METODY STATYSTYCZNE W FINANSACH

Krzysztof Jajuga
Akademia Ekonomiczna we Wrocławiu, Katedra Inwestycji Finansowych i Ubezpieczeń

Wprowadzenie

W ostatnich kilkunastu latach na świecie obserwuje się dynamiczny rozwój narzędzi teoretycznych finansów. Dotyczy to przede wszystkim tych narzędzi, które mają u podstaw zaawansowane metody matematyczne, w tym metody statystyczne. Efektywne zastosowanie tych metod stało się możliwe po pierwsze dzięki rozwojowi efektywnych narzędzi informatycznych (*hardware* i *software*), po drugie dzięki pojawieniu się zbiorów danych finansowych (np. długich szeregów czasowych), w odniesieniu do których metody te można stosować. Z jednej strony, rozwój metod statystycznych i wykorzystanie technologii informatycznych umożliwiły rozwiązanie skomplikowanych problemów praktycznych z zakresu finansów, z drugiej strony wyzwania praktyki finansów spowodowały powstanie i rozwój niektórych metod statystycznych.

W opracowaniu dokonujemy bardzo syntetycznego przeglądu praktycznych problemów finansowych, w których kluczową rolę odgrywa zastosowanie metod statystycznych, wskazując na te metody. Na wstępie jednak poczynimy uwagę terminologiczną, dotyczącą rozróżnienia między metodami statystycznymi i metodami ekonometrycznymi.

Różnica między metodami statystycznymi i ekonometrycznymi dotyczy tu przypadku metod wywodzących się z podejścia stochastycznego. Te metody mają u podstaw jedno z dwóch pojęć:

- rozkład statystyczny;
- proces stochastyczny.

W pierwszym pojęciu wymiar czasu jest do pewnego stopnia ignorowany i główny nacisk położony jest na analizę **struktury** zbioru danych. Zwyczajowo te metody nazywa się **metodami statystycznymi**. W drugim pojęciu główny nacisk położony jest na analizę **dynamiki**, a znacznie mniejszy (jeśli w ogóle) na strukturę. Zwyczajowo te metody nazywa się **metodami ekonometrycznymi**.

W tym opracowaniu pojęcie „metody statystyczne” jest rozumiane szerzej, gdyż oprócz klasycznych metod statystycznych obejmuje również metody ekonometryczne.

Dodajmy jeszcze, że najpełniejsza struktura danych finansowych, jaka może być analizowana za pomocą metod statystycznych, jest to **kostka danych**. Kostka ta jest niczym innym jak wielowymiarowym szeregiem przekrojowo-czasowym i ma 3 wymiary:

- Obiekty (w sumie n);
- Czas (w sumie N);
- Zmienne (w sumie m).

Ogólnie zatem mamy do czynienia z N x n x m zmiennymi losowymi, przy czym dysponuje się tylko jedną obserwacją dla zmiennej losowej, dlatego też czyni się pewne uproszczenia w celu umożliwienia przeprowadzania wnioskowania statystycznego.

Analiza problemów praktycznych występujących w finansach, w których kluczową rolę odgrywa zastosowanie metod statystycznych, pozwala na wyodrębnienie następujących grup zagadnień:

- analiza i modelowanie finansowych szeregów czasowych;
- analiza dochodu i ryzyka inwestycji finansowych;
- modelowanie zależności między zmiennymi finansowymi;
- modelowanie ryzyka rynkowego za pomocą koncepcji Value at Risk (VaR);
- modelowanie ryzyka kredytowego.

Zagadnienia te są syntetycznie omówione w kolejnych częściach opracowania. Jest to oczywiście niepełna lista zagadnień, obejmująca najważniejsze problemy, w których obserwuje się największy rozwój metod.

Analiza i modelowanie finansowych szeregów czasowych

Jest pewna liczba rodzajów finansowych szeregów czasowych, które zazwyczaj podlegają modelowaniu. Można je podzielić ze względu na dwa kryteria.

