ETAPY BUDOWY MODELU EKONOMETRYCZNEGO

1. Sformułowanie modelu

   a. wybór zmiennych: $y, x_1, x_2,...$

   b. wybór postaci matematycznej modelu: liniowa, potęgowa,...

2. Zebranie danych statystycznych (różne źródła)

3. Selekcja zmiennych objaśniających

4. Estymacja parametrów modelu:

   a. parametrów strukturalnych: $a_0, a_1, a_2,...$

   b. parametrów stochastycznych: $s(a_i), s(y), R^2, R$

5. Weryfikacja modelu

MODEL BEZ WERYFIKACJI NIE MA ŻADNEJ WARTOŚCI

NIE NALEŻY KORZYSTAĆ Z PROGRAMÓW KOMPUTEROWYCH NIE DAJĄCYCH MOŻLIWOŚCI WERYFIKACJI

6. Interpretacja modelu

ETAP 1a. WYBÓR ZMIENNYCH

- zmienna objaśniana Y:

- zmienne objaśniające X_i (jak najwięcej dla modelu przyczynowo-skutkowego) z następujących źródeł (w kolejności):
  - — teoria danej dziedziny wiedzy
  - — doświadczenie zleceniodawcy i statystyka
  - — metoda prób i błędów (intuicyjnie)

- wybrane zmienne muszą mieć dużą zmienność (V>30%)

- najczęstszy błąd — *„masło maślane”* prowadzące do związku funkcyjnego i nie dające żadnej informacji o zmiennej objaśnianej

ETAP 1b. WYBÓR POSTACI MATEMATYCZNEJ

- modele przyczynowo-skutkowe — najbardziej zalecane jest równoczesne prowadzenie obliczeń dla dwu postaci:
  - — liniowej $y = \sum a_i x_i + \xi$
  - — potęgowej $y = \prod x_i^{a_i} \varepsilon \qquad \ln y = \sum a_i \ln x_i + \xi$
  - — stosuje się tu modele nieliniowe o narzuconej postaci nieliniowej, których parametry ustala się przez programowanie liniowe lub innymi metodami

- modele tendencji rozwojowej:
  - — funkcja liniowa
  - — proste funkcje nieliniowe
  - — wielomiany
  - — modele kombinowane: trend + wahania okresowe

ETAP 2. GROMADZENIE DANYCH STATYSTYCZNYCH

- rodzaje danych: dane przekrojowe i szeregi czasowe
- źródła danych: roczniki statystyczne, różne działy przedsiębiorstwa, badania marketingowe, wywiady itd.
- wiarygodność danych: do jakiego celu zostały one przygotowane?
- porównywalność danych: inflacja (ceny bieżące a ceny stałe), zmiany procesów technicznych
- zmienność zjawisk: trzeba sprawdzić, czy wybrana w etapie 1a zmienna jest rzeczywiście zmienną losową

$$V_x = \frac{s(x)}{\bar{x}} 100\%$$

V_x musi wynosić co najmniej 30-40%

ETAP 3. SELEKCJA ZMIENNYCH OBJAŚNIAJĄCYCH

KAŻDĄ ZMIENNĄ X WYTYPOWANĄ W ETAPIE 1a TRAKTUJEMY JAKO KANDYDATKĘ NA ZMIENNĄ OBJAŚNIAJĄCĄ

- w modelu nie może być zbyt wielu zmiennych (nieczytelny)
- kandydatka może nie mieć wpływu na zmienną Y
- kandydatka może wnosić prawie tę samą informację o Y co inna kandydatka
- dwie bardzo podobne kandydatki mogą sobie nawzajem przeszkadzać (efekt katalityczny)

Kryteria, jakie musi spełniać kandydatka X_i, aby nadawała się do modelu:

- musi być silnie powiązana ze zmienną Y
- nie może być powiązana z inną kandydatką X_j

Metody selekcji zmiennych objaśniających:

- badanie istotności korelacji
- grafowa
- Hellwiga (pojemności informacji)
- taksonomiczne (clustering)

$r_{x_i x_j}$

Przykład Macierz powiązania zmiennych ze sobą przedstawia tabela (n=20). Jak ją można zinterpretować? Które zmienne są powiązane ze sobą w sposób istotny?

	Y	X_1	X_2	X_3
Y		0,52	0,64	−0,21
X_1	0,52		0,82	−0,18
X_2	0,64	0,82		0,08
X_3	−0,21	−0,18	0,08	

Macierz współczynników korelacji

- **Testowanie istotności współczynnika korelacji**

H_0: $\rho = 0$

H_1: $\rho \neq 0$

Można przeprowadzić testem Studenta (t);

Wallace'a-Snedecora (*R*)

TEST Wallece'a-Snedecora

Fragment tablicy rozkładu Wallece'a-Snedecora

Stopnie swobody	Liczba zmiennych					
	2		3		4	
	0,05	0,01	0,05	0,01	0,05	0,01
8	0,632	0,765	0,726	0,827	0,777	0,860
18	0,444	0,561	0,532	0,633	0,587	0,678
28	0,361	0,463	0,439	0,530	0,490	0,573

Reguła decyzyjna: — jeżeli $|r_{obl}| > R_{tabl}$, odrzucamy H_0 (korelacja istotna)

— jeżeli $|r_{obl}| < R_{tabl}$, przyjmujemy H_0 (brak korelacji)

W przykładzie, jeżeli przyjmiemy $\alpha = 0{,}05$, to $R_{tabl} = 0{,}444$.

	Y	X_1	X_2	X_3
Y		0,52	0,64	–0,21
X_1	0,52		0,82	–0,18
X_2	0,64	0,82		0,08
X_3	–0,21	–0,18	0,08	

Pozostan□zatem tylko trzy istotne powi□zania: $y\text{-}x_1$, $y\text{-}x_2$, $x_1\text{-}x_2$

Test Studenta (t)

H_0: $\rho = 0$

H_1: $\rho \neq 0$

$$r_{kr} = \sqrt{\frac{\frac{t_\alpha^2}{n-2}}{1+\frac{t_\alpha^2}{n-2}}}$$

$r_{x_i x_j}$, $r_{y x_i}$

Reguła decyzyjna: — jeżeli $|r_{ij}| > r_{kr}$, odrzucamy H_0 (korelacja istotna)

— jeżeli $|r_{ij}| < r_{kr}$, przyjmujemy H_0 (brak korelacji)

W przykładzie, jeżeli przyjmiemy $\alpha = 0{,}05$, to $r_{kr} = 0{,}3778$.

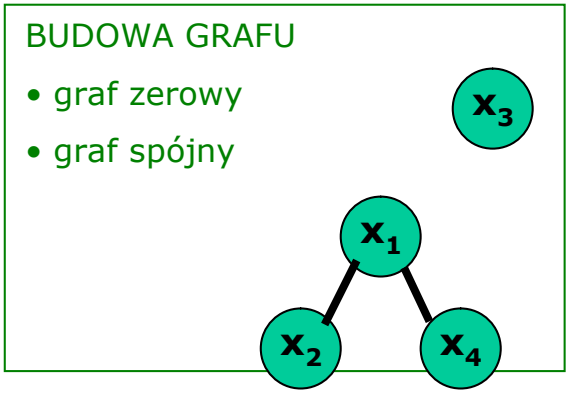
Pozostan□równie□tylko trzy istotne powi□zania: $y\text{-}x_1$, $y\text{-}x_2$, $x_1\text{-}x_2$

Metoda grafowa

- obliczenie macierzy współczynników: $r_{x_i x_j}, r_{y x_i}$

- wyłączenie z macierzy powiązań nieistotnych
- budowa grafu z powiązań nieistotnych
- wybranie zmiennych na podstawie grafu

