Prognozowanie i symulacje
Wykład 5

Cel budowy przyczynowo-skutkowego modelu ekonometrycznego

- Poznanie związków przyczynowo-skutkowych wiążących zmienne (wartość poznawcza – cel analityczny)
- Wyznaczenie przyszłej wartości zmiennej opisującej badane zjawisko (cel prognostyczny)

Etapy budowy modelu ekonometrycznego

określenie celu badania

dobór zmiennych objaśniających

zebranie danych statystycznych

konstrukcja modelu

estymacja parametrów modelu

weryfikacja modelu

zastosowanie modelu

Metody doboru zmiennych objaśniających

- Metoda Hellwiga – metoda integralnej pojemności informacyjnej
- Badanie współliniowości
- Analiza czynnikowa (np. analiza głównych składowych)
- Błąd ex-ante
- Test t-Studenta
- Kryteria informacyjne (Akaike, Hannana-Quinna, bayesowskie Schwarza)

Współliniowość

Zgodnie z założeniami Klasycznego Modelu Regresji Liniowej:

— Zmienne objaśniające powinny być skorelowane ze zmiennymi objaśnianymi

— Zmienne objaśniające nie powinny być skorelowane ze sobą

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_t$$

$$\operatorname{cov}(y, x_1) \neq 0 \quad \operatorname{cov}(y, x_2) \neq 0 \quad \operatorname{cov}(x_1, x_2) = 0$$

Skutki występowania (niepełnej) współliniowości

- Niewielkie zmiany w zbiorze danych skutkują dużymi zmianami w otrzymanych wynikach (estymatorach)
- Oceny parametrów β mają duże średnie błędy szacunków → małe wartości statystyk t-Studenta → częściej wnioskujemy o braku istotności zmiennych objaśniajacych
- Wysokie wartości współczynnika determinacji
- Oceny parametrów strukturalnych mają znaki niezgodne z oczekiwaniami (niezgodne z teorią ekonomiczną)

Skutki występowania (niepełnej) współliniowości

Wariancja estymatora parametru β

$$\operatorname{var}(\beta_j) = \frac{\sigma^2}{(1 - r_{j.}^2)\sum_{i=1}^{n}(x_{ij} - \bar{x})^2}$$

— Rośnie wraz ze wzrostem współczynnika korelacji r_j

— maleje wraz ze wzrostem wariancji j-tej zmiennej

Metoda Hellwiga – metoda integralnej pojemności informacyjnej

Współczynnik służy porównaniu wartości informacyjnej konkurencyjnych modeli liniowych opartych o tę samą zmienną objaśnianą Y i różne kombinacje zmiennych objaśniających.

$$H_l = \sum_{j=1}^{p} \frac{r_j^2}{1 + \sum_{i \neq j} \left| r_{ij} \right|},$$

Gdzie: r_j współczynnik korelacji liniowej pomiędzy zmiennymi Y i X_j

r_{ij} współczynnik korelacji liniowej między zmiennymi X_i i X_j

Współczynnik przyjmuje wartości <0,1>.

Wartość 1 oznacza maksymalną ilość informacji dostarczaną przez wyróżnioną kombinację zmiennych

Variance Inflation Factor

VIF mierzy jaka część wariancji estymatora B jest spowodowana przez to, ze j-ta zmienna nie jest ortogonalna względem pozostałych zmiennych objaśniających modelu – „czynnik rozdęcia wariancji”

$$VIF = \frac{1}{1 - r_{j.}^2}$$

r_j^2 Współczynnik korelacji wielorakiej pomiedzy j-tą zmienną a pozostałymi zmiennymi niezależnymi modelu