1. Zjawisko finansowe.

Wyróżniamy tutaj:

- szeregi czasowe stóp procentowych;
- szeregi czasowe kursów walutowych;
- szeregi czasowe cen akcji;
- szeregi czasowe wartości indeksów giełdowych;
- szeregi czasowe cen towarów giełdowych (energii, metali, produktów rolnych itp.).

2. Modelowana charakterystyka.

Wyróżniamy tutaj następujące szeregi czasowe:

- szeregi czasowe cen *spot* (cen natychmiastowych);

- szeregi czasowe cen *forward* (cen terminowych);
- szeregi czasowe stóp zwrotu;
- szeregi czasowe zmienności.

Różnice między tymi szeregami najprościej jest przedstawić na przykładzie, w którym modelowanym zjawiskiem finansowym jest kurs walutowy.

1. Szereg czasowy kursu walutowego *spot* (natychmiastowego).

Jest to szereg, którego elementami są wartości kursu walutowego w transakcjach walutowych dokonywanych obecnie.

2. Szereg czasowy kursu walutowego *forward* (terminowego).

Jest to szereg, którego elementami są wartości kursu walutowego w transakcjach terminowych, tzn. takich, które są dokonywane w przyszłości (np. za trzy miesiące), lecz cena ustalana jest dziś (oczywiście z reguły ta cena różni się od ceny natychmiastowej).

3. Szereg czasowy stóp zwrotu z rynku walutowego.

Jest to szereg, którego elementami są stopy zwrotu, określane na podstawie cen (zazwyczaj pod uwagę bierze się jedynie ceny *spot*). Stopa zwrotu wyrażona jest w tym wypadku w skali pewnego okresu (dnia, miesiąca, roku) i określa procentowy dochód z transakcji polegającej na zakupie waluty na początku okresu i sprzedaży waluty na koniec okresu. Przy tym stosowane są dwie konwencje wyznaczania stopy zwrotu:

- prosta stopa zwrotu, dana wzorem:

$$R_t = \frac{P_{t+1}}{P_t} - 1$$

- logarytmiczna stopa zwrotu, dana wzorem:

$$R_t = \ln P_{t+1} - \ln P_t$$

Przy tym R oznacza stopę zwrotu, zaś P oznacza kurs walutowy, natomiast indeksy oznaczają kolejne momenty czasowe.

4. Szereg czasowy zmienności kursu walutowego.

Jest to szereg, którego elementami są parametry zmienności – zazwyczaj odchylenia standardowe wyznaczone w odniesieniu do stóp zwrotu (prostych lub logarytmicznych) z rynku walutowego. Oznacza to, że dla każdego momentu wyznaczane jest jedno odchylenie standardowe, zazwyczaj na podstawie danych z okresu poprzedzającego ten moment. Każda z tych wartości może być interpretowana jako przeciętna zmienność stóp zwrotu z transakcji walutowych.

Z powyższych rozważań wynika, że istnieją różne rodzaje szeregów czasowych, które podlegają modelowaniu. Najczęściej modelowaniu podlegają pojedyncze szeregi czasowe. Każdy z nich może być analizowany za pomocą wielu różnych metod statystycznych.

Analiza tych różnych możliwych metod doprowadziła nas do wyróżnienia następujących podstawowych grup:

- proste metody analizy dynamiki;
- metody wywodzące się z koncepcji procesu stochastycznego;
- metody sieci neuronowych.

Należy również zwrócić uwagę, iż w finansach metody te zazwyczaj stosowane są w celach prognostycznych.

1. Proste metody analizy dynamiki.

Do tej grupy zalicza się proste metody statystyczne, których głównym celem jest analiza dynamiki. Będą to przeto głównie różnego rodzaju modele trendu (liniowe i nieliniowe), w tym również modele adaptacyjne. Środowisko finansowe obecnie sięga po zdecydowanie bardziej zaawansowane metody niż metody należące do tej grupy. Wydaje się jednak, że nie należy zapominać o tych prostych metodach, gdyż czasem mogą one prowadzić do dobrych rezultatów z punktu widzenia skuteczności prognozowania.

