Analiza szeregów czasowych

Wykład

Analiza dynamiki

Podstawowe czynniki wywierające wpływ na poziom zjawisk masowych to:

- **Trend** (**tendencja rozwojowa**) - długookresowa tendencja powodująca systematyczny wzrost, lub spadek obserwowanych w szeregu czasowym wielkości.
(Przy braku wyraźnej tendencji wzrostowej, lub spadkowej bywa używane określenie **trend boczny**.).
- **Wahania (zmiany) okresowe** - zakłócenia odznaczające się określoną regularnością, zazwyczaj wynikającą z naturalnej cykliczności zjawisk, lub procesów. Typowe okresy to dni, tygodnie, kwartały, lata.
- **Wahania nieregularne (przypadkowe, losowe)**
o nieprzewidywalnym charakterze.

- $S_1, S_2, \ldots, S_t, \ldots, S_{t+h}$ - ceny akcji

S_t – cena akcji (instrumentu finansowego) w chwili t (np. minucie, dniu, roku …)

- $R_t = ln\frac{S_t}{S_{t-1}}$ - logarytmiczna stopa zwrotu

- $R_t = \frac{S_t - S_{t-1}}{S_{t-1}}$ - stopa zwrotu

Def. Niech $\{X_t\}$ szereg czasowy, $E(X_t^2) < \infty$.

- $\mu_X(t) = E(X_t)$ - funkcja wartości oczekiwanej (średniej) procesu
- $\gamma_X(t,s) = Cov(X_t, X_s) = E\left[\left(X_t - \mu_X(t)\right)\left(X_s - \mu_X(s)\right)\right]$ - funkcja kowariancyjna procesu
- $\{X_t\}$ jest stacjonarny (stacjonarny w słabym sensie), jeśli

(i) $\mu_X(t)$ nie zależy od czasu (stała)

(ii) $\gamma_X(t+h, t)$ nie zależy od t dla każdego h.

Przykłady obliczania i zastosowania funkcji kowariancji

1) Błądzenie losowe $X_t = X_{t-1} + Z_t,\ X_0 = 0$

$$\gamma(t+h,t) = Cov(X_t + Z_{t+1} + \dots + Z_{t+h}, X_t) = Cov(X_t, X_t)$$

$$= Var(Z_1 + \dots + Z_t) = t\sigma^2$$

Funkcja kowariancji zależy od czasu, więc błądzenie losowe nie jest szeregiem czasowym stacjonarnym.

2) Proces ruchomej średniej rzędu 1: $X_t = Z_t + \theta Z_{t-1}$

$$\gamma(t+h,t) = Cov(Z_{t+h} + \theta Z_{t+h-1}, Z_t + \theta Z_{t-1})$$

$$= Cov(Z_{t+h} + \theta Z_{t+h-1}, Z_t) + Cov(Z_{t+h} + \theta Z_{t+h-1}, \theta Z_{t-1})$$

$$\gamma(t+h,t)=\begin{cases}\sigma^2(1+\theta^2), & h=0\\ \sigma^2\theta^2, & h=\pm 1\\ 0, & |h|>1\end{cases},\quad \mu(t)=E(Z_t)=0$$

Proces jest stacjonarny

3) Prognoza liniowa – najprostszy przykład: Niech X_1, X_2 zmienne losowe o wartościach oczekiwanych 0.

$$\min_a E(X_2-aX_1)^2=\frac{E(X_2X_1)}{E(X_1^2)}=\frac{\gamma(2,1)}{\gamma(1,1)}$$

$\widehat{X_2}=\frac{\gamma(2,1)}{\gamma(1,1)}X_1$ – najlepsza średniokwadratowa liniowa prognoza X_2 na podstawie X_1

- $X_t = Z_t, \ \{Z_t\}$ - biały szum (ciąg niezależnych, nieskorelowanych zmiennych losowych o wartości oczekiwanej $E(Z_t) = 0$ i wariancji $Var(Z_t) = \sigma^2$

- $X_t = m_t + Y_t = m_t + S_t + Z_t$

- $X_t = \beta_0 + \beta_1 t + Z_t$ model prostej regresji liniowej, zmienną objaśniającą jest czas (dzień, miesiąc, rok, kwartał, …) , trend jest liniowy

- $X_t = X_{t-1} + Z_t, \ X_0 = 0,$ - błądzenie losowe

- $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$ proces ARMA(p,q) (autoregressive moving average process)

- $X_t = \phi X_{t-1} + Z_t$ proces AR(1)

- $X_t = \phi X_{t-1} + Z_t + \theta Z_{t-1}$ proces ARMA(1,1)

- Model ARCH(q) (nagroda Nobla 2003: R.F. Engle - za metody analizy ekonomicznych szeregów czasowych ze zmienną w czasie wariancją

$$X_t = \sqrt{h_t} Z_t, \qquad h_t = c_0 + c_1 X_{t-1}^2 + \cdots + c_q X_{t-q}^2$$

Warunkowa wariancja (= h_t) zależy od poprzednich wartości procesu

Warunkowa wartość oczekiwana = 0 ($EZ_t = 0$)

Analiza dynamiki

Prognozowanie ma na celu przewidywanie przyszłości w sposób racjonalny z wykorzystaniem metod naukowych.

Ze względu na różne typy wielkości prognozowanych i różne sytuacje, w których prognozy są stosowane, nie ma jednej uniwersalnej metody ich realizacji.

Wybór metody prognozowania na podstawie danych szeregu czasowego zależy od występowania w nim określonego trendu oraz wahań (zakłóceń) o charakterze okresowych.

Metoda prognozowania Boxa-Jenkinsa – etapy konstruowania modelu prognozowania

1. Zidentyfikuj jeden (lub kilka) model dobrze pasujący do empirycznego szeregu czasowego
   (testujemy hipotezy o strukturze korelacyjnej szeregu czasowego)

2. Estymuj parametry modelu
   (etap estymacji podobny jak w regresji)

3. Przeprowadź kontrolę diagnostyczną modelu (sprawdzamy model na podstawie danych historycznych, wybieramy najlepszy)

4. Zastosuj uzyskany model

Analiza dynamiki

Równanie opisujące kształtowanie się określonego zjawiska w zależności od trendu, wahań okresowych i przypadkowych nosi nazwę **modelu wahań w czasie**.

Wyróżnia się dwa modele wahań:

Model addytywny w którym szereg czasowy stanowi sumę trendu, wahań okresowych i wahań przypadkowych

$$\boxed{y_t = f(t) + G(t) + \xi_t}, \qquad t = 1, \ldots ,n \qquad (1)$$

$f(t)$ - funkcja czasu, charakteryzująca tendencję rozwojową szeregu, nazywana **funkcją trendu**,

$G(t)$- funkcja czasu, charakteryzująca wahania okresowe,

ξ_t - zmienna losowa, charakteryzująca efekty oddziaływania wahań przypadkowych na zmienną prognozowaną, o wartości oczekiwanej równej 0 i skończonej wariancji.

W modelu addytywnym przyjmuje się stałą amplitudę wahań okresowych.

