Bank i Kredyt 41 (5), 2010, 101–122

BANK I KREDYT

www.bankikredyt.nbp.pl
www.bankandcredit.nbp.pl

O eksperckich ocenach niepewności w ankietach makroekonomicznych

Halina Kowalczyk*

Nadesłany: 19 kwietnia 2010 r. Zaakceptowany: 25 sierpnia 2010 r.

Streszczenie

Coraz powszechniejszy staje się pogląd, że prognozom makroekonomicznym, zarówno modelowym, jak i eksperckim, powinny towarzyszyć oceny niepewności. Zasadnicza teza niniejszej pracy jest taka, że w przypadku prognoz eksperckich mamy do czynienia z innym typem niepewności niż w przypadku prognoz modelowych. Wymaga to odmiennej interpretacji przedstawianych przez ekspertów rozkładów prawdopodobieństwa. Właściwszy wydaje się opis w kategoriach prawdopodobieństwa subiektywnego, rozumianego jako stopień przekonania eksperta co do przedstawianych hipotez. Dla uzasadnienia powyższej tezy przeanalizowano rozróżnienia pojęciowe dokonywane przy klasyfikacji niepewności i prawdopodobieństwa. Kolejna teza to konieczność uwzględnienia realnych możliwości ekspertów w zakresie kwantyfikacji niepewności w fazie projektowania ankiet makroekonomicznych oraz potrzeba zróżnicowania sposobu opracowywania ankiet w zależności od obszaru wykorzystywania. Problemy zostały przedstawione na przykładzie ankiet makroekonomicznych adresowanych do profesjonalnych prognostów i zastosowań typowych dla banków centralnych (badanie oczekiwań, agregacja opinii, prognozowanie).

Słowa kluczowe: ankiety makroekonomiczne, badanie oczekiwań, niepewność prognoz eksperckich, prawdopodobieństwo subiektywne

JEL: C46, E27

* Narodowy Bank Polski, Instytut Ekonomiczny; e-mail: halina.kowalczyk@nbp.pl.

1. Wstęp

Oceny niepewności są coraz częściej uznawane za niezbędny element prognoz. Celem poniższego opracowania jest znalezienie odpowiedzi na pytania: czy istnieje potrzeba wprowadzania pytań probabilistycznych do ankiet makroekonomicznych badających oczekiwania ekspertów ekonomicznych, czy i w jakim zakresie eksperckie prognozy probabilistyczne mogą spełniać wyznaczone im zadanie, a także czy obecny kształt i stosowane metody opracowywania ankiet makroekonomicznych odpowiadają potrzebom.

Na początku skupimy uwagę na podstawowych dla niniejszej pracy pojęciach niepewności i prawdopodobieństwa. Omówione zostaną podstawowe rozróżnienia i związki pojęciowe.

Rozdział drugi poświęcono problemowi właściwej interpretacji pytań o prawdopodobieństwo w ankietach, dla których wzorcem stała się wprowadzona w 1968 r. przez American Statistical Association i National Bureau of Economic Research w Stanach Zjednoczonych ankieta ASA-NBER, znana od 1990 r. jako Philadelphia Fed Survey of Professional Forecasters (SPF). Zaproponowano interpretację niepewności w terminach subiektywnego (osobowego) prawdopodobieństwa, co powoduje konieczność przyjęcia innych kryteriów oceny poprawności eksperckich rozkładów. Omówiono problemy dotyczące kwantyfikacji niepewności w przypadku ankiet z predefiniowanymi przedziałami. Ponadto przedstawiono propozycję alternatywnych pytań probabilistycznych.

Rozdział trzeci nawiązuje do opublikowanych w ostatnim okresie prac, dotyczących problemu braku spójności prognoz punktowych z odpowiadającymi im prognozami w postaci rozkładów prawdopodobieństwa. Przedstawiono w nim argumenty przeciwko używaniu prognoz punktowych.

W rozdziale czwartym scharakteryzowano potencjalne obszary zastosowań „probabilistycznych" ankiet makroekonomicznych i ograniczenia wynikające ze specyfiki tych obszarów. Zaprezentowano też metodę obiektywizacji subiektywnego prawdopodobieństwa, która z powodzeniem jest wykorzystywana w takich dziedzinach, jak klimatologia, medycyna, badania nad wykorzystaniem energii jądrowej czy badania kosmiczne.

2. Niepewność

Niepewność jest terminem, którego używamy na co dzień. Jak to często bywa z pojęciami lub zjawiskami powszednimi – to, co większość z nas uważa za oczywiste, nie jest już takie proste np. dla filozofów nauki czy praktyków zajmujących się tym zagadnieniem. W literaturze przedmiotu podejmowane są próby klasyfikacji i uwzględnienia różnych wymiarów niepewności (np. Helton i in. 2008). Wyróżnia się zwykle dwa typy niepewności:

- Niepewność typu I (ang. *type I, type A, aleatory, variability*) jest obserwowaną w naturze lub systemie przypadkowością i zmiennością;
- Niepewność typu II to:
  - niepewność wiedzy (ang. *type II, type B, epistemic, Knightian*), odzwierciedlająca wątpliwości wynikające ze świadomości, że nasza wiedza jest niepełna, a możliwości pomiarów i modelowania zjawisk są ograniczone,
  - niepewność lingwistyczna, wiążąca się z niejednoznacznością, względnością lub niemierzalnością pojęć używanych do opisu zjawisk.

Niepewność typu I jest właściwością systemu (jest niejako zewnętrzna w stosunku do obserwatora), a niepewność typu II jest cechą badających go analityków. Całkowitą niepewność można zredukować jedynie przez zmniejszanie niepewności typu II. W praktyce często trudno wydzielić te, wydawałoby się, dobrze teoretycznie zdefiniowane składowe. Niepewność zewnętrzna zależy od naszej percepcji i zrozumienia otaczającego nas świata. Brak wiedzy o zachodzących procesach może powodować, że dane zjawisko interpretujemy początkowo jako naturalnie zmienne albo przypadkowe i dopiero po zastosowaniu dokładniejszego opisu okazuje się, że część niepewności była niepewnością typu II.

Powyższemu podziałowi niepewności można przyporządkować dwie różne koncepcje prawdopodobieństwa. Pierwsza koncepcja mówi o prawdopodobieństwie obiektywnym, a także fizycznym, statystycznym, natomiast druga – o prawdopodobieństwie subiektywnym, inaczej: osobowym.

Prawdopodobieństwo obiektywne jest miarą losowości zjawisk, które istnieją w świecie realnym, a których wystąpienie leży całkowicie lub częściowo poza zasięgiem kontroli człowieka.

Takie pojmowanie prawdopodobieństwa ściśle wiąże się z rachunkiem prawdopodobieństwa i statystyką matematyczną, a więc działami matematyki zajmującymi się badaniem modeli zjawisk przypadkowych.

Zarówno w rachunku prawdopodobieństwa, jak i statystyce matematycznej punktem wyjścia jest to, że każdym zjawiskiem przypadkowym rządzi pewne prawo prawdopodobieństwa. W rachunku prawdopodobieństwa, na podstawie znajomości tego prawa, wnioskuje się o prawdopodobieństwie interesujących nas zdarzeń i obserwowanie zjawiska nie jest konieczne. W przypadku statystyki identyfikuje się i weryfikuje nieznane prawo rządzące zjawiskiem na podstawie obserwacji.

Dla określenia prawdopodobieństwa obiektywnego posługujemy się doświadczeniem myślowym lub statystycznym. Jako przykład doświadczenia myślowego można podać określanie, drogą dedukcji z wykorzystaniem metod kombinatoryki, prawdopodobieństwa zdarzenia polegającego na tym, że w kolejnych pięciu rzutach symetryczną monetą orzeł wystąpi dwukrotnie. Doświadczenie statystyczne wymagałoby w tym wypadku wielokrotnego powtórzenia eksperymentu, polegającego na pięciokrotnym rzucie monetą i obliczeniu, jaki jest udział interesujących nas zdarzeń w ogólnej liczbie prób. W obu wypadkach użyliśmy opisu probabilistycznego, zakładając istnienie prawa prawdopodobieństwa rządzącego zjawiskiem.

Opis deterministyczny wymagałby precyzyjnego określenia: kształtu i rozkładu masy monety, jej początkowego położenia, sposobu trzymania monety i mechaniki rzutu, oporów środowiska oraz uwzględnienia wielu innych czynników. Gdyby nawet taki opis był możliwy, to byłby mało przydatny np. w sytuacjach, gdyby interesowało nas jedynie oszacowanie strat w grze opartej na rzucie monetą.

Podobnie jest z opisem wielu zjawisk przyrodniczych (zmiany pogody, trzęsienia ziemi), procesów biologicznych (mutacje genetyczne, epidemie) oraz zjawisk ekonomicznych, dla których nie istnieją raz na zawsze ustalone prawa natury i wiele zależy od trudnych do przewidzenia zachowań ludzkich. Złożoność wymienionych procesów jest tak duża, że opis deterministyczny, opierający się na rozpoznanych prawach przyczynowych, musi okazać się niewystarczający. Konieczne jest włączenie czynników, które odbieramy intuicyjnie jako losowe, i możliwość pomiaru tej losowości. Z poszukiwaniami ilościowych miar losowości związane było stworzenie przez Laplace'a

na początku XIX w. rachunku prawdopodobieństwa oraz wprowadzenie przez Kołmogorowa w 1933 r. aksjomatycznej definicji prawdopodobieństwa opartej na teorii miary, co zbudowało podstawy całej współczesnej probabilistyki.

