych). To podejście zostało zastosowane w celu wyłonienia czynników w zbiorze danych dotyczących haseł Google. Należy podkreślić, iż zmienne, z których zostają wyłaniane główne składowe, są zmiennymi stacjonarnymi (w analizie wykorzystano stacjonarne przyrosty zmiennych niestacjonarnych).

Temat prognozowania inflacji przy użyciu analizy czynnikowej przybliżył m.in. artykuł Baranowskiego, Leszczyńskiej i Szafrąńskiego [2010], opisujący prognozowanie inflacji z wykorzystaniem modeli czynnikowych w krótkim horyzoncie czasowym. Zdaniem autorów „dynamiczne modele czynnikowe umożliwiają uzyskanie syntetycznej informacji o kształtowaniu się zmienności dużego zbioru danych” [2010, s. 23]. Autorzy podkreślili również, że zastosowanie dynamicznych modeli czynnikowych pozwala na zmniejszenie informacji płynących z wielu determinant, bez istotnej utraty liczby stopni swobody oraz obawy przed otrzymaniem obciążonych prognoz (ze względu na nieuwzględnienie istotnych zmiennych w specyfikacji modelu) [Baranowski, Leszczyńska, Szafrąński, 2010].

Wcześniejszą publikacją, w której opisano problem wykorzystania analizy czynnikowej przy prognozowaniu, jest artykuł Stocka i Watsona [2002]. Auto-

rzy ci zwrócili uwagę na problem prognozowania zjawisk makroekonomicznych przy wielu determinantach w relatywnie krótkich szeregach czasowych. Rozwiązaniem tego problemu jest wyciągnięcie kilku czynników, które charakteryzują zbiór wielu determinant, tym samym upraszczając model prognostyczny. Wspólne czynniki wyciągnięto dla 149 miesięcznych zmiennych makroekonomicznych. Wyniki dla prognozy opartej na analizie czynnikowej porównano z modelem autoregresyjnym (który stanowił punkt odniesienia, benchmark), modelem VAR oraz modelem zmiennej wyprzedzającej (*leading indicator model*). Ocena miernika jakości *ex post* prognozy (MSE) pozwoliła stwierdzić, iż prognozy otrzymane na podstawie wyciągniętych wspólnych czynników generują znacznie niższe błędy w porównaniu z modelami wyjściowymi (wartość wskaźnika MSE niższa o 45% wobec wyjściowego modelem AR).

W przypadku opisywanego badania [Stock, Watson, 2002] przeprowadzana analiza głównych składowych na 32 zmiennych reprezentujących wyszukiwania haseł Google wskazuje, że korzystnym rozwiązaniem byłoby uwzględnienie w modelu prognostycznym trzech pierwszych składowych. Wyjaśniają one bowiem ponad połowę zmienności wszystkich haseł, tym samym stanowią wiarygodną informację na temat wariacji całego zbioru. Wyniki analizy głównych składowych przedstawia tabela 3.

Tabela 3. Komponenty wyłonione w analizie czynnikowej

Komponent	Procent całkowitej (pozostalej) zmienności	Procent skumulowany	Komponent	Procent całkowitej (pozostalej) zmienności	Procent skumulowany
Comp1	35,85%	35,85%	Comp17	0,79%	94,69%
Comp2	14,04%	49,89%	Comp18	0,74%	95,43%
Comp3	7,79%	57,68%	Comp19	0,61%	96,04%
Comp4	6,06%	63,74%	Comp20	0,58%	96,62%
Comp5	5,25%	68,99%	Comp21	0,53%	97,15%
Comp6	4,23%	73,21%	Comp22	0,48%	97,63%
Comp7	3,75%	76,97%	Comp23	0,45%	98,09%
Comp8	2,96%	79,93%	Comp24	0,39%	98,48%
Comp9	2,58%	82,50%	Comp25	0,34%	98,82%
Comp10	2,36%	84,86%	Comp26	0,29%	99,11%
Comp11	2,05%	86,92%	Comp27	0,24%	99,36%
Comp12	1,90%	88,81%	Comp28	0,21%	99,57%
Comp13	1,51%	90,32%	Comp29	0,17%	99,74%
Comp14	1,36%	91,68%	Comp30	0,15%	99,89%
Comp15	1,15%	92,83%	Comp31	0,06%	99,95%
Comp16	1,07%	93,90%	Comp32	0,05%	100,00%

6. Metoda badawcza

Parametry modelu zostały oszacowane metodą najmniejszych kwadratów, sprawdzona została normalność reszt wygenerowanych z modeli prognostycznych za pomocą modelu Jarque'a-Bery, wynik testu pozwolił stwierdzić, iż na poziomie istotności 5% nie ma podstaw do odrzucenia hipotezy zerowej o normalności składników losowych stworzonych modeli.