Analiza dynamiki

Model multiplikatywny w którym szereg czasowy jest iloczynem trendu, wahań okresowych i wahań przypadkowych

$$\boxed{y_t = f(t)\, G(t)\, \xi_t}, \qquad t = 1, \ldots ,n \qquad (2)$$

$f(t)$ - funkcja czasu, charakteryzująca tendencję rozwojową szeregu, nazywana **funkcją trendu**,

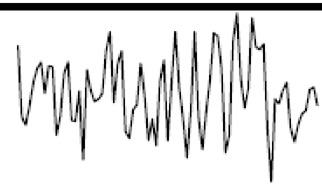
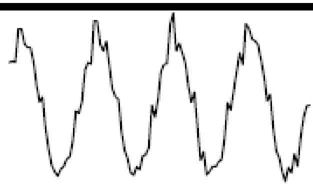
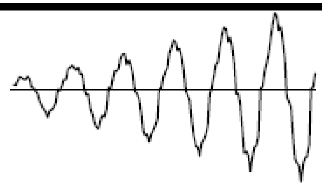
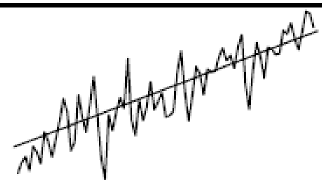
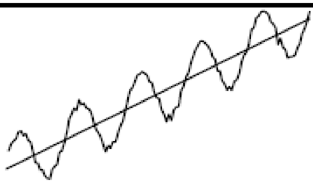
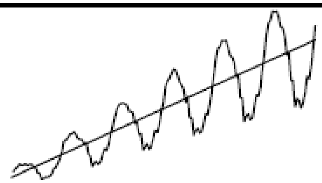
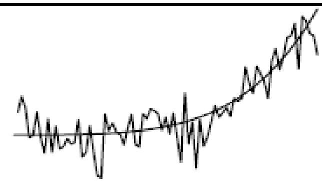
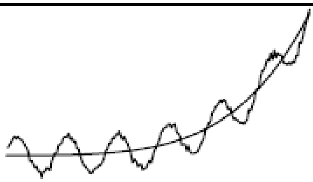
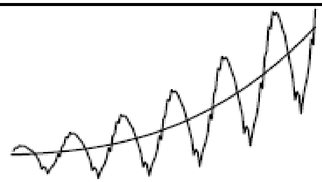
$G(t)$- funkcja czasu, reprezentująca wahania okresowe,

ξ_t - zmienna losowa, charakteryzująca efekty oddziaływania wahań przypadkowych na zmienną prognozowaną, o wartości oczekiwanej równej 1 i skończonej wariancji.

W modelu multiplikatywnym wahania okresowe są proporcjonalne do skali zjawiska

Analiza dynamiki

Podstawę doboru metody prognozowania stanowi prezentowana niżej klasyfikacja Pegel’a

		Sezonowość		
		Brak sezonowości	Sezonowość addytywna	Sezonowość multiplikatywna
Trend	Brak trendu			
	Trend addytywny			
	Trend multiplikatywny			

Analiza dynamiki

Zazwyczaj prognozowanie dotyczy procesów z wyraźnie zaznaczoną tendencją rozwojową. Realizujące je algorytmy wymagają identyfikacji trendu na podstawie posiadanych danych. Zadanie to, określane terminem **wygładzania (wyrównywania) szeregów czasowych,** jest wykonywane przy użyciu dwóch rodzajów metod:

- **mechanicznych** – opartych na średnich ruchomych,
- **analitycznych –**dopasowujących funkcje matematyczne do reprezentacji zmiennej zależnej szeregu.

Metody mechaniczne

Wyrównywanie (wygładzanie) szeregu czasowego za pomocą średnich ruchomych (kroczących) polega na zastępowaniu oryginalnych wyrazów szeregu czasowego przez średnie arytmetyczne obliczoną z nieparzystej, lub parzystej liczby wybranych wyrazów szeregu, które częściowo eliminują zarówno wahania okresowe (periodyczne) jak i przypadkowe.

Wartości wygładzone możemy interpretować jako wartości oczyszczone z zakłóceń losowych.

Jest to najłatwiejsza metoda wyrównywania szeregu czasowego.

Aby wyeliminować wahania okresowe, średnie ruchome powinny być obliczane z takiej liczby wyrazów oryginalnego szeregu, które odpowiadają liczbie pomiarów w cyklu wahań.

Przykładowo, przy rocznym cyklu wahań i miesięcznych danych średnia powinna być obliczana z 12 danych.

Średnie ruchome

Średnie ruchome k-okresowe

Oznaczmy kolejne wartości szeregu czasowego jako

$$y_1, y_2, y_3, \cdots, y_{n-2}, y_{n-1}, y_n \,.$$

Średnie ruchome wyznaczamy różnie w zależności od ich długości (k).

- Gdy k jest nieparzyste (np. $k = 3$), to średnie ruchome wyznacza się następująco

$$\boxed{\bar{y}_2 = \frac{y_1 + y_2 + y_3}{3}, \quad \bar{y}_3 = \frac{y_2 + y_3 + y_4}{3}, \ldots, \bar{y}_{n-1} = \frac{y_{n-2} + y_{n-1} + y_n}{3}}\,.$$

Przy wyznaczaniu średniej ruchomej z trzech okresów ($k = 3$) otrzymujemy szereg wygładzony o dwa wyrazy krótszy (bez wyrazu pierwszego i ostatniego), przy $k = 5$ krótszy o cztery wyrazy itd.

Średnie ruchome

Średnie ruchome k-okresowe

- Gdy k jest parzyste, to wyznaczamy tzw. **średnie ruchome scentrowane**. Np. dla $k = 4$ wzory mają postać

$$\bar{y}_3 = \frac{\frac{1}{2}y_1 + y_2 + y_3 + y_4 + \frac{1}{2}y_5}{4}$$

$$\bar{y}_4 = \frac{\frac{1}{2}y_2 + y_3 + y_4 + y_5 + \frac{1}{2}y_6}{4}$$

....

$$\bar{y}_{n-2} = \frac{\frac{1}{2}y_{n-4} + y_{n-3} + y_{n-2} + y_{n-1} + \frac{1}{2}y_n}{4}$$

