Modele ARMA

- Ważną klasę modeli dynamicznych stanowią modele $ARMA(p, q)$

- Ważną klasę modeli dynamicznych stanowią modele $ARMA(p, q)$
- Modele tej klasy są modelami ateoretycznymi

- Ważną klasę modeli dynamicznych stanowią modele $ARMA(p, q)$
- Modele tej klasy są modelami ateoretycznymi
- Model zjawiska budowany jest w oparciu o statystyczne właściwości danych

- Ważną klasę modeli dynamicznych stanowią modele $ARMA(p, q)$
- Modele tej klasy są modelami ateoretycznymi
- Model zjawiska budowany jest w oparciu o statystyczne właściwości danych
- Modele ARMA są użytecznym narzędziem prognostycznym

- Model ARMA składa się z dwóch części:

$$y_t = \underbrace{\alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p}}_{AR} + \underbrace{\mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}}_{MA}$$

- model autoregresyjny

$$y_t = \mu + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p}$$

- model autoregresyjny

$$y_t = \mu + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p}$$

- model średniej ruchomej

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

$$E(\varepsilon_t) = 0 \quad var(\varepsilon_t) = \sigma^2 \quad E(\varepsilon_t, \varepsilon_s) = 0 \quad \forall t \neq s$$

- składniki ε_t nazywa się innowacjami

- składniki ε_t nazywa się innowacjami
- cześć MA można traktować jako składnik losowy o rozbudowanej strukturze

$$u_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}$$

- składniki ε_t nazywa się innowacjami
- cześć MA można traktować jako składnik losowy o rozbudowanej strukturze

$$u_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}$$

- w takim przypadku podlega on zjawisku autokorelacji

- składniki ε_t nazywa się innowacjami
- cześć MA można traktować jako składnik losowy o rozbudowanej strukturze

$$u_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

- w takim przypadku podlega on zjawisku autokorelacji
- model ARMA można wówczas przedstawić jako

$$y_t = \mu + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + u_t \quad (1)$$

- Jeśli model nie ma części autoregresyjnej to można jego parametry oszacować MNK

- Jeśli model nie ma części autoregresyjnej to można jego parametry oszacować MNK
- Model (1) nie spełnia założeń KRML

- Jeśli model nie ma części autoregresyjnej to można jego parametry oszacować MNK
- Model (1) nie spełnia założeń KRML
- Estymatory MNK nie są zgodne

- Jeśli model nie ma części autoregresyjnej to można jego parametry oszacować MNK
- Model (1) nie spełnia założeń KRML
- Estymatory MNK nie są zgodne
- Oszacowania uzyskuje się wykorzystując MNW lub Nieliniową MNK

- Jeśli model nie ma części autoregresyjnej to można jego parametry oszacować MNK
- Model (1) nie spełnia założeń KRML
- Estymatory MNK nie są zgodne
- Oszacowania uzyskuje się wykorzystując MNW lub Nieliniową MNK
- W przypadku rozbudowanej postaci modelu obliczenia są czasochłonne i źle numerycznie uwarunkowane

- Jeśli model nie ma części autoregresyjnej to można jego parametry oszacować MNK
- Model (1) nie spełnia założeń KRML
- Estymatory MNK nie są zgodne
- Oszacowania uzyskuje się wykorzystując MNW lub Nieliniową MNK
- W przypadku rozbudowanej postaci modelu obliczenia są czasochłonne i źle numerycznie uwarunkowane
- Jest to szczególnie duży problem dla krótkich szeregów

- Jeśli model nie ma części autoregresyjnej to można jego parametry oszacować MNK
- Model (1) nie spełnia założeń KRML
- Estymatory MNK nie są zgodne
- Oszacowania uzyskuje się wykorzystując MNW lub Nieliniową MNK
- W przypadku rozbudowanej postaci modelu obliczenia są czasochłonne i źle numerycznie uwarunkowane
- Jest to szczególnie duży problem dla krótkich szeregów
- Do testowania hipotez można użyć testów W, LR, LM

- Przed estymacją parametrów należy ustalić rząd procesów

- Przed estymacją parametrów należy ustalić rząd procesów
- Metoda Boxa-Jenkinsa

- Przed estymacją parametrów należy ustalić rząd procesów
- Metoda Boxa-Jenkinsa
- Metoda od ogólnego do szczegółowego

