

Mirosław Szreder*

Szum w prognozach

Początek nowego roku obfituje, jak zawsze, w wiele różnych prognoz dotyczących sfery gospodarczej, społecznej czy politycznej. W literaturze naukowej znaczny ich odsetek jest formułowany na podstawie rzetelnych danych statystycznych odnoszących się do minionych okresów czasu, z wykorzystaniem naukowego podejścia do „sztuki i nauki prognozowania”. Ujęte w cudzysłów wyrażenie stanowi drugi człon tytułu znakomitej książki amerykańskiego statystyka N. Silvera pt. *The Signal and the Noise. The Art and Science of Prediction* [Silver, 2012]. Problematyka tej książki nawiązuje – o czym z kolei informuje pierwszy człon jej tytułu – do szumu z przetłumaczonej na język polski monografii D. Kahnemana, O. Sibony’ego i C.R. Sunsteina pt. *Szum, czyli skąd się biorą błędy w naszych decyzjach* [Kahneman i in., 2022]¹. Przypo-

mnijmy, że przez szum statystycy rozumieją istnienie niepożądanych rozbieżności (dyspersji) w ocenach i osądach. W obu tych pracach znajdujemy ciekawe rozważania na temat przyczyn niepowodzeń w prognozowaniu – istnienia szumu w prognozach, mimo coraz większej ilości dostępnych danych statystycznych i innego rodzaju użytecznych informacji. Błędnych prognoz wcale nie ubywa! Wydaje się, że niewystarczające dla poprawy trafności i precyzji prognoz są rosnące możliwości pozyskiwania i przetwarzania dużych zbiorów danych, a także wykorzystania do tych celów elementów sztucznej inteligencji, np. uczenia maszynowego (*machine learning*).

¹ W oryginalnym wydaniu angielskim tytuł ten brzmi: *Noise: A Flaw in Human Judgment* [William Collins, 2021]. Omówienie tej monografii zostało zamieszczone m.in. w numerze 3/2022 „Biuletynu PTE”, por. [Szreder, 2022a], a także w opracowaniu [Szreder, 2022b].

* Prof. dr hab. Mirosław Szreder – Uniwersytet Gdański.

Źródła niepewności w prognozowaniu

Sam proces prognozowania, który jest oparty najczęściej na modelach i danych statystycznych, nie jest łatwy. Można go co prawda uznać za rodzaj wnioskowania statystycznego, którego metody i techniki są od dawna dobrze poznane i rzadko w praktyce zawodzą, z tym jednak zastrzeżeniem, że we wnioskowaniu o przyszłości pojawia się drugie, obok próby statystycznej, źródło niepewności. Z jednej strony mamy niepewność będącą konsekwencją niepełnej wiedzy o przeszłości, którą charakteryzują informacje z próby statystycznej (np. w postaci szeregów czasowych wybranych zmiennych). W większości metod prognozowania ta niepełna wiedza o przeszłości znajduje zastosowanie w mniej lub bardziej trafny wybór odpowiedniego modelu statystycznego (lub ekonometrycznego) służącego do opisu przebiegu interesującego nas zjawiska i do oszacowania parametrów tego modelu. Żaden model, jak wiadomo, nie jest doskonały, stąd pierwszym źródłem niepewności w prognozowaniu są możliwe błędy modelowania tego, co już zaistniało w przeszłości.