Prawdopodobieństwo subiektywne wiąże się z poglądem, że prawdopodobieństwo nie musi odnosić się do możliwych wyników doświadczenia myślowego lub statystycznego, lecz wyraża stopień przekonania danej osoby o możliwości wystąpienia zdarzenia lub stopień przekonania o prawdziwości hipotezy czy wydawanego sądu. Prekursorami takiej koncepcji prawdopodobieństwa byli de Finetti i Ramsey. Tak rozumiane prawdopodobieństwo bywa też określane terminem osobowe (*personal*), np. Savage (1967; 1971), co według Dawida (2004) jest właściwsze, choć rzadziej stosowane.

Przykładami probabilistycznych sądów są: „Uważam za bardzo prawdopodobne, że w przyszłym roku wystąpi recesja" albo: „Prawdopodobieństwo, że inflacja znajdzie się w przedziale 1,5–3,5%, oceniam na 0,7".

Prawdopodobieństwo subiektywne nie jest atrybutem zdarzenia, lecz cechą obserwatora, odzwierciedlającą jego niepewność. Nie jest zatem „elementem" świata realnego. Można tu przytoczyć słynne powiedzenie de Finettiego (1974): „prawdopodobieństwo nie istnieje". Prawdopodobieństwo subiektywne wiąże się ze światem wewnętrznym podmiotu wypowiadającego sąd i jest określane za pomocą miary właściwej temu podmiotowi. Nie ma żadnych podstaw, by zakładać, że określenia typu: „bardzo prawdopodobne" albo „0,7" znaczą to samo dla każdego, kto ich używa.

W związku z tym pojawiają się pytania zarówno o definicję przestrzeni zdarzeń elementarnych, jak też miary probabilistycznej, a w konsekwencji o adekwatność aparatu probabilistycznego do reprezentowania niepewności typu II.

Warto wspomnieć, że powstały i są rozwijane formalizmy alternatywne. W odpowiedzi na problemy związane z niepewnością lingwistyczną Zadeh stworzył teorię zbiorów rozmytych (*fuzzy sets*; por. Zadeh 1965; Bellman, Zadeh 1970). Gdy posługujemy się zbiorami rozmytymi, nie musimy przypisywać dokładnej wartości pojęciom lingwistycznym – możemy je rozumieć jako wartości z pewnego przedziału, np. eksperckie: „0,7" może być interpretowane jako przedział $<0{,}65, 0{,}75>$. Przynależność elementu do zbioru jest określana nie w ramach logiki dwuwartościowej („nie należy" = 0, „należy" = 1), lecz wielowartościowej – przez wprowadzenie funkcji przynależności, która może przyjmować wartości z przedziału $<0, 1>$. Zapewnia to płynne przejście od całkowitej nieprzynależności, przez przynależność częściową, do całkowitej przynależności.

Inną koncepcją, która wydaje się atrakcyjna w sytuacjach, gdy oprócz losowości występuje niepewność typu II, jest teoria Dempstera-Shafera, która wzięła swą nazwę od nazwisk autorów dwóch prac: Dempstera (1968) i Shafera (1976). Jest ona uogólnieniem logiki bayesowskiej określającej sposób modyfikacji prawdopodobieństwa na podstawie nowych faktów.

Na użytek tej pracy pozostaniemy jednak przy założeniu, że niepewność typu II może być reprezentowana za pomocą pojęć teorii prawdopodobieństwa. Świadomość konsekwencji subiektywizmu jest jednak bardzo ważna dla właściwego formułowania i wykorzystania probabilistycznych opinii czy prognoz ekspertów.

W literaturze dotyczącej sposobów uzyskiwania ocen probabilistycznych znajdziemy opisy różnych metod mających na celu lepszy „pomiar" prawdopodobieństwa subiektywnego. Podkreśla się, że bardzo istotny jest sposób zadawania pytań. Na przykład: pytanie wprost o prawdopodobieństwo zdarzenia jest w dużym stopniu obciążone niepewnością lingwistyczną. Może być ona

zmniejszona, jeżeli udzielenie odpowiedzi będzie wymagało porównania – wystąpi wówczas efekt znoszenia się niepewności lingwistycznej dotyczącej każdego z porównywanych zdarzeń.

Zewnętrznym wyrazem prawdopodobieństwa subiektywnego mogą być wybory dokonywane przez ludzi. Jeśli można je wielokrotnie obserwować, to tak mierzone prawdopodobieństwo subiektywne zyskuje interpretację częstościową, której w zasadzie nie ma.

W wielu sytuacjach zachodzi potrzeba obiektywizacji subiektywnych rozkładów eksperckich po to, by na ich podstawie można było wnioskować o prawdopodobieństwie zdarzeń w świecie realnym, jak również by znaleźć wspólny mianownik przy łączeniu opinii czy prognoz.

Kończąc rozważania na temat rozróżnień pojęciowych, dodajmy, że prawdopodobieństwu subiektywnemu nadaje się czasem szersze znaczenie i używa się tego terminu jako synonimu intuicyjnie określanego prawdopodobieństwa obiektywnego.

Należy też zauważyć, że przedstawiona klasyfikacja nie rozwiązuje wszystkich problemów interpretacyjnych (por. Hájek 2009; Gillies 2002).

W kolejnym rozdziale zastanowimy się, o jakie prawdopodobieństwo: obiektywne czy subiektywne (osobowe), są pytani profesjonalni progności uczestniczący w badaniach ankietowych: Systemu Rezerwy Federalnej USA – Survey of Professional Forecasters (SPF), Banku Anglii – Survey of External Forecasters (SEF) oraz Europejskiego Banku Centralnego – Quarterly Survey of Professional Forecasters ECB (SPF ECB). Wydaje się, że właściwa interpretacja tych pytań, odnalezienie ich sensu przez uściślenie, jakiej niepewności oczekuje się od ekspertów, ma zasadnicze znaczenie dla zmniejszenia obaw przed wprowadzaniem pytań probabilistycznych do ankiet makroekonomicznych. Być może dla przezwyciężenia awersji prognostów do prognoz probabilistycznych (prognozy punktowe są w tej grupie wyraźnie preferowane), konieczne okaże się też odejście od wzorców pytań oraz sposobów prezentacji wyników, wyznaczonych przed laty przez twórców ankiety SPF.

3. Problemy z interpretacją pytań o prawdopodobieństwo w ankietach makroekonomicznych

Analizując formularze ankiet SPF, SEF i SPF ECB, dojdziemy do wniosku, że nie do końca jest jasne, jakie są oczekiwania wobec ekspertów. Postawione tam pytania o rozkład prawdopodobieństwa mogą być niejednoznacznie interpretowane przez respondentów. Na przykład w ankiecie Banku Anglii tabela dotycząca inflacji zatytułowana jest następująco: „Probability distribution of 12-month CPI inflation over the medium term". Poniżej umieszczono objaśnienie, które wydaje się sprzeczne z tytułem: „The probabilities of these alternative forecasts should of course add up to 100, as indicated". Pojawiają się wątpliwości, co ekspert powinien zrobić:

1. Czy przedstawić swoją aproksymację rozkładu prawdopodobieństwa inflacji?
2. Czy podać rozkład opisujący błąd punktowej prognozy inflacji?
3. Czy opisać swoją niepewność dotyczącą hipotez o ukształtowaniu się inflacji w podanych przedziałach?

Pytanie eksperta o rozkład prawdopodobieństwa inflacji (pkt 1) byłoby uzasadnione jedynie wówczas, gdybyśmy mogli przyjąć założenie, że przyszła inflacja (w ustalonym punkcie czasowym) podlega pewnemu nieznanemu prawu prawdopodobieństwa, które ekspert powinien ziden-

tyfikować. Odpowiedź na tak rozumiane pytanie byłaby obarczona niepewnością drugiego rzędu – niepewnością oceny niepewności. Wymagałoby to dalszego opisywania tej niepewności, co byłoby bardzo problematyczne.

Pytanie o błąd prognozy punktowej (pkt 2) rodziłoby wątpliwości, czy można ufać samoocenie eksperta w zakresie popełnianych przez niego błędów.

Sensowniej jest zatem przyjąć interpretację ostatnią (pkt 3) – ekspert powinien określić swoją niepewność dotyczącą różnych hipotez co do przyszłych wartości wskaźnika inflacji. Wydaje się, że jest to również zgodne z intencjami twórców ankiety ASA-NBER (Zarnowitz, Lambros 1983), która stała się wzorem dla wszystkich wymienionych wcześniej ankiet.

Mamy więc do czynienia z pytaniem o niepewność typu II, wymagającą opisu w kategoriach subiektywnego prawdopodobieństwa. Przyjęcie takiej interpretacji powoduje, że zmieniają się kryteria oceny poprawności rozkładu. Rozkład jest dobry, jeżeli poprawnie odzwierciedla przekonania eksperta. Jeżeli ekspert jest przekonany o prawdziwości pewnej hipotezy, to stopień jego przekonania powinien znaleźć odzwierciedlenie w przypisaniu jej wysokiego prawdopodobieństwa bez obaw o to, czy w przyszłości okaże się ona prawdziwa czy fałszywa. Jeżeli nie jest pewny, to powinien przypisać prawdopodobieństwa bardziej równomiernie większej liczbie przedziałów. Przedstawiany przez niego rozkład nie musi być prawdziwy w jakimś obiektywnym sensie i nie powinien być tak oceniany. Obiektywizacja rozkładu może okazać się bardzo istotna w przypadku pewnych zastosowań, ale powinna być zadaniem przetwarzających ankiety, jeśli takie są ich potrzeby, a nie ekspertów.