W badaniu dobrano zmienne kontrolne dla każdego komponentu CPI, kierując się kryterium obniżenia mierników dokładności *ex post* prognoz MAE oraz RMSE, które zostały oszacowane dla indeksów inflacji rok do roku. Kryterium to stanowi podstawę doboru zmiennych do modelu prognostycznego. Dodatkowo sprawdzono istotność wybranych zmiennych (za pomocą testu istotności parametru). Warto zaznaczyć, że modele są szacowane na podstawie rozszerzającej się próby, natomiast istotność modelu sprawdzana jest dla całej próby. Oznacza to, iż brak istotności zmiennych nie wpływa negatywnie na własności stworzonych modeli, a sprawdzenie istotności pełni wyłącznie funkcję informacyjną.

Rozważana metoda prognozowania poszczególnych komponentów CPI stanowi prognozę typu *direct*, poza próbę, na której estymowane są parametry (*out-of-sample*). Prognoza poza próbę jest dobrym odwzorowaniem rzeczywistego procesu prognostycznego, gdyż parametry modelu estymujemy na podstawie danych dostępnych na dzień sporządzenia prognozy. Dla przykładu, formułując prognozę na podstawie modelu autoregresyjnego w dniu 15 kwietnia 2017 r., znamy wartość wskaźnika CPI w marcu 2017 r. i wcześniejsze. Zgodnie z przyjętym podejściem zbudowany model prognostyczny powinien korzystać z próby kończącej się w marcu 2017 r. W kolejnym miesiącu poznamy dodatkowo wartość wskaźnika CPI w kwietniu 2017 r., w związku z tym, przygotowując prognozę, dokonamy reestymacji modelu na bazie próby wydłużonej o tę obserwację.

Wartości błędów wyznaczono na podstawie danych z próby, korzystając z rekursywnego schematu prognozowania (*recursive scheme*), który zakłada niestałą liczbę obserwacji okresu weryfikacji dla różnych horyzontów czasowych. Oznacza to, że liczba obserwacji różni się w przypadku horyzontu czasowego równego jeden oraz dwa miesiące. Okres weryfikacji dla horyzontu czasowego $h = 1$ wynosi 33 miesiące (od kwietnia 2014 r. do grudnia 2016 r.). Z kolei dla horyzontu czasowego $h = 2$ okres weryfikacji wynosi 32 miesiące (od maja 2014 r. do grudnia 2016 r.). Prognozy dla poszczególnych komponentów wyznaczono za pomocą następujących modeli (tabela 4).

Tabela 4. Modele prognostyczne

Prognoza	$h = 1$	$h = 2$
Prognoza wzbogacona o hasła Google (prognoza4)	$\hat{\pi}_{(i),t+1} = \hat{\alpha}_0 + \hat{\beta}x_t + \hat{\gamma}g_t$	$\hat{\pi}_{(i),t+2} = \hat{\alpha}_0 + \hat{\beta}x_t + \hat{\gamma}g_t$
Prognoza zawierająca wyłonione główne składowe (prognoza5)	$\hat{\pi}_{(i),t+1} = \hat{\alpha}_0 + \hat{\beta}x_t + \hat{\delta}c_t$	$\hat{\pi}_{(i),t+2} = \hat{\alpha}_0 + \hat{\beta}x_t + \hat{\delta}c_t$

Objaśnienia:

$\pi_{(i),t}$ – łańcuchowy indeks cen dla i -tej grupy dóbr i usług (podstawa: miesiąc $_{t-1} = 1$),

α_0 – wyraz wolny,

β, γ, δ –wektory parametrów strukturalnych (oszacowania parametrów oznaczono daszkiem nad symbolem parametru),

x – wektor wybranych zmiennych kontrolnych,

g – wektor zmiennych zawierający wybrane zapytania z serwisu Google,

c –wektor zmiennych zawierający wybrane główne składowe.