Z drugiej zaś strony, prognozowanie dotyczy przyszłości, a więc każde założenie o rozwoju danego zjawiska według dotychczasowego wzorca, zgodnego ze skonstruowanym i oszacowanym modelem, obarczone jest sporą dozą niepewności. Ta niepewność, odnosząca się do nieznannej przyszłości, dotyczy ponadto wszelkich okoliczności, które w różnym stopniu mogą wpływać na prognozowaną zmienną. Kluczowym zagadnieniem we wnioskowaniu o kształtowaniu się danej zmiennej w przyszłości jest to, czy jej przewidywane wielkości mieszczą się w zakresie obserwacji z przeszłości, a także, czy doświadczenia minionych okresów stanowią adekwatny opis mechanizmu kształtowania się tej zmiennej nie tylko w próbie, ale także poza próbą. Najtrudniejszym wyzwaniem jest oczywiście to, czy poza zakresem próby można ten mechanizm uznać za nadal aktualny? Nate Silver [2012] podkreśla w swojej pracy, że problem wykraczania poza próbę (*out-of-sample problem*) jest odpowiedzialny za szereg poważnych błędów w prognozowaniu wielu zjawisk. Autor ilustruje istotę tego zagadnienia między innymi nietrafnymi przewidywaniami agencji ratingowej Moody's z początku kryzysu finansowego 2008 roku. Modele agencji Moody's oszacowane zostały na podstawie szeregów czasowych sięgających wczesnych lat 80. „Problem polegał na tym, że od roku 1980 do wczesnych lat XXI wieku ceny domów w USA były albo stabilne, albo podlegały trendowi rosnącemu. W tych okolicznościach założenie o tym, że kredyt hipoteczny jednego właściciela domu ma niewiele wspólnego z kredytem innego właściciela domu, było prawdopodobnie uzasadnione. Nic jednak, co zawarte było w danych o przeszłości

w podanym przedziale czasu, nie byłoby w stanie opisać tego, co stałoby się na rynku, gdyby ceny domów zaczęły spadać jednocześnie (*in tandem*). Upadek rynku kredytów hipotecznych był zdarzeniem spoza próby (*an out-of-sample event*), a oszacowane modele okazały się bezużyteczne do prognozowania ryzyka upadku w tych warunkach” (s. 44). Podobną opinię wyraził Alan Greenspan jesienią 2008 roku²: „Gdyby modele te były poprawnie dopasowane do danych historycznych, obejmujących także okresy załamania gospodarczego, określiłyby one wymogi kapitałowe na znacznie wyższym poziomie, a świat finansowy byłby teraz w znacznie lepszym stanie”.

Błędowi wykraczania oszacowań poza zakres próby towarzyszy zwykle zbyt silne przekonanie badaczy o tym, że wiedza o fragmencie przeszłości w kształtowaniu się określonego zjawiska jest wiedzą o istocie tego zjawiska. Abstrahowanie od uwarunkowań charakterystycznych dla określonego – w pewien sposób zawsze specyficznego – przedziału czasu zamazuje różnicę między tym, co badacz wie, a tym, co wydaje mu się, że wie. Chodzi więc o upowszechnioną przez norweskiego naukowca Terjega Avena różnicę pomiędzy tym, co:

- wiemy, że wiemy (*known knowns*),
 - wiemy, że nie wiemy (*known unknowns*),
 - nie wiemy, że nie wiemy (*unknown unknowns*),
- por. [Aven, 2014, s. 12].

Zdaniem N. Silvera zwiększające się zasoby dostępnych informacji, a do tego często prezentowanie danych statystycznych z dużą dokładnością (wspomniana agencja Moody's podawała do ostatnich dni przed kryzysem wyliczenia z dokładnością do dwóch miejsc po przecinku) sprzyjają przeoczeniu różnicy między tym, co wiemy, że nie wiemy, a tym, co nie wiemy, że nie wiemy.

Błędy w modelowaniu zjawisk z przeszłości

Prognozowanie oparte na statystycznych i ekonometrycznych modelach (*model-based forecasting*) jest bardziej cenione w nauce niż prognozowanie subiektywne, odwołujące się do opinii ekspertów lub oparte na zbiorze pewnych heurystyk. Nie oznacza to jednak, że prognozy oparte na modelach są zawsze trafne. Błędy takich prognoz mogą mieć swoje pierwotne źródło – jak już wcześniej wspomniano – w modelowaniu przeszłości interesujących badacza zmiennych i zjawisk. Silver [2012] wskazuje w tym

² „Had instead the models been fitted more appropriately to historic periods of stress, capital requirements would have been much higher and the financial world would be in far better shape today, in my judgment”, <https://www.wsj.com/articles/BL-REB-2187> (dostęp 2023-01-03).

