KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Projekt Innowacyjna SGH = wiedza + doświadczenie współfinansowany
ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego

Materiał dla studentów

Wprowadzenie do modeli ARMA/ARIMA

(na przykładzie zwrotów z instrumentów finansowych)

(studium przypadku)

Nazwa przedmiotu: *ekonometria finansowa I (22204), analiza szeregów czasowych i prognozowanie (13201);*

Kierunek studiów: *Finanse i rachunkowość, Metody ilościowe w ekonomii i systemy informacyjne*
Studia I stopnia/studia II stopnia

Opracowała:
dr hab. Ewa M. Syczewska, Instytut Ekonometrii, Kolegium Analiz Ekonomicznych SGH

Warszawa, 2011

I. Informacje wstępne

Geneza modeli autoregresji i średniej ruchomej (AR, MA oraz ARMA) sięga metod stosowanych przez inżynierów. Monografia Boxa i Jenkinsa[1] wydana w latach 70-tych XX w. stanowi kompendium wiedzy o metodach modelowania szeregów czasowych, m.in. przy użyciu tego typu modeli. Modele ARMA należą do najbardziej efektywnych metod prognozowania szeregów czasowych. Można je stosować do szeregów jednowymiarowych, stacjonarnych, a otrzymywane prognozy są krótkookresowe.

Model ARMA jest narzędziem prognozowania, może być również podstawą konstrukcji modeli ARCH/GARCH: jeśli wykorzystamy go jako pierwsze równanie opisujące zachowanie szeregu, i uzupełnimy o równanie wyrażające zmienną wariancję szeregu.

Aby zastosować model do danego szeregu obserwacji zmiennej ekonomicznej, należy najpierw sprawdzić jej stacjonarność i ewentualnie sprowadzić do zmiennej stacjonarnej metodą obliczania przyrostów. Dla niestacjonarnych zmiennych finansowych takich jak notowania akcji, indeksów giełdowych, kursów walutowych często zamiast przyrostów stosuje się przyrosty logarytmów (zwroty logarytmiczne), gdyż są interpretowane jako zwroty z jednodniowej hipotetycznej inwestycji w dany instrument.

Zwroty logarytmiczne są zmienną stacjonarną, można do nich dopasować odpowiedni model ARMA, następnie należy przetestować występowanie efektu ARCH, i ewentualnie uzupełnić model o drugie równanie dla warunkowej wariancji.

Ćwiczenie polegające na wyszukaniu najlepszej wersji modelu ARMA dla danego szeregu zwrotów zmiennej finansowej będzie przeprowadzone przy użyciu pakietu **gretl**. Oprócz narzędzi takich jak test niestacjonarności ADF, korelogram, model regresji liniowej szacowany metodą najmniejszych kwadratów, pakiet ten zawiera zestaw metod modelowania szeregów czasowych, w tym model ARMA/ARIMA. Model taki szacowany jest metodą największej wiarygodności przy założeniu normalności rozkładu składnika losowego.

Przypomnimy teraz potrzebne pojęcia i wzory.

1. **Funkcja autokorelacji i autokorelacji cząstkowej z próby:**

   Nieznana wartość oczekiwana i wariancja stacjonarnego procesu może być szacowana na podstawie wzorów: $\bar{x} = T^{-1}\sum_{i=1}^{T} x_i$, $s^2 = T^{-1}\sum_{t=1}^{T}(x_t - \bar{x})^2$.

   Ocena ***współczynnika korelacji*** zmiennych x_t, x_{t-k} jest równa

   $r_k = \frac{\sum_{t=k+1}^{T}(x_t-\bar{x})(x_{t-k}-\bar{x})}{Ts^2}$, k=1,2,…; T – liczba obserwacji.

   Współczynniki korelacji z próby tworzą funkcję autokorelacji z próby, ACF (ang. *autocorrelation function*). Współczynnik korelacji większy co do modułu od $2T^{-0,5}$ jest statystycznie istotny.

   Współczynniki korelacji cząstkowej mierzy korelację zmiennych x_t, x_{t-k} bez wpływu korelacji zmiennych pośrednich. Wyznaczany jest na podstawie regresji zmiennej względem jej opóźnień do rzędu *k* włącznie, ocena parametru przy zmiennej opóźnionej o *k* jest równa ocenie współczynnika korelacji cząstkowej rzędu *k*.

[1] Box, G.E.P., G.M. Jenkins, Time Series Analysis: Forecasting and Control, San Francisco: Holden Day, 1976.

Współczynniki korelacji cząstkowej tworzą funkcję autokorelacji cząstkowej z próby (ang. *partial autocorrelation function*, PACF).