Średnie ruchome

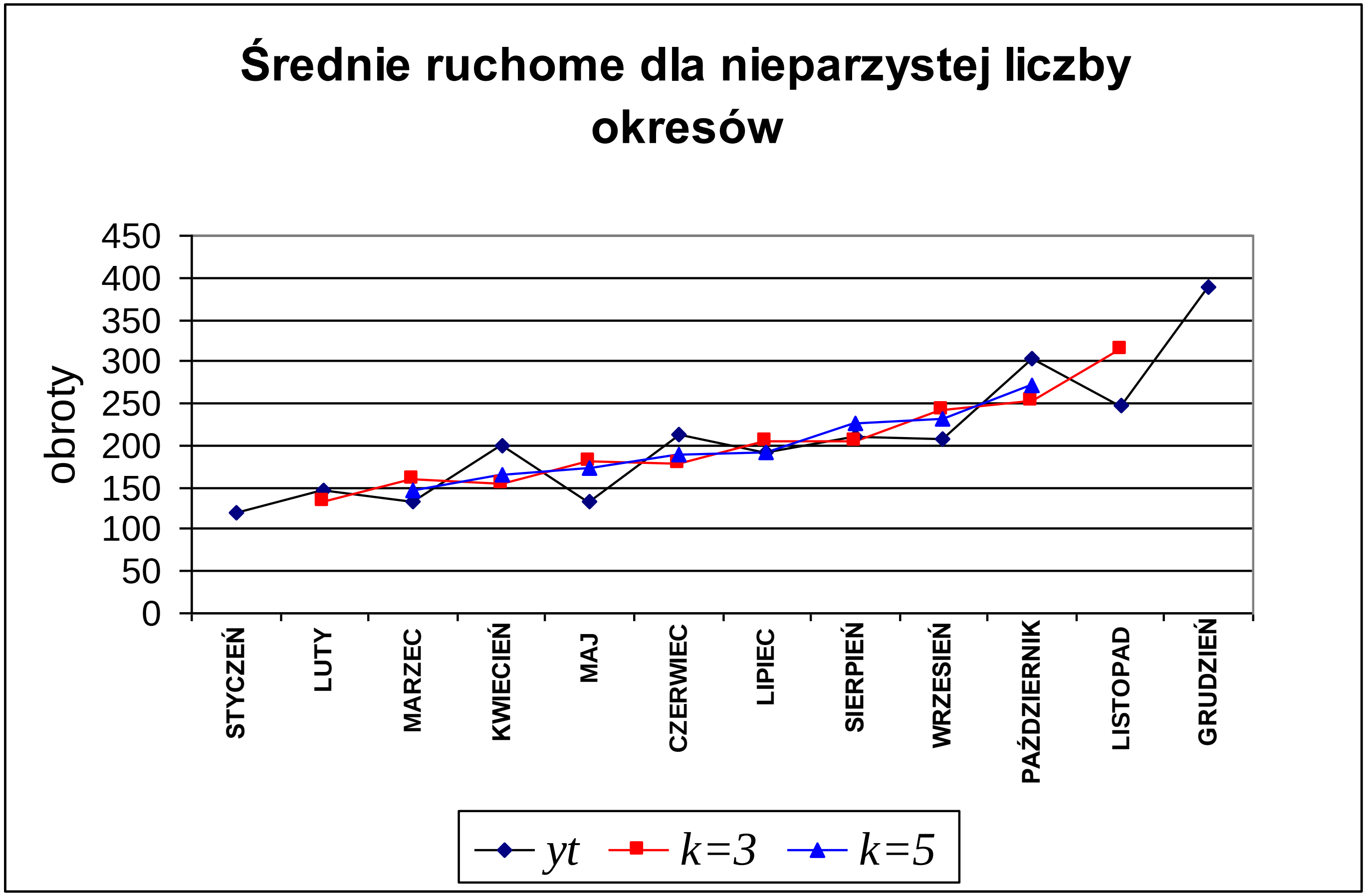
Przykład

Druga kolumna tabeli przedstawia obroty osiągnięte przez pewną firmę na przestrzeni roku. W kolejnych kolumnach zapisano wartości średnich ruchomych. Dodatkowo są one prezentowane na załączonych wykresach

okres	obroty	średnie ruchome			
		nieparzyste		parzyste	
t	y_t	$k=3$	$k=5$	$k=4$	$k=6$
STYCZEŃ	120				
LUTY	146	133			
MARZEC	132	159	146	151	
KWIECIEŃ	200	155	164	161	163
MAJ	132	181	174	177	174
CZERWIEC	212	179	189	185	186
LIPIEC	192	205	191	196	201
SIERPIEŃ	211	204	225	217	219
WRZESIEŃ	209	241	232	236	244
PAŹDZIERNIK	303	253	272	265	
LISTOPAD	247	313			
GRUDZIEŃ	390				

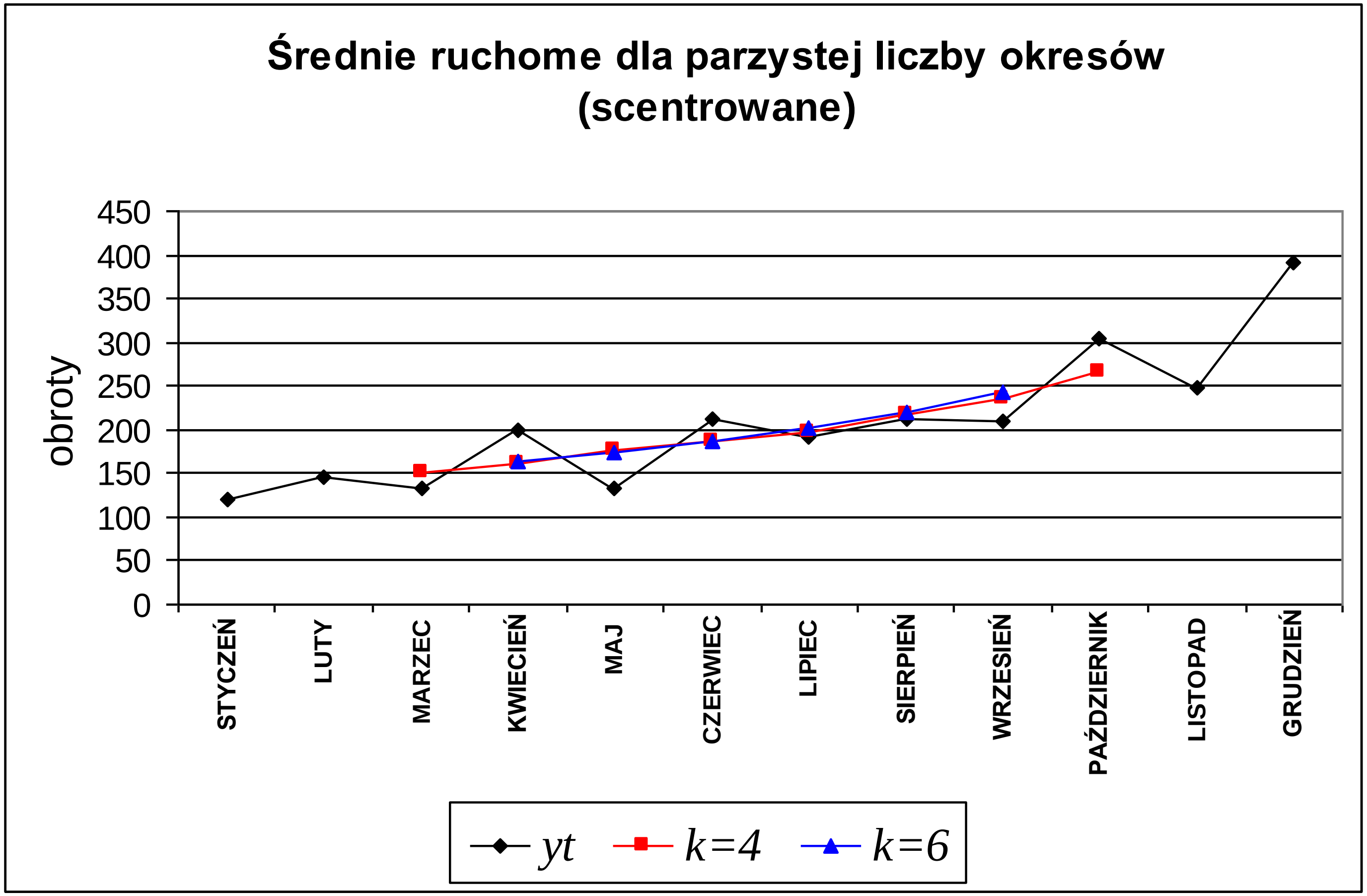
Średnie ruchome

Przykład



Średnie ruchome

Przykład



Wygładzanie analityczne

Wygładzanie analityczne (model liniowy)