kontekście na często występujący w praktyce błąd nadmiernego dopasowania modelu do rzeczywistości (*overfitting error*). Z jednej strony badacz dąży do jak najlepszego dopasowania modelu do opisywanej rzeczywistości i wykorzystuje do tego celu odpowiednie, dobrze znane statystykom miary dobroci dopasowania (m.in. współczynniki determinacji), z drugiej zaś – gdy zbyt dużą część zmienności zjawiska chce wyjaśnić modelem – naraża się na ryzyko pokrycia przez model nie tylko tego, co istotne – sygnału (trwałych prawidłowości w rozwoju zjawiska), lecz także szumu (losowej zmienności). Innymi słowy, nadmierne dążenie do maksymalizacji miar dopasowania modelu do zaobserwowanej w przeszłości rzeczywistości skutkować może uzyskaniem skomplikowanego analitycznie modelu, w którym przypadkowe, specyficzne dla danej próby czasowej wahania mieszać się będą z trwałymi tendencjami w rozwoju danego zjawiska. W konsekwencji model taki może się okazać mało użyteczny do prognozowania, gdyż zbyt silnie powiązany będzie z tym, co jednorazowe i nietrwale w przeszłości. Jeszcze poważniejsze konsekwencje stosowania w nauce tego typu nadmiernie dopasowanych modeli dostrzega Silver [2012, s. 167]: „Gdy model dopasowany jest także do szumu, ma on potencjał do powodowania szkód w nauce”.

Powszechność błędu nadmiernego dopasowania zdaje się współcześnie rosnać, na co wpływ mają coraz lepsze programy optymalizacyjne oraz inne zaawansowane oprogramowanie statystyczne, a także – co nie jest bez znaczenia – absolutyzowanie przez niektórych badaczy kryterium dobroci dopasowania w doborze modelu prognostycznego. Błąd ten występuje w praktyce częściej od błędu zbyt słabego dopasowania modelu (*underfitting*). Zwiększanie liczby parametrów modelu w celu poprawy jego dopasowania nie rodzi obecnie prawie żadnych problemów obliczeniowych. A do tego dostępność danych statystycznych także sprzyja dążeniu, by model był w stanie wyjaśnić niemal całą zmienność obserwowanego fragmentu rzeczywistości. W rezultacie dość często włączony zostaje do modelu również szum.

Skonstruowaniu i oszacowaniu adekwatnego modelu nie służy ponadto niepewność tkwiąca w samych danych, a ściślej wątpliwości, czy dostępne dane są zweryfikowane i uznane za ostateczne. Typową praktyką urzędów statystycznych wielu krajów na świecie jest poddawanie wcześniej ogłoszonych danych kontroli i rewizji. Dotyczy to w szczególności ogłaszanych z częstotliwością miesięczną bądź kwartalną danych ekonomicznych, np. zmian w PKB czy wysokości inflacji. Rewizja danych i wskaźników ekonomicznych bywa nieraz kilkukrotna i znacząca, co podważa wiarygodność modeli oszacowanych na podstawie wstępnych, utrzymujących się w obiegu niekiedy dość długo danych. Pierwsze oficjalne komunikaty w USA

informowały o spadku PKB w trzecim kwartale kryzysowego roku 2008 o 3,8%, podczas gdy ostatecznie spadek ten wyniósł prawie 9%. W okresie od roku 1965 do 2009 rządowe szacunki amerykańskiego PKB w ujęciu kwartalnym były korygowane przeciętnie o 1,7 pkt proc., por. [Silver, 2012; s. 194].