Przykład grafu



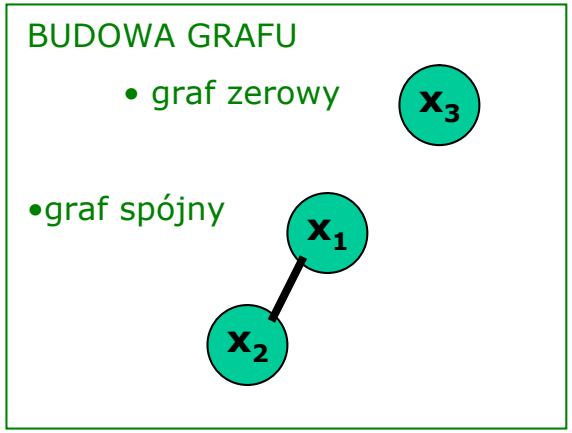
DO MODELU WYBIERA SIĘ ZMIENNE:

- każdy graf zerowy (jeśli $|r_{yx}|>0,1$)
- po jednej reprezentantce grafu spójnego;

- reprezentantką grafu jest zmienna, która ma najwięcej powiązań z innymi kandydatkami
- jeżeli kilka zmiennych ma tą samą maksymalną liczbę powiązań, wybiera się tą, która jest najsilniej powiązana ze zmienną Y (max r_{yx})
- jeżeli graf jest rozległy, może mieć dwie reprezentantki, ale muszą one leżeć na przeciwległych stronach grafu

Przykład cd. Wybierzemy zmienną x_2 (reprezentantka grafu spójnego) i zmienną x_3 (graf zerowy)

	Y	X_1	X_2	X_3
Y		0,52	0,64	-0,21
X_1	0,52		0,82	-0,18
X_2	0,64	0,82		0,08
X_3	-0,21	-0,18	0,08	



Metoda Hellwiga

- wypisujemy wszystkie możliwe kombinacje kandydatek;

jest ich $l=2^m-1$

- obliczamy pojemność indywidualną nośnika informacji (dla każdej zmiennej w każdej kombinacji)

$$h_{lj} = f(r_{yx_j}, r_{x_i x_j})$$

$$h_{lj} = \frac{r_{y,j}^2}{1 + \sum_{i \neq j} |r_{ij}|}$$

$i,j = 1,2, \ldots m$ $\quad l = 2m-1$ $\quad m$ – ilość kandydatek

r_j – wsp. korelacji j-tej kandydatki ze zmienną objaśnianą

r_{ij} – wsp. korelacji i-tej i j-tej zmiennej

- obliczamy pojemność całkowitą dla każdej kombinacji $H_l = \sum_j h_{lj}$

- wybieramy kombinację o największej pojemności $H_l = \max$

Metody taksonomiczne

Taksonomia wrocławska (cluster analysis) to metoda grupowania obiektów (zmiennych) w grupy jednorodne pod względem *n* cech (wymiarów) ilościowych. Podstawą grupowania jest odległość euklidesowa, która w przypadku zmiennych

$$d_{ij} = f(r_{x_i x_j})$$

WSZYSTKIE METODY TO SELEKCJA WSTĘPNA

ETAP 4. ESTYMACJA PARAMETRÓW MODELU

Cel etapu: wyznaczenie parametrów strukturalnych i stochastycznych

Estymacja: szacowanie parametrów populacji na podstawie próbki

Metody estymacji: analiza regresji i korelacji (KMNK i inne)

Założenia dla KMNK

1. Zmienne losowe są zmiennymi nie powiązanymi ze sobą (nie występuje współliniowość)
2. Składnik losowy ξ jest zmienną losową: $E(\xi)=0$; $V^2=const$ (stała wariancja, niezależna od zmiennej *x* lub *t)*
3. Składnik losowy ξ nie jest powiązany ze zmiennymi objaśniającymi
4. Wartości reszt u_i są niezależne od siebie $u_i = y_i - \hat{y}_i$
5. *m* - liczba zmiennych objaśniających;

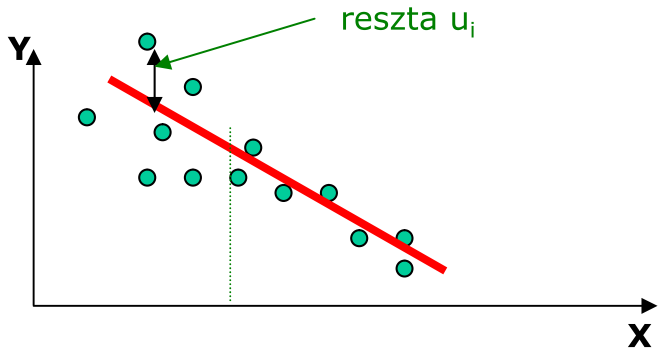
   n – liczność próby: m<n

Skutki niedotrzymania za□□e□i □rodki zaradcze

1. Model nieprzydatny; niekiedy absurdalny (□le uwarunkowane dane)

   Stosuje si□ du¿¹ próbk□(*n*>100), regresj□grzbietow¹ (*ridge regression*)

2. Lewa cz□ê□ãzbioru ma du¿¹ wariancj□, a prawa — wariancj□ma□¹ Stosuje si□specjalny wariant MNK z korekt¹ na ró□ne wariancje



$u_t = y_t - \hat{y}_t$	u_{t-1}
u_1	—
u_2	u_1
u_3	u_2
u_4	u_3
u_5	u_4

3. Je□li reszty u_i s¹ ze sob¹ powi¹zane (skorelowane) tzn. □e wyst□puje *autokorelacja sk□adnika losowego* (najcz□□ciej zjawisko wyst□puje przy szeregach czasowych). Oznacza to, □e istnieje istotna zale□no□□:

$$u_t = f(u_{t-k}) \qquad t = 1,2,...$$

Przyczyny autokorelacji:
zak□ócenia (dodatnie lub ujemne) w jednym okresie wp□ywaj¹ na poziom zjawiska w nast□pnych okresach

Wyst□powanie autokorelacji powoduje nieprzydatno□□modelu

4. Sk□adnik losowy jest skorelowany ze zmienn¹ objaniaj□c¹, wtedy gdy zosta□a pomini□ta jaka□ wa□na zmienna – przyczyna. Model taki nie ma □adnej warto□ci; trzeba dba□o jak najwy□szy wspó□czynnik determinacji (R^2>0,9)

ETAP 5. WERYFIKACJA MODELU

Cele: 1. opis rzeczywisto□ci (populacji generalnej)

2. dok□adna (ostateczna) selekcja zmiennych objaśniaj□cych

3. poznanie sk□adnika losowego (spe□nienie za□o□e□ KMNK)

Narz□dzia: hipotezy i testy statystyczne

Metodyka: ka□de równanie oddzielnie; weryfikacja obejmuje 9 etapów (od najmniej do najbardziej pracoch□onnego)

WYKAZ ETAPÓW WERYFIKACJI MODELU

5.1. Badanie istotno□ci korelacji

5.2. Badanie wyrazisto□ci modelu

5.3. Badanie istotno□ci parametrów

5.4. Badanie symetrii sk□. losowego

5.5. Badanie losowo□ci sk□. losowego

5.6. Badanie stacjonarno□ci sk□. los.

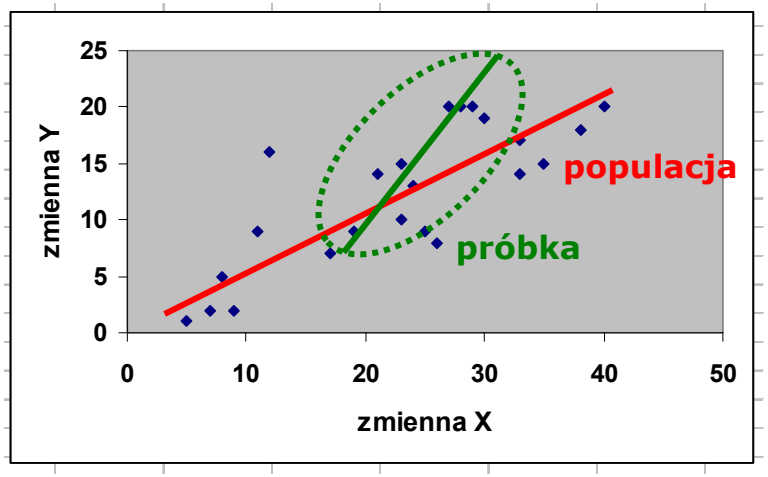
5.7. Badanie warto□ci oczekiwanej sk□. los.