2. **Test Boxa-Pierce'a i Ljung-Boxa**[2],[3]

W celu sprawdzenia łącznej istotności współczynników korelacji można zastosować test Boxa-Pierce'a lub lepiej Ljung-Boxa. Przy założeniu, że szereg jest procesem białego szumu, Box i Pierce pokazali, że statystyka $Q^*(k) = T\sum_{i=1}^{k} r_i^2$ ma rozkład asymptotyczny χ^2 o k stopniach swobody. Lepiej funkcjonuje statystyka zmodyfikowana przez Ljung i Boxa:

$$Q(k) = T(T+2)\sum_{i=1}^{k}(T-i)^{-1}r_i^2$$

również o asymptotycznym rozkładzie $\chi^2(k)$.

3. **Model autoregresji, model średniej ruchomej, model ARMA**

W ***modelu autoregresji*** wartości zmiennej są objaśniane jej opóźnionymi wartościami. Jest to model jednorównaniowy, dynamiczny i symptomatyczny, postaci:

$x_t = a_1x_{t-1} + a_2x_{t-2} + \cdots + a_px_{t-p} + \varepsilon_t$ czyli

$(1 – a_1L + a_2L^2 + \cdots + a_pL^p)x_t = \varepsilon_t$, gdzie L oznacza operator opóźnień.

Wielomian $1 - a_1L + a_2L^2 + \cdots + a_pL^p$ = A(L) określa własności szeregu. Jeśli jest podzielny przez (1–L)=Δ, to szereg jest niestacjonarny z powodu występowania pierwiastka jednostkowego.

***Model średniej ruchomej* , MA** (ang. *moving average*) wyraża wartości zmiennej jako funkcję opóźnionych wartości (stacjonarnego) składnika losowego:

$x_t = \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \cdots + b_q\varepsilon_{t-q}$ czyli $x_t = B(L)\varepsilon_t$.

Połączeniem tych dwóch składowych jest model mieszany ARMA:

$A(L)x_t = B(L)\varepsilon_t$ rzędu p,q.

Model AR można oszacować metodą najmniejszych kwadratów, model MA oraz mieszany – metodą największej wiarygodności.

[2] Box, G.E.P. i Pierce, D.A. (1970), Distribution of Residual Autocorrelations in Autoregressive Moving Average Time Series Models, Journal of the American Statistical Association, 65, 1509-1526.

[3] Ljung, G.M. i Box, G.E.P. (1978), On a Measure of Lack of Fit in Time Series Models, Biometrika, 65, 297-303.

4. **Kryteria informacyjne**

Do porównywania wariantów modeli i do wyboru najlepszego z nich można wykorzystać kryteria informacyjne: Akaike, Schwarza i Hannana-Quinna.[4],[5],[6]
Spośród różnych wariantów modeli objaśniających tę samą zmienną wybieramy ten, dla którego wybrane kryterium ma najmniejszą wartość.

Dla modeli ARMA(p,q) kryteria mają postać:

a) Kryterium Akaike: $\ln(\hat{\sigma}^2) + 2(1 + p + q)/T$

b) Kryterium Schwarza: $\ln(\hat{\sigma}^2) + (1 + p + q)\ln(T)/T$

c) Kryterium Hannana-Quinna: $\ln(\hat{\sigma}^2) + 2(1 + p + q)\ln(\ln(T))/T$,

gdzie T oznacza liczbę obserwacji wykorzystanych do estymacji modelu, a $\hat{\sigma}^2$ oznacza estymator wariancji dla metody największej wiarygodności.

5. **Metodologia Boxa i Jenkinsa**

Metodologia Boxa i Jenkinsa polega na porównaniu funkcji ACF i PACF dla badanego szeregu stacjonarnego z postaciami tych funkcji dla modeli AR(p), MA(q) itp. o znanej liczbie parametrów.

Teoretyczny model AR(p) ma funkcję PACF równą zeru dla opóźnień większych niż *p* (tzn. współczynniki autokorelacji cząstkowej dla opóźnień większych niż *p* są nieistotne). Funkcja ACF maleje (jeśli tempo zmniejszania się współczynników przy wzroście liczby opóźnień jest powolne, może to sugerować że szereg jest niestacjonarny). Teoretyczny model MA(q) ma funkcję ACF równą zeru dla opóźnień większych niż q, natomiast funkcja PACF jest malejąca (MA jako suma składników stacjonarnych jest stacjonarne).

Model mieszany ARMA(p.q) ma funkcję ACF podobną do funkcji dla procesu AR(p) ale od miejsca *q–p*, zaś funkcja PACF dla k > *p–q* zachowuje się jak funkcja dla procesu MA(q).

