Modelowanie ekonometryczne

Kamil Skoczylas

Kamilskoczylas@wp.pl

1. Wstęp

Otaczający nas świat to zbiór różnych zjawisk. W zależności od zainteresowań człowiek staje się obserwatorem niektórych z nich. Zjawiska fizyczne czy chemiczne z reguły mogą być opisane za pomocą jednego równania, które uznawane jest za aksjomat. Modele ekonometryczne różnią się znacząco od równań, ponieważ bardzo rzadko w stu procentach pokrywają się z rzeczywistością. Wyrażane mogą być za pomocą różnych postaci funkcyjnych, jednak najczęściej spotykana jest postać liniowa. Popularność tego podejścia opisu wynika głównie z prostoty interpretacji. Ekonometria korzysta z dorobku innych nauk takich jak statystyka, informatyka czy matematyka. Przedrostek „eko" został przyjęty ze względu na miejsce, w którym dany zbiór technik opisywania rzeczywistości jest stosowany czyli ekonomii. Zjawisko które jest opisywane nazywane jest zmienna zależną bądź objaśnianą. Składniki modelu, które opisują daną rzecz znane są jako zmienne niezależne bądź objaśniające. Przy obecnych technikach predykcji zmiennych częściej używane są te drugie terminy, ponieważ wartości na podstawie których są określane zmienne objaśniane mogą być modelowane.

2. Pojęcia używane w modelowaniu ekonometrycznym

Korelacja - często nazywana współzależnością, jest główną miarą zależności zmiennej objaśnianej od zmiennej objaśniającej. Najpopularniejszym narzędziem jest Współczynnik korelacji liniowej Pearsona, jego wartość zależy od kowariancji oraz odchyleń standardowych zmiennych, które są badane. Korelacja jest wykorzystywana przy budowie macierzy korelacji, która daje ogólny obraz zależności występujących pomiędzy wektorem zmiennej objaśnianej a macierzą zmiennych objaśniających

Współczynnik Pearsona jest wyrażony wzorem:

$$r_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Autokorelacja – funkcja przypisująca wartość współczynnika korelacji Pearsona dla szeregu czasowego opóźnionego o k jednostek czasu względem tego samego szeregu czasowego. Na jej podstawie możemy stwierdzić występowanie zależności wartości obecnej od wartości, która wystąpiła wcześniej. Jest to jedno z podstawowych narzędzi wykorzystywanych przy budowie modeli dynamicznych, których główną cechą jest występowanie opóźnień w szeregu czasowym.

Regresja – metoda szacowania wartości oczekiwanej zmiennej objaśnianej za innych zmiennych. Estymacja (parametrów regresji) to element wnioskowania statystycznego, która pozwala uzyskiwać uogólnione wyniki z badania próby losowej. Wyniki zazwyczaj różnią się od wartości rzeczywistych, a różnice w wartościach nazywa się resztami. We wzorze funkcji opisującej jakąś zależność reszty są wyrażone jako czynnik losowy. Najczęściej wykorzystywaną metodą regresji jest metoda najmniejszych kwadratów, która działa w taki sposób, aby reszty modelu wynikające z różnic pomiędzy wartością teoretyczną a wartością rzeczywistą były minimalne. Reszty są podnoszone do drugiej potęgi, aby suma nie została zaburzona przez redukcję wyrazów o przeciwnych znakach.

$$S = \sum_{i=1}^{n} {e_i}^2 = min$$

Autoregresja – metoda prognozowania statystycznego, w której wartość szeregu czasowego X_{n+1} zależy od wartości przeszłych tego samego szeregu wraz ze współczynnikami modelu wynikającymi z regresji, określanych jako $a_0X_n + a_0X_{n-1} + \cdots + a_kX_{n-k} + c + \varepsilon$. Wyraz wolny c może świadczyć o obecności trendu, gdy $c \neq 0$. ε jest to błąd modelu.