Interpretacja rozkładów eksperckich w kategoriach prawdopodobieństwa subiektywnego (osobowego) jest dosyć powszechna wśród reprezentantów różnych dziedzin nauki i techniki wykorzystujących wiedzę ekspercką. W literaturze ekonometrycznej tymczasem dominują prace, które identycznie traktują eksperckie oceny niepewności i miary niepewności uzyskiwane metodami statystycznymi, np. drogą modelowania zmienności wskaźników makroekonomicznych czy badania historycznych błędów prognoz modelowych. Stawiają więc znak równości między różnymi typami niepewności i każą ekspertom konkurować z modelami statystycznymi używanymi do opisu skomplikowanych zjawisk obciążonych niepewnością typu I.

Zastanówmy się teraz nad wymaganiami stawianymi respondentom ankiet SPF, SEF, SPF ECB w zakresie kwantyfikacji niepewności. Wszystkie te ankiety makroekonomiczne należą do klasy ankiet z predefiniowanymi przedziałami. Zakres możliwych (tu trzeba zaznaczyć: możliwych zdaniem przeprowadzających ankietę) wartości zmiennych jest dzielony na ustaloną liczbę przedziałów o zadanej z góry długości. Zadaniem prognosty jest przypisanie im prawdopodobieństwa.

Niezależnie od sensu, jaki zostałby nadany ankietowemu pytaniu, sprostanie oczekiwaniom przeprowadzających ankietę wydaje się co najmniej bardzo trudne. Nawet gdybyśmy się zgodzili, że najwłaściwsza jest interpretacja pytań w kategoriach prawdopodobieństwa subiektywnego, to nadal pozostałyby wątpliwości, czy podany przez eksperta rozkład dobrze reprezentuje jego przekonania. Duża liczba wąskich przedziałów powoduje, że przypisane im prawdopodobieństwa mogą być w znacznym stopniu obciążone niepewnością lingwistyczną. Można ją zmniejszyć w sytuacji, gdy istnieje możliwość określania prawdopodobieństwa subiektywnego przez porównywanie prawdopodobieństwa zdarzeń, o czym była mowa w rozdziale 2. Ekspert musiałby zatem określić wzajemne relacje pomiędzy prawdopodobieństwami wszystkich przedziałów i przedstawić rozkład, w którym byłyby one zachowane. Liczba koniecznych porównań jest w tym wypadku zbyt duża.

Definiowanie przedziałów przez przeprowadzających ankietę może też spotęgować, rozpoznaną przez psychologów, tendencję do „zakotwiczania" ocen. Ankietowani mogą sugerować się podanym zakresem wartości, a także tym, który z przedziałów zakresu jest przedziałem centralnym.

Problem może być rozwiązany przez zastąpienie pytań o prawdopodobieństwo pytaniami o granice przedziałów o podanym prawdopodobieństwie, czyli pytaniami o kwantyle. Jeżeli np. ekspert zapytany o kwantyle rzędu 0,05; 0,50 i 0,95 (inaczej: centyle 5, 50, 90) udzieli następującej odpowiedzi:

$$x_5 = 0{,}5 \quad x_{50} = 1{,}5 \quad x_{95} = 4{,}0$$

będzie to oznaczało, że:

1) medianą jego subiektywnego rozkładu jest: 1,5, tzn. że uznaje on za jednakowo prawdopodobne wystąpienie wartości wyższych i niższych od 1,5;

2) 90-procentowym centralnym przedziałem wiarygodności jest 0,5–4;

3) równie mało prawdopodobne ($p = 0{,}05$) jest wystąpienie wartości mniejszych od 0,5, jak większych od 4;

4) jednakowo prawdopodobne ($p = 0{,}45$) są przedziały: 0,5–1,5 oraz 1,5–4.

Zwróćmy uwagę na możliwość porównywania prawdopodobieństwa zdarzeń, co jest istotne dla zmniejszenia niepewności lingwistycznej.

4. Czy na pewno istnieje potrzeba zastępowania prognoz punktowych prognozami probabilistycznymi?

W ostatnim okresie ukazało się wiele publikacji dotyczących spójności przedstawianych przez uczestników ankiet makroekonomicznych prognoz punktowych z ich prognozami probabilistycznymi (Engelberg, Manski, Williams 2009a; 2009b; Boero, Smith, Wallis 2008a; 2008b). Badania stały się możliwe dzięki dostępności indywidualnych (anonimowych) odpowiedzi ankietowych. Zawierają one zarówno prognozy punktowe, jak i rozkłady prawdopodobieństwa, które powinny im odpowiadać. Zwykle zakładano, że prognoza punktowa to wartość oczekiwana przedstawianego rozkładu. Badania wykazały, że jest inaczej. Eksperci różnie transformują swoje subiektywne rozkłady prawdopodobieństwa na wartości punktowe. Nie można zakładać, że są to wartości oczekiwane. Bywają bardziej zgodne z dominantami lub medianami, ale też nierzadko są którymś z preferowanych kwantyli. Engelberg, Manski i Williams (2009a) ostrzegają decydentów przed podejmowaniem decyzji na podstawie prognoz punktowych.

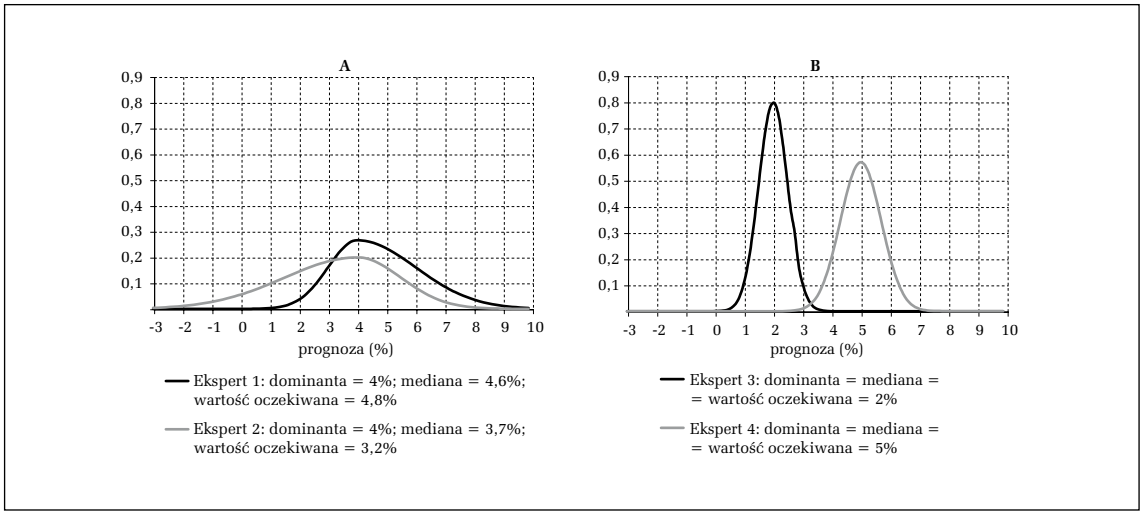
To, czy ekspert posługuje się wartością oczekiwaną, medianą czy dominantą, byłoby nieistotne, gdyby jego subiektywny rozkład prawdopodobieństwa był symetryczny – wszystkie wymienione parametry są wtedy identyczne. Wydaje się jednak, że asymetria jest bardziej naturalną cechą rozkładów odzwierciedlających przekonania eksperta. Prawdopodobieństwo scenariuszy prowadzących do wzrostu danego wskaźnika będzie na ogół inne niż tych, które powodowałyby jego spadek.

Im większa asymetria rozkładu, tym bardziej wartość oczekiwana, mediana i dominanta będą od siebie oddalone. Jeżeli nie znamy stosowanych przez poszczególnych ekspertów strategii

formowania prognoz, to nie możemy ich porównywać. Zbieżność prognoz punktowych nie musi oznaczać zgodności opinii, tak jak rozbieżność nie musi być oznaką zróżnicowania. Traci też sens traktowanie średniej lub mediany prognoz jako konsensusu, a dyspersji jako miary niepewności ekspertów.

Dla zilustrowania tych interpretacyjnych problemów załóżmy, że u podstaw eksperckich prognoz punktowych leżą rozkłady prawdopodobieństwa opisane funkcjami gęstości przedstawionymi na wykresach 1A i 1B.

Wykresy 1A–1B
Eksperckie rozkłady prawdopodobieństwa i ich charakterystyki punktowe



Dominanta, mediana i wartość oczekiwana rozkładu eksperta nr 1 wynoszą odpowiednio: 4%, 4,6%, 4,8%. W przypadku eksperta nr 2 dominanta jest identyczna: 4%, natomiast pozostałe charakterystyki punktowe są inne: mediana 3,7%, a wartość oczekiwana 3,2%. Gdyby obaj eksperci w odpowiedzi na pytanie ankietowe przedstawiali swoje dominanty, moglibyśmy mówić o pełnym konsensusie (dyspersja = 0). Z kolei gdyby każdy z nich uważał, że to wartość oczekiwana najlepiej odzwierciedla jego subiektywny rozkład prawdopodobieństwa, dyspersja byłaby duża. Jeszcze inaczej byłoby, gdyby jeden z ekspertów używał dominanty, a drugi wartości średniej czy mediany.

Porównanie wykresów 1A i 1B pokazuje z kolei, że relacje między dyspersją ocen punktowych a niepewnością ekspertów nie są oczywiste. Duża zgodność prognoz punktowych (załóżmy, że są to dominanty) nie zawsze świadczy o silnym przekonaniu ekspertów co do przedstawianych wartości (wykres 1A) i odwrotnie – silne przekonanie nie musi iść w parze z niewielkim zróżnicowaniem opinii (wykres 1B).