5.8. Badanie autokorelacji sk□. losowego

5.9. Badanie normalno□ci sk□. losowego

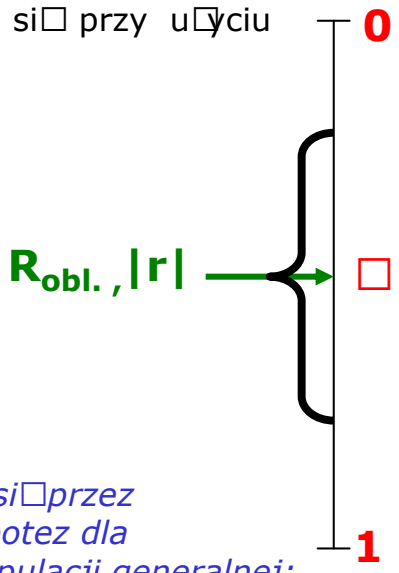
ETAP 5.1. Badanie istotno□ci korelacji

Celem etapu jest sprawdzenie, czy istnieje w populacji generalnej powi□zanie pomi□dzy zmienn□ Y i wszystkimi zmiennymi objaśniaj□cymi

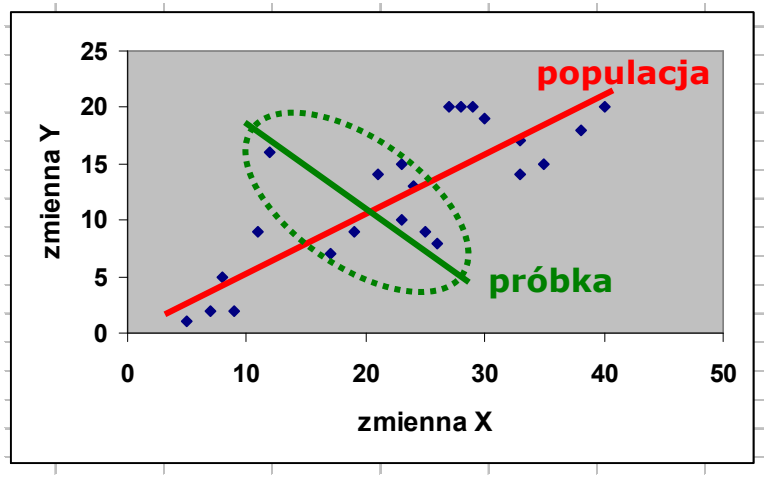


Przedzia□ ufno□ci dla nieznanego wspó□czynnika korelacji ρ dla populacji buduje si□ przy u□yciu b□□du

$$s_R = \sqrt{\frac{1-R^2}{n-k}}$$



Istotno□□ korelacji weryfikuje si□ przez postawienie nast□puj□cych hipotez dla wspó□czynnika korelacji dla populacji generalnej:



$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

Brak korelacji, nie ma powi□zania...

Korelacja istotna, jest powi□zanie...

Hipotezy te można weryfikować trzema równoważnymi testami:

- testem *t* Studenta (tylko dla regresji dwu zmiennych)
- testem *F* Fishera
- testem *R* Wallace'a-Snedecora

UWAGA!

W przypadku regresji wielorakiej, gdy liczba zmiennych objaśniających jest duża w porównaniu z liczbą obserwacji (n), współczynnik determinacji R^2 może dawać zawyżoną ocenę stopnia wyjaśnienia zmienności zmiennej objaśnianej; dlatego wprowadzono

skorygowany współczynnik determinacji $\mathbf{R_a^2}$ (i korelacji):

$$\mathbf{R_a^2 = 1 - \left(\frac{n-1}{n-k}\right)\left(\frac{SSE}{TOSS}\right) = 1 - (n-1)\left(\frac{s^2}{TOSS}\right)}$$

k – ilość parametrów w modelu regresji

$$\mathbf{R_a^2 = 1 - \frac{\frac{\sum(\hat{y}_i - \bar{y})^2}{n-k}}{\frac{\sum(y_i - \bar{y})^2}{n-1}}}$$

- *adjusted coefficient of multiple determination* (wydruki komputerowe)
- jeśli k jest małe, nie ma większej różnicy pomiędzy normalnym a skorygowanym R^2

TEST STUDENTA

$$t_{obl} = r\sqrt{\frac{n-2}{1-r^2}} = \frac{r}{s_r} \qquad t_{tabl} = t_{\alpha/2}\{n-2\}$$

TEST FISHERA

$$F_{obl} = \frac{MSTR}{MSE} = \frac{R^2}{1-R^2}\frac{n-k}{k-1} \qquad F_{tabl} = F_{\alpha}\{k-1, n-k\}$$

Źródło zmienności	*Liczba stopni swobody*	*Suma kwadratów*	*Średni kwadrat*	*Statystyka F*
Model (czynniki)	k–1	SSTR	MSTR	$F_{obl} = \frac{MSTR}{MSE}$
Błąd (reszta)	n–k	SSE	MSE	
Razem	n–1	SSTO		

TEST WALLACE'A-SNEDECORA

$$R_{obl} = \sqrt{R^2} \qquad R_{tabl} = R_{\alpha}\{k, n-k\}$$

Zmienna (czynnik)	Wartość oszacowana	Błąd oszacowania	Statystyka t_{obl}	Rzeczywisty poziom istotności P
Wyraz wolny	a_0	$s(a_0)$	$t(a_0)$	$P(a_0)$
Czynnik X_1	a_1	$s(a_1)$	$t(a_1)$	$P(a_1)$
Czynnik X_2	a_2	$s(a_2)$	$t(a_2)$	$P(a_2)$

Współczynniki: determinacji R^2, zbieżności φ^2, błąd resztowy s(y) i inne

Odczyt R_{tabl} z tablicy testu *R* Wallace'a-Snedecora

Stopnie swobody	Liczba zmiennych					
	2		3		4	
	0,05	0,01	0,05	0,01	0,05	0,01
8	0,632	0,765	0,726	0,827	0,777	0,860
18	0,444	0,561	0,532	0,633	0,587	0,678
28	0,361	0,463	0,439	0,530	0,490	0,573

Wnioski rozkładu *R* Wallace'a-Snedecora:

- im wyższy poziom istotności, tym niższe R_{tabl}
- im większa liczba zmiennych w modelu, tym wyższe R_{tabl}
- im wyższa liczba stopni swobody (większa próbka), tym niższe R_{tabl}
- tablica *R* powstała z przeliczenia tablic *t* oraz *F* (odwrócenie wzorów)

Tablica testu *R* jest najszybszym i najwygodniejszym narzędziem do weryfikacji istotności korelacji

Reguła decyzyjna (podsumowanie etapu 5.1.):

jeżeli $R_{obl} > R_{tabl}$, model jest poprawny, można przejść do etapu 5.2

jeżeli $R_{obl} < R_{tabl}$, model jest niepoprawny, trzeba zmienić albo zestaw zmiennych objaśniających albo jego postać matematyczną

