**Politechnika Opolska**

**Wydział Elektrotechniki, Automatyki i Informatyki**



**Projekt z ekonometrii**

Kierunek studiów: **Informatyka** Prowadzący:

Rok akademicki: **2013/2014** dr inż. Arkadiusz Gardecki

Rok studiów: **I** Autor:

Semestr: **I** Wojciech Bierda

**1. Cel projektu**

Celem projektu było stworzenie dla wybranych danych statystycznych modelu ekonometrycznego, dokonanie na podstawie wybranych współczynników i hipotez jego weryfikacji. Sprawdzono także, czy zmienne objaśniające istotnie oddziałują na zmienną endogeniczną oraz czy dopasowanie oszacowanego modelu do danych empirycznych jest wystarczające. Na końcu wykonano graficzne porównanie rzeczywistego przebiegu zmiennej objaśnianej z symulacyjnym przebiegiem tej zmiennej dla danych z przeszłości oraz porównanie wyników uzyskanych z modelu z następnym okresem funkcji predykcyjnej modelu.

**2. Dane statystyczne**

W modelu wykorzystano następujące dane statystyczne.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Rok** | **Y** | **X1** | **X2** | **X3** | **X4** |
| 1990 | 71,00 | 42,40 | 102,96 | 1591,70 | 4490,00 |
| 1991 | 70,69 | 33,80 | 175,60 | 2155,39 | 4200,00 |
| 1992 | 71,21 | 32,00 | 289,73 | 2996,10 | 3950,00 |
| 1993 | 71,69 | 27,90 | 390,43 | 4050,50 | 3845,00 |
| 1994 | 71,80 | 31,60 | 525,02 | 5459,00 | 3710,00 |
| 1995 | 72,01 | 38,10 | 690,92 | 7938,00 | 3496,00 |
| 1996 | 72,35 | 39,40 | 874,30 | 9981,00 | 3522,00 |
| 1997 | 72,72 | 42,60 | 1065,76 | 12144,00 | 3295,00 |
| 1998 | 73,11 | 45,20 | 1232,69 | 14211,00 | 2888,00 |
| 1999 | 73,16 | 42,00 | 1697,12 | 15913,37 | 2670,00 |

Tabela 2.1. Dane statystyczne wykorzystane w modelu

Gdzie:

Y - średnia długość życia ludności w Polsce,   
X1 - ilość rozwodów w tysiącach,   
X2 - przeciętne wynagrodzenie miesięczne brutto w zł.,   
X3 - PKB na 1 mieszkańca w zł.,   
X4 - całkowita emisja głównych zanieczyszczeń powietrza (dwutlenek siarki, dwutlenek azotu) w tys. ton.

**3. Wyznaczenie współczynników korelacji**

Ocena siły liniowej zależności pomiędzy zmienną objaśnianą Y, a zmiennymi objaśniającymi Xi została dokonana przy pomocy współczynnika korelacji ri określonego wzorem:



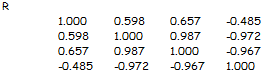
Po wyliczeniu poszczególnych współczynników korelacji otrzymano następujący wektor R0:



Następnie obliczone zostały współczynniki korelacji między potencjalnymi zmiennymi objaśniającymi. Do tego celu wykorzystano poniższy wzór:



Po wstawieniu odpowiednich elementów do wzoru otrzymano wynikową macierz współczynników korelacji R.



**4. Wybór zmiennych objaśniających**

Wyboru zmiennych objaśniających dokonano metodą analizy grafów dla poziomu istotności α=0,1.

Przed przejściem do właściwej analizy grafów należało wyznaczyć wartość krytyczną współczynnika korelacji, definiowaną wzorem:

*r*\* ,

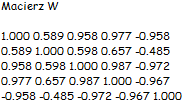
gdzie:

n – liczba obserwacji,

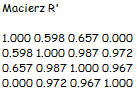
- wartość odczytana z tablic t-Studenta dla danego poziomu istotności.

Dla liczby obserwacji n = 10 (8 stopni swobody) oraz odczytanej wartości = 1,86 wyznaczono *r*\* = 0,549. Następnie skonstruowano macierz *W* postaci:

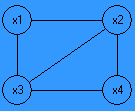
,

.

W kolejnym kroku w macierzy *W* współczynniki o wartości *r\** dla *i* ≠ j zastąpiono zerami. Po tej operacji otrzymano macierz *R*’, będącą odpowiednikiem dolnej części macierzy *W* (zawierającej wartości macierzy *R)*.



Analizując macierz wg. metody analizy grafów otrzymano następujący graf:



Następnie przystąpiono do wyboru odpowiedniej zmiennej objaśniającej. Wybrano zmienną X3, ponieważ wierzchołek ją reprezentujący posiada najwyższy stopień, podobnie jak wierzchołek X2, jednak zmienna X3 jest silniej skorelowana ze zmienną objaśnianą.

**5. Oszacowanie parametrów strukturalnych modelu**

Po wybraniu zmiennej objaśniającej dokonano oszacowania parametrów strukturalnych modelu stosując Metodę Najmniejszych Kwadratów. Ogólny wzór określający model wygląda następująco:

.

Do wyznaczenia parametrów *a*i posłużył wzór:

.

Aby możliwe było skorzystanie z powyższego wzoru wyznaczono macierze *X* oraz *Y*. Macierz *X* sprowadza się do wybrania kolumny reprezentującej wartości zmiennej X3 z tabeli zawierającej dane statystyczne i dopisania do niej kolumny jednostkowej. Natomiast macierz *Y* to kolumna reprezentująca wartości zmiennej Y.