- Przed estymacją parametrów należy ustalić rząd procesów
- Metoda Boxa-Jenkinsa
- Metoda od ogólnego do szczegółowego
- Kryteria informacyjne

- Jeżeli procesy są prawidłowo ustalone to reszty są białym szumem

- Jeżeli procesy są prawidłowo ustalone to reszty są białym szumem
- testem niezależności reszt jest statystyka Ljunga-Boxa

$$Q = n(n+2)\sum_{k=1}^{m} \frac{\hat{\rho}_k}{T-k} \xrightarrow{D} \chi^2_{m-p-q}$$

gdzie

$$\hat{\rho}_k = \frac{(T-k)\sum(y-\bar{y})\sum(y_{t-k}-\bar{y})}{T^{-1}\sum(y-\bar{y})^2}$$

- Funkcja autokowariancji pokazuje zależność między y_t a poprzednimi wartościami

- Funkcja autokowariancji pokazuje zależność między y_t a poprzednimi wartościami
- Dla szeregu stacjonarnego wartości zależą wyłącznie od odległości między obserwacjami

$$E(y_t - E(y_t))(y_{t-k} - E(y_{t-k})) = \gamma_k \quad k = 0, 1, 2, \ldots$$

- Funkcja autokowariancji pokazuje zależność między y_t a poprzednimi wartościami
- Dla szeregu stacjonarnego wartości zależą wyłącznie od odległości między obserwacjami

$$E(y_t - E(y_t))(y_{t-k} - E(y_{t-k})) = \gamma_k \quad k = 0, 1, 2, \ldots$$

- autkowariancje są nieunormowane

- Funkcja autokowariancji pokazuje zależność między y_t a poprzednimi wartościami
- Dla szeregu stacjonarnego wartości zależą wyłącznie od odległości między obserwacjami

$$E(y_t - E(y_t))(y_{t-k} - E(y_{t-k})) = \gamma_k \quad k = 0, 1, 2, \ldots$$

- autkowariancje są nieunormowane
- Funkcja autokorelacji (ACF) (**A**uto**c**orrelation **F**unction)

$$\rho_k = \frac{cov(y_t, y_{t-k})}{var(y_t)}$$

- Funkcja autokowariancji pokazuje zależność między y_t a poprzednimi wartościami
- Dla szeregu stacjonarnego wartości zależą wyłącznie od odległości między obserwacjami

$$E(y_t - E(y_t))(y_{t-k} - E(y_{t-k})) = \gamma_k \quad k = 0, 1, 2, \ldots$$

- autkowariancje są nieunormowane
- Funkcja autokorelacji (ACF) (**A**uto**c**orrelation **F**unction)

$$\rho_k = \frac{cov(y_t, y_{t-k})}{var(y_t)}$$

- $\rho_k \in [-1, 1]$

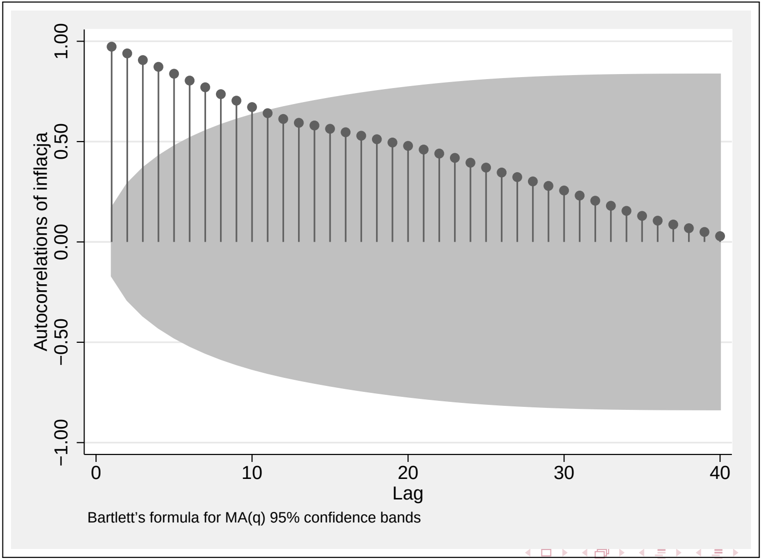
- Funkcja autokowariancji pokazuje zależność między y_t a poprzednimi wartościami
- Dla szeregu stacjonarnego wartości zależą wyłącznie od odległości między obserwacjami