Minimalna wartość VIF=1,

VIF>10 wskazuje na „rozdęcie wariancji” czyli na występowanie współliniowości

```
Ocena współliniowości VIF(j) - czynnik rozdęcia wariancji
VIF (Variance Inflation Factors) - minimalna możliwa wartość = 1.0
Wartości > 10.0 mogą wskazywać na problem współliniowości - rozdęcia
wariancji

         PKB    2,201
         SRD    2,201
```

$$AK_t = \beta_0 + \beta_1 PKB_{t1} + \beta_2 SRD_{t2} + \varepsilon_t$$

```
VIF(j) = 1/(1 - R(j)^2), gdzie R(j) jest współczynnikiem korelacji
wielorakiej
pomiędzy zmienną 'j' a pozostałymi zmiennymi niezależnymi modelu.
```

```
Ocena współliniowości VIF(j) - czynnik rozdęcia wariancji
VIF (Variance Inflation Factors) - minimalna możliwa wartość = 1.0
Wartości > 10.0 mogą wskazywać na problem współliniowości - rozdęcia
wariancji

          SI  133,290
          SO  133,290
```

$$PKB_t = \beta_0 + \beta_1 SI_{t1} + \beta_2 SO_{t2} + \varepsilon_t$$

```
VIF(j) = 1/(1 - R(j)^2), gdzie R(j) jest współczynnikiem korelacji
wielorakiej
pomiędzy zmienną 'j' a pozostałymi zmiennymi niezależnymi modelu.
```

SI	SO	
1,0000	0,9962	SI
	1,0000	SO

Dobór zmiennych na podstawie błędu ex-ante

- Wyznaczamy błąd ex-ante dla modelu ze wszystkimi potencjalnymi zmiennymi objaśniającymi
- Wyznaczamy błędy ex-ante dla wszystkich możliwych kombinacji zmiennych objaśniających
- Wybieramy tę kombinację, która daje najmniejszy błąd ex-ante
- Metoda bardzo czasochłonna/pracochłonna – przy dużej liczbie zmiennych objaśniających szacowana jest duża liczba modeli, aby wyznaczyć błędy ex-ante, (pośrednio prognozy) należy znać wartości zmiennych objaśniajacych w okresie prognozowanym

Wartości zmiennych objaśniających w okresie prognozowanym

Ustalenie na poziomie **planowanym**

Wyznaczenie i **ekstrapolacja trendów** wartości zmiennych objaśniających

Zbudowanie **modeli przyczynowo-opisowych** oddzielnie dla każdej zmiennej objaśniającej

Symulacje – zbiór prognoz dla zmiennej objaśnianej odpowiadających różnym możliwym wartościom zmiennych objaśniających

Prognozowanie na podstawie jednorównaniowego modelu przyczynowo-skutkowego

Zależność przyczynowo-skutkowa pomiędzy zmiennymi opisana jest następującym modelem:

$$y_t = \sum_{i=0}^{k} \beta_i x_{it} + \varepsilon_t, \quad t = 1, \ldots, n$$

Abu wyznaczyć prognozy, przyjmujemy że wartości zmiennych objaśniających w okresie prognozowanym (t>n) wynoszą odpowiednio:

$$x_{0t}^*, \ldots, x_{kt}^*$$

Oraz spełnione są założenia stochastyczne modelu

Prognozowanie na podstawie jednorównaniowego modelu przyczynowo-skutkowego

W takim przypadku prognoza przedstawia się następująco:

$$y_t^P = E\left(y_t \middle| x_{0t}^*, .., x_{kt}^*\right) = E\left(\sum_{i=0}^k \beta_i x_{it}^* + \varepsilon_t\right)$$