2. Metody wywodzące się z koncepcji procesu stochastycznego.

Metody te w ostatnich kilkunastu latach zyskały bardzo dużą popularność w zastosowaniach finansowych. Punktem wyjścia w rozwoju tych metod jest koncepcja procesu stochastycznego w czasie dyskretnym. Liczba metod, a właściwie modeli, należących do tej grupy jest bardzo duża. Jak się okazuje, istotna część proponowanych modeli może być zapisana w następującej ogólnej postaci (por. Tsay (2002)):

$$X_t = g(F_{t-1}) + \sqrt{h(F_{t-1})}\varepsilon_t$$
$$g(F_{t-1}) = \mu_t = E(X_t | X_{t-1},....)$$
$$h(F_{t-1}) = \sigma_t^{\;2} = V(X_t | X_{t-1},....)$$

W tej ogólnej postaci model składa się z dwóch części: warunkowej średniej oraz warunkowej wariancji (ściślej: warunkowego odchylenia standardowego). Termin „warunkowy” oznacza tutaj, że analizowany jest rozkład zmiennej losowej, będącej częścią składową procesu stochastycznego, w momencie *t*, pod warunkiem rozkładu w przeszłych momentach.

Jeśli weźmiemy pod uwagę fakt, że każda część składowa może być modelem liniowym lub nieliniowym, wówczas otrzymujemy następujące grupy:

- Modele liniowe średniej i liniowe wariancji; ta grupa zawiera takie modele, jak: stacjonarne modele szeregów czasowych (ARMA), niestacjonarne modele zintegrowanych szeregów czasowych (ARIMA), modele sezonowe (SARIMA), niestacjonarne modele ułamkowo zintegrowanych szeregów czasowych (ARFIMA), modele trendowo-niestacjonarne itp.
- Modele liniowe średniej i nieliniowe wariancji; ta grupa zawiera modele wymienione powyżej, w których część składowa wariancji jest modelowana osobno, za pomocą takich modeli, jak: ARCH, GARCH, SV, CHARMA, GARCH-M itp.

- Modele nieliniowe średniej i nieliniowe wariancji; ta grupa zawiera m.in. modele: TAR, STAR, SETAR, dwuliniowe, przełącznikowe modele Hamiltona itp.

Najpopularniejszym modelem spośród wyżej wymienionych jest ten, w którym warunkowa średnia modelowana jest za pomocą procesu autoregresji, zaś warunkowa wariancja za pomocą procesu GARCH. Szczegółowy przegląd modeli wywodzących się z koncepcji procesu stochastycznego zawiera m.in. praca Tsaya (2002).

3. Metody sieci neuronowych.

Jest to klasyczne zastosowanie sieci neuronowych, w którym zmienna na wyjściu sieci neuronowej to zmienna w okresie *t*, zaś zmienne na wejście są to zmienne z przeszłych okresów. Zadaniem sieci neuronowej jest wykorzystanie danych historycznych zawartych w szeregu czasowym do aproksymacji obecnej wartości zmiennej za pomocą przeszłych wartości tej zmiennej.

Analiza dochodu i ryzyka inwestycji finansowych

Wprowadzenie

Analiza dochodu i ryzyka jest to klasyczne zagadnienie finansowe, które zostało sformułowane na gruncie teorii portfela (por. Markowitz (1952)). W sposób najbardziej ogólny zagadnienie to może być opisane następująco:

Należy stworzyć portfel instrumentów finansowych, np. akcji, w taki sposób, że dochód z tego portfela jest jak największy, zaś ryzyko tego portfela jest jak najmniejsze.

Przy tym z reguły ryzyko rozumiane jest jako możliwość osiągnięcia dochodu różniącego się od spodziewanego dochodu.