$$E(y_t - E(y_t))(y_{t-k} - E(y_{t-k})) = \gamma_k \quad k = 0, 1, 2, \ldots$$

- autkowariancje są nieunormowane
- Funkcja autokorelacji (ACF) (**A**uto**c**orrelation **F**unction)

$$\rho_k = \frac{cov(y_t, y_{t-k})}{var(y_t)}$$

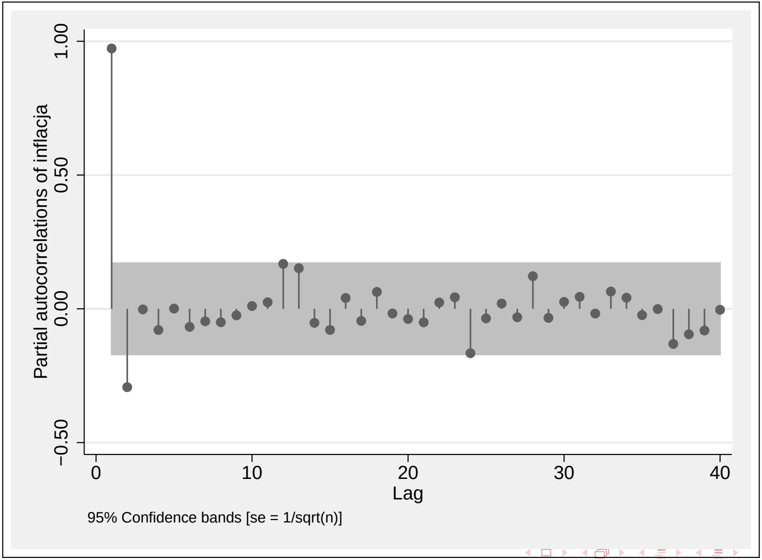
- $\rho_k \in [-1, 1]$
- $\rho_k = \rho_{-k}$


Autocorrelations of inflacja
1.00
0.50
0.00
−0.50
−1.00
0
10
20
30
40
Lag
Bartlett's formula for MA(q) 95% confidence bands

- Funkcja cząstkowej autokorelacji (**P**artial **A**uto**c**orrelation **F**unction) pokazuje zależność między y_t a poprzednimi wartościami, pomijając wpływ pośrednich opóźnień

- Funkcja cząstkowej autokorelacji (**P**artial **A**uto**c**orrelation **F**unction) pokazuje zależność między y_t a poprzednimi wartościami, pomijając wpływ pośrednich opóźnień
- Liczbowo jest równa oszacowaniu współczynnika ρ_k w modelu:

$$y_t = \mu + \rho_1 y_{t-1} + \ldots + \rho_k y_{t-k} + \varepsilon_t$$


Partial autocorrelations of inflacja
1.00
0.50
0.00
−0.50
0
10
20
30
40
Lag
95% Confidence bands [se = 1/sqrt(n)]

```
                                          -1       0       1 -1       0       1
LAG   AC       PAC      Q      Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1    0.9731   0.9731   127.85  0.0000           |-------           |-------
2    0.9395  -0.2928   247.95  0.0000           |-------         --|
3    0.9062  -0.0021   360.56  0.0000           |-------           |
4    0.8727  -0.0791    465.8  0.0000           |------            |
5    0.8383   0.0009   563.66  0.0000           |------            |
6    0.8041  -0.0675   654.44  0.0000           |------            |
7    0.7708  -0.0466    738.5  0.0000           |------            |
8    0.7365  -0.0502   815.87  0.0000           |-----             |
9    0.7041  -0.0245   887.16  0.0000           |-----             |
10   0.6721   0.0104   952.64  0.0000           |-----             |
11   0.6416   0.0245   1012.8  0.0000           |-----             |
12   0.6129   0.1681   1068.2  0.0000           |----              |-
```