Problemowi przydatności dyspersji prognoz punktowych jako przybliżonej miary niepewności poświęca się wiele uwagi (oprócz wspomnianych już prac np. Clements 2008; Giordani, Soderlind 2003). Można odnieść wrażenie, że zdania są równo podzielone. Dopiero głębsza analiza prowadzi do wniosku, że zasadniczą przyczyną tego zróżnicowania jest fakt, że część badaczy zajmuje się zależnością między dyspersją eksperckich prognoz a niepewnością typu I, odzwierciedlającą

zmienność zjawisk makroekonomicznych. W takim kontekście przydatność dyspersji jest często potwierdzana (np. Giordani, Soderlind 2003). Inni z kolei badają związki dyspersji z niepewnością typu II, która towarzyszy ekspertom przy formułowaniu prognoz. Tu przydatność dyspersji jest kwestionowana (np. Boero, Smith, Wallis 2008a).

5. Czy prognozy probabilistyczne rozwiążą dotychczasowe problemy?

Podstawowym problemem w analizie prognoz punktowych jest niejednoznaczność tego, co wyrażają. Eksperci przyjmują różne taktyki „streszczania" swoich subiektywnych rozkładów, co sprawia, że ich punktowe predykcje mogą być nieporównywalne. Wprowadzenie ankiet probabilistycznych umożliwiłoby wyliczenie potrzebnych parametrów w jednolity sposób. Analizowanie oczekiwań profesjonalnych prognostów byłoby oparte na bardziej wiarygodnych wskaźnikach.

Drugi problem to brak informacji o niepewności towarzyszącej ekspertom, jeżeli podawane są tylko wartości punktowe. Prognozy probabilistyczne umożliwiłyby obliczanie różnych miar niepewności i wybranie właściwej dla kontekstu badawczego czy decyzyjnego.

Bez względu na to, czy pytania probabilistyczne sformułujemy w sposób tradycyjny, tzn. gdy będą to pytania o prawdopodobieństwo ustalonych przedziałów, czy zastąpimy je pytaniami o kwantyle, oba problemy będą rozwiązane. Pozostaną jednak ograniczenia wynikające z faktu, że eksperckie rozkłady są subiektywne. Ważna jest świadomość tych ograniczeń. Pozwoli to właściwie interpretować wyniki ankiet, jak również zastosować sposoby przetwarzania i prezentacji adekwatne do celów, którym ankiety mają służyć.

W kolejnych podrozdziałach przyjrzymy się specyfice potencjalnych zastosowań, próbując odpowiedzieć na pytania:

– jakiego typu ograniczenia występują w poszczególnych obszarach?

– czy dotychczasowy sposób opracowywania i przedstawiania wyników ankiet typu SPF, polegający na agregacji histogramów, w każdym z tych obszarów jest właściwy?

– co można zrobić, by lepiej wykorzystać ankiety?

5.1. Badanie oczekiwań

Wiele cennych sugestii, dotyczących badania oczekiwań grup ekspertów ekonomicznych za pomocą ankiet probabilistycznych, znajdziemy w kolejnej pracy Engelberga, Manskiego i Williamsa (2009b). Autorzy, którzy wcześniej zakwestionowali sens posługiwania się prognozami punktowymi, tym razem przedstawili argumenty przeciwko wnioskowaniu o oczekiwaniach na podstawie analizy zmian wartości punktowych charakterystyk rozkładów zagregowanych. Zbadali dane na temat oczekiwań, pochodzące z ankiet SPF przeprowadzanych w USA w latach 1992–2006. Okazało się, że heterogeniczność ekspertów ujawnia się nie tylko w kontekście prognoz punktowych. Część osób, przedstawiając swoje subiektywne rozkłady, stale koncentrowała podawane prawdopodobieństwa w niewielkiej liczbie przedziałów. Inną grupę cechowała z kolei nadmierna ostrożność. Przejawiało się to w dopuszczeniu możliwości ukształtowania się wskaźników inflacji i wzrostu w relatywnie większej liczbie przedziałów. Progności, którzy byli niepewni w jednej an-

kiecie, pozostawali też niepewni w innych. Również ci, którzy spodziewali się np. wysokiej inflacji, w późniejszych ankietach byli podobnego zdania. Tendencja okazała się niezależna od czasu i prognozowanej zmiennej – wynikała ewidentnie z indywidualnych cech prognostów.

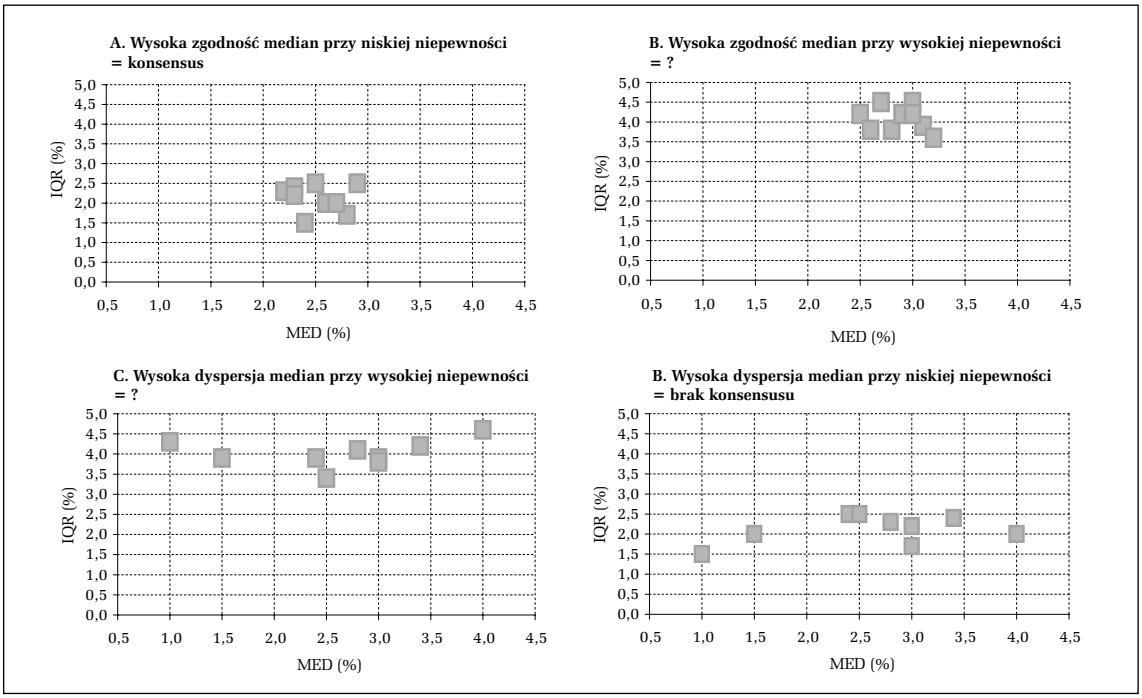
Badania wykazały dodatkowo, że skład panelu podlegał nieustannym zmianom. Nie mogło to pozostać bez wpływu na proporcje między liczbami „pewnych" i „niepewnych", „optymistów" i „pesymistów", a w konsekwencji na zmienność zagregowanych miar oczekiwań.

Obserwowane zmiany oczekiwań mogą więc być pozorne. Analiza oczekiwań na podstawie zagregowanych szeregów czasowych mogłaby być podstawą wnioskowania statystycznego jedynie w sytuacji, gdyby członkowie panelu byli losowo rekrutowani ze stabilnej populacji potencjalnych prognostów i gdyby ich późniejsza, zdarzająca się nieobecność w badaniach ankietowych była statystycznie niezależna od ich cech.

Jeżeli powyższe warunki nie są spełnione, Engelberg, Manski i Williams (2009b) rekomendują badanie zmian parametrów rozkładów indywidualnych. Jako wygodny graficzny sposób przedstawiania informacji zawartych w indywidualnych prognozach probabilistycznych proponują tzw. *scatter graphs* (por. wykresy 2A–2D). Każdy punkt odpowiada jednemu uczestnikowi badania ankietowego. Odcięta punktu to mediana (MED) subiektywnego rozkładu eksperta, a rzędna to rozstęp kwantylowy (IQR). Rozrzut punktów (dyspersja median) wzdłuż osi odciętych pozwala stwierdzić, w jakim stopniu progności są zgodni w swoich sądach. Usytuowanie w pionie odzwierciedla wewnętrzną niepewność ekspertów. Przesunięcie zbioru punktów w górę lub w dół wskazuje na wzrost lub spadek niepewności całej grupy

Wykresy 2A–2D

Mediany (MED) i rozstępy kwantylowe (IQR) indywidualnych prognoz probabilistycznych



Uwaga: o konsensusie możemy mówić, gdy zgodności median towarzyszy mała niepewność.

Engelberg, Manski i Williams (2009b) nie zakładają zmiany dotychczasowego formatu pytań ankiety SPF. Przybliżone wartości median i rozstępy kwantylowe, w ich metodzie są obliczane na podstawie indywidualnych histogramów. Obliczenia takie nie byłyby potrzebne w przypadku zaproponowanych w rozdziale trzecim pytań o kwantyle. Dane potrzebne do sporządzenia *scatter graphs* pochodziłyby bezpośrednio od uczestników ankiety.

5.2. Podejmowanie decyzji na podstawie opinii ekspertów

Tradycyjny sposób opracowywania ankiet typu SPF, przez tworzenie zagregowanych histogramów, okazał się mało przydatny do badania oczekiwań. Ta trywialna, mechaniczna agregacja może być jednak bardzo uzasadniona w sytuacjach, gdy decydent chce wykorzystać opinie ekspertów, ale nie ma żadnej wiedzy o ich cechach i kompetencjach.