Zagadnienie teorii portfela jest to zatem zagadnienie decyzyjne, wyrażone w dwóch następujących problemach, rozwiązywanych za pomocą metod optymalizacji warunkowej:

- minimalizacja ryzyka, przy zadanym poziomie dochodu;
- maksymalizacja dochodu, przy zadanym poziomie ryzyka.

Kluczową rolę w tych zagadnieniach odgrywa określenie dochodu i ryzyka. To właśnie tutaj stosowane są metody statystyczne, konkretnie metody analizy rozkładu statystycznego. Metody te stanowią istotną część analizy portfelowej, czyli praktycznego zastosowania teorii portfela na rynku finansowym.

Rozkład statystyczny, który jest analizowany, jest to rozkład stóp zwrotu (np. stóp zwrotu akcji), przy czym mogą to być zarówno proste stopy zwrotu, jak i logarytmiczne stopy zwrotu. W praktyce często rozkład ten określany jest na podstawie szeregów czasowych stóp zwrotu, co oznacza, że tak naprawdę zakładamy implicite, że obserwacje pochodzące z kolejnych momentów można traktować jako pochodzące z populacji o tym samym rozkładzie. W praktyce o tym założeniu się z reguły nie mówi, przechodząc nad tym do porządku dziennego.

W zagadnieniu portfela mamy do czynienia z rozkładem wielowymiarowym, jest to rozkład stóp zwrotu składników (np. akcji) wchodzących w skład portfela. Jeśli na przykład mamy do czynienia z portfelem złożonym z akcji 10 spółek, wówczas oczywiście rozpatrujemy rozkład 10-wymiarowy. Jednak na początku rozpatrzymy przypadek jednowymiarowy, tzn. taki, w którym każdy składnik portfela analizowany jest osobno.

Analiza jednowymiarowego rozkładu stopy zwrotu

Rozkład stopy zwrotu jest podstawą do określenia dochodu i ryzyka. Klasyczne podejście (zaproponowane jeszcze przez Markowitza) jest tu bardzo proste i naturalne, mianowicie:

- dochód określony jest jako oczekiwana stopa zwrotu, czyli wartość oczekiwana rozkładu stopy zwrotu;
- ryzyko określone jest jako odchylenie standardowe (ewentualnie wariancja) stopy zwrotu, czyli odchylenie standardowe rozkładu stopy zwrotu.

Pozwala to na sformułowanie następującej zasady.

Wyznaczanie charakterystyk dochodu i ryzyka w klasycznym podejściu sprowadza się do estymacji wartości oczekiwanej i odchylenia standardowego rozkładu stopy zwrotu.

W najprostszym zatem ujęciu:

- miarą dochodu jest średnia arytmetyczna stóp zwrotu;
- miarą ryzyka jest odchylenie standardowe stóp zwrotu.

Takie ujęcie jest oczywiście uzasadnione w najprostszej sytuacji, która może wystąpić, mianowicie takiej, gdy rozkład stóp zwrotu jest normalny. Wtedy wystarczy zwykła estymacja wartości oczekiwanej i odchylenia standardowego. W praktyce jednak jest to raczej rzadziej spotykana sytuacja.

Zatem zagadnienie określania dochodu i ryzyka należy traktować bardziej ogólnie, co prowadzi do następującej zasady:

Wyznaczanie charakterystyk dochodu i ryzyka sprowadza się do estymacji parametru położenia (miara dochodu) i parametru skali (miara ryzyka) rozkładu stopy zwrotu.

W takim przypadku jako miarę dochodu można stosować dowolny estymator parametru położenia, np. medianę stóp zwrotu, modalną stóp zwrotu, czy też tzw. punkt środkowy (*midrange* – średnia arytmetyczna wartości maksymalnej i minimalnej). Z punktu widzenia teorii finansów bardzo istotną rolę odgrywa również tzw. geometryczna stopa zwrotu, dana wzorem:

$$R = [(1+R_1)(1+R_2)...(1+R_N)]^{1/N} - 1$$

Podobnie jako miarę ryzyka można stosować dowolny estymator parametru skali (rozrzutu), np. odchylenie przeciętne, odchylenie międzykwartylowe (lub odchylenie ćwiartkowe), rozstęp, średnie odchylenie od mediany, czy też medianę bezwzględnych odchyleń od mediany.