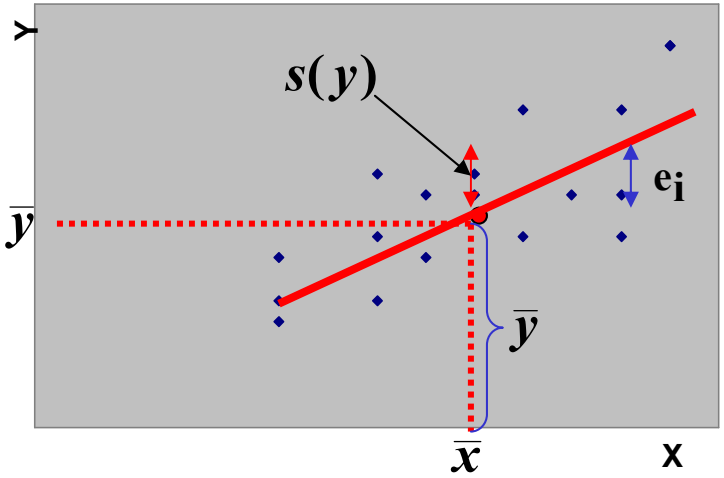
ETAP 5.1. OBOWIĄZUJE DLA WSZYSTKICH MODELI
ETAP 5.1. PRZEPROWADZA SIĘ DLA TRANSFORMATY (dla postaci pierwotnej można uzyskać R>1)

Rola współczynnika determinacji R^2

- korelacja może być istotna przy małym R i bardzo małym R^2
- małe R^2 oznacza niski stopień wyjaśnienia rzeczywistości i stanowi zagrożenie dla modelu
- należy dążyć (poprzez odpowiedni dobór zmiennych-przyczyn i postaci matematycznej modelu) do jak największego R^2 (dla postaci pierwotnej)
- wysoka wartość R^2 świadczy o dobrym poznaniu badanego zjawiska
- wysoka wartość R^2 bardzo często wynika jednak ze złego dobrania zmiennych objaśniających

ETAP 5.2. Badanie wyrazisto□ci modelu

Celem etapu jest kontrola rozrzutu danych



Wyrazisto□□modelu dana jest wzorem

$$V_{obl} = \frac{s(y)}{\bar{y}} 100\ \%$$

Wspó□czynnik zmienno□ci losowej V_{obl}<30%

(w przeciwnym przypadku rozrzut danych jest zbyt du□y)

Uwaga: gdy $\bar{y}$ jest bliskie 0 trudno□ci w ustaleniu czy model poprawny czy niepoprawny

ETAP 5.2. OBOWI□ZUJE DLA WSZYSTKICH MODELI (ale nie ma on charakteru statystycznego)

ETAP 5.2. PRZEPROWADZA SI□DLA POSTACI PIERWOTNEJ

ETAP 5.3. Badanie istotno□ci parametrów (wspó□czynników) modelu

Celem etapu jest sprawdzenie:

- czy poszczególne zmienne objaśniające mają□istotny wpływ na zmienną□objaśnianą?
- czy zmienne objaśniające są□wybrane prawidłowo?
- czy wyraz wolny różni się□istotnie od zera?

W etapie 5.3. następuje ostateczna selekcja zmiennych objaśniających:

• jeżeli wszystkie a_i okażą się istotne, model jest poprawny:
— model przyczynowo-skutkowy: do interpretacji
— model tendencji rozwojowej: do etapu 5.4

• jeżeli choćjedno a_i okaże się nieistotne, model jest niepoprawny i wymaga poprawy przez usunięcie nieistotnych zmiennych:
— zmienne należy usuwać po jednej (ze względu na efekt katalityczny)
— usuwa się zawsze zmienną o najniższej wartości $|t(a_i)|$ [max $P(a_i)$]
— usunięcie ostatniej zmiennej nieistotnej kończy proces selekcji kandydatek na zmienne,
— selekcja nie jest ostateczna, gdyż zawsze istnieje możliwość zamiany zmiennych, które są powiązane ze sobą

ETAP 5.3. OBOWIĄZUJE DLA WSZYSTKICH MODELI

ETAP 5.3 PRZEPROWADZA SIĘ DLA TRANSFORMATY

ETAP 5.4. Badanie symetrii składnika losowego

Badanie symetrii: dla n>30 test *z* (r-d normalny); dla n<30 test *t*-Studenta

$$H_0: \frac{m}{n} = \frac{1}{2}$$

$$H_1: \frac{m}{n} \neq \frac{1}{2}$$

$$t_{obl} = \frac{\left|\frac{m}{n} - \frac{1}{2}\right|}{\sqrt{\frac{\frac{m}{n}\left(1 - \frac{m}{n}\right)}{n-1}}} \qquad t_{\alpha,\nu=n-1}$$

m – liczba reszt dodatnich (lub ujemnych)
n - liczność próby

Brak symetrii wymaga zmiany matematycznej postaci modelu

ETAP 5.4. OBOWIĄZUJE DLA MODELI TENDENCJI ROZWOJOWEJ

ETAP 5.4. PRZEPROWADZA SIĘ DLA POSTACI PIERWOTNEJ

Zmienna (czynnik)	Warto□□ oszacowana	B□□d oszacowania	Statystyka t_{obl}	Rzeczywisty po-ziom istotno□ci P
Wyraz wolny	a_0	$s(a_0)$	$t(a_0)$	$P(a_0)$
Czynnik X_1	a_1	$s(a_1)$	$t(a_1)$	$P(a_1)$
Czynnik X_2	a_2	$s(a_2)$	$t(a_2)$	$P(a_2)$
Czynnik X_3	a_3	$s(a_3)$	$t(a_3)$	$P(a_3)$
Wspó□czynniki: determinacji R^2, zbie□no□ci φ^2, b□□d resztowy s(y) i inne				

Istotno□□parametrów a_i mo□na sprawdza□(tak jak problem □redniej dla populacji) na dwa sposoby:

- konstruuj□c przedzia□ufno□ci dla nieznanej warto□ci α_i
- weryfikuj□c hipotez□— H_0: $\alpha_i=0$ wobec H_1: $\alpha_i \neq 0$

$$P(a_i - t_{\alpha/2}\{n-k\}s(a_i) < \alpha_i < a_i + t_{\alpha/2}\{n-k\}s(a_i) = 1-\alpha$$

W praktyce stosuje si□hipotezy:

Badanie istotno□ci przeprowadza si□dla ka□dego α_i oddzielnie: $t_{obl}(a_i) = \frac{a_i - 0}{s(a_i)}$ $t_{tabl} = t_{\alpha/2}\{n-k\}$

- je□eli $|t_{obl}(a_i)| > t_{tabl}(a_i)$, odrzucamy hipotez□zerow□; parametr jest istotny z b□□dem równym co najwy□ej α
- je□eli $|t_{obl}(a_i)| < t_{tabl}(a_i)$, nie ma podstaw do odrzucenia hipotezy zerowej; parametr jest nieistotny

lub

- porównujemy rzeczywisty poziom istotno□ci $P(a_i)$ z przyj□tym *a'priori* α

Je□eli $P(a_i)$ jest mniejsze od α odrzucamy H0

Odrzucaj□c H_0 → ZMIENNA X_i MA WP□YW NA ZMIENN□Y

ETAP 5.5. Badanie losowo□ci sk□adnika losowego

Badanie losowo□ci przeprowadza si□testem *t*-Studenta lub testem serii

Test serii:

$$\mathbf{H_0 : \xi_t \text{ jest skladnikiem losowym} \quad H_1 : \xi_t \text{ nie jest losowy}}$$

Warto□ciom $u_t>0$ nadajemy symbol *a* (liczba symboli *a*: n_1), warto□ciom $u_t<0$ symbol *b* (liczba symboli *b*: n_2). Otrzymujemy podci□gi czyli serie z kolejnych symboli *a* lub *b*. Liczb□wszystkich serii (podci□gów) oznaczmy jako *K*. W tablicy liczby serii (dla n_1; n_2; α) odczytujemy liczb□krytyczn□K_α.