Proces stochastyczny- jest to funkcja zmiennych losowych(X) oraz nielosowego argumentu, nazwijmy , którym jest czas(t). Funkcję tę można więc zapisać jako X_t[1].

Średnia ruchoma – również metoda predykcyjna, kryjąca w swojej nazwie prostą koncepcję. Wybór rodzaju średniej zależy od badacza, najczęściej wykorzystywana jest średnia arytmetyczna. Średnia jest liczona dla określonej liczby obserwacji i przesuwana po szeregu czasowym, tak aby ilość obserwacji wchodzących do średniej pozostała taka sama. Zazwyczaj szereg jest dzielony na próbę uczącą i testową w celu weryfikacji odchyleń od wartości rzeczywistych. Metodę tę czasem nazywa się „Rolling Window".

---

[1] Osińska M. Ekonometria finansowa, wydawnictwo PWE , Warszawa 2006

Proces stacjonarny – jest to proces stochastyczny charakteryzujący się stałą w czasie wartością oczekiwaną oraz stałą wariancją i kowariancją. Jeśli któreś założenie nie jest spełnione proces jest niestacjonarny, wtedy nie ma ściśle określonego kierunku zmian, a wnioskowanie na podstawie takiego szeregu zazwyczaj prowadzi do błędów. Nie stanowi to problemu, ponieważ proces niestacjonarny może zostać sprowadzony do stacjonarnego za pomocą modelu, który sprawi, że powyższe założenia o stacjonarności będą spełnione.

Biały szum – wywodzi się ze stacjonarności procesu stochastycznego, taki też jest. W przypadku większości dynamicznych modeli ekonometrycznych, własności białego szumu są pożądane dla reszt modeli, a więc jego własności w czasie to: wartość oczekiwana równa 0, wariancja równa wariancji oraz kowariancja składników resztowych równa 0.

3. Prezentacja wybranych modeli ekonometrycznych

Klasyczny model regresji liniowej jest używany do modelowania wartości zmiennej objaśnianej, której funkcja ma postać liniową, a jej parametry są estymowane na podstawie zależności zachodzących wewnątrz zbioru elementów przyjętych do szacowania zmiennej objaśniającej włączając tą zmienną. Model jest akceptowany, gdy jest zweryfikowana jego poprawność. Weryfikację przeprowadza się za pomocą określonych narzędzi takich jak: współczynnik determinacji, który mówi o poziomie w jakim model dopasowuje się do zmiennej objaśnianej, współczynnik zbieżności, który mówi w jakim stopniu model nie wyjaśnia zmiennej objaśniającej. Na tym etapie przeprowadza się także testy istotności parametrów modelu takich jak na przykład test F-Fishera-Snedecora. Weryfikację modelu poprzedza jego budowa, która powinna uwzględniać takie elementy jak: rozsądny dobór elementów poparty logicznym powiązaniem, choć dopuszczalne są odstępstwa; eliminacja zmiennych quasi-stałych, czyli takich które charakteryzują się niską zmiennością, a więc nie wnoszą zbyt wiele informacji do modelu; analiza korelacji, której problem był poruszany wcześniej; metodę doboru zmiennych, ja w szczególności polecam metodę Hellwiga, stworzoną we Wrocławiu ze względu na jej kompleksowość; estymację parametrów, choć metod jest o wiele więcej. Za wszystkimi jednak stoi podobna koncepcja. Jego postać określona jest wzorem:

$$\hat{y} = a_0 + a_1 x_1 + a_j x_j + \cdots + a_m x_m,$$

gdzie:

- j=1,2,…,m

- $\hat{y}$ – postać teoretyczna

- a_m- parametry strukturalne

- x_m- wartości empiryczne zmiennych objaśniających

Modele Klasy ARIMA

Autoregressive integrated moving average, tak właśnie wygląda rozbicie ARIMA na słowa. „AR" jest to autoregresja, której problem był poruszany wcześniej; „I" odpowiada początkowemu krokowi różnicowania, który odpowiada za dopasowanie modelu do procesów niestacjonarnych tak jakby proces był stacjonarny; „MA" jest to średnia ruchoma czy też krocząca, czyli taka która zmienia się wraz z czasem, wyznacza wartość na podstawie n ostatnich obserwacji, tak wyjaśniane było to wcześniej.