- Polega na graficznej analizie przebiegu funkcji ACF i PACF

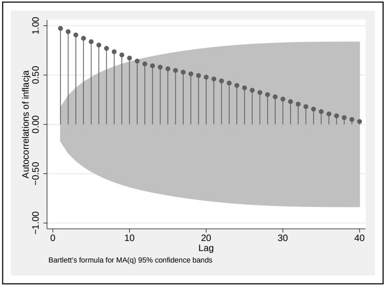
- Polega na graficznej analizie przebiegu funkcji ACF i PACF
- Wielkość p oraz q można oszacować na podstawie przebiegu funkcji ACF i PACF

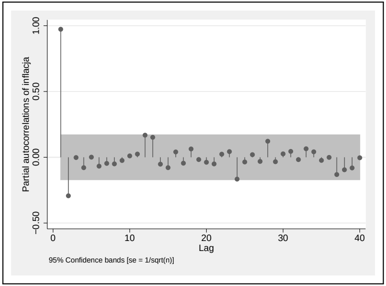
- Polega na graficznej analizie przebiegu funkcji ACF i PACF
- Wielkość p oraz q można oszacować na podstawie przebiegu funkcji ACF i PACF
- Urywająca się funkcja ACF wskazuje na rząd procesu MA

- Polega na graficznej analizie przebiegu funkcji ACF i PACF
- Wielkość p oraz q można oszacować na podstawie przebiegu funkcji ACF i PACF
- Urywająca się funkcja ACF wskazuje na rząd procesu MA
- Urywająca się funkcja PACF wskazuje na rząd procesu AR

- Polega na graficznej analizie przebiegu funkcji ACF i PACF
- Wielkość p oraz q można oszacować na podstawie przebiegu funkcji ACF i PACF
- Urywająca się funkcja ACF wskazuje na rząd procesu MA
- Urywająca się funkcja PACF wskazuje na rząd procesu AR
- Jeżeli obie funkcję maleją bardzo powoli należy szereg zróżnicować

- Polega na graficznej analizie przebiegu funkcji ACF i PACF
- Wielkość p oraz q można oszacować na podstawie przebiegu funkcji ACF i PACF
- Urywająca się funkcja ACF wskazuje na rząd procesu MA
- Urywająca się funkcja PACF wskazuje na rząd procesu AR
- Jeżeli obie funkcję maleją bardzo powoli należy szereg zróżnicować
- Regularne skoki mogą wskazywać na obecność sezonowości


Autocorrelations of inflacja
1.00
0.50
0.00
−0.50
−1.00
0
10
20
30
40
Lag
Bartlett's formula for MA(q) 95% confidence bands


Partial autocorrelations of inflacja
1.00
0.50
0.00
−0.50
0
10
20
30
40
Lag
95% Confidence bands [se = 1/sqrt(n)]

- W praktyce wykresy są trudne do jednoznacznej interpretacji

- W praktyce wykresy są trudne do jednoznacznej interpretacji
- Dodatkowy problemem jest brak zgodności oszacowań PACF

- W praktyce wykresy są trudne do jednoznacznej interpretacji
- Dodatkowy problemem jest brak zgodności oszacowań PACF
- Wskazane jest posłużenie się formalnymi metodami

- W praktyce wykresy są trudne do jednoznacznej interpretacji
- Dodatkowy problemem jest brak zgodności oszacowań PACF
- Wskazane jest posłużenie się formalnymi metodami
- Celem jest wybór modelu o małej liczbie parametrów

Przykład

Porównanie statystyk różnych modeli

Kryterium	ARMA(2,0)	ARMA(2,1)	ARMA(1,1)	ARMA(2,2)
AIC	102.2005	127.7031	105.7842	107.6906
BIC	113.7317	142.1171	117.3154	122.1046
LL	-47.10025	-58.85155	-48.89209	-48.84532

- Powszechność sezonowości

- Powszechność sezonowości
- Brak jej uwzględnienia może powodować autokorelację

- Powszechność sezonowości
- Brak jej uwzględnienia może powodować autokorelację
- Zmienne zero-jedynkowe

- Powszechność sezonowości
- Brak jej uwzględnienia może powodować autokorelację
- Zmienne zero-jedynkowe
- Sezonowe wyrównywanie szeregów (X11, TRAMO-SEATS)

- Powszechność sezonowości
- Brak jej uwzględnienia może powodować autokorelację
- Zmienne zero-jedynkowe
- Sezonowe wyrównywanie szeregów (X11, TRAMO-SEATS)
- Różnicowanie sezownowe