Wygładzanie szeregu czasowego polega tutaj na oszacowaniu liniowej funkcji trendu

$$\boxed{\hat{y}_t = at + b}$$

Współczynniki a i b wyliczamy na podstawie danych z szeregu czasowego stosując następujące wzory

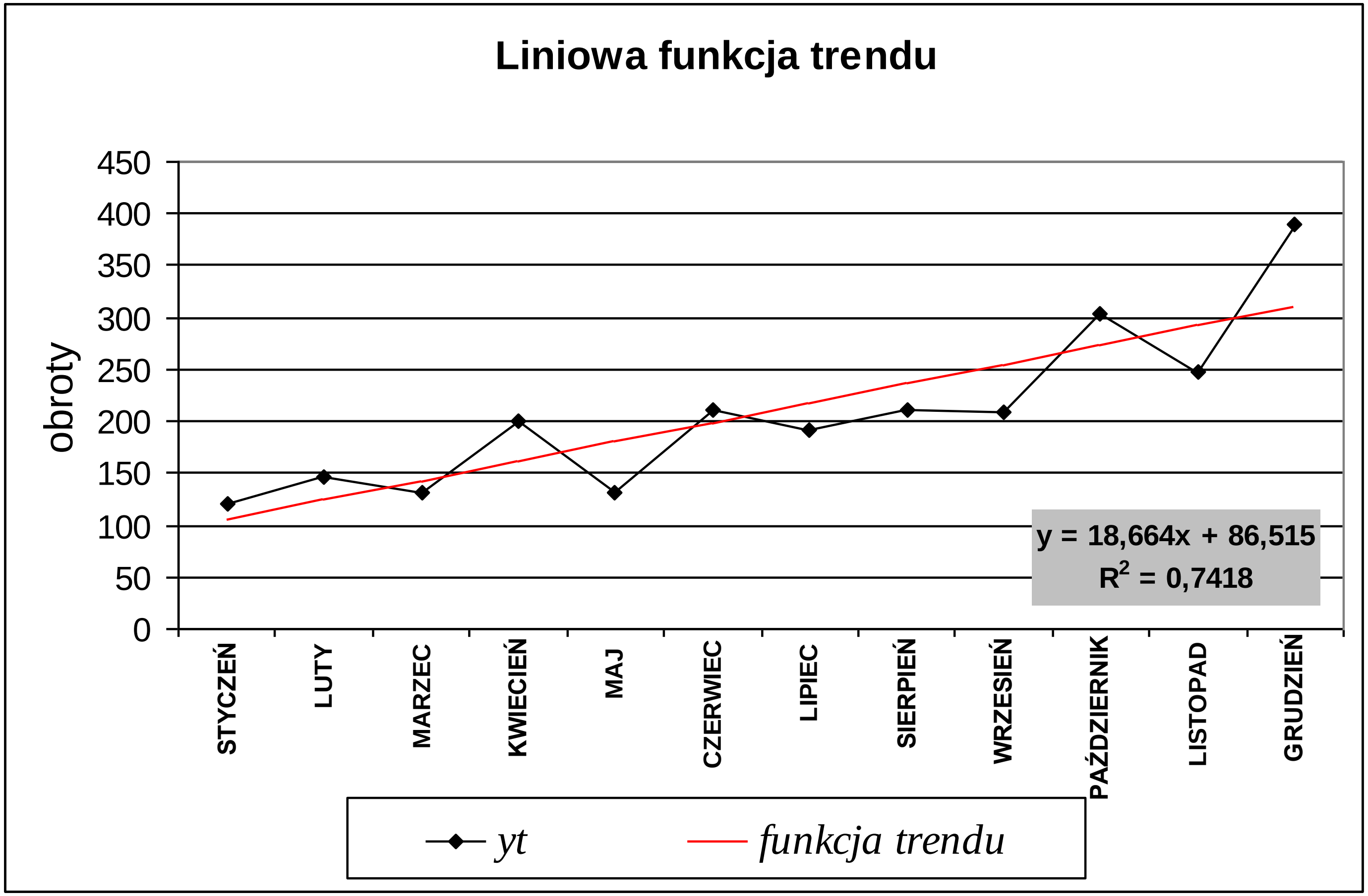
$$\boxed{a = \frac{\sum_{t=1}^{n} \left(t - \bar{t}\right)\left(y_t - \bar{y}\right)}{\sum_{t=1}^{n} \left(t - \bar{t}\right)^2}, \qquad b = \bar{y} - a\bar{t}}$$

współczynnik a reprezentuje stały przyrost (dodatni, bądź ujemny) wartości zmiennej prognozowanej w kolejnych jednostkach czasu,

współczynnik b – określa poziom zjawiska w okresie wyjściowym (tzn. dla $t = 0$).

Wygładzanie analityczne

Przykład



Wygładzanie analityczne

Miarą dopasowania modelu do wartości rzeczywistych badanej zmiennej jest **współczynnik determinacji**

$$R^2 = \frac{\sum_{t=1}^{n} \left(\hat{y}_t - \bar{y}\right)^2}{\sum_{t=1}^{n} \left(y_t - \bar{y}\right)^2}$$

Im wartość R^2 jest bliższa 1, tym dopasowanie modelu do danych empirycznych jest lepsze.

W prezentowanym przykładzie $R^2 = 0{,}74$. Oznacza to, że liniowa funkcja trendu w 74% opisuje (wyjaśnia) zmiany wielkości obrotów firmy na przestrzeni roku.

Wygładzanie analityczne

Inne funkcje trendu

- wielomian stopnia drugiego (parabola)

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2, \quad \alpha_2 \neq 0$$

- funkcja wykładnicza:

$$y_t = e^{\alpha + \tilde{\beta} t},$$

$$y_t = \alpha \beta^t,$$

- funkcja potęgowa:

$$y_t = \alpha t^{\beta}, \quad \beta > 1$$

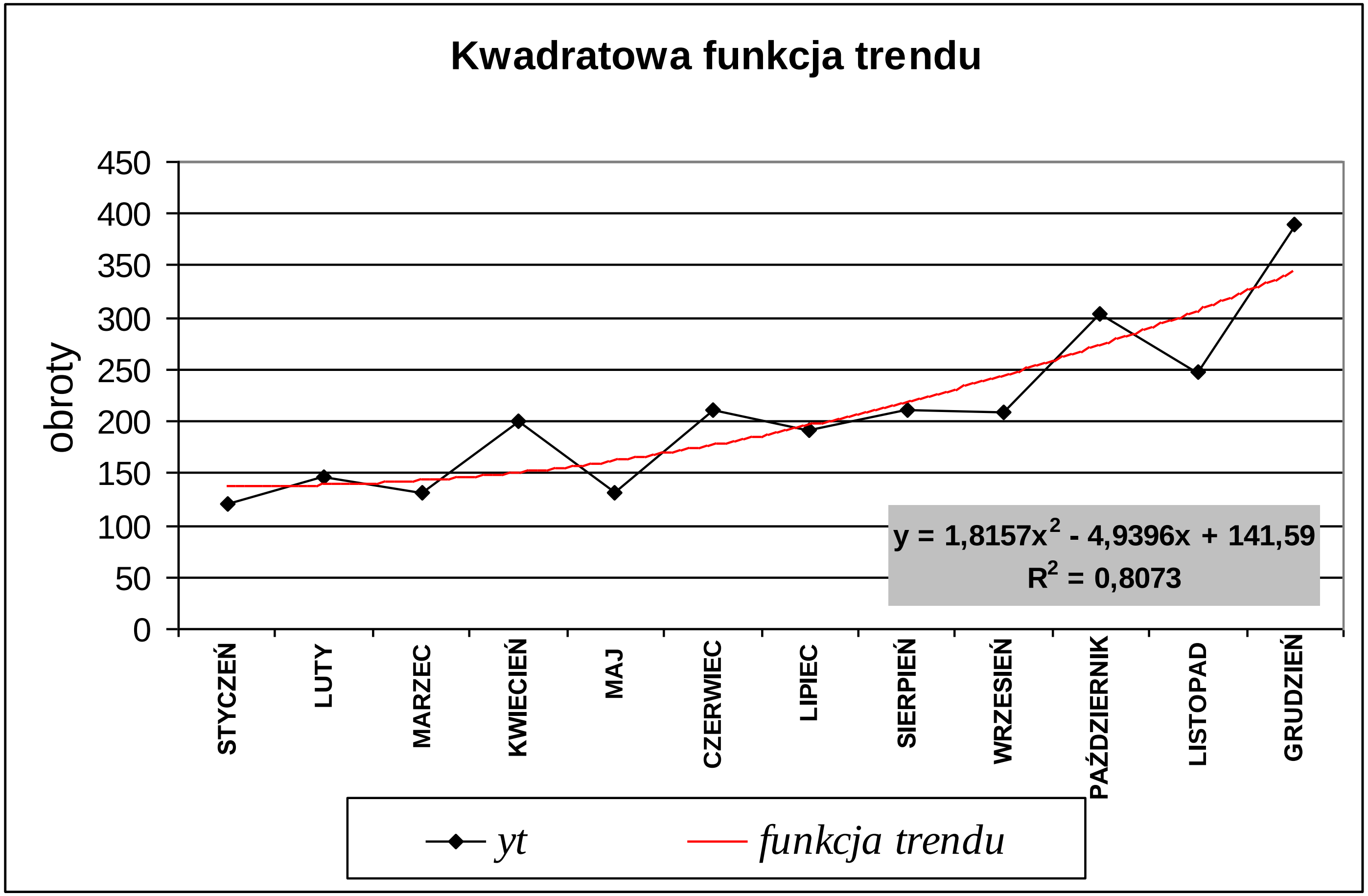
- funkcja logarytmiczna:

$$y_t = \alpha + \beta \ln t$$

Wykres kwadratowej funkcji trendu dla danych przykładowych jest prezentowany na kolejnym slajdzie. Wartość współczynnika determinacji 0,81 wskazuje na lepsze dopasowanie modelu kwadratowego, w porównaniu z funkcją liniową

Wygładzanie analityczne

Przykład



Metody prognozowania

Metoda średniej ruchomej prostej *(Moving Average Model)*
Wykorzystuje się średnią danych z k ostatnich okresów do prognozowania wartości w kolejnym okresie

$$\boxed{\hat{y}_t^* = \frac{1}{k}\sum_{i=t-k}^{t-1} y_i}$$

k –liczba okresów branych pod uwagę w obliczeniach.
Wada wszystkie wartości historyczne mają takie same wagi

Metody prognozowania

Opracowano wiele metod prognozowania wykorzystujących szeregi czasowe.
W niniejszej prezentacji zostały omówione wybrane metody, typowe dla różnych typów szeregów.
Metoda średniej ruchomej prostej *(Moving Average Model)*
Wykorzystuje się średnią danych z k ostatnich okresów do prognozowania wartości w kolejnym okresie

$$\boxed{\hat{y}_t^* = \frac{1}{k} \sum_{i=t-k}^{t-1} y_i}$$

k –liczba okresów branych pod uwagę w obliczeniach.
Jeżeli $k = 1$ otrzymujemy tzw. **metodę naiwną** *(Random Walk Model)*

$$\hat{y}_t^* = y_{t-1}$$

prognozowana wielkość w następnym okresie jest taka sama jak w ostatnim (najbliższym aktualnemu)

Metody prognozowania

Jeżeli k = 1 otrzymujemy tzw. **metodę naiwną** *(Random Walk Model);*

$$\boxed{\hat{y}_t^* = y_{t-1}}$$

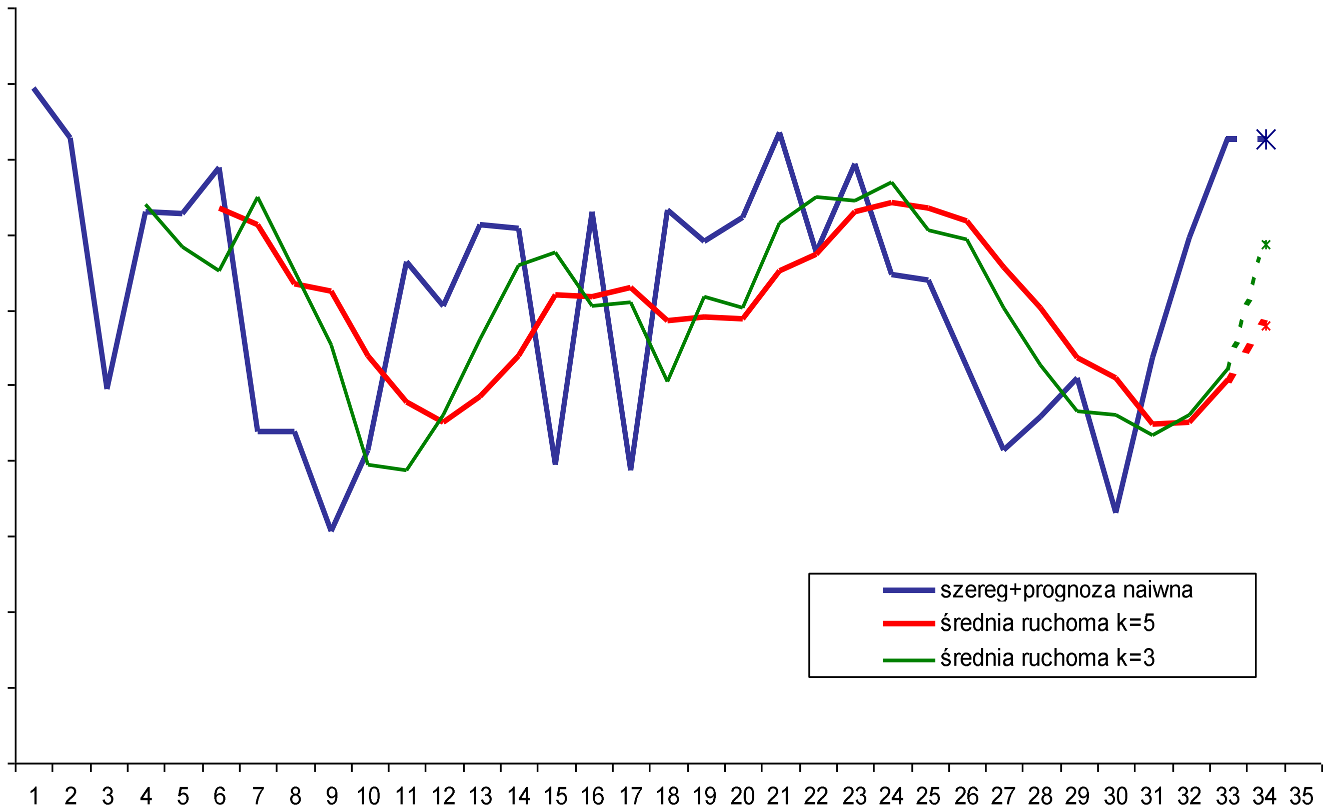
prognozowana wartość w następnym okresie jest taka sama jak w ostatnim (najbliższym aktualnemu)