Analiza inwestycji finansowych nie poprzestaje na analizie dochodu i ryzyka, podobnie jak analiza rozkładów statystycznych nie poprzestaje na analizie parametrów położenia i parametrów skali. Jak wiadomo, w analizie rozkładu statystycznego często interesuje nas również analiza skośności i analiza spłaszczenia. Jak się okazuje, takie podejście ma uzasadnienie w analizie inwestycji finansowych, gdyż rozkłady stóp zwrotu nierzadko charakteryzują się skośnością oraz są leptokurtyczne (występują w nich grube ogony). Dodatkowo może do tego dojść analiza obserwacji nietypowych (*outliers*).

Oznacza to, że analiza inwestycji finansowych przeprowadzana metodami analizy rozkładu statystycznego obejmuje z reguły następujące zadania:

- estymacja parametrów położenia;
- estymacja parametrów skali (rozrzutu, rozproszenia);
- estymacja parametrów skośności;
- estymacja parametrów spłaszczenia (kurtozy);
- analiza obserwacji nietypowych.

Analiza wielowymiarowego rozkładu stóp zwrotu

W zagadnieniu tworzenia portfela bardzo ważną rolę odgrywa jednak rozkład wielowymiarowy, gdzie liczba wymiarów jest równa liczbie składników (np. spółek) portfela. Wynika to z następujących podstawowych zasad stosowanych w klasycznej teorii portfela:

- dochód portfela określony jest jako oczekiwana stopa zwrotu portfela, czyli ważona średnia oczekiwanych stóp zwrotu składników portfela, które z kolei są składowymi wektora średnich rozkładu wielowymiarowego stóp zwrotu;
- ryzyko portfela określone jest jako odchylenie standardowe (ewentualnie wariancja) stopy zwrotu portfela, które jest liniową funkcją elementów macierzy kowariancji rozkładu wielowymiarowego stóp zwrotu (odchyleń standardowych oraz współczynników korelacji).

Takie ujęcie jest oczywiście uzasadnione w najprostszej sytuacji, która może wystąpić, mianowicie takiej, gdy rozkład wielowymiarowy stóp zwrotu jest wielowymiarowym rozkładem normalnym, lub szerzej: gdy rozkład ten należy do klasy wielowymiarowych rozkładów eliptycznie symetrycznych. Wtedy wystarczy klasyczna estymacja wektora średnich i macierzy kowariancji. W praktyce jednak ta sytuacja może występować niezbyt często.

Zatem zagadnienie określania dochodu i ryzyka portfela należy traktować bardziej ogólnie, co prowadzi do następującej zasady:

Wyznaczanie charakterystyk dochodu i ryzyka portfela sprowadza się do estymacji wektora położenia i macierzy rozrzutu wielowymiarowego rozkładu stopy zwrotu.

Oczywiście do tego mogą dojść bardziej skomplikowane zagadnienia analizy rozkładu, takie jak: analiza skośności wielowymiarowej, analiza kurtozy wielowymiarowej, analiza wielowymiarowych obserwacji nietypowych, w odniesieniu do których stosuje się mniej znane narzędzia statystyczne.

Modelowanie zależności między zmiennymi finansowymi

Innym, dość szerokim zagadnieniem finansowym rozwiązywanym za pomocą metod statystycznych, jest modelowanie zależności między zmiennymi finansowymi. W sposób bardzo ogólny zagadnienie to można zapisać następująco:

$$Y = f(X_1, X_2, ..., X_m, \varepsilon)$$

Jak widać zatem jest to klasyczne zagadnienie modelowania zależności jednej zmiennej (zwanej objaśnianą) od zbioru innych zmiennych (zwanych objaśniającymi) z uwzględnieniem składnika losowego. Oznacza to, że właściwie można tu stosować dowolne metody analizy regresji, ewentualnie metody sieci neuronowych. Z punktu widzenia zagadnień finansowych istotne jest natomiast to, że:

- najczęściej jako zmienna objaśniana występuje cena lub stopa zwrotu;
- jako zmienne objaśniające często występują zmienne, które można interpretować jako czynniki ryzyka;
- pochodne cząstkowe zmiennej objaśnianej względem zmiennych objaśniających interpretowane są jako współczynniki wrażliwości;
- ze względów interpretacyjnych najczęściej przyjmowana jest liniowa postać funkcji określającej zależność.