Różnicowanie – idea używana przy założeniu błądzenia losowego, które ma postać:

$$X_t = X_{t-1} + Z_t$$

X_t-błądzenie losowe

X_{t-1}- poprzednia wartość błądzenia losowego

Z_t- biały szum, wartość losowa

Działanie „integrated", czyli procesu różnicowania:

$$Z_t = X_t - X_{t-1}$$

Proces staje się stacjonarny dopiero wtedy, gdy jest spełnione są założenia o stałej w czasie średniej i wariancji, widać zatem, że proces różnicowania pozwala na stworzenie procesu stacjonarnego na podstawie procesu niestacjonarnego.

Modeli klasy ARIMA używa się w celu zrozumienia danych lub też ich predykcji. Kombinacja powyższych 3 metod pozwala na uchwycenie zależności zachodzących w szeregu czasowym. Z reguły oczekiwane działanie wygląda tak, że wybierając odpowiedni model decydujemy się na wpasowanie modelu do danych, następnie szacowana jest zależność w czasie na podstawie autoregresji, a ostatecznie wyniki wygładzane są za pomocą średniej ruchomej. W praktyce jednak badacze często wykonują symulację polegającą na dopasowaniu modelu do kryterium informacyjnego Akaikego, tak aby wartość współczynnika była najmniejsza. Skutkuje to dobraniem postaci, która jest optymalna pod względem złożoności modelu oraz jego wartości informacyjnej.

Występują różne odmiany modeli ARIMA. Najpopularniejszym spośród nich jest ARMA, czyli taki który nie ma możliwości wpasowania się do danych. Kolejnym jest SARIMA uwzględniający sezonowość w procesie. VARIMA, używana jako modelowanie wielu scenariuszy przy pomocy modelu, można to określić jako metoda symulacji Monte Carlo na podstawie otrzymanego modelu. Jeśli chcemy wychwycić zależność długiego okresu najlepiej jest użyć modelu FARIMA, który zwany jest też ARFIMA. Ogólny wzór na model ARIMA to:

$$\alpha(B)Y_t = \beta(B)\varepsilon_t \ ,$$

Gdzie:
$\alpha(B) = 1 - \alpha_1 B - \cdots - \alpha_p B_p$ – operator regresji rzędu p,
$\beta(B) = 1 - \beta_1 B - \cdots - \beta_p B^q$ – operator średniej ruchomej rzędu q,
ε_t - reszty modelu, czynnik losowy, proces białego szumu
co można przedstawić jako:

$$Y_t = \ \alpha(B) = 1 - \alpha_1 B - \cdots - \alpha_p B_p + \varepsilon_t \ - \beta_1 B \varepsilon_{t-1} \ - \cdots - \beta_p B^q \varepsilon_{t-q}$$

Modele klasy ARCH

Modele klasy ARCH są to modele autoregresji z warunkową heteroskedastycznością. Pojęcie heteroskedastyczności to odwrotność homoskedastyczności, która jest definiowana jako posiadanie tej samej, skończonej wariancji przez wszystkie zmienne losowe w ciągu lub wektorze zmiennych losowych. W modelach ARCH zakłada się, że wariancja błędu losowego danego okresu jest określana przez błędy losowe z poprzednich okresów. Modele te stosuje się przy procesach stacjonarnych o zmiennej wariancji. Uniwersytet Ekonomiczny we Wrocławiu może poszczycić się wybitnym specjalistą z zakresu modelowania ekonometrycznego a szczególnie modeli klasy ARCH, Profesor nadzwyczajny UE Krzysztof Piontek napisał na ich temat dużą liczbę prac, rozważając zastosowanie ich różnych form.