Pokażemy, jakie są teoretyczne podstawy agregacji histogramów, i spróbujemy odnaleźć probabilistyczną treść tej metody.

Niech i oznacza numer eksperta. Przedstawiany przez niego histogram możemy traktować jako wynik dyskretyzacji jego ciągłego subiektywnego rozkładu o dystrybuancie $F_i(x)$. Histogramowi i-tego eksperta odpowiada zbiór par $\{(s_k, p_{ik})\}$, gdzie s_k jest środkiem przedziału $< x_{k-1}, x_k)$, a pole prostokąta o podstawie $< x_{k-1}, x_k)$ jest równe prawdopodobieństwu $p_{ik} = F_i(x_k) - F_i(x_{k-1})$.

W wyniku uśrednienia prawdopodobieństw dla każdego przedziału w grupie liczącej N ekspertów otrzymamy dla przedziału $< x_{k-1}, x_k)$ prawdopodobieństwo:

$$p_{Ak} = \frac{1}{N}\sum_{i=1}^{N} p_{ik} = \frac{1}{N}\sum_{i=1}^{N} F_i(x_k) - \frac{1}{N}\sum_{i=1}^{N} F_i(x_{k-1}) \tag{1}$$

co po wprowadzeniu oznaczenia:

$$F_A(x) = \frac{1}{N}\sum_{i=1}^{N} F_i(x) \tag{2}$$

jest równoważne:

$$p_{Ak} = F_A(x_k) - F_A(x_{k-1}) \tag{3}$$

$F_A(x)$ jest zatem dystrybuantą odpowiadającą histogramowi zagregowanemu, opisanemu zbiorem par $\{(s_k, p_{Ak})\}$, i wobec ciągłości dystrybuant składowych jest ciągła. Jeżeli F_i są różniczkowalne, to również F_A jest różniczkowalna i funkcja gęstości rozkładu zagregowanego $f_A(x)$ będzie dana wzorem:

$$f_A(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x) \tag{4}$$

gdzie $f_i(x)$ to funkcja gęstości rozkładu i-tego eksperta.

Sprowadzimy tę gęstość do postaci, która umożliwi jej łatwiejszą probabilistyczną interpretację. W tym celu przeprowadzimy randomizację: indeks i będzie traktowany jako wartość dyskretnej zmiennej losowej I o rozkładzie określonym wzorem:

$$P(I = i) = p_i = 1/N, \text{ dla } i = 1, 2, ..., N \tag{5}$$

Dodatkowo zdefiniujemy zmienną losową Z, która w zależności od losowego indeksu I staje się zmienną losową o rozkładzie f_i. Zmienną scharakteryzowaną przez rozkład f_i nazwiemy X_i. Przy takich oznaczeniach:

$$Z = \begin{cases} X_1 \text{ gdy } I = 1, \\ X_2 \text{ gdy } I = 2, \\ \ldots \\ X_N \text{ gdy } I = N \end{cases} \tag{6}$$

Pamiętając o tym, że I jest zmienną losową, obliczymy dystrybuantę zmiennej losowej Z:

$$F_Z(x) = P(Z < x) = \sum_{i=1}^{N} P(I = i, Z < x) = \sum_{i=1}^{N} P(I = i, X_i < x) \tag{7}$$

Możemy założyć niezależność losowego indeksu I od zmiennych X_i (byłoby to nieprawdziwe w przypadku I i Z). Wtedy:

$$F_Z(x) = \sum_{i=1}^{N} p_i P(X_i < x) \tag{8}$$

i odpowiednio dla gęstości prawdopodobieństwa:

$$f_Z(x) = \sum_{i=1}^{N} p_i f_i(x) \tag{9}$$

Jak widać, rozkład zagregowany, o dystrybuancie i gęstości opisanych odpowiednio wzorami (2) i (4), można utożsamiać z rozkładem zmiennej Z, opisanej wzorami (5) i (4). Należy on do klasy rozkładów nazywanych mieszaninami lub mieszankami rozkładów. O mieszaninach, jako rozkładach otrzymywanych drogą randomizacji, pisze np. Feller (1978, cz. 2).

Wprowadzając losowy indeks I i zmienną Z, pokazaliśmy, że zagregowany histogram opisuje rozkład możliwych wyników związanych z następującym hierarchicznym schematem losowania. W pierwszym kroku wybieramy eksperta z N-elementowego zbioru ekspertów (bo nie wiemy, czyją opinię należy wybrać). W drugim kroku, na podstawie rozkładu wylosowanego eksperta, losujemy wartość interesującej nas zmiennej.

Jeżeli istnieją podstawy do różnego „ważenia" opinii ekspertów, może to być zrealizowane przez zróżnicowanie prawdopodobieństw p_i, które odgrywają rolę wag w mieszaninie rozkładów ($p_i \in (0,1)$ oraz $\sum_{i=1}^{N} p_i = 1$).

Przykładowy sposób konstrukcji wag zostanie opisany w podrozdziale 5.4.

5.3. Wnioskowanie o konsensusie

Pisaliśmy już o tym, że wnioskowanie o konsensusie na podstawie prognoz punktowych, o których nie wiemy, jaki mają związek z subiektywnym rozkładem eksperta, jest bardzo ryzykowne. Duże wątpliwości budzi też określanie konsensusu na podstawie momentów lub funkcji momen-

tów zagregowanego rozkładu, gdyż – jak to określali Giordani i Soderlind (2003) – są one wielkościami hybrydowymi, obejmującymi niemożliwy do identyfikacji wpływ niepewności ekspertów i dzielącej ich różnicy zdań. Hybrydowy charakter wariancji rozkładu zagregowanego łatwo dostrzec, gdy przeprowadzimy jej dekompozycję.

Rozkład zagregowany jest, jak pamiętamy, następującą kombinacją liniową rozkładów indywidualnych:

$$f_A(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x) \tag{10}$$

Momenty rzędu k zmiennej losowej X o rozkładzie $f_A(x)$ są średnimi arytmetycznymi momentów μ_{ki} odpowiadających rozkładom $f_i(x)$, gdyż:

$$\mu_{kA} = E(X^k) = \int_{-\infty}^{+\infty} x^k f_A(x) = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{+\infty} x^k f_i(x)dx = \frac{1}{N}\sum_{i=1}^{N}\mu_{ki} \tag{11}$$

Dla $k = 1$ i $k = 2$ otrzymamy:

$$\mu_{1A} = E(X) = \frac{1}{N}\sum_{i=1}^{N}\mu_{1i}, \quad \mu_{2A} = E(X^2) = \frac{1}{N}\sum_{i=1}^{N}\mu_{2i} \tag{12}$$

Wariancja może być wyrażona za pomocą momentu rzędu pierwszego i drugiego. Dla rozkładu zagregowanego mamy:

$$\sigma_A^2 = \mu_{2A} - \mu_{1A}^2 \tag{13}$$

i analogicznie dla rozkładu i-tego eksperta:

$$\sigma_i^2 = \mu_{2i} - \mu_{1i}^2 \tag{14}$$

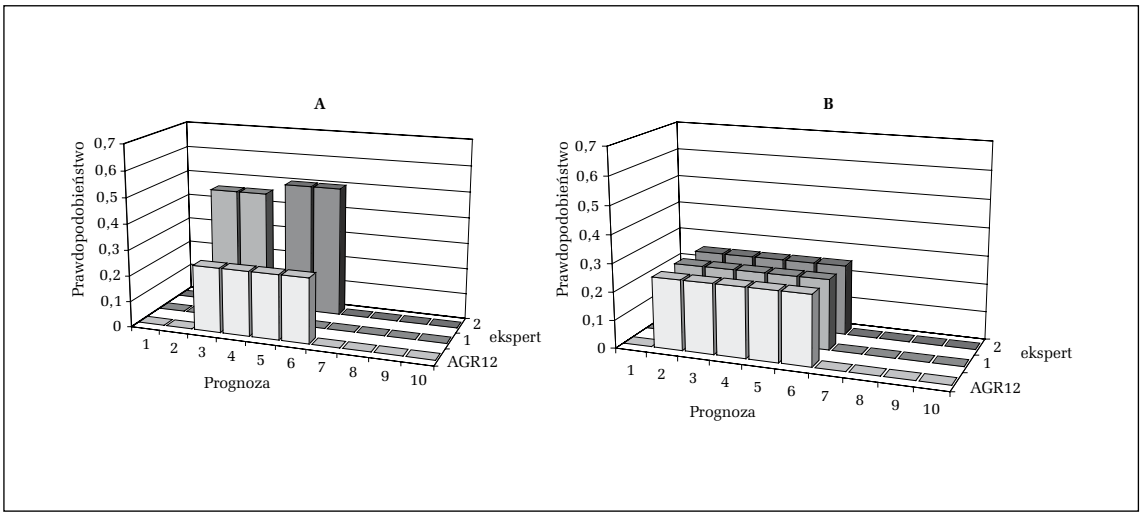
Za Boero, Smith i Wallis (2008a) przeprowadzimy następującą dekompozycję wariancji rozkładu zagregowanego:

$$\sigma_A^2 = \mu_{2A} - \mu_{1A}^2 = \frac{1}{N}\sum_{i=1}^{N}(\sigma_i^2 + \mu_{1i}^2) - \mu_{1A}^2 = \frac{1}{N}\sum_{i=1}^{n}\sigma_i^2 + \frac{1}{N}\sum_{i=1}^{n}(\mu_{1i} - \mu_{1A})^2 \tag{15}$$