Przykład

Porównanie statystyk różnych modeli

Kryterium	SARMA(2,0)	SARMA(1,1)	SARMA(1,0)
AIC	182.0381	182.9793	189.6979
BIC	193.1881	194.1292	198.0603
LL	-87.01905	-87.48964	-91.84893

- Model ARMA może być wykorzystany jako narzędzie prognostyczne

- Model ARMA może być wykorzystany jako narzędzie prognostyczne
- Wartość w kolejnym okresie zależy wyłącznie od znanych wartości

$$y_{T+1} = \mu + \alpha_1 y_T + \ldots + \alpha_p y_{T-p+1} + \varepsilon_{T+1} + \theta_1 \varepsilon_T + \ldots + \theta_q \varepsilon_{T-q+1}$$

gdzie T oznacza ostatnią obserwację w próbie

- Model ARMA może być wykorzystany jako narzędzie prognostyczne
- Wartość w kolejnym okresie zależy wyłącznie od znanych wartości

$$y_{T+1} = \mu + \alpha_1 y_T + \ldots + \alpha_p y_{T-p+1} + \varepsilon_{T+1} + \theta_1 \varepsilon_T + \ldots + \theta_q \varepsilon_{T-q+1}$$

gdzie T oznacza ostatnią obserwację w próbie

- Zakłada się że reszty dla przyszłych okresów są równe 0

- Model ARMA może być wykorzystany jako narzędzie prognostyczne
- Wartość w kolejnym okresie zależy wyłącznie od znanych wartości

$$y_{T+1} = \mu + \alpha_1 y_T + \ldots + \alpha_p y_{T-p+1} + \varepsilon_{T+1} + \theta_1 \varepsilon_T + \ldots + \theta_q \varepsilon_{T-q+1}$$

gdzie T oznacza ostatnią obserwację w próbie

- Zakłada się że reszty dla przyszłych okresów są równe 0
- Prognoza y_{T+1} jest równa

$$\hat{y_{T+1}} = \hat{\mu} + \alpha_1 \hat{y_T} + \ldots + \alpha_p \hat{y_{T-p+1}} + \theta_1 e_T + \ldots + \theta_q e_{T-q+1}$$

- Model ARMA może być wykorzystany jako narzędzie prognostyczne
- Wartość w kolejnym okresie zależy wyłącznie od znanych wartości

$$y_{T+1} = \mu + \alpha_1 y_T + \ldots + \alpha_p y_{T-p+1} + \varepsilon_{T+1} + \theta_1 \varepsilon_T + \ldots + \theta_q \varepsilon_{T-q+1}$$

gdzie T oznacza ostatnią obserwację w próbie

- Zakłada się że reszty dla przyszłych okresów są równe 0
- Prognoza y_{T+1} jest równa

$$\hat{y_{T+1}} = \hat{\mu} + \alpha_1 \hat{y_T} + \ldots + \alpha_p \hat{y_{T-p+1}} + \theta_1 e_T + \ldots + \theta_q e_{T-q+1}$$

- prognozy oblicza się rekurencyjnie

- sensowne prognozy z modelu ARMA najwyżej na $max\{p, q\}$ okresów

- sensowne prognozy z modelu ARMA najwyżej na $max\{p, q\}$ okresów
- dla kolejnych okresów przyjmują wartość

$$y^{\star} = \frac{\mu}{1 - \sum \alpha_i}$$

```
ARIMA regression

Sample:  1998m1 - 2008m12                    Number of obs    =       132
                                             Wald chi2(2)     =  25062.40
Log likelihood = -47.10025                   Prob > chi2      =    0.0000

------------------------------------------------------------------------------
             |                 OPG
    inflacja |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
inflacja     |
       _cons |   8.235239   4.280912     1.92   0.054    -.155194    16.62567
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   1.409584   .0487239    28.93   0.000    1.314087    1.505081
         L2. |  -.4130118     .04851    -8.51   0.000   -.5080896    -.317934
-------------+----------------------------------------------------------------
      /sigma |    .338333   .0105082    32.20   0.000    .3177374    .3589287
------------------------------------------------------------------------------
```