Początkowo realizowana jest prognoza badanych cen dla horyzontu jednego miesiąca z wykorzystaniem wcześniej odsezonowanych szeregów. W kolejnym kroku badania do prognozy zostaje dołączony odpowiedni wzorzec sezonowości dla każdego szeregu i na tej podstawie budowana jest prognoza inflacji komponentów CPI w ujęciu miesiąc do miesiąca. Następnie prognozy miesiąc do miesiąca przekształca się na prognozy w ujęciu rok do roku (metoda opisana m.in. w artykule [Baranowski i in., 2010, s. 23]). Procedura ta jest następnie powtarzana dla horyzontu czasowego wynoszącego dwa miesiące.

Zestawienie zawarte w tabeli 5 stanowi spis wraz z objaśnieniem prognoz wygenerowanych w niniejszym badaniu.

Tabela 5. Prognozy wygenerowane w badaniu

Prognoza	Opis prognozy
1	2
Prognoza oparta na modelu autoregresyjnym (prognoza1)	Pierwszym typem prognoz jest prognoza oparta na modelu autoregresyjnym (AR), która stanowi punkt odniesienia (benchmark) dla kolejnych prognoz. Otrzymane mierniki dokładności <i>ex post</i> prognoz są porównywane z miernikami dokładności otrzymanymi na podstawie prognozy AR, a ich mniejsze wartości oznaczają poprawę trafności zbudowanej prognozy. Prognoza autoregresyjna przyjmuje postać (odpowiednio dla horyzontu jednego miesiąca oraz dwóch miesięcy): $\hat{\pi}_{(i),t} = \alpha_0 + \alpha_1\pi_{(i),t-1} + \varepsilon_t$ $\hat{\pi}_{(i),t} = \alpha_0 + \alpha_1\pi_{(i),t-2} + \varepsilon_t$
Prognoza kontrolna (prognoza2)	Kolejnym krokiem badania jest dobranie zmiennych kontrolnych dla każdego komponentu CPI, kierując się kryterium redukcji wartości mierników dokładności <i>ex post</i> prognoz MAE i RMSE, które zostały wyznaczone dla indeksów inflacji rok do roku. Kryterium to stanowi podstawę doboru zmiennych do modelu prognostycznego. Doboru zmiennych kontrolnych dokonuje się osobno dla prognozy o horyzoncie jednego miesiąca i dla prognozy o horyzoncie dwóch miesięcy, stosując odpowiednio zmienne kontrolne opóźnione o jeden okres i dwa okresy. Zgodnie z intuicją prognoza kontrolna powinna poprawiać trafność modelu prognostycznego, bowiem dodaje do specyfikacji cenne informacje o zewnętrznych czynnikach oddziałujących na prognozowany szereg

cd. tabeli 5

1	2
Błądzenie losowe (<i>random walk</i>) inaczej nazywane prognozą naiwną lub prostą (prognoza3)	Proces błądzenia losowego (<i>random walk</i>) jest przykładem procesu generowania danych (<i>data generating process</i> – DGP), który zakłada, że wartość bieżąca stanowi wartość skumulowanych impulsów otrzymanych w poprzednich okresach. Wszystkie zakłócenia (reprezentowane przez składnik losowy) oddziałują w równym stopniu na bieżącą wartość zmiennej endogenicznej
Prognoza wzbogacona o hasła Google (prognoza4)	Innowacyjnym podejściem niniejszego badania jest dołączanie do specyfikacji modeli prognostycznych zmiennych prezentujących wyszukiwania hasła w serwisie Google. Na tym etapie badania następuje sprawdzenie, czy (a jeśli tak to które z nich) obniżają błędy prognoz $h = 1$ oraz $h = 2$ względem modelu benchmarkowego – odpowiednio metodą <i>direct forecasting</i> AR(1) i AR(2) – oraz w porównaniu do prognozy kontrolnej, która w specyfikacji uwzględnia jedynie zmienne kontrolne dające najniższe błędy prognoz. Wyszukiwania hasła Google są opóźniane o jeden i o dwa okresy odpowiednio dla horyzontu jednego miesiąca i dwóch miesięcy. W celu obniżenia błędów prognoz zastosowano zarówno metodę dołączania po jednym hasle Google do zmiennych kontrolnych, jak i poszukiwano kombinacji hasła, która daje najniższe błędy prognozy
Prognoza zawierająca wyłonione główne składowe (prognoza5)	Biorąc pod uwagę zblizoną strukturę oraz właściwości zmiennych dotyczących wyszukiwań hasła Google, poddano je analizie czynnikowej. Wyłonione w tym procesie główne składowe poszerzają specyfikację kolejnej prognozy. Aby prognoza uwzględniająca wspólne czynniki mogła zostać uznana za zasadną, powinna ona obniżyć błędy w porównaniu do modelu benchmarkowego, a także generować niższe błędy w porównaniu do modelu uwzględniającego wybrane hasła Google

Celem badania jest wykazanie, że zarówno prognozy uwzględniające dane z serwisu Google Trends, jak i prognozy zawierające główne składowe cechują się niższymi błędami niż bazowy model autoregresyjny.