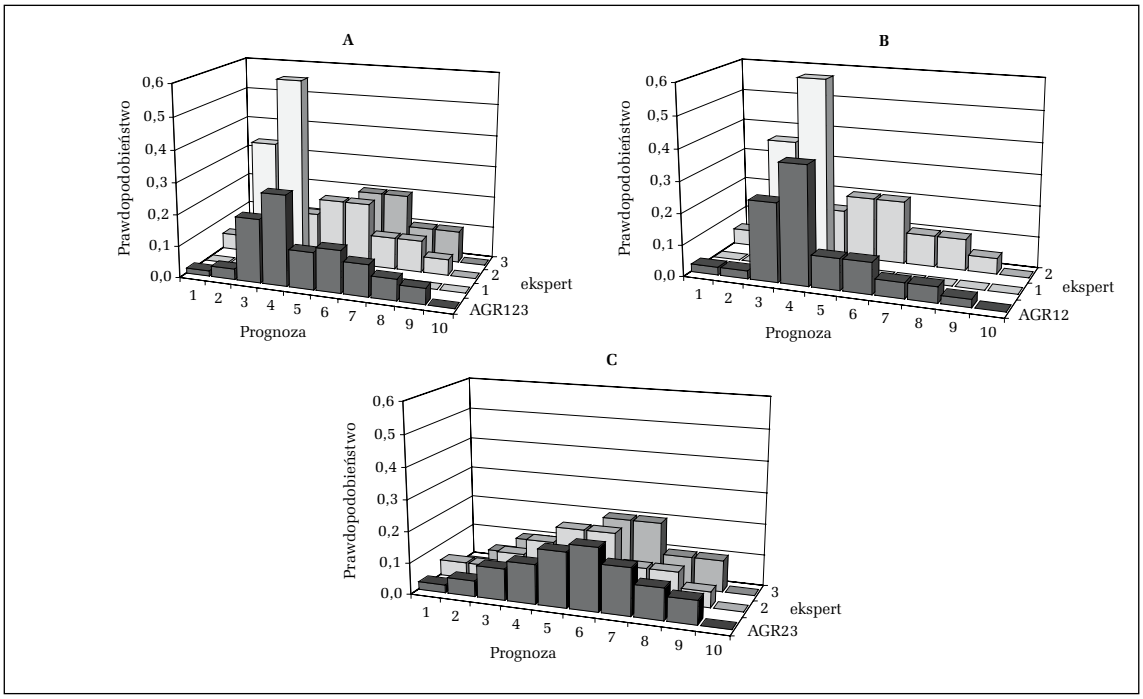
Pierwszy składnik otrzymanej równości jest średnią wariancją rozkładów indywidualnych, a więc średnią niepewnością osobową. Drugi to średnie odchylenie kwadratowe indywidualnych wartości oczekiwanych od wartości oczekiwanej zagregowanego rozkładu. Jest to więc niepewność wynikająca z różnicy zdań. Jak widać, wariancja zagregowanego rozkładu nie daje podstaw do wnioskowania o konsensusie lub niepewności z osobna.

Dla zilustrowania innych problemów związanych z wnioskowaniem o konsensusie na podstawie zagregowanych rozkładów przedstawimy graficznie pewne hipotetyczne sytuacje. Wykresy 3A i 3B przedstawiają proces agregacji histogramów dwóch ekspertów. W obu przypadkach otrzymujemy identyczne rozkłady zagregowane, choć w pierwszym indywidualne rozkłady wskazują na brak konsensusu (pusta część wspólna), w drugim natomiast świadczą o pełnym konsensusie.

Wykresy 3A – 3B
Przykłady agregacji histogramów dwóch ekspertów: A – różniące się rozkłady indywidualne,
B – identyczne rozkłady indywidualne



Wykresy 4A – 4C
Wynik agregacji w zależności od składu grupy ekspertów: A – dominujący wpływ eksperta nr 1;
B i C – nieobecność jednego z ekspertów



Uwagi:
Rozkłady zagregowane pokazano na pierwszym planie:
AGR123 (wykres A) – obecni wszyscy eksperci, AGR12 (wykres B) – nieobecność eksperta nr 3, AGR23 (wykres C) – nieobecność eksperta nr 1.

Wykres 4A pokazuje, jak konsensus może być zniekształcony przez to, że wariancja rozkładu jednego z trzech ekspertów jest dużo mniejsza niż w przypadku pozostałych. Wykresy 4B i 4C są ilustracją wpływu zmian w składzie grupy prognostów. Nieobecność eksperta nr 1 w badaniu ankietowym miałaby zupełnie inny skutek niż nieobecność eksperta nr 2 lub eksperta nr 3.

5.4. Wnioskowanie o obiektywnym prawdopodobieństwie zdarzeń

Opinie ekspertów bardzo często są potrzebne do tego, by oszacować obiektywne prawdopodobieństwo niepewnych zdarzeń. W takich sytuacjach konieczne będzie uwzględnienie możliwych obciążeń ocen eksperckich, a także tego, że każdy z ekspertów posługuje się swoją własną miarą przy określaniu prawdopodobieństwa.

Wiele wskazówek znajdziemy w literaturze. Problemy adekwatnej reprezentacji i właściwego wykorzystywania subiektywnych opinii ekspertów od wielu lat są przedmiotem zainteresowania badaczy z różnych dziedzin nauki. Pierwsze prace pojawiły się w latach 60. i 70. (np. Winkler 1967; 1971; Ellsberg 1961; Savage 1967; 1971; Hogarth 1975). Warto też wspomnieć o późniejszych pracach: Kadane, Winkler (1988) oraz Kadane, Wolfson (1998).

To, na ile poprawnie ludzie są w stanie formułować probabilistyczne oceny i jak przebiega ten proces, od dawna jest też analizowane przez psychologów. Badania zapoczątkowane przez Kahnemana i Tversky'ego (1974), a później prowadzone przez innych kognitywistów wykazały, że ludzie nieświadomie stosują uproszczone reguły wnioskowania i przy udzielaniu odpowiedzi o niepewność mają skłonność do posługiwania się pewnym ograniczonym zestawem heurystyk, czyli operacji mentalnych o charakterze skrótów myślowych. W wielu przypadkach są one efektywne, ale bywają też błędne, niezgodne z logiką i prawami prawdopodobieństwa.

Heurystyka reprezentatywności to zawyżanie prawdopodobieństwa scenariuszy definiowanych za pomocą większej liczby szczegółów (mają one wtedy więcej reprezentatywnych cech). Rachunek prawdopodobieństwa nakazywałby w takiej sytuacji obliczyć prawdopodobieństwo koniunkcji zdarzeń, co dałoby efekt przeciwny.

Heurystyka dostępności oznacza wydawanie opinii na podstawie tego, jak łatwo można przywołać z pamięci przykłady zdarzeń podobnych. Pamięciowa dostępność zależy od czasu i miejsca analogicznych sytuacji, od tego, jakie wzbudziły emocje i jakie wywołują skojarzenia. Ludzie mogą obawiać się pewnych sytuacji, co wpływa na ich opinie.

Heurystyka zakotwiczenia to formułowanie opinii drogą modyfikacji oceny wynikającej z arbitralnie przyjętego punktu startowego. Korekta jest zwykle niewystarczająca. Tendencją do zakotwiczania ocen Lahiri i Wang (2006) tłumaczyli zaniżanie prawdopodobieństwa zdarzeń, które występują rzadko (na przykład recesja).

Posługiwanie się heurystykami może być przyczyną niezgodności oczekiwań ekspertów z obiektywnym prawdopodobieństwem. Oznacza to, że eksperckie rozkłady mogą mieć zmniejszoną wartość predyktywną, co nie znaczy, że są niepoprawne. Jedynym kryterium oceny subiektywnego rozkładu powinno być to, na ile wiernie odtwarza on rzeczywiste przekonania eksperta. W literaturze silnie akcentowana jest potrzeba stosowania miar, które mogłyby to zapewnić (tzw. *proper scoring rules*).

Istnienie heurystyk należy uwzględnić zarówno w sposobie zadawania pytań o wielkości niepewne (stąd wcześniejsze uwagi dotyczące predefiniowanych przedziałów w ankietach makroekonomicznych), jak i w fazie analizowania wyników ankiet.

Kolejne ograniczenia dotyczące możliwości bezpośredniego wykorzystywania rozkładów subiektywnych do oceny prawdopodobieństwa niepewnych zdarzeń wynikają z faktu (podkreślaliśmy to wielokrotnie), że eksperci kwantyfikują swoją niepewność i mogą to robić w różny sposób. Jedni są zbyt pewni, inni nadmiernie ostrożni w formułowaniu swoich hipotez. Będzie to prowadziło do niedoszacowania lub przeszacowania niepewności zdarzeń. Może być też źródłem problemów z agregacją opinii, co pokazaliśmy wcześniej, przy omawianiu konsekwencji łączenia rozkładów o różniących się znacznie wariancjach. Agregacja rozkładów subiektywnych na potrzeby oceny prawdopodobieństwa obiektywnego może wymagać zróżnicowania wag przypisywanych indywidualnym subiektywnym rozkładom w mieszaninie rozkładów.

Przykład metody obiektywizacji subiektywnych rozkładów

Ciekawą metodę konstrukcji wag zaproponował Cooke (1991). Jej istotą jest obliczanie wag w mieszaninie rozkładów z wykorzystaniem dwóch wskaźników charakteryzujących eksperta: kalibracji (*calibration score*) i informatywności (*information score*). Są one obliczane na podstawie zbioru tzw. zmiennych kalibrujących (*calibration variables*), których prawdziwe wartości są znane przeprowadzającym badanie opinii, natomiast nie są znane ekspertom poproszonym o przedstawienie swoich subiektywnych rozkładów. W przypadku ankiet makroekonomicznych rolę takich zmiennych mogą odgrywać historyczne prognozy ekspertów.