Jeśli porównamy zachowanie funkcji ACF i PACF dla naszego szeregu z teoretycznymi funkcjami dla różnych wariantów *p*, *q*, możemy oszacować ten wariant, który wydaje się najlepszy, następnie sprawdzić zachowanie reszt oszacowanego modelu (powinny nie wykazywać autokorelacji, tzn. współczynniki ACF nieistotne, co można sprawdzić również np. testem Ljung-Boxa).

[4] Akaike, H. (1974), A New Look at the Statistical Model Identification, I.E.E.E. Transactions on Automatic Control, 19, 716-723.
[5] Schwarz, G. (1978), Estimating the Dimension of a Model, Annals of Statistics, 6, 461-464.
[6] Hannan, E.J. i B.G. Quinn (1979), The Determination of the Order of an Autoregression, Journal of the Royal Statistical Society, B, 41, 190-195.

6. **Redukcja modelu od maksymalnego (P,Q)**.

   Można wybrać maksymalne możliwe dla danego szeregu wartości P i Q, a następnie szacować wszystkie warianty ARMA(p,q) dla $0 \leq p \leq P, 0 \leq q \leq Q$, wybrać wariant modelu minimalizujący wartość kryterium informacyjnego.

II. Harmonogram/scenariusz realizacji/kolejność działań

1. Studenci oceniają i analizują cechy charakterystyczne wybranych szeregów czasowych obserwacji zmiennych oraz ich przyrostów, sporządzają wykresy funkcji autokorelacji i autokorelacji cząstkowej dla zmiennych oraz formułują wstępne wnioski co do jakościowych cech tych wykresów.
2. Analiza istotności współczynników korelacji i korelacji cząstkowej dla poszczególnych zmiennych – na podstawie wykresów oraz przy wykorzystaniu testu Ljung-Boxa. Studenci porównują wyniki otrzymane dla zmiennych różnego typu.
3. Na podstawie wykresów funkcji ACF i PACF należy zaproponować maksymalną możliwą liczbę opóźnień P oraz Q w części AR i MA proponowanego modelu – dla zmiennej lub dla jej przyrostów (a w przypadku zmiennych finansowych – dla zwrotów logarytmicznych).
4. Estymacja kilku wariantów modeli ARMA(p,q) – dla liczby opóźnień p nie przekraczającej P oraz liczby opóźnień q nie przekraczającej Q. Należy zwrócić uwagę na:
   a. Istotność parametrów przy zmiennych opóźnionych
   b. Pierwiastki wielomianów A(L) oraz B(L) – czy są co do modułu większe od 1, co zapewnia odpowiednie własności tych wielomianów oraz całego procesu ARMA
   c. Wartości kryteriów informacyjnych Akaike, Schwarza i Hannana-Quinna dla poszczególnych wariantów modeli.
5. Metoda postępowania polega na wybraniu modelu ARMA(p.q) dla szeregu (ewentualnie dla różnic szeregu, lub dla zwrotów logarytmicznych w przypadku szeregu finansowego) na podstawie ACF i PACF z próby.
   a) Jeśli funkcja PACF ma na początku istotne składniki, a potem „ucięcie" dla jakiejś wartości *p*, a autokorelacja dla pierwszego opóźnienia jest dodatnia, dodajemy składniki autoregresyjne. Miejsce „ucięcia" odpowiada rzędowi modelu AR.
   b) Jeśli funkcja ACF ma na początku istotne składniki, a potem „ucięcie" dla jakiejś wartości q, i/lub autokorelacja rzędu pierwszego jest ujemna, należy wprowadzić składniki części MA do rzędu wskazanego przez „ucięcie".
   c) W modelu mieszanym AR i MA mogą wpływać na siebie nawzajem.
6. Studenci formułują wnioski co do wyboru najlepszego z kilku wybranych wariantów modelu ARMA, zapisują postać oszacowaną modelu.
7. Dodatkowym elementem dla zmiennych finansowych (tzn. dla zwrotów logarytmicznych) jest sprawdzenie, czy dla reszt oszacowanego modelu występuje efekt ARCH grupowania wariancji. Można w tym celu zastosować test Engle'a.

8. Moderator przedstawia wyniki analogicznej procedury wyboru modelu ARMA na podstawie metodologii Boxa i Jenkinsa oraz kryteriów informacyjnych dla omawianych na początku zmiennych (WIG, kursy walutowe itp.)
9. Studenci przedstawiają wyniki otrzymane dla badanych przez siebie zmiennych, porównują warianty modeli w zależności od rodzaju danych, częstotliwości obserwacji, długości szeregu czasowego i cech badanej zmiennej.