Kolejną istotną kwestią stało się zagwarantowanie stabilności prezentowanego modelu – żeby był on odporny na zmiany okresu weryfikacji prognozy. W tym celu zaprezentowano prognozy w dwóch horyzontach czasowych ($h = 1$, $h = 2$) dla trzech okresów weryfikacji: 33 miesiące (kwiecień 2014 r. – grudzień 2016 r.), 24 miesiące (styczeń 2015 r. – grudzień 2016 r.) oraz 12 miesięcy (styczeń 2016 r. – grudzień 2016 r.). Ostatnim etapem badania jest graficzne zaprezentowanie najlepszej wygenerowanej prognozy (prognozy o najmniejszych błędach MAE oraz RMSE) na tle empirycznej inflacji komponentu CPI.

7. Wyniki

7.1. Transport

Najmniejsze błędy prognoz MAE oraz RMSE inflacji komponentu transport (w ujęciu rok do roku) przyjmuje model prognostyczny uwzględniający następujące zmienne (tabela 6).

Tabela 6. Specyfikacja modelu prognostycznego (CPI_tran_rr)

$h = 1$	$h = 2$
Jako zmienne kontrolne: • opóźniona o jeden okres dynamika cen ropy	Jako zmienne kontrolne: • opóźniona o dwa okresy dynamika cen ropy
Wybrane hasła z serwisu Google Trends: • opóźnione o jeden okres wyszukiwanie hasła „podwyżka” (google5)	Wybrane hasła z serwisu Google Trends: • opóźnione o dwa okresy wyszukiwanie hasła „inflacja w Polsce” (google3), • opóźnione o dwa okresy wyszukiwanie hasła „cena baryłki ropy” (google9)

W kolejnym kroku obliczane zostają wartości mierników dokładności *ex post* prognoz MAE (tabela 7) oraz RMSE (tabela 8) dla stworzonego modelu.

Tabela 7. Porównanie wartości miernika MAE poszczególnych prognoz (CPI_tran_rr)

MAE	04.2014 – 12.2016 (33 obs.)		01.2015 – 12.2016 (24 obs.)		01.2016 – 12.2016 (12 obs.)	
	$h = 1$	$h = 2$	$h = 1$	$h = 2$	$h = 1$	$h = 2$
	prognoza1 (benchmark)	0,965	1,821	1,047	1,805	1,171
prognoza2	0,855	1,593	0,924	1,554	1,167	1,876
prognoza3	1,353	2,250	1,463	2,257	1,546	2,377
prognoza4	0,839	1,516	0,994	1,493	1,181	1,879
prognoza5	0,928	1,676	0,969	1,609	1,254	2,002
prognoza4/benchmark	86,90%	83,26%	87,58%	82,74%	100,91%	93,22%
prognoza5/benchmark	96,12%	92,04%	92,56%	89,19%	107,13%	99,31%

Tabela 8. Porównanie wartości miernika RMSE poszczególnych prognoz (CPI_tran_rr)

RMSE	04.2014 – 12.2016 (33 obs.)		01.2015 – 12.2016 (24 obs.)		01.2016 – 12.2016 (12 obs.)	
	$h = 1$	$h = 2$	$h = 1$	$h = 2$	$h = 1$	$h = 2$
	prognoza1 (benchmark)	1,299	2,293	1,426	2,270	1,623
prognoza2	1,259	1,926	1,387	1,849	1,752	2,185
prognoza3	1,763	2,805	1,895	2,798	2,122	3,118
prognoza4	1,240	1,890	1,373	1,836	1,735	2,213
prognoza5	1,379	2,249	1,482	2,013	1,827	2,421
prognoza4/benchmark	95,46%	82,42%	96,28%	80,88%	106,90%	85,80%
prognoza5/benchmark	106,16%	98,08%	103,93%	88,68%	112,57%	101,34%

Analizując wyznaczone wartości miernika MAE (tabela 7), stwierdzono, że uzasadnione jest dodanie do specyfikacji modelu zmiennych kontrolnych (ponieważ zmniejsza to błędy prognozy w porównaniu do sytuacji, gdy zmienne kontrolne nie są uwzględnione). Co więcej, dodanie zarówno zmiennych związanych z hasłami Google, jak i głównych składowych poprawia trafność prognoz. Pomijając wyniki otrzymane w przypadku prognozy dla $h = 1$ przy okresie weryfikacji 12 miesięcy (kiedy prognozy uwzględniające główne składowe oraz hasła Google generują większe błędy niż benchmark), wszystkie błędy są niższe niż te wygenerowane przez model autoregresyjny. Dalsza analiza wartości mier-

nika MAE pozwala stwierdzić, że najmniejsze błędy generuje prognoza dla okresu weryfikacji 24 miesięcy (horyzont czasowy $h = 2$).