Upraszczając, można powiedzieć, że kalibracja jest miarą statystycznej zgodności eksperckich ocen z obserwacjami, a informatywność mierzy stopień koncentracji eksperckich rozkładów.

Zaletą metody Cooke'a jest to, że opinie ekspertów mogą, ale nie muszą być formułowane w postaci pełnych rozkładów. Wystarcza podanie kilku kwantyli. Zazwyczaj eksperci są pytani o kwantyle rzędu: 0,05, 0,50, 0,95 lub 0,05, 0,25, 0,50, 0,75, 0,95. W pierwszym przypadku oznacza to pytanie o medianę i centralny 90-procentowy przedział wiarygodności, w drugim o medianę oraz przedziały 50- i 90-procentowe. Przedziały nie muszą być symetryczne względem mediany.

Dane ekspertów wykorzystywane są do konstrukcji wskaźników kalibracji i informatywności w następujący sposób.

Kalibracja. Załóżmy, że ekspert został poproszony o podanie, dla każdej z prognozowanych zmiennych X_k, $k = 1, 2 \ldots, K$, wartości trzech kwantyli odpowiedniego rozkładu prawdopodobieństwa:

$$\begin{aligned}
&\text{kwantyla rzędu 0,05: } x_{5k}\text{: } F(x_{5k}) = P(X_k < x_{5k}) = 0,05\\
&\text{kwantyla rzędu 0,50 (mediany): } x_{50k}\text{: } F(x_{50k}) = P(X_k < x_{50k}) = 0,5\\
&\text{kwantyla rzędu 0,95: } x_{95k}\text{: } F(x_{95k}) = P(X_k < x_{95k}) = 0,95
\end{aligned} \tag{16}$$

Wartości tych kwantyli dzielą zbiór możliwych wartości zmiennej X_k na cztery przedziały o prawdopodobieństwach równych kolejno:

$$\begin{aligned}
&p_1 = 0,05,\ p_2 = 0,45,\ p_3 = 0,45,\ p_4 = 0,05\\
&p_1 = P(X_k < x_{5k}) = 0,05, \quad p_2 = P(x_{5k} \le X_k < x_{50k}) = 0,45, \quad p_3 = P(x_{50k} \le X_k < x_{95k}) = 0,45\\
&p_4 = P(X_k \ge x_{95k}) = 1 - \Pr ob(X_k < x_{95k}) = 0,05;\ \ k = 1, 2..., K
\end{aligned} \tag{17}$$

Ogólnie: n kwantyli generuje $n + 1$ przedziałów. Na prawdopodobieństwa $\{p_j\}_{j=1,2,3,4}$ możemy też spojrzeć inaczej: p_j jest prawdopodobieństwem przypisanym przez eksperta zdarzeniu, że prognozowana przez niego zmienna znajdzie się w swoim j-tym przedziale kwantylowym. Prawdopodobieństwa $\{p_1, p_2, p_3, p_4\}$ definiują więc rozkład dyskretnej zmiennej losowej, nazwijmy ją R, której wartościami są numery przedziałów kwantylowych. Rozkład zmiennej losowej R oznaczymy jako f_p $\{R; p_1, p_2, p_3, p_4\}$.

W momencie, gdy mamy możliwość poznania realizacji $x_1, x_2, ..., x_K$ zmiennych $X_1, X_2, ..., X_K$, o które pytaliśmy eksperta, tworzymy rozkład f_r $\{R; r_1, r_2, r_3, r_4\}$ na podstawie liczby realizacji, które znalazły się pomiędzy kolejnymi kwantylami prognozowanych zmiennych. Dokładniej: r_1 jest stosunkiem liczby realizacji, które znalazły się w swoim pierwszym przedziale kwantylowym, do liczby wszystkich realizacji; r_2 jest zdefiniowane analogicznie dla drugich przedziałów kwantylowych itd.

Kalibracja wymaga określenia statystycznej wiarygodności hipotezy, że realizacje pochodzą z rozkładu zgodnego z kwantylami podanymi przez eksperta. Byłoby idealnie, gdyby 5% realizacji zmiennych użytych do kalibracji wpadło do pierwszego przedziału kwantylowego, 45% do drugiego, 45% do trzeciego i 5% do czwartego. Rozkłady f_r i f_p byłyby wówczas identyczne.

Zwróćmy uwagę, że rozkład f_p jest taki sam dla wszystkich ekspertów (wszyscy określają analogiczne kwantyle). Rozkład f_r zależy od eksperta (granice przedziałów kwantylowych są różne dla różnych zmiennych i różnych ekspertów).

Używaną przez Cooke'a miarą rozbieżności rozkładów f_p i f_r jest entropia względna (odległość Kullbacka-Leiblera) określona wzorem:

$$I(f_r, f_p) = \sum_{j=1}^{4} r_j \ln\left(\frac{r_j}{p_j}\right) \tag{18}$$

Zwróćmy uwagę na to że f_r reprezentuje dane rzeczywiste, a f_p jest rozkładem hipotetycznym.

Wskaźnik kalibracji (*calibration score*) Cooke definiuje następująco:

$$Cal_score = 1 - F_{\chi_n^2}(2K \cdot I(f_r, f_p)) \tag{19}$$

gdzie $F_{\chi_n^2}$ jest dystrybuantą χ^2 rozkładu o liczbie stopni swobody n równej liczbie kwantyli, o które jest pytany ekspert (w naszym przykładzie $n = 3$), a K jest liczbą zmiennych użytych do kalibracji.

Wskaźnik kalibracji przyjmuje wartość maksymalną równą 1, gdy rozkład f_r pokrywa się z rozkładem f_p (entropia względna, mierząca rozbieżność, jest wtedy minimalna i wynosi 0). W sytuacji gdy r_i znacznie odbiegają od p_i, wskaźnik kalibracji będzie bliski zera.

Informatywność. Drugą wielkością wchodzącą w skład kryterium oceny prognozy jest wskaźnik informatywności (*information score*). Spośród wszystkich rozkładów określonych na danym przedziale rozkładem o najmniejszej zawartości informacyjnej (największej entropii) jest rozkład równomierny. Im większa zatem będzie entropia względna podawanego przez eksperta rozkładu w stosunku do rozkładu równomiernego (im bardziej te rozkłady będą rozbieżne w sensie odległości Kullbacka-Leiblera), tym lepiej.

Proces obliczania wskaźnika informatywności przebiega następująco.

1. Dla każdej z K prognozowanych przez eksperta zmiennych konstruowany jest jej przedział „określoności". Rozpiętość przedziału jest dobierana tak, by zawierał podane przez wszystkich ekspertów wartości kwantyli oraz wartość zrealizowaną, a ponadto był dłuższy od najwęższego przedziału o tej własności o m%. Najczęściej przyjmuje się $m = 10$ (tzw. *10% overshoot rule*). Następnie wyznaczany jest rozkład bazowy, względem którego liczona będzie entropia, czyli rozkład równomierny na skonstruowanym wcześniej przedziale. Oznaczymy go przez u_k. Rozkład bazowy k-tej zmiennej jest taki sam dla wszystkich ekspertów.

2. Do podanych przez eksperta wartości kwantyli dla zmiennej k dopasowywana jest (zgodnie z zasadą maksymalnej entropii, przy ograniczeniach danych kwantylami) funkcja gęstości prawdopodobieństwa, którą oznaczymy przez g_k.

„Nota" za informatywność (*Information Score*), jaką uzyska ekspert, jest entropią względną obliczoną dla zbioru zmiennych użytych do kalibracji:

$$Inf_score = \frac{1}{K}\sum_{k=1}^{K} I(g_k, u_k) \tag{20}$$

Wagi. Na podstawie wskaźników kalibracji i informatywności, uzyskanych przez każdego eksperta, wyznaczane są wagi, które następnie wykorzystuje się do agregacji opinii dotyczących właściwych zmiennych. Nie uwzględnia się zwykle ekspertów, którzy nie spełniają założonego minimum dotyczącego wskaźnika kalibracji.

Waga określająca udział rozkładu i-tego eksperta w mieszaninie rozkładów obliczana jest w następujący sposób:

$$w(i) = Cal_score(i) \cdot Inf_score(i) \tag{21}$$

Metoda Cooke'a jest stosowana przede wszystkim w sytuacjach, gdy trafność opinii eksperckich nie może być łatwo zweryfikowana za pomocą obserwacji. W procesie kalibracji ekspertów wykorzystywane są w związku z tym pytania niebędące przedmiotem bezpośredniego zainteresowania. W przypadku eksperckich prognoz makroekonomicznych sytuacja jest łatwiejsza. Dotyczą one zazwyczaj wielkości, które mogą być zaobserwowane w pewnej, nieodległej przyszłości. Umożliwia to skonstruowanie wag będących funkcjami eksperckich subiektywnych rozkładów prawdopodobieństwa i obserwacji, a następnie użycie tych wag w kolejnych okresach prognostycznych.