Jeżeli k =t-1 dostajemy **metodę średniej arytmetycznej** *(Simple Mean Forecasting Model, long-term mean model)*

$$\boxed{y_{T+1}^* = \frac{1}{T}\sum_{t=1}^{T} y_t}$$

prognozowana wartość jest równa średniej ze wszystkich dotychczasowych obserwacji

Metody prognozowania



Metody prognozowania

Metoda średniej ruchomej ważonej
(Weighted Moving Average Model)

$$\boxed{\hat{y}_t^* = \sum_{i=1}^{k} y_{t-k-1+i} w_i}$$

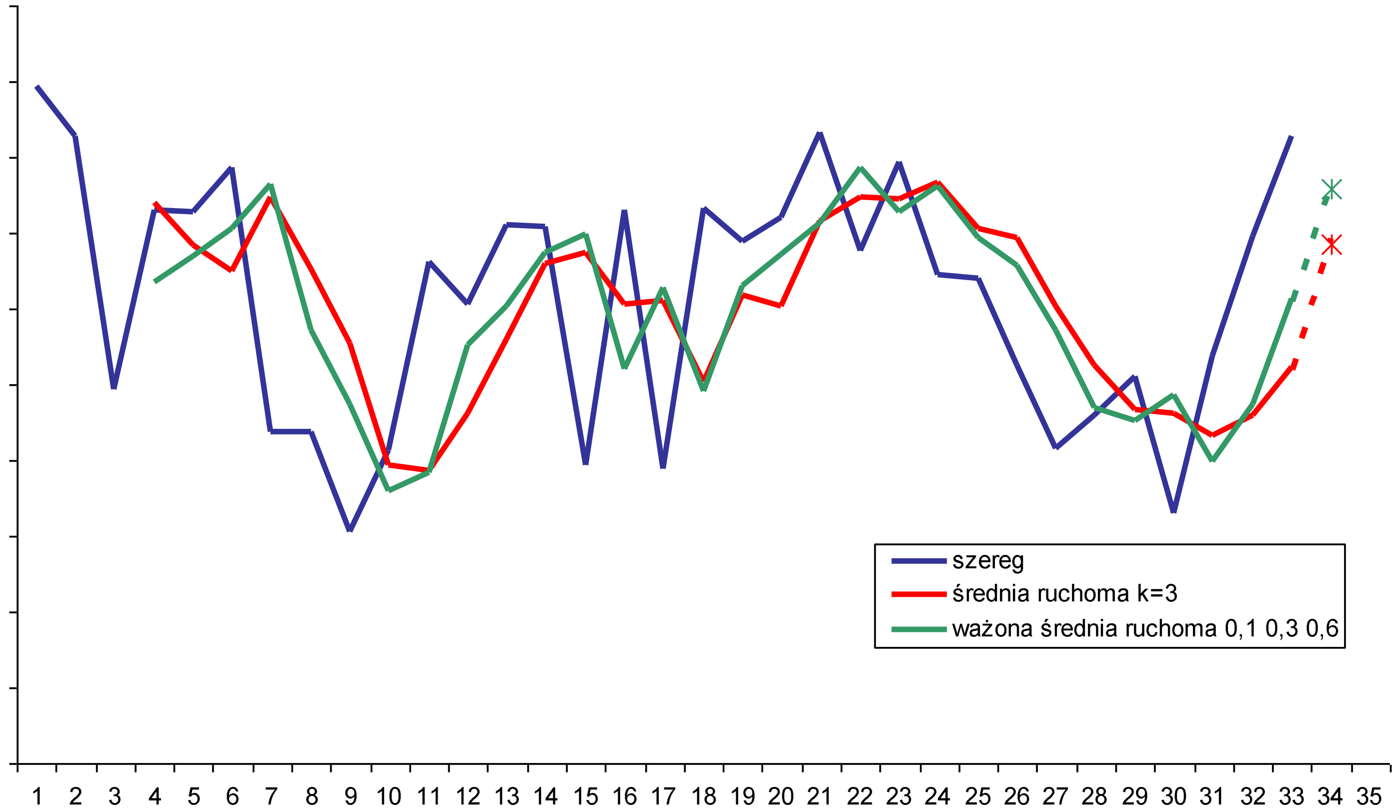
$w_1, w_2, \ldots, w_k$ – współczynniki wagowe,
$w_1 < w_2 < \ldots < w_k$ oraz $w_1 + w_2 + \ldots + w_k = 1$

często wykorzystywane wagi, to wagi liniowe:

$$w_t = \frac{2t}{T(T+1)}$$ dla $t = 1, 2, \ldots, T$,

np. $T = 3 \Rightarrow$ $w_1 = \frac{2}{12}$ $w_2 = \frac{4}{12}$ $w_3 = \frac{6}{12}$

Metody prognozowania



Metody prognozowania

Metoda wygładzania wykładniczego (*Exponential smoothing*)
Jest to rozwinięcie techniki średniej ruchomej (ważonej) w której wagi wyliczane są zgodnie z funkcją wykładniczą. Powoduje wykładnicze „postarzanie" informacji.