III. Opis przypadku/sytuacji

Przygotowany jest przykładowy zbiór danych (podczas testowania kejsu: plik z danymi dla indeksów SP500 oraz WIG20, dane dzienne, notowania zamknięcia, oraz dla kursów walutowych USD/PLN, EUR/PLN).

Należy wyznaczyć zwroty logarytmiczne poprzez dodanie odpowiednich zmiennych do bazy danych odpowiednim poleceniem w gretl. (W razie potrzeby, jeśli nie sprawdzono tego na poprzednich zajęciach, należy sprawdzić czy zmienna jest stacjonarna/niestacjonarna, czy zwroty logarytmiczne są stacjonarne, np. testem ADF.)

Celem jest wybór i oszacowanie modelu ARMA dla zwrotów logarytmicznych/ dla przyrostów zmiennej makroekonomicznej.

W gretl estymacja modelu ARMA/ARIMA wywoływana jest odpowiednim poleceniem z menu **Modele-> Modele szeregów czasowych -> ARMA**

A. *Jakie jest zachowanie zmiennej? Zachowanie jej przyrostów lub zwrotów logarytmicznych? Opisać pokrótce cechy jakościowe badanego szeregu.*

Polecenie **Zmienna → Korelogram** powoduje wyznaczenie zarówno wykresu słupkowego funkcji ACF i PACF, jak i tabeli zawierającej w kolejnych kolumnach współczynniki autokorelacji, korelacji cząstkowej, i w ostatniej kolumnie – wartości statystyki Q(k), k=1,2,…

B. *Jaki jest wygląd wykresu ACF? Czy sugeruje powolne wygasanie wartości współczynników? Czy „ucięcie" po kilku pierwszych składnikach?*
C. *Jaki jest wygląd wykresu PACF?*
D. *Co można sądzić o postaci modelu ARMA na podstawie funkcji ACF i PACF?*

Wariant: Studenci mogą podzielić się na dwie grupy: pierwsza przyjmuje maksymalny rząd opóźnień P i Q, następnie szacuje modele ARMA(p,q) o nie większej od (P,Q) liczbie opóźnień, porównując wartości kryteriów informacyjnych. Dla ostatecznie wybranego modelu bada własności reszt, ACF i test Q dla reszt.
Druga grupa startuje od modelu ARMA(p.q) o niewielkiej liczbie opóźnień, sprawdza zachowanie reszt modelu, na podstawie wykresów ACF i PACF dla reszt ewentualnie poszerza zestaw opóźnień części AR i MA, następnie sprawdza własności reszt itd.

E. Jaka jest zaproponowana postać modelu? Czy wygląd wykresów ACF i PACF dla reszt modelu świadczą o dobrym doborze postaci ARMA?

Dla wybranego modelu trzeba sprawdzić, czy jest stabilny i odwracalny – w **gretl** w oknie z wynikami estymacji modelu ARMA podane są oprócz ocen parametrów, błędów szacunku itp. również wartości pierwiastków wielomianów A(L) i B(L). Jeśli w części AR występuje pierwiastek równy 1 (część rzeczywista = 1, część urojona =0), oznacza to, że szereg był niestacjonarny (wtedy procedura estymacji może nie osiągnąć zbieżności) i należało jeszcze raz policzyć przyrosty. Jeśli pierwiastki są mniejsze co do modułu od 1, proces jest stacjonarny a model powinien mieć dobre własności.
Dodatkowo dla zmiennych finansowych, charakteryzujących się zmiennością wariancji, trzeba sprawdzić czy nie występuje efekt ARCH (grupowania wariancji). W pakiecie **gretl**

a) W oknie z wynikami estymacji modelu ARMA wybieramy zapisywanie reszt, następnie dla kwadratów reszt szacujemy ich regresję względem wyrazu wolnego i kwadratów reszt opóźnionych. Testem F sprawdzamy łączną istotność opóźnionych kwadratów reszt. Hipoteza zerowa o braku istotności oznacza brak efektu ARCH.
b) Szybszy sposób: w oknie z wynikami estymacji wybieramy: Testowanie → Efekt ARCH.

(Taka wersja testu nazywa się testem Engle'a efektu ARCH.)[7]

F. Czy reszty wybranego modelu ARMA cechuje zmienność wariancji (efekt grupowania wariancji, efekt ARCH)? Jeśli tak, model ARMA będzie równaniem dla średniej, które uzupełnia się o równanie dla wariancji, tworząc model ARCH.

IV. Wymagane rezultaty pracy i ich forma

Rezultatem pracy będzie krótki (kilkustronicowy) raport z opisem procedury budowy modelu, zawierający odpowiedzi na punkty A.–F.

[7] Engle, R.F. (1988), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica, 96, 1988, 893–920.