- Prognoza na 1 okres naprzód

$$y_{T+1} = \alpha_1 y_T + \alpha_2 y_{T-1} = 1.41 * 4.53 - 0.41 * 4.67 = 4.47$$

- Prognoza na 1 okres naprzód

$$y_{T+1} = \alpha_1 y_T + \alpha_2 y_{T-1} = 1.41 * 4.53 - 0.41 * 4.67 = 4.47$$

- Prognoza na 2 okresy naprzód

$$y_{T+2} = \alpha_1 y\hat{_{T+1}} + \alpha_2 y_T = 1.41 * 4.47 - 0.41 * 4.53 = 4.45$$

- Prognoza na 1 okres naprzód

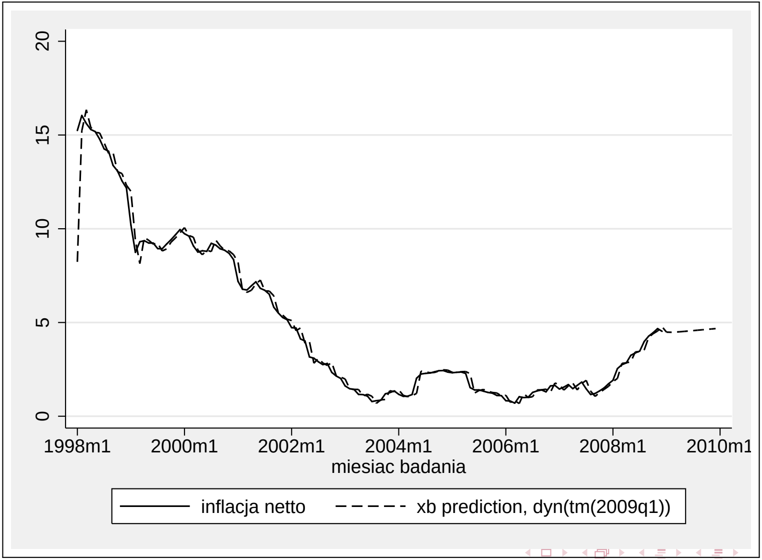
$$y_{T+1} = \alpha_1 y_T + \alpha_2 y_{T-1} = 1.41 * 4.53 - 0.41 * 4.67 = 4.47$$

- Prognoza na 2 okresy naprzód

$$y_{T+2} = \alpha_1 \hat{y_{T+1}} + \alpha_2 y_T = 1.41 * 4.47 - 0.41 * 4.53 = 4.45$$

- Prognoza na 3 okresy naprzód

$$y_{T+3} = \alpha_1 \hat{y_{T+2}} + \alpha_2 \hat{y_{T+1}} = 1.41 * 4.45 - 0.41 * 4.47 = 4.44$$


20
15
10
5
0
1998m1
2000m1
2002m1
2004m1
2006m1
2008m1
2010m1
miesiac badania
inflacja netto
xb prediction, dyn(tm(2009q1))

- Ze względu na ateoretyczną naturę procesu ARMA, czasem wykorzystuje się go w połączeniu z innym modelem

- Ze względu na ateoretyczną naturę procesu ARMA, czasem wykorzystuje się go w połączeniu z innym modelem
- Wówczas ma on postać

$$y_t = X_t\beta + u \qquad u_t \sim ARMA(p, q)$$

```
ARIMA regression
Sample:  1998m2 - 2008m12                     Number of obs     =       131
                                              Wald chi2(4)      =  19031.20
Log likelihood = -38.16698                    Prob > chi2       =    0.0000
------------------------------------------------------------------------------
             |                 OPG
    inflacja |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       stopa |
         --. |  -.2230583   .1066216    -2.09   0.036    -.4320329   -.0140838
         L1. |  -.1902509   .1413774    -1.35   0.178    -.4673454    .0868436
       _cons |   14.29238   5.359286     2.67   0.008      3.78837    24.79638
-------------+----------------------------------------------------------------
ARMA      ar |
         L1. |   1.386976     .06362    21.80   0.000     1.262284    1.511669
         L2. |  -.3891893   .0640723    -6.07   0.000    -.5147686     -.26361
-------------+----------------------------------------------------------------
      /sigma |   .3163894   .0122967    25.73   0.000     .2922883    .3404905
==============================================================================
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+----------------------------------------------------------------
       ARMA  |    131           .   -38.16698      6     88.33396    105.5851
------------------------------------------------------------------------------
```