Fakt, iż w przypadku komponentu transport prognozy z hasłami Google generują niższe błędy niż prognozy uwzględniające analizę czynnikową potwierdza analiza wartości miernika RMSE (tabela 8). W takiej sytuacji autorka sugeruje skorzystanie z prognozy uwzględniającej poszczególne hasła. Stworzone modele są stabilne – odporne na zmiany okresu weryfikacji prognozy.

7.2. Ceny paliw i smarów do prywatnych środków transportu

Najmniejsze błędy prognozy inflacji komponentu paliwa i smary do prywatnych środków transportu (w ujęciu rok do roku) można otrzymać, uwzględniając w specyfikacji modelu prognostycznego następujące dane (tabela 9).

Tabela 9. Specyfikacja modelu prognostycznego (CPI_pal_rr)

$h = 1$	$h = 2$
Jako zmienne kontrolne: <ul style="list-style-type: none"> • opóźniona o jeden okres zmienna objaśniana, • opóźniona o jeden okres dynamika cen ropy 	Jako zmienne kontrolne: <ul style="list-style-type: none"> • opóźniona o dwa okresy zmienna objaśniana, • opóźniona o dwa okresy dynamika cen ropy
Wybrane hasła z serwisu Google Trends: <ul style="list-style-type: none"> • pierwsza różnica wyszukiwania hasła „cena” (google7), • opóźnione o jeden okres wyszukiwanie hasła „cena ropy” (google6), • opóźnione o jeden okres wyszukiwanie hasła „tanie paliwo” (google14) 	Wybrane hasła z serwisu Google Trends <ul style="list-style-type: none"> • opóźnione o jeden i dwa okresy wyszukiwanie hasła „cena ropy” (google6), • opóźnione o dwa okresy wyszukiwanie hasła „tanie paliwo” (google14)

W kolejnym kroku obliczane są błędy prognoz MAE (tabela 10) oraz RMSE (tabela 11) wygenerowanych prognoz.

Tabela 10. Porównanie wartości miernika MAE poszczególnych prognoz (CPI_pal_rr)

MAE	04.2014 – 12.2016		01.2015 – 12.2016		01.2016 – 12.2016	
	(33 obs.)		(24 obs.)		(12 obs.)	
	$h = 1$	$h = 2$	$h = 1$	$h = 2$	$h = 1$	$h = 2$
prognoza1 (benchmark)	1,604	3,118	1,754	3,216	2,037	3,621
prognoza2	1,354	2,432	1,566	2,512	2,141	3,223
prognoza3	2,281	3,726	2,473	3,766	2,753	4,087
prognoza4	1,358	2,319	1,557	2,397	2,109	3,135
prognoza5	1,497	2,677	1,676	2,679	2,285	3,448
prognoza4/benchmark	84,67%	74,39%	88,78%	74,54%	103,54%	86,57%
prognoza5/benchmark	93,30%	85,87%	95,58%	83,30%	112,17%	95,23%

Tabela 11. Porównanie wartości miernika RMSE poszczególnych prognoz (CPI_pal_rr)

RMSE	04.2014 – 12.2016		01.2015 – 12.2016		01.2016 – 12.2016	
	(33 obs.)		(24 obs.)		(12 obs.)	
	$h = 1$	$h = 2$	$h = 1$	$h = 2$	$h = 1$	$h = 2$
prognoza1 (benchmark)	2,158	3,808	2,366	3,863	2,715	4,137
prognoza2	1,980	2,990	2,234	3,017	2,918	3,636
prognoza3	2,988	4,635	3,178	4,656	3,703	5,347
prognoza4	1,991	2,941	2,237	3,074	2,910	3,683
prognoza5	2,142	3,510	2,353	3,261	3,017	3,977
prognoza4/benchmark	92,26%	77,23%	94,55%	79,58%	107,18%	89,03%
prognoza5/benchmark	99,26%	92,17%	99,45%	84,42%	111,12%	96,13%