Przedstawimy teraz kilka modeli finansowych, stosowanych często w praktyce, w których zastosowanie ma przedstawiona powyżej funkcja.

1. Model jednowskaźnikowy Sharpe'a i współczynnik beta.

Model Sharpe'a ma następującą postać (por. Sharpe (1963)):

$$R = \alpha + \beta R_M + \varepsilon$$

W modelu tym zmienną objaśnianą jest stopa zwrotu akcji spółki, zaś zmienną objaśniającą stopa zwrotu indeksu rynku, zazwyczaj indeksu giełdowego. W modelu tym współczynnikiem kierunkowym jest właśnie współczynnik beta. Wskazuje on, o ile (w przybliżeniu) zmieni się stopa zwrotu akcji (lub portfela akcji), gdy stopa zwrotu indeksu rynku wzrośnie o jednostkę (jeden punkt procentowy).

2. Model wyceny arbitrażowej APT.

Model ten, zaproponowany przez Rossa (por. Ross (1976)), ma następującą postać:

$$R = \beta_0 + \beta_1 F_1 + \beta_2 F_2 + ... + \beta_m F_m + \varepsilon$$

W tym modelu zmienną objaśnianą jest stopa zwrotu akcji spółki, zaś zmiennymi objaśniającymi pewne inne zmienne, zwane czynnikami ryzyka. Współczynniki przy zmiennych objaśniających są to oczywiście współczynniki wrażliwości. Określają, jak stopa zwrotu akcji reaguje na zmiany poszczególnych czynników ją kształtujących, przy założeniu, że pozostałe czynniki się nie zmieniają.

Przy wyznaczaniu powyższego modelu możliwe są dwie sytuacje:

- czynniki ryzyka zostały zidentyfikowane za pomocą wiedzy merytorycznej – wówczas mamy do czynienia ze zwykłym modelem regresji;
- czynniki ryzyka nie są znane – wówczas stosowane są metody analizy czynnikowej; należy jednak zaznaczyć, iż wtedy mogą się pojawić kłopoty z interpretacją otrzymanych wyników.

3. Model liniowy współczynnika zabezpieczenia dla kontraktu terminowego.

Współczynnik zabezpieczenia dla kontraktu terminowego (*hedge ratio*) jest również miarą wrażliwości. Współczynnik ten w zagadnieniach finansowych stosowany jest w przypadku zabezpieczania instrumentu finansowego (np. akcji) kontraktem terminowym (zwłaszcza kontraktem *futures*). Omawiany model liniowy, w którym współczynnik zabezpieczenia jest dany jako współczynnik kierunkowy, ma następującą postać:

$$S = \alpha + \beta F + \varepsilon$$

gdzie:

S – cena instrumentu podstawowego (cena *spot*),

F – cena kontraktu terminowego (cena *futures*).

Z powyższego wzoru wynika, iż współczynnik zabezpieczenia wskazuje, o ile (w przybliżeniu) zmieni się cena instrumentu podstawowego, gdy cena kontraktu terminowego wzrośnie o jednostkę. Może on być estymowany za pomocą standardowej procedury, tzn. analizy regresji zastosowanej w odniesieniu do historycznych cen instrumentu podstawowego i historycznych cen kontraktu terminowego.