$\mathbf{P(K \le K_\alpha) = \alpha}$ Gdy $K > K\alpha$ nie ma podstaw do odrzucenia H_0

$\mathbf{P(K > K_\alpha) = 1 - \alpha}$ Gdy $K <= K\alpha$ odrzucamy H_0

Brak losowo□ci wymaga zmiany matematycznej postaci modelu

ETAP 5.5. OBOWI□ZUJE DLA MODELI TENDENCJI ROZWOJOWEJ

ETAP 5.5. PRZEPROWADZA SI□DLA POSTACI PIERWOTNEJ

ETAP 5.6. Badanie stacjonarno□ci sk□adnika losowego

W etapie tym sprawdzamy niezale□no□□wariancji sk□adnika losowego od zmiennej objа□niaj□cej *t* (warunek stosowalno□ci KMNK)

Bada si□istotno□□współczynnika korelacji reszt i zmiennej czasowej *t* $\mathbf{r_{u_t,t}}$

Stosuje si□test istotno□ci *t* (slajd: badanie współczynnika korelacji)

Przyczyny braku stacjonarno□ci:

- niew□a□ciwa posta□analityczna modelu
- niew□a□ciwa metoda szacowania parametrów strukturalnych modelu

ETAP 5.6. OBOWI□ZUJE DLA MODELI TENDENCJI ROZWOJOWEJ

ETAP 5.6 PRZEPROWADZA SI□DLA TRANSFORMATY

ETAP 5.7. Badanie wartości oczekiwanej składnika losowego

Z założeń KMNK wynika, że parametry składnika losowego:

$$\mathbf{E(\xi) = 0} \qquad \mathbf{\{\sum u_i = 0\}} \qquad \mathbf{V(\xi) = const}$$

Po retransformacji do postaci pierwotnej mamy nowe reszty u', dla których

$$\sum \mathbf{u_i'} \neq 0 \qquad \mathbf{\overline{u}_i'} \neq 0$$

Celem etapu jest sprawdzenie, czy odchylenie od „*0*" nie jest zbyt duże (służy do tego test Studenta)

ETAP 5.7. OBOWIĄZUJE DLA MODELI TENDENCJI ROZWOJOWEJ

ETAP 5.7. PRZEPROWADZA SIĘ DLA POSTACI PIERWOTNEJ

ETAP 5.8. Badanie autokorelacji składnika losowego

Składnik losowy nie jest czysto losowy, lecz zależy od wskaźnika *i*, czyli zmienne losowe ξ_i są zależne od poprzednich wartości ξ_{i-1}.

Autokorelacja to korelacja wartości zmiennej z jej wartościami z okresów wcześniejszych o jeden lub więcej okresów.
Na ogół autokorelację można wyrazić w postaci relacji: $\xi_i = f(\xi_{i-1}, \xi_{i-2}, ..., \xi_{i-\tau})$ $\quad u_t = f(u_{t-k})$ $\quad t = 1,2,...$

W praktyce przyjmuje się, że funkcja *f* jest funkcją liniową, a maksymalne opóźnienie czasowe wynosi jeden lub dwa.

Estymator współczynnika autokorelacji r_1 (rzędu pierwszego):

$$r_1 = \frac{\sum_{i=2}^{n} (u_i - \overline{u}_i)(u_{i-1} - \overline{u}_{i-1})}{\sqrt{\sum_{i=2}^{n} (u_i - \overline{u}_i)^2 \sum_{i=2}^{n} (u_{i-1} - \overline{u}_{i-1})^2}}$$

Przyczyny autokorelacji:

- zakłócenia (dodatnie lub ujemne) w jednym okresie wpływają na poziom zjawiska w następnych okresach: skutki niektórych zdarzeń rozciągają się na wiele okresów (natura procesów gospodarczych, społecznych);
- psychologia i sposób podejmowania decyzji, na które duży wpływ mają zdarzenia z najbliższej przeszłości;
- niepoprawna postać funkcyjna modelu (model nie uwzględnia cykliczności zjawiska, aproksymacja zależności nieliniowej przez funkcję liniową);
- wadliwa struktura dynamiczna modelu: w roli zmiennej objaśniającej nie występuje – a powinna – opóźniona zmienna objaśniana; brak opóźnionych zmiennych niezależnych lub zmiennej czasowej;
- pominięcie w modelu istotnej zmiennej objaśniającej (reszty mogą układać się seriami, nawet mieć tendencję do stałego zwiększania swej wartości bezwzględnej);
- interpolacja, wygładzanie oraz agregacja (np. przekształcanie danych miesięcznych w kwartalne).

Badanie autokorelacji można przeprowadzić:

- testem R istotności korelacji $\mathbf{r_{u_t,u_{t-k}}}$
- testem Durbina-Watsona

Test Durbina-Watsona służy do sprawdzenia hipotezy: $\mathbf{H_0: \rho_1 = 0}$ $\mathbf{H_1: \rho_1 < 0 \ lub \ H_1: \rho_1 > 0}$

Statystyka d:
$$\mathbf{d} = \frac{\sum_{i=2}^{n}(u_i - u_{i-1})^2}{\sum_{i=2}^{n} u_i^2}$$

$\mathbf{d = 2(1-r_1)}$

Z relacji wynika, że d [0,4]:

jeżeli $r_1=0$ to d=2 (brak autokorelacji)

jeżeli $r_1=1$ to d=0 (silna autokorelacja dodatnia)

jeżeli $r_1=-1$ to d=4 (silna, ujemna autokorelacja)

Rozkład statystyki *d* przy założeniu, że H_0 jest prawdziwa i składniki losowe mają rozkład normalny N(0; □) zależy od liczby obserwacji *n* oraz liczby zmiennych objaśniających i $d \in \langle d_L; d_U \rangle$.

Wartości krytyczne d_L i d_U zawiera tablica testu Durbina-Watsona dla poziomu istotności α.

Reguła decyzyjna: $\mathbf{H_0: \rho_1 = 0}$ $\mathbf{H_1: \rho_1 < 0}$

- jeżeli d d_L to H_0 odrzucamy
- jeżeli $d_L < d < d_U$?
- jeżeli d d_U nie ma podstaw do odrzucenia H_0

Reguła decyzyjna: $\mathbf{H_0: \rho_1 = 0}$ $\mathbf{H_1: \rho_1 > 0}$

- jeżeli d $4 - d_L$ to H_0 odrzucamy
- jeżeli $4 - d_U < d < 4 - d_L$?
- jeżeli d $4 - d_U$ nie ma podstaw do odrzucenia H_0

W przypadku stwierdzenia autokorelacji mamy trzy możliwości:

- usunąć przyczyny autokorelacji;
- zastosować procedury estymacji w warunkach autokorelacji, aby zapewnić wysoką efektywność estymatorów (usuwanie autokorelacji);
- pozostać przy KMNK godząc się z mniejszą efektywnością estymatorów.

ETAP 5.8. OBOWIĄZUJE DLA MODELI TENDENCJI ROZWOJOWEJ
ETAP 5.8. PRZEPROWADZA SIĘ DLA TRANSFORMATY

Etap 5.9. Badanie normalności składnika losowego

Celem etapu jest stwierdzenie, czy reszty mają rozkład normalny

Stosuje się znane ze statystyki testy nieparametryczne:

- λ - Kołmogorowa

lub

- test χ^2

Analiza reszt — oddzielny dział analizy regresji i korelacji

ETAP 6. INTERPRETACJA MODELU

Celem etapu jest wydobycie z modelu całej nowej wiedzy, której nie widać „gołym okiem”

Interpretacja modelu przyczynowo-skutkowego

Polega na określeniu wpływu poszczególnych czynników na badane zjawisko:

- podział wszystkich czynników na trzy grupy

  A — czynniki nie mające wpływu na Y

  B — czynniki mające wpływ na Y i wprowadzone do modelu

  C — czynniki mające wpływ na Y, ale nie występujące w modelu

- ocena jakościowa wpływu czynników B i C
- ocena ilościowa wpływu czynników B i C

Ocena jakościowa

Na podstawie znaków stojących przy współczynnikach r oraz a_i

Ocena ilościowa

- **Model potęgowy – FUNKCJA PRODUKCJI COBBA-DOUGLASA** $P = \alpha_0 X_1^{\alpha 1} X_2^{\alpha 2} e^{\gamma} \xi$

$\alpha_1, \alpha_2 > 0$

α_1
α_2 **Współczynniki elastyczności produkcji względem X_1 i X_2**

$K = \alpha_1 + \alpha_2$ **Efekt skali produkcji**

Funkcja produkcji to specjalny model, określający zależność pomiędzy produkcją (P), a czynnikami produkcji: majątkiem produkcyjnym (X_1) i nakładami pracy żywej (X_2).