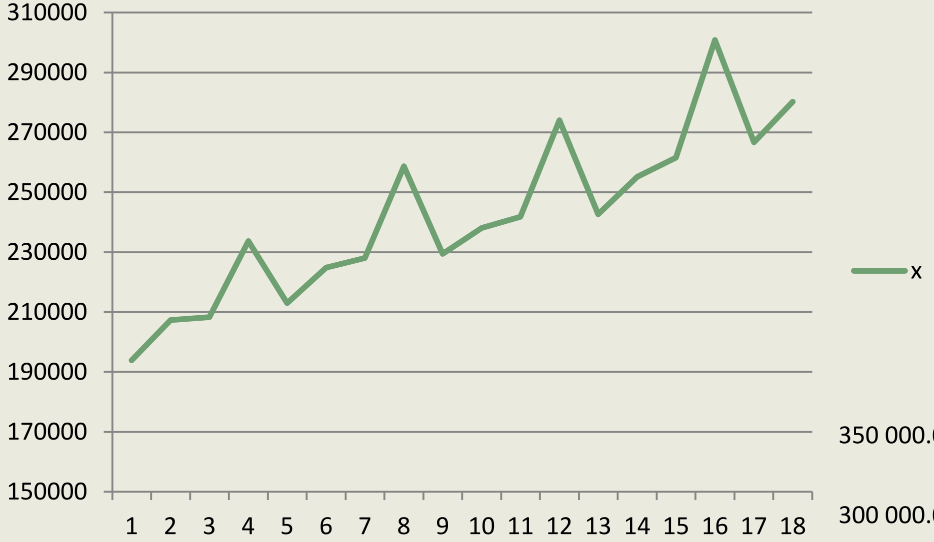
$$= \sum_{i=0}^k E(\beta_i) x_{it}^* + E(\varepsilon_t) = \sum_{i=0}^k E(\beta_i) x_{it}^*$$

Wartości parametrów strukturalnych β nie są znane, znamy tylko <u>oceny</u> parametrów strukturalnych β^.

Jednak jeżeli spełnione są założenia stochastyczne modelu, KMNK jest najlepszym nieobciążonym estymatorem liniowym nieznanych parametrów β, wówczas:

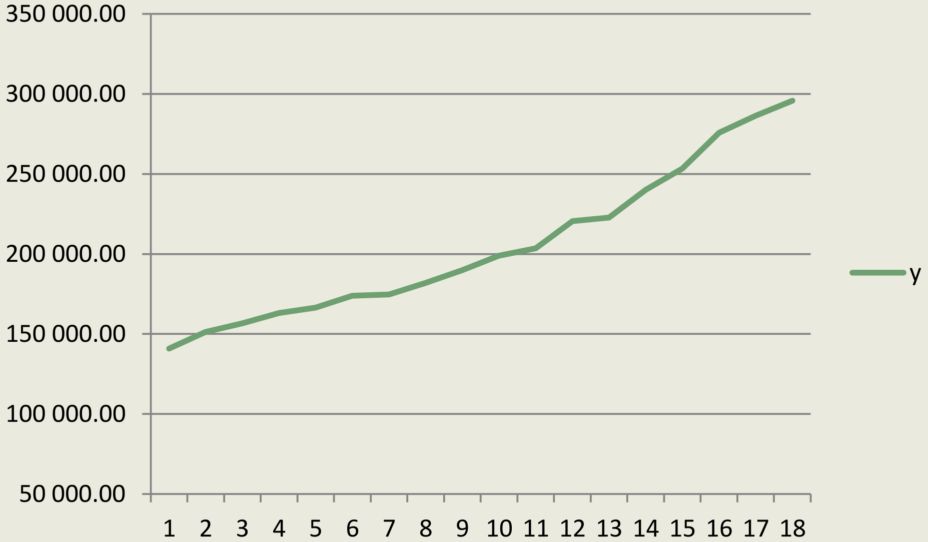
$$y_t^P = \sum_{i=0}^k \beta_i x_{it}^*$$