Modele ARCH dobrze sprawdzają się przy prognozowaniu krótkoterminowym oraz analizie ryzyka. Reszty modeli mogą być ponownie modelowane w celu usunięcia z nich takich właściwości jak autokorelacja, grube ogony czy rozkład reszt inny niż normalny czy gromadzenie zmienności. Przy wykorzystywaniu modeli ARCH występują efekty tego też modelu. Model z resztami ARCH może być aproksymacją innego modelu bez efektu ARCH, dzieje się tak jeśli do modelowania weźmiemy złe zmienne bądź błędną specyfikę modelu. Błędy występują też, gdy dane nie są dostosowane ze względu na czas ekonomiczny oraz na czas kalendarzowy. Ostatnim efektem jest skupianie danych, które występuje, jeśli dane napływają w pakietach.

Najbardziej popularnym modelem ARCH jest GARCH, w którym „G" reprezentuje generalizację modelu GARCH. Polega ona na tym, że końcowe zależności modelu są bardziej oparte na bezpośrednich wartościach poprzedzających okres prognostyczny. Dzięki temu sprawdza się lepiej przy opisie rozkładów o grubych ogonach. Dodatkowo jest bardziej przydatny, ponieważ prognozowanie procesów finansowych na krótki okres nie jest rozwiązaniem większości problemów, GARCH z powodzeniem radzi sobie z dużymi rzędami opóźnień w czasie[2]. Dodatkowo szacowanie wariancji warunkowej h_t przy ARCH prowadzi często do złamania założenia co do nieujemności h_t. Model GARCH ma następującą postać:

$y_t = x'_t\xi + \varepsilon_t$,

gdzie:

x'_t – wektor zmiennych niezależnych

ξ-ksi, wektor parametrów

t = 1,2,…,T.

Dalsze własności zarezerwowane dla tego modelu jak i modelu ARCH, są dość mocno skomplikowane i jako, że w potrzebnych pojęciach nie napisałem o rozkładzie normalnym o wariancji warunkowej, tutaj zakończę rozważania na temat postaci modelu.

4. Podsumowanie

Modelowanie ekonometryczne jest bardzo przydatnym obszarem zainteresowań. Nie wykluczone, że proces decyzyjny w sztucznej inteligencji nie będzie oparty właśnie na podstawie tej metody. Osobiście uważam, że nie ma lepszego sposobu wnioskowania niż przez model ekonometryczny. Nie zdajemy sobie sprawy ale w codziennych czynnościach kierujemy się właśnie takim schematem jaki możemy stworzyć za pomocą modelu. Podejmując decyzję o zakupie przedmiotu pod uwagę bierzemy cenę, jakość, dostępność a następnie przypisujemy wagi odpowiednim cechom, które nas interesują, a następnie wybieramy najkorzystniejszą dla nas. Tak, ekonometria otacza nas. To, że szereg czasowy możemy opisać za pomocą licz, to tylko początek, możemy do tego użyć wartości „prawda i fałsz" „„Tak, Nie, może, ciężko

[2] Krawczyk T. Modelowanie ryzyka inwestycyjnego, wydawnictwo CeDeWu Sp. z o.o., Warszawa 2013

stwierdzić” i innych, które jesteśmy w stanie wyodrębnić i stworzyć model decyzyjny. Wielu naukowców pokłada w tej dziedzinie nauki duże nadzieje. Aktualnie dostęp do wiedzy na temat modelowania jest bardzo prosty. Możemy korzystać z pomocy tutorów z całego świata. W celu sprawnego tworzenia modelu warto jest nauczyć się, któregoś z analitycznych języków programowania.

Mam nadzieję, że opis powyższych metod decyzyjno-prognostycznych był opisem ciekawym i pełnym nowych informacji.