Metoda estymacji parametrów strukturalnych funkcji Cobba-Douglasa (jak dla modelu potęgowego) to MNK

Graficznym obrazem funkcji jest krzywa wypukła do początku układu współrzędnych X_1, X_2. Poruszając się po krzywej P otrzymujemy ten sam wolumen produkcji P przy różnych kombinacjach czynników-nakładów X1 i X2.

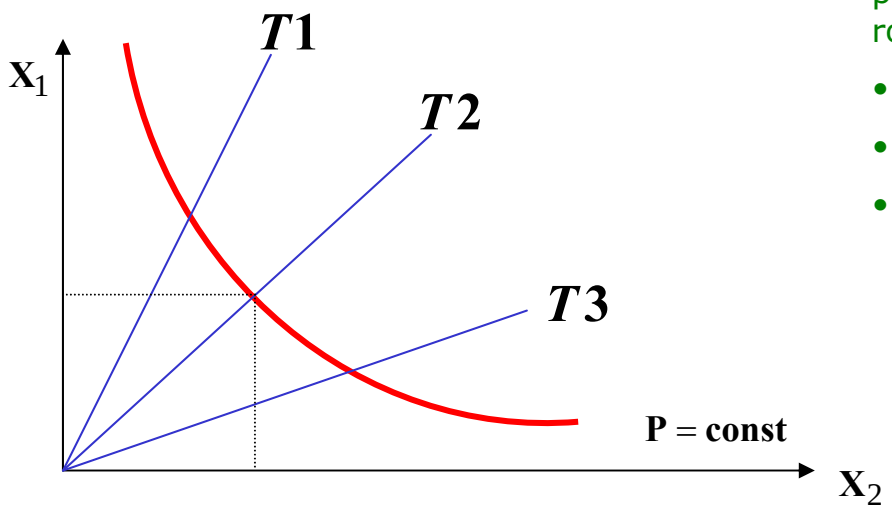
Ustalanie różnych proporcji czynników X1 i X2 w celu otrzymania tej samej wielkości produkcji P jest możliwe ze względu na wzajemną substytucję tych czynników.

Współczynnik elastyczności substytucji: $\varepsilon = \frac{\Delta X_1}{X_1} / \frac{\Delta X_2}{X_2}$

Elastyczność ε to współczynnik określający reakcję zmiany jednej zmiennej na inną zmienną — na przykład *elastyczność cenowa popytu* (jak popyt zmienia się wraz z ceną towaru).

Z ekonomicznego punktu widzenia elastyczność produkcji ε określa o ile należy zwiększyć nakłady na środki trwałe X1, aby nakłady pracy X2 zmniejszyć o jednostkę (przy stałym poziomie produkcji P).

Izolinie funkcji produkcji

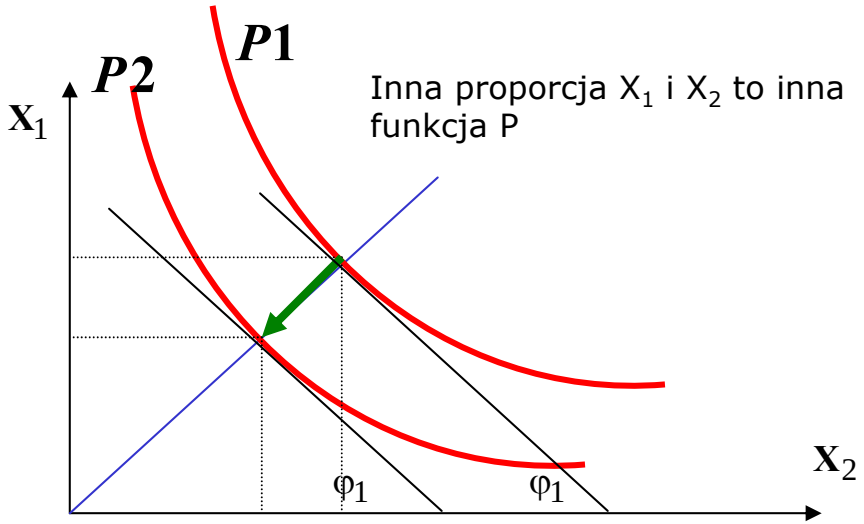


Wartość produkcji *P* można osiągnąć przy różnych proporcjach pomiędzy majątkiem X1 i nakładami siły roboczej X2:

- T1— technika kapitałochłonna, np. pełna automatyzacja
- T2 — poziom techniki
- T3 — niski poziom techniki, produkcja pracochłonna

Efekt poprawy zarządzania bez zmiany techniki (lepsze wykorzystanie ludzi i sprzętu):

- zmniejszenie kapitałochłonności
- zmniejszenie pracochłonności
- zwiększenie produkcji (przejście z funkcji P1 na P2)



Inna proporcja X_1 i X_2 to inna funkcja P

Krańcowa stopa substytucji $\mathbf{tg\varphi_1 = R}$

Przykład. Zinterpretuj zweryfikowaną funkcję produkcji Cobba-Douglasa

$$P_i = 45{,}2 M_i^{0{,}65} Z_i^{0{,}73} e^{0{,}04} \xi$$

- wartości 45,2 nie interpretuje się
- wartość 0,65 to współczynnik elastyczności majątkowej (kapitałowej) produkcji: zmiana majątku o 10% daje średnio zmianę produkcji o 6,5% (przy stałym zatrudnieniu = pozostałe czynniki cateris paribus)
- wartość 0,73 to współczynnik elastyczności zatrudnieniowej produkcji: zmiana zatrudnienia o 10% daje zmianę produkcji średnio o 7,3% (przy stałym majątku)
- suma obu współczynników K=0,65+0,73=1,38 to współczynnik skali produkcji:
  - jeżeli K<1, to firma rozwija się ekstensywnie (szybsze tempo wzrostu czynników produkcji aniżeli przyrostu produkcji)
  - jeżeli K>1, to firma rozwija się intensywnie (szybsze tempo przyrostu produkcji aniżeli wzrostu czynników produkcji)
- wartość γ = 0,04 to współczynnik postępu organizacyjnego:
  - jeżeli γ >O, to mamy miejsce postęp organizacyjny
  - jeżeli γ <0, to mamy miejsce regres organizacyjny

MODEL TENDENCJI ROZWOJOWEJ

Model tendencji rozwojowej to konstrukcja teoretyczna (równanie lub układ równań) opisująca kształtowanie się określonego zjawiska jako funkcji:

- zmiennej czasowej t
- wahań okresowych (periodycznych)
- wahań przypadkowych (nieregularnych).

Czyli na zmienność zjawiska w czasie ma wpływ:

- tendencja rozwojowa (trend)
- wahania typu okresowego
- wahania typu przypadkowego (losowego).