Wygładzanie wykładnicze

$$\boxed{\hat{y}_t^* = \alpha y_{t-1} + (1-\alpha)\hat{y}_{t1}}$$

α parametr wygładzania

α przeważnie znajduje się między 0,05 i 0,5

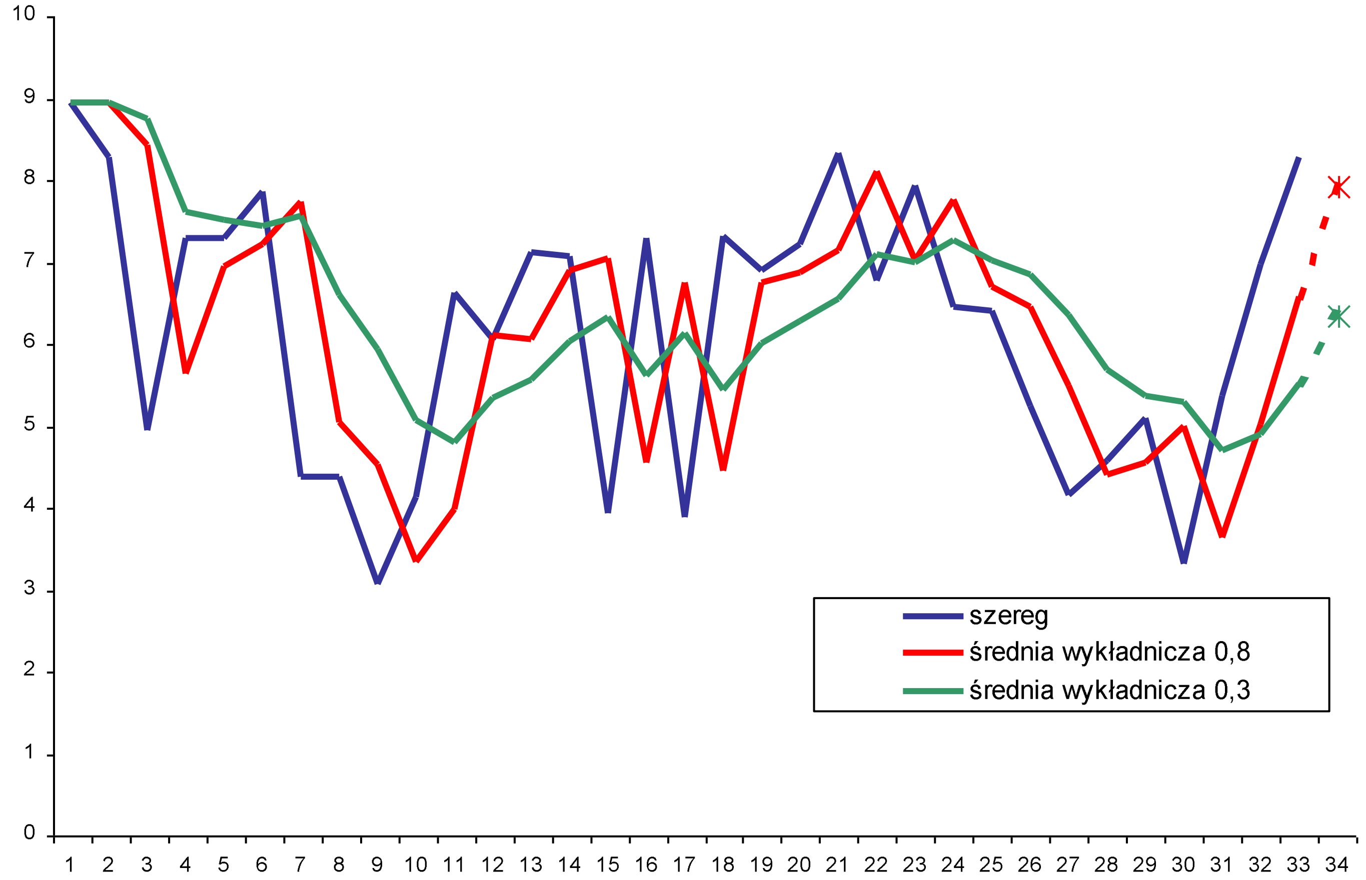
α niskie=> silniej uwzględnia dawniejsze dane,

α wysokie => silniej uwzględnia ostatnie dane,

gdy α rośnie => spada istotność starszych okresów,

gdy $\alpha = 1$ => (prognozowanie naiwne).

Zaleta: potrzeba niewielkiej liczby danych z przeszłości


10
9
8
7
6
5
4
3
2
1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34
szereg
średnia wykładnicza 0,8
średnia wykładnicza 0,3

Metody analityczne

Przyszłą (prognozowaną wartość zmiennej uzyskuje się przez ekstrapolację funkcji trendu, tj. przez podstawienie do modelu w miejsce zmiennej niezależnej numeru momentu (lub okresu) - T, na który wyznacza się prognozę

$$\boxed{y_T^* = f(T), \quad T > n}$$

Jest to prognoza punktowa.

Wykorzystanie modelu analitycznego jest uzasadnione w przypadku stabilności (w czasie) czynników kształtujących poziom badanego zjawiska

Model trendu liniowego

$$\hat{y}_t = a_0 + a_1 t$$

Do oceny precyzji prognozy wykorzystuje się **błąd prognozy** „ex ante”, który w przypadku liniowej funkcji trendu jest określony wzorem

$$v_T = S \cdot \sqrt{\frac{(T-\bar{t})^2}{\sum_{t=1}^{n}(t-\bar{t})^2} + \frac{1}{n} + 1}$$

gdzie S - **odchylenie standardowe reszt** określone wzorem

$$s = \sqrt{\frac{1}{n-(m+1)}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$$

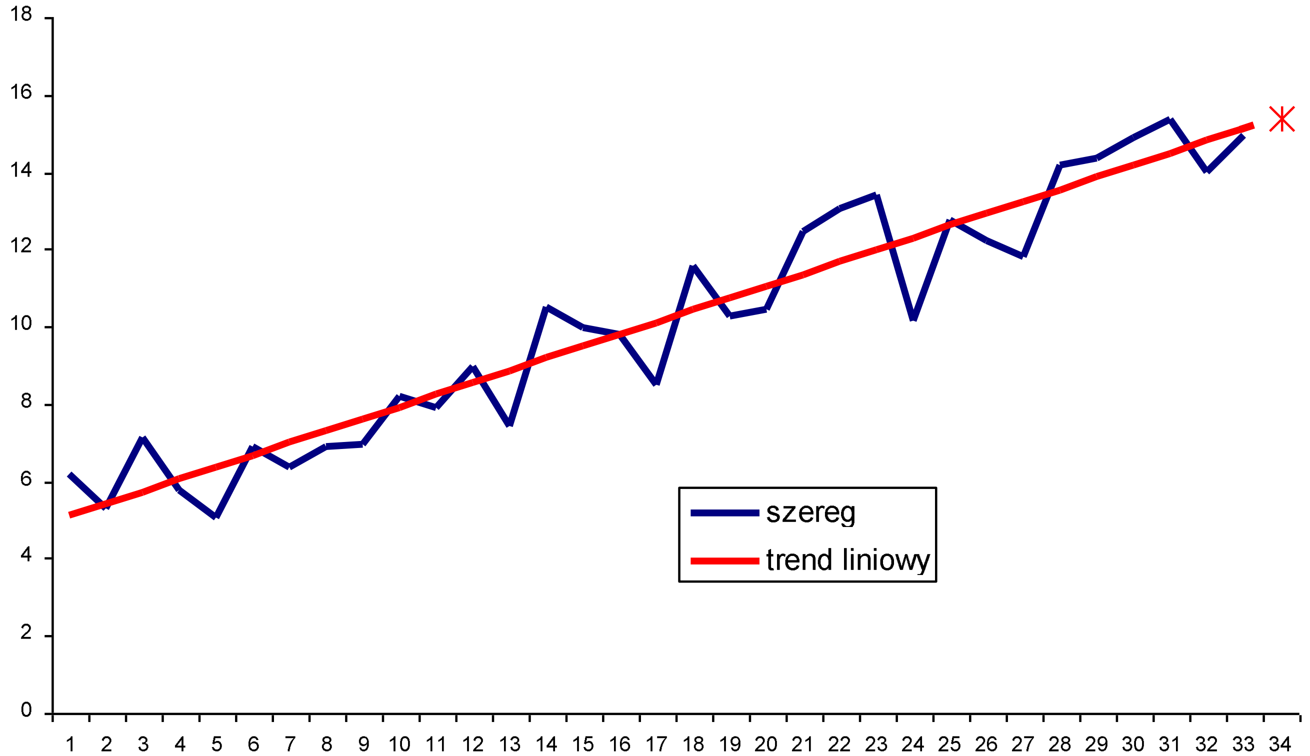
y_t - obserwowana wartość zmiennej Y w momencie (lub okresie t),

$\hat{y}_t$ - teoretyczna wartość zmiennej Y

n - liczba obserwacji

m - liczba zmiennych objaśniających

Metody prognozowania



Metody prognozowania

Jeżeli w procesie weryfikacji modelu można przyjąć hipotezę o normalnym rozkładzie reszt modelu, to przedział ufności, na poziomie ufności $1-\alpha$, dla prognozowanej wartości szeregu (**przedział prognozy**) ma postać

$$\boxed{[y_T^* - uv_T, \quad y_T^* + uv_T]}$$

Współczynnik *u* jest kwantylem rzędu (1 – *a*/2) rozkładu *t*-Studenta o (*n*-2) stopniach swobody, lub rozkładu normalnego (gdy *n* >30).