$$y_t = \beta_0 + \beta_1 x_{t1} + \varepsilon_t$$



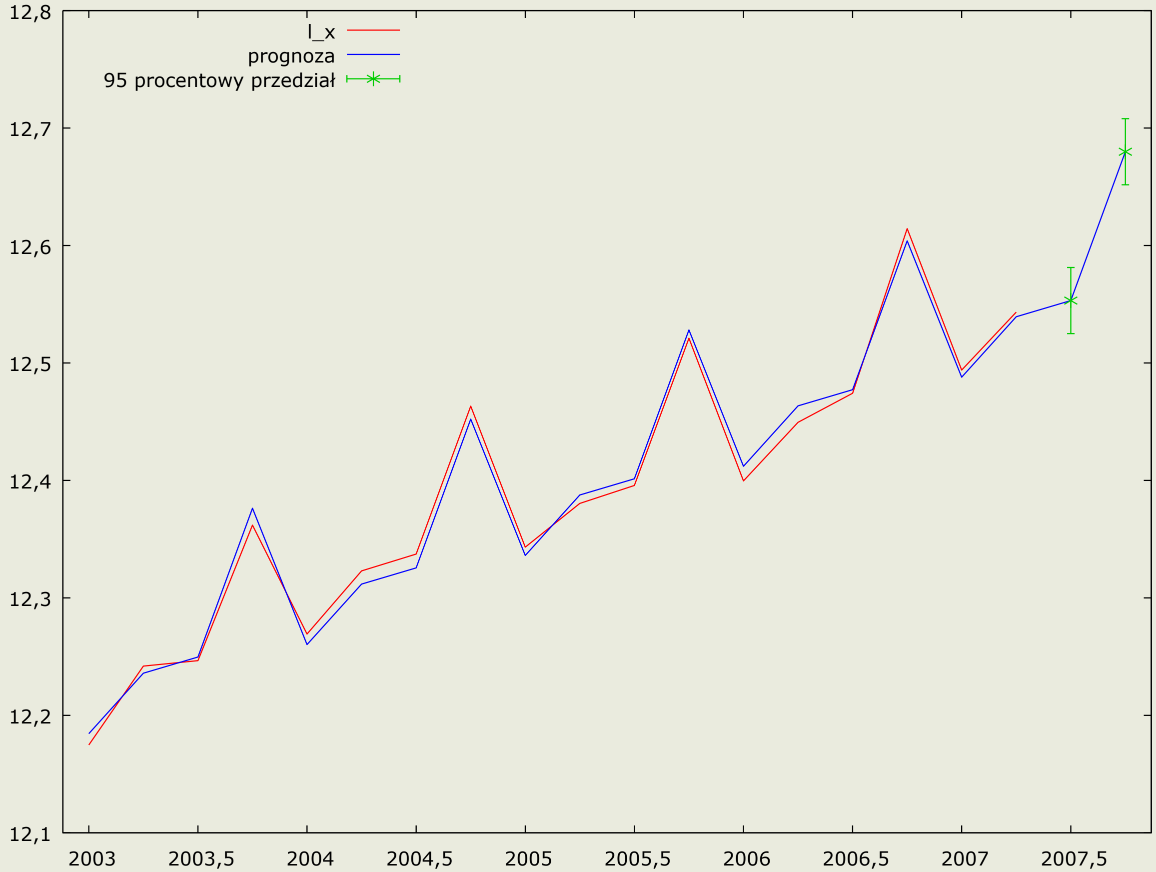
Model trendu z sezonowością

Model przyczynowo-skutkowy →



```
Ordinary Least Squares Estimation
*******************************************************************************
 Dependent variable is LNX
 18 observations used for estimation from 1 to 18
*******************************************************************************
 Regressor              Coefficient       Standard Error         T-Ratio[Prob]
 A                         12.2142            .0053324             2290.5[.000]
 T                         .018967            .4925E-3            38.5155[.000]
 SR1                      -.048705            .0042719           -11.4012[.000]
 SR2                      -.016201            .0042719            -3.7924[.002]
 SR3                      -.021415            .0045999            -4.6555[.000]
*******************************************************************************
 R-Squared                        .99347   R-Bar-Squared                   .99146
 S.E. of Regression              .010789   F-stat.    F(  4,  13)  494.2496[.000]
 Mean of Dependent Variable      12.3907   S.D. of Dependent Variable      .11673
 Residual Sum of Squares        .0015133   Equation Log-likelihood        58.9137
 Akaike Info. Criterion          53.9137   Schwarz Bayesian Criterion     51.6877
 DW-statistic                     1.0479
*******************************************************************************
                               Diagnostic Tests
*******************************************************************************
*    Test Statistics  *        LM Version         *         F Version         *
*******************************************************************************
*                     *                           *                           *
* A:Serial Correlation*CHSQ(   4)=   7.2438[.124]*F(   4,   9)=   1.5153[.277]*
*                     *                           *                           *
* B:Functional Form   *CHSQ(   1)=   .58035[.446]*F(   1,  12)=   .39979[.539]*
*                     *                           *                           *
* C:Normality         *CHSQ(   2)=   1.7389[.419]*       Not applicable       *
*                     *                           *                           *
* D:Heteroscedasticity*CHSQ(   1)= .6655E-4[.993]*F(   1,  16)= .5915E-4[.994]*
*******************************************************************************
```