Dzięki analizie wartości mierników MAE oraz RMSE (tabele 10 i 11) dla prognozy inflacji badanego komponentu CPI stwierdzono, że dodanie do specyfikacji prognozy zmiennych kontrolnych zmniejsza błędy prognozy w porównaniu do sytuacji, gdy zmienne kontrolne nie są uwzględnione. Wyjątek stanowi tylko prognozowanie komponentu przy okresie weryfikacji 12 obserwacji z horyzontem wynoszącym jeden miesiąc. Analiza wyników z tabel 10 i 11 wskazuje, że zaproponowany model uwzględniający wyszukiwania Google, obarczony jest niższymi błędami niż model benchmarkowy – AR(1) dla $h = 1$ i AR(2) dla $h = 2$. Zaproponowane modele generują większe błędy niż model benchmarkowy jedynie dla okresu weryfikacji 12 miesięcy przy horyzoncie czasowym jednego miesiąca. Najbardziej zadawalające wyniki (najmniejsze błędy) otrzymano dla okresu weryfikacji 33 miesięcy przy horyzoncie czasowym dwóch miesięcy. Również model uwzględniający wspólne czynniki cechuje się niższymi błędami niż model benchmarkowy, jednak błędy tego modelu przewyższają błędy modelu prognoza4. Oznacza to, że model uwzględniający konkretne hasła ma lepsze własności prognostyczne. Stworzone modele są stabilne – odporne na zmiany okresu weryfikacji prognozy.

7.3. Transport pasażerski powietrzny

Prognoza szeregu reprezentującego powyżej wymieniony komponent CPI charakteryzuje się dużymi błędami, nie daje satysfakcjonujących wyników. Związane jest to z dużą zmiennością cen biletów lotniczych, wysoką częstotliwością tych zmian (nawet w ujęciu godzinowym) oraz wieloma czynnikami, które determinują kształtowanie się ceny biletów. Najmniejsze błędy prognozy inflacji analizowanego komponentu (w ujęciu rok do roku) otrzymano, uwzględniając w specyfikacji następujące zmienne (tabela 12).

Tabela 12. Specyfikacja modelu prognostycznego (CPI_lot_rr)

$h = 1$	$h = 2$
Jako zmienne kontrolne: <ul style="list-style-type: none"> opóźniona o jeden okres dynamika kursu EUR/PLN 	Jako zmienne kontrolne: <ul style="list-style-type: none"> opóźniona o dwa okresy dynamika kursu EUR/PLN
Wybrane hasła z serwisu Google Trends: <ul style="list-style-type: none"> pierwsza różnica wyszukiwania hasła „cena” (google7), opóźnione o jeden okres wyszukiwanie hasła „długi weekend” (google16), opóźnione o jeden okres wyszukiwanie hasła „Ryanair tanie loty” (google19) 	Wybrane hasła z serwisu Google Trends: <ul style="list-style-type: none"> opóźnione o dwa okresy wyszukiwanie hasła „promocja bilety lotnicze” (google18) opóźnione o dwa okresy wyszukiwanie hasła „Ryanair tanie loty” (google19)

Stworzone modele prognostyczne cechują się następującymi błędami prognoz (tabele 13 i 14).

Tabela 13. Porównanie wartości miernika MAE poszczególnych prognoz (CPI_lot_rr)

MAE	04.2014 – 12.2016 (33 obs.)		01.2015 – 12.2016 (24 obs.)		01.2016 – 12.2016 (12 obs.)	
	$h = 1$	$h = 2$	$h = 1$	$h = 2$	$h = 1$	$h = 2$
	prognoza1 (benchmark)	4,124	6,369	4,377	7,459	5,118
prognoza2	4,089	6,206	4,363	7,155	5,070	8,240
prognoza3	14,611	11,342	12,914	11,837	12,436	13,223
prognoza4	4,088	5,807	4,190	6,854	4,855	7,726
prognoza5	4,151	6,397	4,478	7,440	5,075	8,234
prognoza4/benchmark	99,14%	91,17%	95,72%	91,90%	94,87%	84,27%
prognoza5/benchmark	100,66%	100,43%	102,31%	99,75%	99,16%	89,81%

Tabela 14. Porównanie wartości miernika RMSE poszczególnych prognoz (CPI_lot_rr)