, 

Mając macierze *X* oraz *Y* przystąpiono do wyznaczenia współczynników *a*1 oraz *a2*. Kolejno wyznaczono macierz *XT*, którą pomnożono przez macierz *X*. Otrzymaną w wyniku macierz odwrócono i pomnożono przez iloczyn macierzy *XT* i *Y*. Po wykonaniu tych operacji otrzymano macierz *a*.











Po podstawieniu wartości macierzy *a* do ogólnego wzoru opisującego model otrzymujemy:



**3. Weryfikacja modelu**

Weryfikacja modelu została przeprowadzona przy pomocy wybranych współczynników i hipotez.

Na początku za pomocą następującego wzoru wyznaczona została wariancja składnika resztowego:

,



Następnie obliczono odchylenie standardowe składnika resztowego reprezentowane wzorem:

,  


Kolejnym krokiem było określenie błędów średnich szacunku parametrów. W tym celu wykorzystano wzór:

,

następnie po policzeniu pierwiastka kwadratowego z wartości diagonalnych uzyskanej macierzy otrzymano wartości:

, .

Kolejnym wyznaczanym współczynnikiem był współczynnik zbieżności określony wzorem:

,

.

Następnym wyznaczonym współczynnikiem był współczynnik determinacji określony wzorem:



.

Na końcu wyznaczono współczynnik zmienności losowej *V*:

,

.

Ostatecznie po weryfikacji wzór modelu wygląda następująco:



**4. Statystyka t-Studenta**

Statystyka t-Studenta została wykorzystana w celu sprawdzenia, czy zmienne objaśniające istotnie oddziałują na zmienną endogeniczną.

Przyjęto hipotezę postaci:

,

oraz hipotezę alternatywną postaci:

.

Dla statystyki testowej danej wzorem:

,

otrzymano następujące wartości:

,.

Dla poziomu istotności =0,1 oraz *n*=8 (liczba stopni swobody) z tablic testu t-Studenta została odczytana wartość 1,85955.

Dla poziomu istotności =0,1 dokonano następujących porównań:

, bo 6,4>1,85955

oraz

, bo 485,172>1,85955.

Z wykonanych porównań wynika, że należy odrzucić hipotezę *H0* i za słuszną uznać hipotezę alternatywną.

**5. Statystyka F rozkładu Fischera Snedecora**

Po wykonaniu sprawdzenia przy pomocy statystki t-Studenta wykonano sprawdzenia, czy dopasowanie oszacowanego modelu do danych empirycznych jest wystarczające. Wykorzystano w tym celu statystykę F rozkładu Fischera Snedecora. Na początku zostały określone hipotezy:

,

.

Korzystając ze statystyki określonej wzorem:

,

gdzie *R2=*1-= 1-0,500 = 0,500,

otrzymano *F* = 8.

Następnie z tablic Fischera Snedecora odczytano wartość krytyczną statystyki dla poziomu istotności 0,1 oraz stopni swobody m1=1 i m2=8 równą *F*α = 3,46. Kolejnym krokiem było wykonanie porównania:

>*Fα* , bo 8>3,46.

Z wykonanego porównania wynika, że należy przyjąć hipotezę *H*1.

**6. Graficzna prezentacja wyników**

Dla otrzymanego wzoru określającego model ekonometryczny wykonano porównanie rzeczywistego przebiegu zmiennej objaśnianej z symulacyjnym przebiegiem tej zmiennej. Otrzymane wyniki przedstawiono w poniższej tabeli. Dodatkowo zaprezentowano wartość otrzymaną w kolejnym kroku (okresie) funkcji predykcyjnej dla danych z 2000 roku.

|  |  |  |
| --- | --- | --- |
| **Rok** | **Rzeczywisty Y** | **Symulacyjny Y** |
| 1990 | 71,00 | 71,05444 |
| 1991 | 70,69 | 71,16718 |
| 1992 | 71,21 | 71,33532 |
| 1993 | 71,69 | 71,5462 |
| 1994 | 71,80 | 71,8279 |
| 1995 | 72,01 | 72,3237 |
| 1996 | 72,35 | 72,7323 |
| 1997 | 72,72 | 73,1649 |
| 1998 | 73,11 | 73,5783 |
| 1999 | 73,16 | 73,91877 |
| 2000 | 73,78 | 74,1982 |

Tabela 6.1. Porównanie rzeczywistego Y z symulacyjnym

Graficzne porównanie rzeczywistego Y z symulacyjnym wygląda następująco.

Wykres 6.1. Porównanie rzeczywistego Y z symulacyjnym

Przedstawioną na wykresie wartość dla roku 2000 wyliczono podstawiając wartość PKB na   
1 mieszkańca w roku 2000, czyli 17310,50 zł (zmienna X3) do wzoru:



Otrzymano y = 7111,6942.

**7. Wnioski**

Wykorzystując statystykę t-Studenta można było skontrolować, czy zmienne objaśniające wpływają na zmienną endogeniczną. Ponieważ okazało się, że hipoteza alternatywna jest prawdziwą, dlatego można było stwierdzić, iż zmienna X3 wpływa na zmienną endogeniczną. Do podobnej konkluzji można było dojść po przeprowadzeniu rozkładu Fischera Snedecora, gdyż również została wybrana hipoteza alternatywna.

Poprawność modelu potwierdza również graficzne porównanie rzeczywistego przebiegu zmiennej objaśnianej z symulacyjnym. Dobrze on przedstawia jego właściwe działanie. Wartości predykcyjne są bliskie realnym. Nawet w przypadku roku 2000 błąd przewidywania wyniósł 0,5%.