$$\ln x_t = 12.214 + 0.019t - 0.049SR1 - 0.016SR2 - 0.021SR3 + \hat{\varepsilon}_t$$



Single Equation Static Forecasts

****

Based on OLS regression of LNX on:

A	T	SR1	SR2	SR3

18 observations used for estimation from 2003Q1 to 2007Q2

****

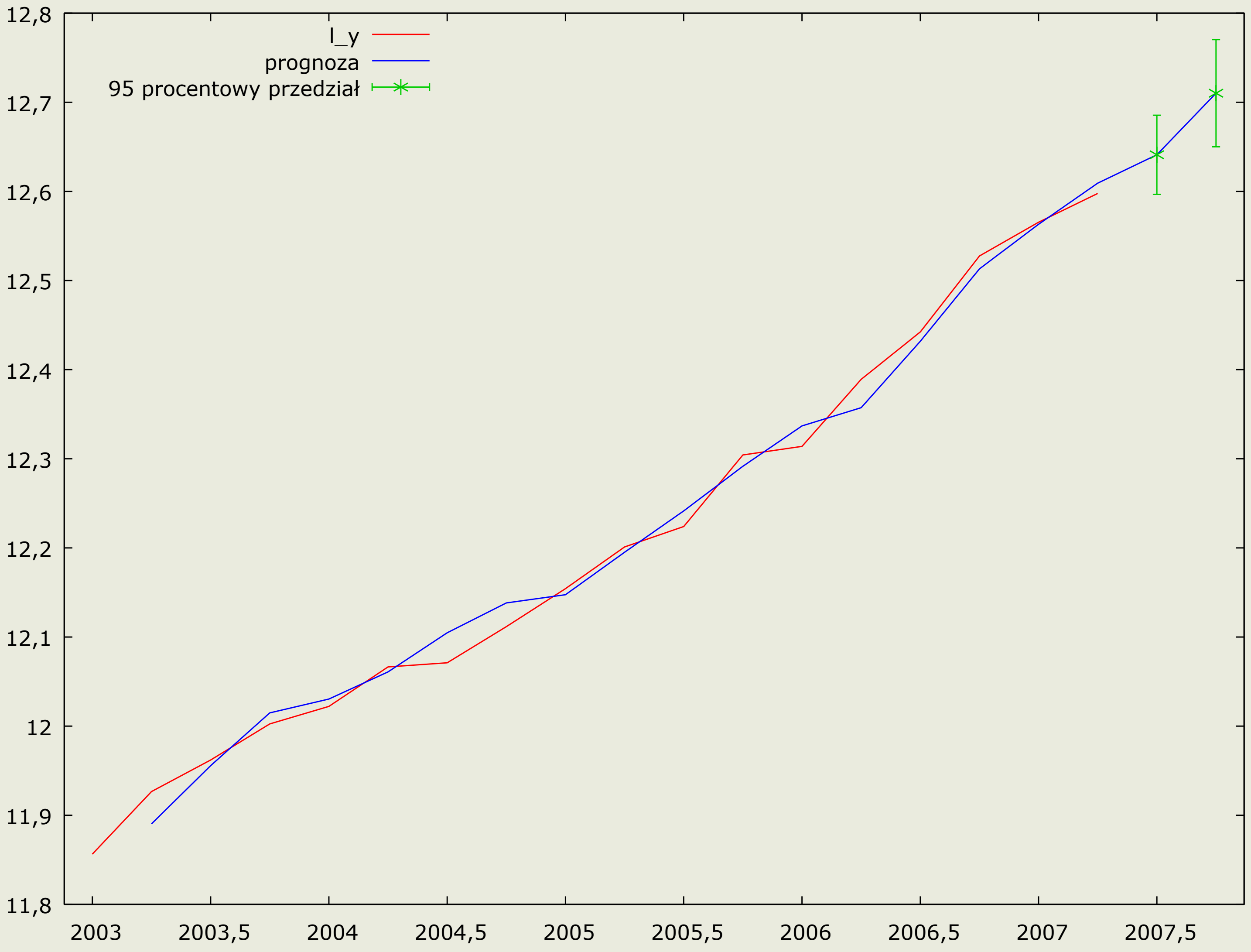
Observation	Actual	Prediction	Error	S.D. of Error
19	*NONE*	12.5531	*NONE*	.013029
20	*NONE*	12.6798	*NONE*	.013029