RMSE	04.2014 – 12.2016 (33 obs.)		01.2015 – 12.2016 (24 obs.)		01.2016 – 12.2016 (12 obs.)	
	$h = 1$	$h = 2$	$h = 1$	$h = 2$	$h = 1$	$h = 2$
	prognoza1 (benchmark)	5,295	8,444	5,610	9,509	6,611
prognoza2	5,262	7,970	5,571	8,889	6,536	10,417
prognoza3	18,698	13,510	15,166	14,039	13,360	16,229
prognoza4	5,312	7,735	5,510	8,687	6,278	9,866
prognoza5	5,372	8,162	5,718	9,184	6,581	10,481
prognoza4/benchmark	100,32%	91,60%	98,22%	91,36%	94,96%	84,09%
prognoza5/benchmark	101,45%	96,66%	101,93%	96,58%	99,55%	89,34%

Pomimo problemów w prognozowaniu komponentu związanego z transportem pasażerskim powietrznym (w ujęciu rok do roku) widoczne jest, że dołączenie do specyfikacji danych z serwisu Google Trends poprawia prognozy (tabele 13 i 14). Prognoza oparta na analizie czynnikowej gwarantuje mniejsze wartości błędów w niektórych przypadkach (okres weryfikacji 33 miesiące ($h = 1$, $h = 2$) oraz 24 miesiące dla horyzontu czasowego jednego miesiąca. Autorka sugeruje więc skorzystanie z prognozy uwzględniającej poszczególne hasła w celu przewidywania przyszłego poziomu niniejszego komponentu CPI. Stworzona pro-

gnoza jest stabilna, a najbardziej zadowalające wyniki gwarantuje prognoza z okresem weryfikacji 33 miesiące z horyzontem czasowym dwóch miesięcy (analiza wartości miernika MAE, tabela 13) oraz 24 miesiące z horyzontem czasowym dwóch miesięcy (błędy RMSE, tabela 14).

Podsumowanie

Artykuł miał na celu sprawdzenie, czy dane udostępniane przez serwis Google Trends mogą przyczynić się do obniżenia błędów prognoz trzech wybranych komponentów CPI: transportu, transportu pasażerskiego powietrznego oraz paliw i smarów do prywatnych środków transportu. W tym celu przeprowadzono prognozy poza próbę (*out-of-sample*) o horyzoncie czasowym jednego oraz dwóch miesięcy dla każdego z badanych szeregów. Posługiwano się danymi miesięcznymi, obejmującymi okres stabilnej inflacji, od stycznia 2003 r. do grudnia 2016 r. (168 obserwacji).

Stworzone modele autoregresyjne, które w swojej specyfikacji uwzględniają dodatkowo wybrane zapytania Google, i takie, które oparte są o główne składowe wyciągnięte dzięki analizie czynnikowej. Analiza czynnikowa została zastosowana w badaniu w celu zredukowania dużej ilości zmiennych (32 zmienne dotyczące zapytań Google) oraz zidentyfikowania czynników, które w najlepszy sposób oddadzą własności tego zbioru danych.

Dobierając zmienne do specyfikacji modeli prognostycznych, kierowano się kryterium minimalizacji błędów MAE oraz RMSE (w tym stworzenia modeli, których błędy są niższe od modelu benchmarkowego – odpowiednio dla horyzontu czasowego jednego miesiąca AR(1) oraz AR(2) dla horyzontu czasowego dwóch miesięcy). Małe błędy prognoz punktowych świadczą bowiem o ich dużej trafności, a tym samym mówią o dobrym odwzorowaniu rzeczywistości za pomocą modelu.

Prognozowane były jedne z najbardziej zmiennych komponentów CPI, dla komponentów: paliwo, transport oraz bilety lotnicze, odchylenia standardowe wynoszą odpowiednio 9,8 p.p., 5,58 p.p. i 18.34 p.p., co przełożyło się na stosunkowo wysokie wartości mierników MAE oraz RMSE. Ograniczenie to nie wpływa jednak na główny wniosek z badania, tj. poprawę trafności dzięki zastosowaniu danych Google Trends. Wyniki przeprowadzonego badania empirycznego potwierdzają, iż dodanie do specyfikacji modelu prognostycznego danych udostępnianych przez serwis Google Trends zmniejsza błędy prognoz inflacji. Warto podkreślić, że wykorzystane dane są pomocne nawet przy prognozowaniu tak skomplikowanych prognostycznie szeregów. Autorka stwierdza, że błędy

wygenerowane przez prognozy z hasłami Google są niższe niż błędy otrzymane na podstawie prognoz wykorzystujących analizę czynnikową. Autorka wzięła pod uwagę trzy pierwsze główne składowe, wyjaśniają one bowiem ponad połowę zmienności całego zbioru danych Google. Badanie można rozszerzyć, uwzględniając w modelu inną liczbę głównych składowych. Dalszą poprawę trafności prognoz mogłoby zapewnić zastosowanie innej klasy modeli, np. modeli nieliniowych czy modeli uwzględniających powiązania między zmiennymi makroekonomicznymi (np. VAR).