6. Zakończenie

Pomimo powszechnej świadomości, że prognozowanie pojedynczych wartości zmiennych makroekonomicznych ma ograniczoną przydatność, większość ankiet makroekonomicznych przeprowadzanych przez banki centralne zawiera pytania jedynie o wartości punktowe (por. Kokoszczyński, Łyziak 2009). Dzieje się tak prawdopodobnie dlatego, że w odczuciu zarówno ankietowanych, jak i decydentów lepsze są niepewne prognozy punktowe niż niepewne rozkłady prawdopodobieństwa. Dla przezwyciężenia awersji do prognoz probabilistycznych (co wydaje się ważne w świetle

ostatnich badań) konieczne jest nadanie właściwego sensu pytaniom ankietowym. Powszechne wykorzystywanie modeli ekonometrycznych w prognozowaniu powoduje, że rozkłady eksperckie traktujemy podobnie jak te, które są wynikiem analizy szeregów czasowych, tyle że tworzone subiektywnie. Tymczasem istnieje zasadnicza różnica pomiędzy określaniem rozkładu obiektywnego prawdopodobieństwa w sposób subiektywny a wyrażaniem opinii za pomocą prawdopodobieństwa subiektywnego, rozumianego jako stopień przekonania eksperta co do prawdziwości hipotez. Nie możemy oczekiwać, że eksperci okażą się lepsi od modeli w prognozowaniu zmienności otoczenia makroekonomicznego, tak jak od modeli nie oczekujemy formułowania hipotez.

Wykorzystanie subiektywnych rozkładów prawdopodobieństwa wiąże się z wieloma problemami natury teoretycznej i praktycznej. Waga tych problemów jest różna w zależności od obszaru zastosowań. W badaniu oczekiwań najważniejsze jest to, by rozkłady jak najwierniej odzwierciedlały wiedzę i przekonania ekspertów, a także by w procesie przetwarzania nie zgubić tych informacji.

W sytuacjach, gdy opinie eksperckie są potrzebne tylko po to, by sformułować rozkłady wstępne, które mogą być następnie poprawione na podstawie obserwacji, tak silnie akcentowana w niniejszej pracy potrzeba odróżniania prawdopodobieństwa subiektywnego i prawdopodobieństwa subiektywnie określanego nie musi być aż tak istotna. Zastosowanie wnioskowania bayesowskiego, tzn. wielokrotne zastosowanie reguły warunkowania (twierdzenia Bayesa) w miarę dostępności nowych danych prowadzi w granicy do prawdopodobieństwa niezależnego od wyboru rozkładu początkowego. Podejście bayesowskie jest obecnie coraz częściej postrzegane jako remedium na problemy związane z niepewnością parametrów modeli wykorzystywanych do opisu zjawisk ekonomicznych. Jest też proponowane jako sposób na ograniczanie niepewności typu II przez zastosowanie tzw. bayesowskiego łączenia wiedzy, *Bayesian model averaging* – BMA (por. np. Hoeting i in. 1999).

Nie sposób oczywiście rozwiązać wszystkich problemów, które towarzyszą procesom przebiegającym w warunkach niepewności. W przypadku procesów decyzyjnych istnieje zazwyczaj potrzeba wypełnienia luki pomiędzy tym, co już wiemy, o czym mówią dane i modele skonstruowane na podstawie dotychczasowych doświadczeń, a tym, co jest konieczne do dokonania właściwego wyboru. Trudno przewidzieć, co się może wydarzyć, jakie będą konsekwencje podejmowanych decyzji, jak mogą się zmienić zachowania ludzi itd. Wydaje się, że wobec tego typu niepewności bardzo ważna jest umiejętność korzystania z opinii ekspertów. Jest rzeczą oczywistą, że im bardziej złożona będzie sytuacja, tym mniej pewni swoich ocen będą eksperci. Subiektywne rozkłady prawdopodobieństwa stwarzają możliwość wyrażenia tej niepewności i uwzględnienia jej w procesie decyzyjnym.

Bibliografia

Bellman R., Zadeh L. (1970), Decision-Making in a Fuzzy Environment, *Management Science*, 17 (4), B141–B164.

Boero G., Smith J., Wallis K. (2008a), Uncertainty and disagreement in economic prediction: the Bank of England Survey of External Forecasters, *Economic Journal*, 118, 1107–1127.

Boero G., Smith J., Wallis K. (2008b), Evaluating a three-dimensional panel of point forecasts: the Bank of England Survey of External Forecasters, *International Journal of Forecasting*, 24, 354–367.

Clements M. (2008), Consensus and uncertainty: Using forecast probabilities of output declines, *International Journal of Forecasting*, 24, 76–86.

Cooke R. (1991), *Experts in Uncertainty: Expert Opinion and Subjective Probability in Science*, Oxford University Press.

Dawid A. (2004), Probability, Causality and the Empirical World: A Bayes-de Finetti-Popper-Borel Synthesis, *Statistical Science*, 19, 44–57.

Dempster A. (1968), A Generalization of Bayesian Inference, J*ournal of the Royal Statistical Society*, Series B, 30 (2), 205–247.

Engelberg J., Manski Ch., Williams J. (2009a), Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters, *Journal of Business & Economic Statistics*, 27, 30–41.

Engelberg J., Manski Ch., Williams J. (2009b), *Assessing the Temporal Variation of Macroeconomic Forecasts by a Panel of Changing Composition*, http://faculty.wcas.northwestern.edu.

Ellsberg D. (1961), Risk, Ambiguity, and the Savage Axioms, *Quarterly Journal of Economics*, 75, 643–669.

Feller W. (1978), *Wstęp do rachunku prawdopodobieństwa*, PWN, Warszawa.

de Finetti B. (1974), *Theory of Probability*, John Wiley & Sons, New York.

French S. (1995), Uncertainty and Imprecision: Modelling and Analysis, *Journal of the Operational Research Society*, 46, 70–79.

Gillies D. (2002) *Philosophical Theories of Probability*, Routledge, London.

Giordani P., Soderlind P. (2003), Inflation forecast uncertainty, *European Economic Review*, 47, 1037–1059.

Hájek A. (2009), Interpretations of Probability, w: E.N. Zalta (red.), *The Stanford Encyclopedia of Philosophy,* http://plato.stanford.edu/archives/spr2009/entries/probability-interpret/.

Helton J., Johnson J., Oberkampf W., Sallaberry C. (2008), *Representation of Analysis Results Involving Aleatory and Epistemic Uncertainty*, SANDIA Report, http://www.osti.gov/bridge.

Hogarth R. (1975), Cognitive Processes and the Assessment of Subjective Probability Distributions, *Journal of the American Statistical Association*, 70 (350), 271–289.

Hoeting J., Madigan D., Raftery A., Volinsky C. (1999), Bayesian Model Averaging: A Tutorial, *Statistical Science*, 14, 382–401.

Kadane J., Winkler R. (1988), Separating Probability Elicitation From Utilities, *Journal of the American Statistical* Association, 83 (402), 357–363.

Kadane J., Wolfson L. (1998), Experiences in Elicitation, *The Statistitian*, 47 (1), 3–19.

Kokoszczyński R., Łyziak T. (2009), *The use of different measures of inflation expectations in monetary policy making, CCBS-NBP survey results*, materiał zaprezentowany podczas CCBS Chief Economists' Workshop on „New challenges in assessing and managing inflation expectations", 19–20 maja, Bank Anglii.

Lahiri K., Wang J. (2006) Subjective Probability Forecasts for Recession, *Bussiness Economics*, 41 (2), 26–37.

Tversky A., Kahneman D. (1974), Judgement Under Uncertainty: Heuristics and Biases, *Science*, 185, 1124–1131.

Savage L. (1967), Implications of Personal Probability for Induction, *The Journal of Philosophy*, 64, 593–607.

Savage L. (1971), Elicitation of Personal Probabilities and Expectations, *Journal of the American Statistical Association*, 66 (336), 783–801.

Shafer G. (1976), *A Mathematical Theory of Evidence*, Princeton University Press.

Winkler R. (1967), The Assessment of Prior Distributions in Bayesian Analysis, *Journal of the American Statistical Association,* 62 (319), 776–800.

Winkler R. (1971), Probabilistic Prediction: Some Experimental Results, *Journal of the American Statistical Association*, 66 (336), 675–685.

Zadeh L. (1965), Fuzzy sets, *Information and Control*, 8, 338–353.

Zarnowitz A., Lambros L. (1983), *Consensus and uncertainty in economic prediction*, NBER Working Paper, 1171.

Podziękowania

Autorka składa podziękowania Tomaszowi Łyziakowi, Ryszardowi Kokoszczyńskiemu, Andrzejowi Kocięckiemu i Ewie Stanisławskiej, którzy mieli istotny wpływ na powstanie i kształt niniejszej publikacji. Wyrazy wdzięczności za cenne uwagi kieruje pod adresem anonimowych Recenzentów.

Expert assessment of uncertainty in macroeconomic surveys

Abstract

The view that macroeconomic forecasts, both model-based and expert-made, should be accompanied by uncertainty assessment is gaining more and more acceptance. The principal thesis of the present paper ascertains that in the case of expert forecasts we deal with a different type of uncertainty than in the case of model forecasts. This requires a different interpretation of probability distributions submitted by experts. A more appropriate approach seems to consist in a description made in terms of subjective probability, understood as the degree of expert conviction with respect to the submitted hypotheses. In order to justify the above thesis, notional differences were analyzed relevant to the process of classifying uncertainty and probability. Another postulated thesis is the necessity to take into consideration the extent to which experts can adequately quantify uncertainty at the stage of designing macroeconomic surveys and to differentiate the way surveys are processed depending on the field where they are used. Those problems have been presented on the example of macroeconomic surveys targeted at professional forecasters and uses that are typical for central banking (studying expectations, aggregating opinions, forecasting).

Keywords: macroeconomic surveys, studying expectations, uncertainty of expert forecasts, subjective probability