****

```
Ordinary Least Squares Estimation
*******************************************************************************
 Dependent variable is LNY
 17 observations used for estimation from 2003Q2 to 2007Q2
*******************************************************************************
 Regressor              Coefficient       Standard Error         T-Ratio[Prob]
 A                         -1.8029            .73094             -2.4665[.027]
 LNY(-1)                    .91523          .046870             19.5272[.000]
 LNX                       .23214          .093172              2.4915[.026]
*******************************************************************************
 R-Squared                        .99202   R-Bar-Squared                   .99088
 S.E. of Regression              .020687   F-stat.    F(  2,  14)  870.0630[.000]
 Mean of Dependent Variable      12.2283   S.D. of Dependent Variable      .21661
 Residual Sum of Squares        .0059914   Equation Log-likelihood        43.4586
 Akaike Info. Criterion          40.4586   Schwarz Bayesian Criterion     39.2088
 DW-statistic                     1.7874   Durbin's h-statistic      .44679[.655]
*******************************************************************************
                               Diagnostic Tests
*******************************************************************************
*    Test Statistics  *        LM Version        *         F Version         *
*******************************************************************************
*                     *                          *                           *
* A:Serial Correlation*CHSQ(   4)=  .038101[1.00]*F(   4,  10)= .0056157[1.00]*
*                     *                          *                           *
* B:Functional Form   *CHSQ(   1)=   1.3526[.245]*F(   1,  13)=   1.1237[.308]*
*                     *                          *                           *
* C:Normality         *CHSQ(   2)=   .28478[.867]*      Not applicable       *
*                     *                          *                           *
* D:Heteroscedasticity*CHSQ(   1)=   1.1114[.292]*F(   1,  15)=   1.0493[.322]*
*                     *                          *                           *
* E:Predictive Failure*CHSQ(   1)=  .010669[.918]*F(   1,  14)=  .010669[.919]*
*******************************************************************************
```

$$\ln y_t = -1{,}803 + 0.915 \ln y_{t-1} + 0.232 \ln x_t + \hat{\varepsilon}_t$$

```
Single Equation Dynamic Forecasts
*******************************************************************************
 Based on  OLS regression of LNY on:
 A                LNY(-1)        LNX
 17 observations used for estimation from 2003Q2 to 2007Q2
*******************************************************************************
 Observation         Actual        Prediction        Error        S.D. of Error
  19                 12.6435          12.6411       .0024469          .023689
  20                  *NONE*          12.7101         *NONE*          .036310
*******************************************************************************
           Summary statistics for single equation dynamic forecasts
*******************************************************************************
 Based on 1 observations from 19 to 19
 Mean Prediction Errors      .0024469   Mean Sum Abs Pred Errors     .0024469
 Sum Squares Pred Errors     .5987E-5   Root Mean Sumsq Pred Errors  .0024469
 Predictive failure test   F(   1,  14)=  .010669[.919]
*******************************************************************************
```



l_y
prognoza
95 procentowy przedział
12,8
12,7
12,6
12,5
12,4
12,3
12,2
12,1
12
11,9
11,8
2003
2003,5
2004
2004,5
2005
2005,5
2006
2006,5
2007
2007,5

Dziękuję za uwagę!

Aleksandra.Kordalska@pg.edu.pl