Niewątpliwą zaletą danych udostępnianych przez serwis Google Trends jest fakt, iż zapewniają one dostęp do bieżących statystyk. Co więcej, dane zapewniają badaczom dużą elastyczność ich wykorzystania. Zdaniem autorki rosnąca popularność niniejszych danych wynika m.in. z faktu, iż stanowią one narzędzie uniwersalne. W przypadku gdy dobrane zostaną inne hasła, przedstawiona metoda może pozwolić na wyznaczenie prognoz pozostałych zmiennych makroekonomicznych. Co więcej, zaproponowany model można rozszerzać w celu poprawy trafności prognoz samej inflacji. Ze względu na widoczną poprawę prognoz wzbogaconych o zapytania z Google Trends można rozważyć rozszerzenie tego badania na pozostałe komponenty CPI.

Literatura

- Acel A.D. (2005), *Statystyka w zarządzaniu*, Wydawnictwo Naukowe PWN, Warszawa.
- Askitas N., Zimmermann K.F. (2009), *Google Econometrics and Unemployment Forecasting*, "Applied Economics Quarterly", Vol. 55, No. 2, s. 107-120.
- Baranowski P., Leszczyńska A., Szafrąński G. (2010), *Krótkookresowe prognozowanie inflacji z użyciem modeli czynnikowych*, „Bank i Kredyt”, Vol. 41, nr 4.
- Baranowski P., Mazurek M., Nowakowski M., Raczko M. (2010), *Czy dezagregacja indeksu cen poprawia prognozy polskiej inflacji?* „Przegląd Statystyczny”, R. LVII, s. 17-33.
- Choi H., Varian H. (2012), *Predicting the Present with Google Trends*, "The Economic Record", Vol. 88, Iss. 1, s. 2-9.
- Gradzewicz M., Hagemeyer J., Hałka A., Baranowski P., Jankiewicz Z., Kołodziejczyk D., Leszczyńska A., Macias P., Niechciał M., Popowski P., Puchalska K. (2013), *Strukturalne uwarunkowania inflacji*, „Materiały i Studia”, nr 297.
- Guzman G. (2011), *Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations*, "The Journal of Economic and Social Measurement", Vol. 36, No. 3, s. 119-167.
- Kwiatkowski E., Milewski R. (2005), *Podstawy ekonomii*, Wydawnictwo Naukowe PWN, Warszawa.

Podstawka M. (2010), *Finanse*, Wydawnictwo Naukowe PWN, Warszawa.

Preis T., Reith D., Stanley E. (2010), *Complex Dynamics of Our Economic Life on Different Scales: Insights from Search Engine Query Data*, "Philosophical Transactions of the Royal Society A", Vol. 368, No. 1933, s. 5707-5719.

Schmidt T., Vosen S. (2011), *Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends*, "Journal of Forecasting", Vol. 30, No. 6, s. 565-578.

Sławiński A. (2011), *Polityka pieniężna*, Wydawnictwo C.H. Beck, Warszawa.

Stock J.H., Watson M.W. (2002), *Forecasting Using Principal Components from a Large Number of Predictors*, "Journal of the American Statistical Association", Vol. 97, Iss. 460, s. 1167-1179.

[www 1] <http://www.internetlivestats.com/internet-users/> (dostęp: 25.03.2017).

[www 2] <http://www.ranking.pl> (dostęp: 11.02.2017).

FORECASTING TRANSPORT INFLATION USING GOOGLE TRENDS

Summary: Inflation forecasts determine the monetary policy and can be treated as a way of raising society's awareness of the fact that it needs certain adjustments. Experts continuously seek for the adequate variables that affect the level of inflation. In this paper the author examines whether Google Trends improve forecast of three chosen CPI components related to transport. The Author created a prognostic model for each of the components. Created models include: independent variables such as oil price and rate of excise duty, chosen variables together with statistics provided by Google Trends or principal components (explaining 32 google variables' volatility) accordingly.

Keywords: inflation, forecasting, CPI, principal components, Google Trends.