Modelowanie ryzyka rynkowego za pomocą koncepcji Value at Risk (VaR)

Przy analizie dochodu i ryzyka inwestycji finansowych wskazywaliśmy już, iż ryzyko inwestycji może być mierzone za pomocą parametrów rozrzutu rozkładu statystycznego stóp zwrotu. Obecnie przedstawimy zbliżone zagadnienie, w którym rozpatrywany jest pomiar ryzyka rynkowego, na które narażony jest podmiot gospodarczy (np. instytucja finansowa bądź przedsiębiorstwo). Przy tym ryzyko rynkowe jest to ryzyko wynikające ze zmian wartości pewnego indeksu ryzyka, np. zmian cen. Miary ryzyka rynkowego mają u podstaw właśnie rozkład statystyczny indeksu ryzyka. Są to (podobnie jak przy analizie ryzyka inwestycji finansowych) miary rozrzutu lub tzw. miary zagrożenia. Określa się je również na podstawie rozkładu statystycznego, jako funkcje kwantyli (statystyk porządkowych) rozkładu.

Są to przede wszystkim miary wywodzące się z koncepcji Value at Risk (VaR) – tzw. wartości zagrożonej. Wartość zagrożona jest to miara potencjalnej straty (zmniejszenia wartości). Jest ona określona jako możliwa strata (zmniejszenie wartości), która powstanie w danym okresie – przy zadanym z góry poziomie ufności (zazwyczaj 95% lub 99%).

Formalnie Value at Risk określa się za pomocą wzoru:

$$P(W \leq W_0 - VaR) = \alpha$$

gdzie:

W_0 – obecna wartość instytucji;

W – wartość instytucji na końcu rozpatrywanego okresu, formalnie jest to zmienna losowa;

α – poziom tolerancji (prawdopodobieństwo bliskie 0, z reguły 0,01 lub 0,05);

1-α – poziom ufności (prawdopodobieństwo bliskie 1, z reguły 0,99 lub 0,95).

Z powyższego wzoru wynika, że wartość zagrożona jest niczym innym, jak różnicą między obecną wartością, a kwantylem wyznaczonym dla rozkładu wartości. Dodajmy tu jeszcze, że do wyznaczania Value at Risk można wykorzystać rozkład stopy zwrotu (zamiast rozkładu wartości).

Z tych rozważań wynika również, iż do wyznaczenia VaR niezbędne jest stosowanie metod wyznaczania kwantyli rozkładu statystycznego.

Modelowanie ryzyka kredytowego

Ryzyko kredytowe jest drugim, obok ryzyka rynkowego, podstawowym rodzajem ryzyka finansowego. Jest ono określane jako ryzyko wynikające z możliwości niedotrzymania warunków przez drugą stronę kontraktu finansowego. Najprostszym (ale nie jedynym) przypadkiem jest ryzyko wynikające z faktu, iż podmiot zaciągający kredyt może nie spłacić kwoty tego kredytu i nie zapłacić odsetek.

Metody statystyczne odgrywają bardzo istotną rolę w pomiarze i modelowaniu ryzyka kredytowego. Ogół metod, a właściwie modeli stosowanych w tym zagadnieniu można podzielić na dwie grupy:

- modele niedotrzymania warunków (*default models*), które mają za zadanie ocenę prawdopodobieństwa niedotrzymania warunków bądź przydzielenie ocenianego podmiotu do konkretnej klasy odzwierciedlającej możliwość niedotrzymania warunków przez ten podmiot;
- modele rynkowe (*market models*), które mają za zadanie oszacowanie straty wynikającej z możliwego niedotrzymania warunków.

W tym opracowaniu zajmujemy się jedynie tą pierwszą grupą modeli, gdyż w tej chwili w Polsce jest ona zdecydowanie częściej stosowana. Do podstawowych metod zaliczanych do tej grupy należą:

- metody skoringowe;
- metody analizy dyskryminacyjnej;
- metody sieci neuronowych.