MODEL ADDYTYWNY $$Y_t = F(t) + G(t) + \xi(t)$$

MODEL MULTIPLIKATYWNY $$Y_t = F(t) \cdot G(t) \cdot \xi(t)$$

gdzie:
Yt - poziom badanego zjawiska
F(t) – funkcja trendu
G(t) – funkcja wahań okresowych
$\xi(t)$ - składnik losowy o rozkładzie normalnym, $E(\xi)=0$, $V(\xi)=const$

Te trzy części trzeba zidentyfikować, a potem złożyć razem w model:
- addytywny (jeśli amplituda wahań jest stała)
- multiplikatywny (jeśli amplituda wahań rośnie lub maleje regularnie)

WYZNACZANIE TRENDU

Metody mechaniczne: $y_t = f(t) + \xi_t$

- wskaźniki bezwzględne (stały przyrost = postęp arytmetyczny = funkcja liniowa)

$$\bar{x}_{chr} = \frac{\frac{1}{2}x_1 + ... + \frac{1}{2}x_n}{n-1}$$

- wskaźniki względne (przyrost o stały procent = postęp geometryczny = funkcja potęgowa)

- średnia ruchoma (wygładzanie danych linią łamaną) $\bar{x}_g = \sqrt[n]{\Pi x_i}$

Średnia ruchoma: mając szereg czasowy $y_1, y_2, ..., y_n$, przyjmujemy długość kroku k = 3, lub 5, lub 7 itd. i liczymy

$$\bar{y}_{k/2+0,5} = \frac{y_1 + ... + y_k}{k}$$ a następnie $$\bar{y}_{k/2+1,5} = \frac{y_2 + ... + y_{k+1}}{k}$$ itd.

dla k=3: $$\bar{y}_2 = \frac{y_1 + y_2 + y_3}{3} \quad \bar{y}_3 = \frac{y_2 + y_3 + y_4}{3}$$ itd.

Dla k=5 $$\bar{y}_3 = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5} \quad \bar{y}_4 = \frac{y_2 + y_3 + y_4 + y_5 + y_6}{5}$$ itd.

Gdy k jest liczbą parzystą to uzyskujemy tzw. średnie scentrowane

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

CECHY ŚREDNIEJ RUCHOMEJ:

- TRACI SIĘ k-1 DANYCH
- IM WIĘKSZE k, TYM BARDZIEJ SZTYWNY TREND

Przykład. Wyznacz trend metodą średniej ruchomej dla następujących danych:

okres t	1	2	3	4	5	6	7	8
wartości y_t	5	4	7	6	8	10	9	8

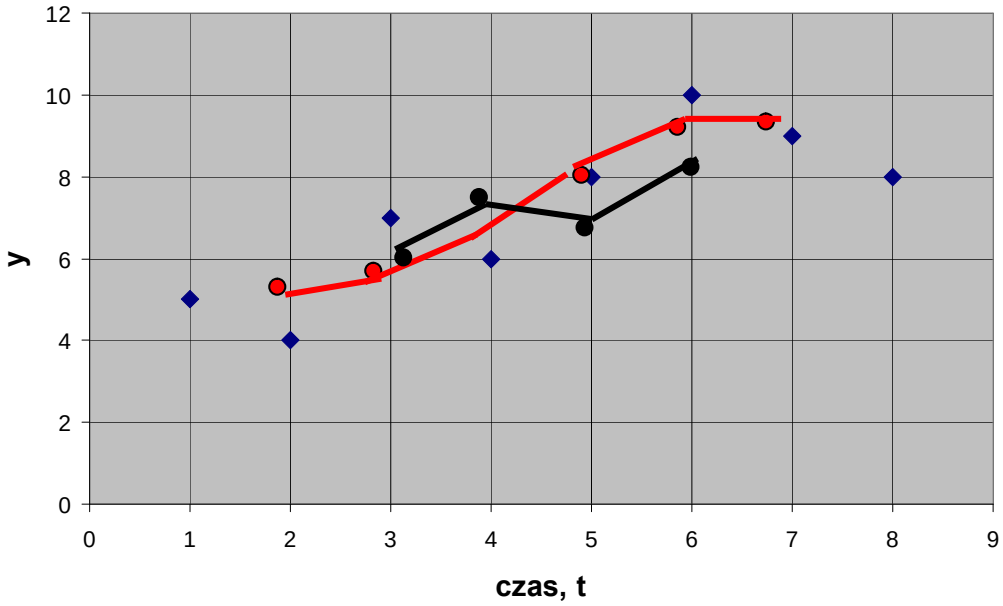
k=5

$$\bar{y}_3 = \frac{30}{5} = 6$$

$$\bar{y}_4 = \frac{35}{5} = 7$$

$$\bar{y}_5 = \frac{39}{5} = 6{,}6$$

$$\bar{y}_6 = \frac{41}{5} = 8{,}2$$



k=3

$$\bar{y}_2 = \frac{5+4+7}{3} = 5{,}33$$

$$\bar{y}_3 = \frac{4+7+6}{3} = 5{,}67$$

$$\bar{y}_4 = \frac{7+6+8}{3} = 7{,}00$$

$$\bar{y}_5 = \frac{6+8+10}{3} = 8{,}00$$

$$\bar{y}_6 = \frac{8+10+9}{3} = 9{,}00$$

$$\bar{y}_7 = \frac{10+9+8}{3} = 9{,}00$$

Wygładzanie wykładnicze

Dla dowolnego momentu *t* operatorem wyrównania rzędu pierwszego dla szeregu y_t jest wyrażenie:

$$\mathbf{S_t} = \alpha \cdot \mathbf{y_t} + (1-\alpha) \cdot \mathbf{S_{t-1}}$$
$$0 < \alpha < 1$$

α - stała wygładzania

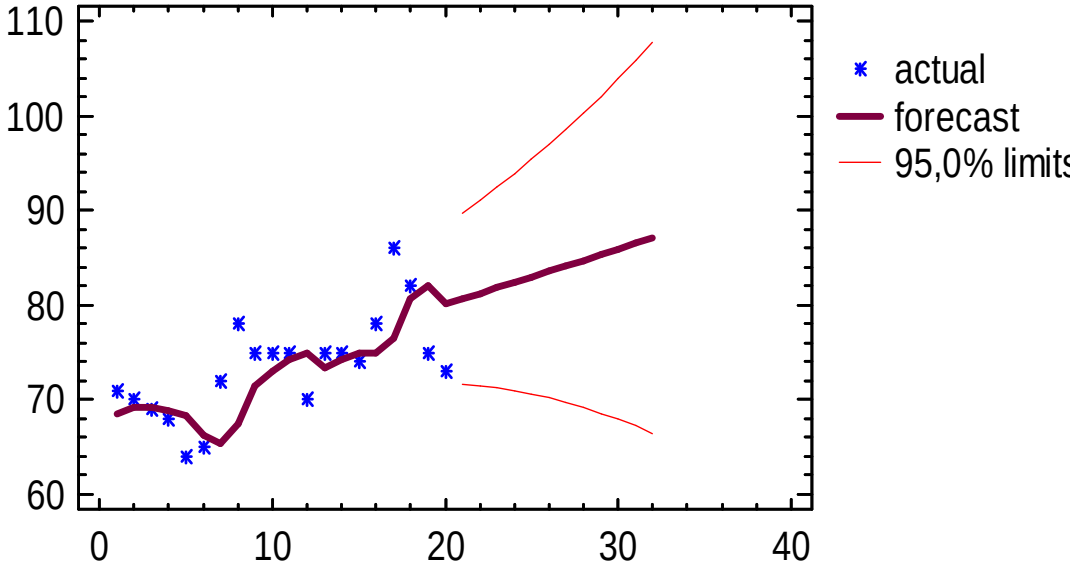
$\mathbf{S_1 = y_1}$

Ogólnie operator wygładzania można zapisać wyrażeniem: $\mathbf{S_t} = \alpha \sum_{\mathbf{i}=0}^{\mathbf{t}-2} (1-\alpha)^{\mathbf{t}} \mathbf{y_{t-i}} + (1-\alpha)^{\mathbf{t}-1} \mathbf{y_1}$

t	y_t	$\hat{y}_t$
1	71,0	71,0
2	70,0	69,1
3	69,0	69,2
4	68,0	68,9
5	64,0	68,3
6	65,0	66,3
7	72,0	65,3
8	78,0	67,5
9	75,0	71,5
10	75,0	73,1
11	75,0	74,2
12	70,0	74,9
13	75,0	73,4
14	75,0	74,3
15	74,0	74,9
16	78,0	74,9
17	86,0	76,4
18	82,0	80,7
19	75,0	82,0
20	73,0	80,1