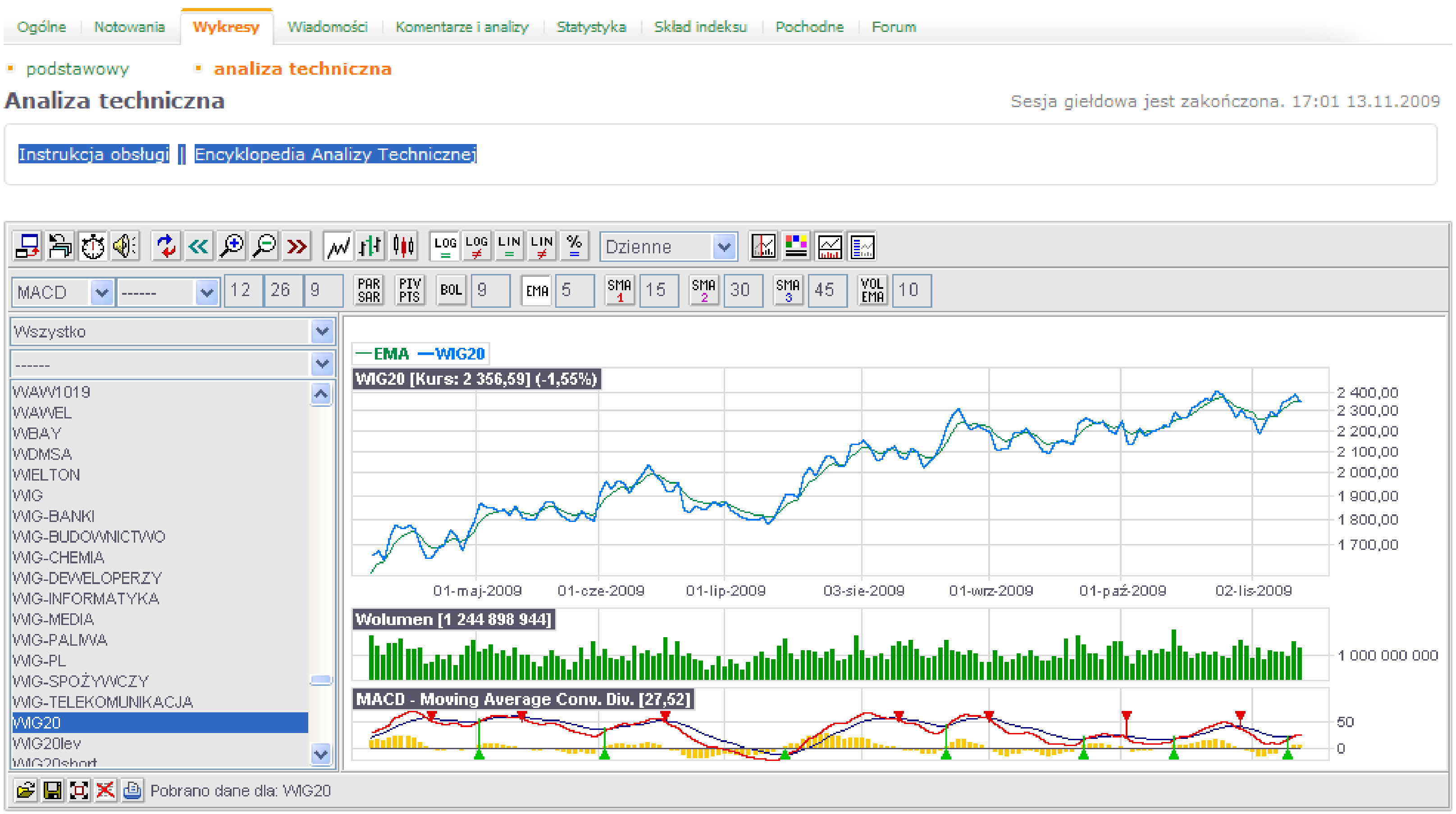
Analiza techniczna

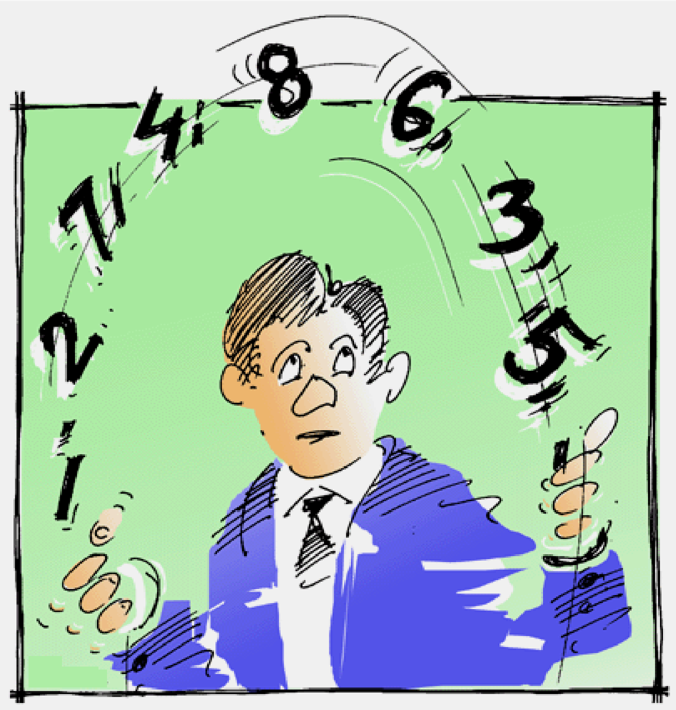
Metody badania szeregów czasowych mają duże znaczenie praktyczne. Przykładem jest **Analiza techniczna** - zbiór wyspecjalizowanych metod analizy szeregów czasowych mających na celu prognozę przyszłych cen (kursów) papierów wartościowych, walut czy surowców na podstawie cen historycznych.

Kolejny slajd przedstawia widok okna aplikacji umożliwiającej za pośrednictwem Internetu analizę w trybie interaktywnym kursów akcji spółek notowanych na Warszawskiej Giełdzie Papierów Wartościowych.

Analiza techniczna

http://gielda.onet.pl/wig20,18489,102,2,215,1,profile-wykresy-analiza





Dziękuję za uwagę