1. Metody skoringowe.

W tych metodach na początku określa się (na podstawie wiedzy merytorycznej) czynniki determinujące prawdopodobieństwo niedotrzymania warunków. Następnie wybiera się pewną funkcję, w której zmiennymi objaśniającymi są te czynniki, natomiast zmienną objaśnianą jest taka zmienna zero-jedynkowa, której wartości przyjmują wartość 0 w przypadku niedotrzymania warunków, zaś 1 w wypadku dotrzymania warunków. Jest kilka możliwych metod skoringowych, przy czym tak naprawdę można tu stosować każdy model regresji, w którym zmienna objaśniana jest zmienną zero-jedynkową. Najprostszy jest **model liniowego prawdopodobieństwa** następującej postaci:

$$Z = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m + \beta_0 + \varepsilon$$

gdzie:

Z – zmienna zero-jedynkowa, określająca niedotrzymanie warunków;

X_i – i-ty czynnik determinujący niedotrzymanie warunków.

Okazuje się, że wartość oczekiwana zmiennej po lewej stronie tego modelu jest równa prawdopodobieństwu niedotrzymania warunków, co jest zaletą tego modelu. Jednak jest też istotna wada: prawdopodobieństwa otrzymane w wyniku zastosowania tego modelu mogą przyjąć wartość spoza przedziału [0;1].

Drugim często stosowanym modelem jest **model logitowy**, pozbawiony wyżej wymienionej wady. W tym modelu zastosowana jest zmienna znormalizowana, a postać modelu jest następująca:

$$Z^* = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m + \beta_0 + \varepsilon$$
$$Z^* = (1 + e^{-Z})^{-1}$$

Oczywiście można stosować również inne modele z dyskretną zmienną objaśnianą, jak model probitowy.

Wszystkie omawiane modele szacowane są z reguły za pomocą klasycznych procedur analizy regresji na podstawie danych z przeszłości dotyczących podmiotów, które dotrzymywały i nie dotrzymywały warunków umowy.

2. Analiza dyskryminacyjna.

Jak wiadomo, analiza dyskryminacyjna polega na przydzieleniu obiektów należących do tzw. próby rozpoznawanej do z góry zdefiniowanych klas, przy czym klasy te określone są na podstawie danych z przeszłości, zawartych w tzw. próbie uczącej. Metody analizy dyskryminacyjnej, standardowo zaliczane do metod statystycznej analizy wielowymiarowej, mogą być przeto wykorzystane do oceny ryzyka kredytowego, poprzez wyznaczenie dwóch klas: jednej zawierającej podmioty niedotrzymujące warunków umowy, zaś drugiej zawierającej podmioty dotrzymujące warunków (możliwe jest uwzględnienie jeszcze trzeciej klasy, neutralnej, zawierającej podmioty, co do których nie można podjąć jednoznacznej decyzji). Jednym z pierwszych przykładów zastosowania analizy dyskryminacyjnej w ocenie ryzyka kredytowego jest znany model Altmana (por. Altman (1968)).

3. Sieci neuronowe.

W tych metodach wykorzystana jest dokładnie ta sama idea co w analizie dyskryminacyjnej. Jednak jako funkcja dyskryminacyjna stosowana jest tu (jak zwykle w sieciach neuronowych) bardzo skomplikowana funkcja nieliniowa, będąca funkcją aproksymującą na podstawie danych z przeszłości zależność między niedotrzymaniem warunków umowy przez dany podmiot (jest to zmienna zero-jedynkowa na wyjściu sieci neuronowej) od czynników wpływających na ten fakt (są to zmienne na wejściu sieci neuronowej). Czasami zmienną na wyjściu może być również prawdopodobieństwo niedotrzymania warunków.

Literatura

1. Altman E.I. (1968), Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, Journal of Finance, 23, s. 589-609.
2. Markowitz H., Portfolio selection, Journal of Finance, 7, s. 77-91.
3. Ross S.A. (1976), The arbitrage theory of capital asset pricing. Journal of Economic Theory, 13, s. 341-360.
4. Sharpe W. (1963), A simplified model for portfolio analysis, Management Science, 19, s. 277-293.
5. Tsay R.S. (2002), Analysis of financial time series, Wiley, New York.