Model: Brown's linear exp. smoothing with *alpha = 0,2*

Wygładzanie wykładnicze trendu dla α=0,2



B. Metody analityczne

B1

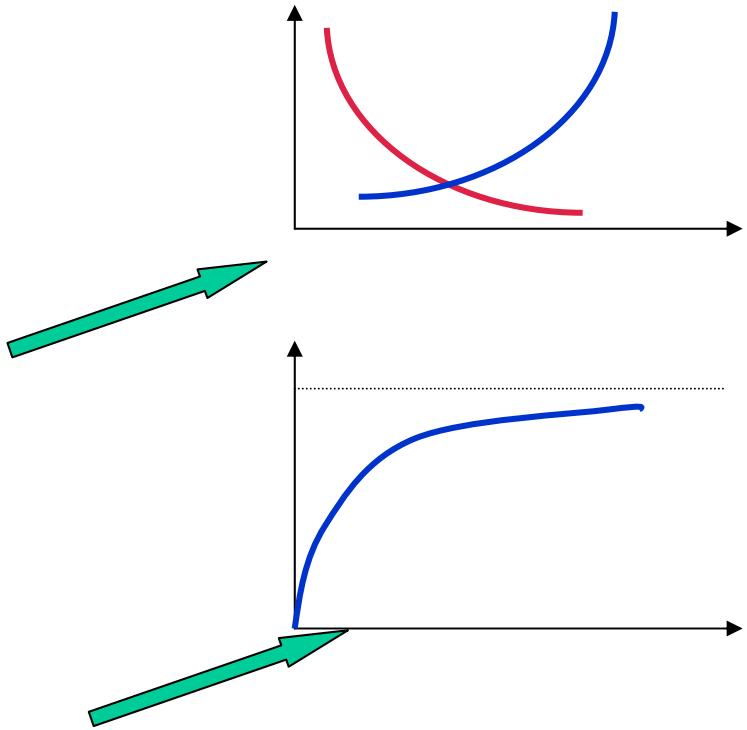
Funkcja liniowa $y_t = \alpha + \beta \cdot t$

Funkcja potęgowa $y_t = \alpha \cdot t^{\beta}$

Funkcja wykładnicza $y_t = \alpha \cdot \beta^{t}$

Funkcja logistyczna $y_t = \frac{\alpha}{1 + \beta \cdot e^{-\gamma \cdot t}}$

Wielomian I rzędu $y_t = \alpha + \beta \cdot t + \gamma \cdot t^2$



Parametry strukturalne funkcji oszacowywane MNK

Mierniki stopnia dopasowania modelu do danych empirycznych:

- Błąd przeciętny (Mean Error)

$$\mathbf{M.E.} = \frac{1}{\mathbf{n}} \sum (\mathbf{y_t} - \hat{\mathbf{y}}_\mathbf{t})$$

- Średni błąd kwadratowy (Mean Square Error)

$$\mathbf{M.S.E.} = \frac{1}{\mathbf{n}} \sum (\mathbf{y_t} - \hat{\mathbf{y}}_\mathbf{t})^2$$

- Średni błąd absolutny (Mean Absolute Error)

$$\mathbf{M.A.E.} = \frac{1}{\mathbf{n}} \sum |\mathbf{y_t} - \hat{\mathbf{y}}_\mathbf{t}|$$

n – liczba reszt

B2. Wyrównywanie szeregów czasowych za pomocą wielomianów Lagrange'a

Zakłada się, że w szeregu czasowym nie występują silne wahania regularne.

$Y_t=f(t)$ zastępujemy wielomianem stopnia p.

Wtedy interpolacyjny wielomian Lagrange'a:

$$y_t = f(t_0)\frac{(t-t_1)(t-t_2)...(t-t_n)}{(t_0-t_1)(t_0-t_2)...(t_0-t_n)} + f(t_1)\frac{(t-t_0)(t-t_2)...(t-t_n)}{(t_1-t_0)(t_1-t_2)...(t_1-t_n)} + ...$$

$$+ f(t_n)\frac{(t-t_0)(t-t_1)...(t-t_{n-1})}{(t_n-t_0)(t_n-t_1)...(t_n-t_{n-1})}$$

Liczba członów wielomianu i stopień zależy od liczby wyrazów szeregu.

Metoda bardzo pracochłonna

Wyznaczanie waha□okresowych

(dobowych, tygodniowych, miesi□cznych, kwartalnych, rocznych, wieloletnich)

Przez wahania okresowe (sezonowe) nale□y rozumie□powtarzaj□ce si□z roku na rok w tych samych jednostkach kalendarzowych do□□regularne zmiany ilo□ciowe.

Cechy waha□okresowych:

- roczny cykl z podokresami miesi□cznymi, kwartalnymi, p□rocznymi
- systematyczne powtarzanie si□waha□w ka□dym roku
- okre□lona regularno□□(sta□e cykle zmian powtarzaj□cych si□)

DWIE METODY: — metoda wska□ników sezonowo□ci
— metoda harmoniczna

Wszystkie metody maj□jeden cel: uzyskanie przeci□tnego obrazu jednego cyklu

METODA WSKAŹNIKÓW SEZONOWOŚCI

- obliczenie surowych wskaźników sezonowości (wyeliminowanie trendu):
- obliczenie oczyszczonych wskaźników sezonowości (wyeliminowanie wahań losowych):

Najprostszym sposobem wyodrębnienia wahań sezonowych jest metoda oparta na średnich okresach jednoimiennych.

Wskaźniki sezonowości oblicza się wg wzoru: $$S_i = \frac{\bar{y}_i \cdot d}{\sum_{i=1}^{d} \bar{y}_i}$$

gdzie:

Si – wskaźnik sezonowości dla i-tego podokresu (zwykle w %)

$\bar{y}_i$ – średnia arytmetyczna dla jednoimiennych podokresów

d – liczba podokresów (podokresy miesięczne d=12; kwartalne d=4; półroczne d=2)

$$\sum_{i=1}^{12} S_i = 1200 \qquad \sum_{i=1}^{4} S_i = 400 \qquad \sum_{i=1}^{2} S_i = 200$$

Wskaźniki spełniające powyższe relacje to oczyszczone wskaźniki sezonowości.
Wskaźniki nie spełniające tych relacji to surowe wskaźniki sezonowości.

Współczynnik korygujący *k:* $$k = \frac{d}{\sum_{i=1}^{d} S_i}$$ pozwala sprowadzić surowe wskaźniki do oczyszczonych wg reguły:

Suma skorygowanych wskaźników: $$S_i^k = k \cdot S_i$$

Metoda harmoniczna (szeregi Fouriera)

Wahania okresowe przedstawia si jako sum okre lonej liczby drga harmonicznych (sinusoid i cosinusoid) przesuni tych w fazie, lecz o jednakowym okresie

$$y_t = a_0 + a_1 t + \sum_{i=1}^{m/2} b_{1i} \sin \frac{2\pi}{m} it + \sum_{i=1}^{m/2} b_{2i} \cos \frac{2\pi}{m} it + \xi$$

Ogólnie, w przypadku *m* obserwacji liczba harmonik nie przekracza m*/2*.

Zmiany przebiegu funkcji okresowej dobrze daje si opisa za pomoc kilku pocz tkowych harmonik (*i* – numer harmoniki).