Prognozowanie bez modelu

O szumie i tendencyjności piszą też w swojej głównej książce noblista Daniel Kahneman i współautorzy [2022]. Jako specjalista z zakresu psychologicznych aspektów podejmowania decyzji, błędów poznawczych i heurystyk, jakimi się na co dzień posługujemy, D. Kahneman zwraca uwagę na zdolności jednostki i grupy osób do formułowania wysokiej jakości prognoz. Dylemat, czy modele, czy człowiek lepiej sprawdzają się w prognozowaniu, autor rozstrzyga jednoznacznie na podstawie własnych badań i analiz innych naukowców: „Ludzie są gorsi od modeli statystycznych pod wieloma względami, a jedną z ich najważniejszych słabości jest to, że cechuje ich szum” (s. 141). Czyli zmienność osądów, spowodowana wieloma czynnikami, które szczegółowo autorzy analizują, czyni nas bardziej chwiejnymi w prognozowaniu w stosunku do sformalizowanych modeli. Agregacja indywidualnych prognoz, ich uśrednianie lub grupowe dyskusowanie mogą część indywidualnych błędów zmniejszać. W szczególności uśrednianie niezależnych sądów przyczynia się do redukcji szumu. Ludzkie błędy poznawcze powodują bowiem, że część osób może przeszacowywać oczekiwane efekty w przyszłości, a część te same efekty niedoszacowywać. Mniej pewnym w tym kontekście rozwiązaniem jest grupowe dyskusowanie indywidualnych ocen i prognoz. W sytuacjach, gdy mało profesjonalne jest zarządzane taką grupą dyskusyjną, może się okazać, że dyskusje te „dodają więcej błędu w formie tendencyjności (*bias*), niż redukują w formie szumu” [Kahneman i in., 2012; s. 320]. W dyskusjach takich perswazja silnych jednostek lub postrzeganych jako autorytety może skłonić innych uczestników do jednokierunkowej zmiany swoich pierwotnych osądów. Dlatego w przypadku wcześniej wspomnianego mechanizmu uśredniania ocen ważne jest, aby były one formułowane w pełni niezależnie.

Kahneman et al. [2022] przywołują także badania wskazujące na to, że dowolny model liniowy, jeżeli stosowany jest konsekwentnie, radzi sobie przeciętnie lepiej niż ludzie formułujący osądy prognostyczne. Nie oznacza to jednak – co podkreślają autorzy – że dowolny model pokona człowieka w każdym wyzwaniu prognostycznym (s. 146). Istnieją bowiem osoby – nazwane przez autorów „superprognostami” (*superforecasters*) – które cechuje wysoki poziom „aktywnej otwartości umysłowej” (*active open-minded thinking*) i które osiągają wyjątkowo dobre rezultaty

w prognozowaniu. Jednym z wyróżników wspomnianej otwartości umysłowej jest zwracanie uwagi na dane, które są sprzeczne z przekonaniami danej osoby. A więc bez emocjonalnego napięcia osoby o tych przymiotach są gotowe zmienić lub zaktualizować własne osądy pod wpływem nowej informacji. Nie jest to, przynajmniej, ani łatwe, ani powszechne. W podsumowaniu autorzy stwierdzają, że osoby o dużych zdolnościach prognostycznych są przywiązane do następującego cyklu myślenia: „spróbuj, ponieś porażkę, przeanalizuj, wprowadź poprawki i spróbuj jeszcze raz” (s. 315). Cykl ten nie jest niczym innym jak dobrze znanym matematykom i statystykom sposobem aktualizacji wiedzy w oparciu o twierdzenie Bayesa. W rzeczywistości więc prognozowanie takie nie jest w pełni niezależne od modelu, który do prognozowania, podobnie jak człowiek, wprowadza szum.

Bibliografia

- Aven T. [2014]. *Risk, Surprises and Black Swans: Fundamental Ideas and Concepts in Risk*, New York: Routledge.
- Kahneman D., Sibony O., Sunstein C.R. [2022]. *Szum, czyli skąd się biorą błędy w naszych decyzjach*, tłum. P. Szymczak, Poznań: Wydawnictwo Media Rodzina.
- Silver N. [2012]. *The Signal and the Noise. The Art and Science of Prediction*, Penguin Books.
- Szreder M. [2022a]. Tendencyjność i rozproszenie osądów. Refleksje na temat książki D. Kahnemana, O. Sibony’ego i C.R. Sunsteina pt. *Szum, czyli skąd się biorą błędy w naszych decyzjach*, Media Rodzina, 2022. *Biuletyn PTE*, nr 3(98), s. 69–74.
- Szreder M. [2022b]. Noise and bias – some controversies raised by the book ‘Noise: A Flaw in Human Judgment’, written by Daniel Kahneman, Olivier Sibony, Cass R. Sunstein. *Przegląd Statystyczny*, nr 1(69). ■