

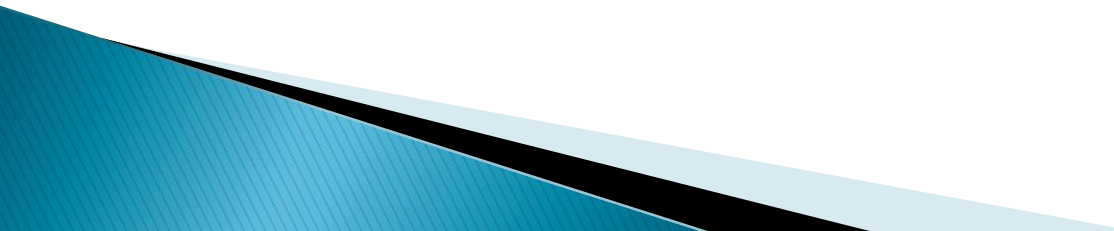
Metoda Zmiennych Instrumentalnych

Stanisław Cichocki

Natalia Nehrebecka

Wykład 2

Plan wykładu

- ▶ 1. Założenia KMRL
 - ▶ 2. MNK dla losowych zmiennych niezależnych
 - ▶ 3. Równoczesność
 - ▶ 4. Estymator MZI
 - ▶ 5. Testy diagnostyczne w MZI
- 

Plan wykładu

- ▶ 1. Założenia KMRL
- ▶ 2. MNK dla losowych zmiennych niezależnych
- ▶ 3. Równoczesność
- ▶ 4. Estymator MZI
- ▶ 5. Testy diagnostyczne w MZI

Założenia klasycznego modelu regresji liniowej

- ▶ 1. Związek pomiędzy zmienną zależną a zmiennymi niezależnymi opisany jest równaniem:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_K x_{Ki} + \varepsilon_i \quad i = 1, 2, 3, \dots, n$$

- ▶ 2. Zmienne objaśniające $x_{2i}, x_{3i}, \dots, x_{Ki}$ są nielosowe dla $i = 1, 2, 3, \dots, n$
- ▶ 3. Wartość oczekiwana błędu losowego jest równa zeru:

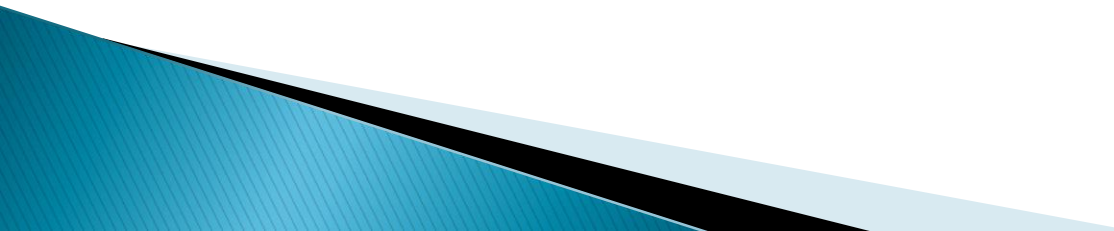
$$E(\varepsilon) = \mathbf{0}$$

- ▶ 4. Zaburzenia losowe ε są **sferyczne**. Oznacza to, że warunkowa macierz wariancji-kowariancji wektora zaburzeń przy danej macierzy X ma postać:

$$\text{Var}(\varepsilon) = \sigma^2 \mathbf{I}$$

- ▶ gdzie \mathbf{I} oznacza macierz jednostkową.

Plan wykładu

- ▶ 1. Założenia KMRL
 - ▶ 2. MNK dla losowych zmiennych niezależnych
 - ▶ 3. Równoczesność
 - ▶ 4. Estymator MZI
 - ▶ 5. Testy diagnostyczne w MZI
- 

MNK dla losowych zmiennych niezależnych

- ▶ Do tej pory przyjmowaliśmy założenie o nielosowości zmiennych objaśniających. Jest to sensowne w zastosowaniach eksperymentalnych. Jednak w badaniach empirycznych jest to trudne do obrony.
- ▶ Możemy wprowadzić analogiczny do KMRL układ założeń w przypadku losowej macierzy zmiennych objaśniających:

1. Model ma postać liniową:

$$y = X\beta + \varepsilon$$

2. Warunkowa wartość oczekiwania błędu losowego jest równa zero:

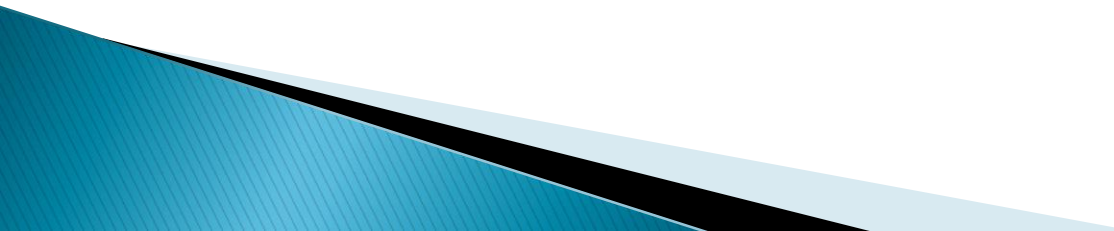
$$E(\varepsilon|X) = 0$$

MNK dla losowych zmiennych niezależnych

- ▶ 3. Warunkowa macierz wariancji i kowariancji ε nie zależy od X i jest równa:

$$\text{Var}(\varepsilon|X) = \sigma^2 I$$

Plan wykładu

- ▶ 1. Założenia KMRL
 - ▶ 2. MNK dla losowych zmiennych niezależnych
 - ▶ 3. Równoczesność
 - ▶ 4. Estymator MZI
 - ▶ 5. Testy diagnostyczne w MZI
- 

Równoczesność

- ▶ O występowaniu problemu **równoczesności (simultaneity)** mówimy, gdy nieprawdziwe jest założenie:

$$Cov(x_i, \varepsilon_i) = 0$$

- ▶ W przypadku występowania równoczesności estymator MNK **nie będzie zgodny**.
- ▶ Zmienną objaśniającą **skorelowaną** z błędem losowym będziemy nazywać **endogeniczną**.
- ▶ Zmienną objaśniającą **nieskorelowaną** z błędem losowym będziemy nazywać **egzogeniczną**.

Równoczesność

- ▶ Możemy wyróżnić **dwie przyczyny** równoczesności:
 - a) problem zmiennych pominiętych
 - b) sprzężenia zwrotne między zmiennymi

Równoczesność

▶ Przykład:

Model:

$$(1) \quad \ln(placa) = \beta_0 + \beta_1 educ + \beta_2 plec + \beta_3 wiek + \beta_4 wiek^2 + \beta_5 zdolnosc + \varepsilon$$

A szacujemy:

$$(2) \quad \ln(placa) = \beta_0 + \beta_1 educ + \beta_2 plec + \beta_3 wiek + \beta_4 wiek^2 + u$$

Błąd losowy w (2): $u = \beta_5 zdolnosc + \varepsilon$

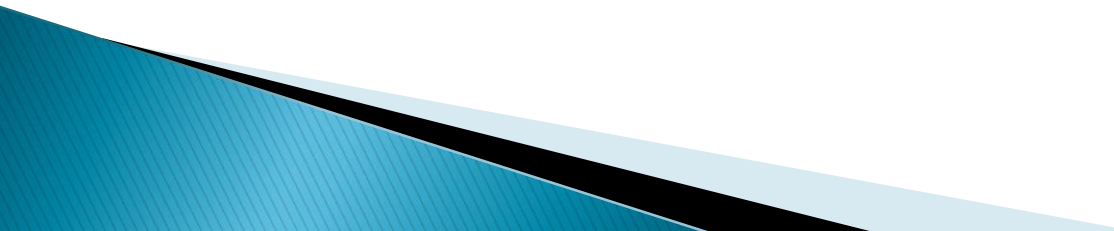
Równoczesność

- ▶ Przykład:

Więc:

$$\begin{aligned} Cov(educ, u) &= Cov(educ, \beta_5 zdolnosc + \varepsilon) \\ &= \beta_5 Cov(educ, zdolnosc) \neq 0 \end{aligned}$$

Plan wykładu

- ▶ 1. Założenia KMRL
 - ▶ 2. MNK dla losowych zmiennych niezależnych
 - ▶ 3. Równoczesność
 - ▶ 4. Estymator MZI
 - ▶ 5. Testy diagnostyczne w MZI
- 

Estymator MZI

- ▶ **Metoda Zmiennych Instrumentalnych** umożliwia uporanie się z problemem równoczesności.
- ▶ Estymacja za pomocą estymatora MZI jest możliwa, jeśli jesteśmy w stanie znaleźć **zmienne instrumentalne** (instrumenty, IV - instrumental variables), które są **nieskorelowane z błędem losowym**, ale są **skorelowane ze zmiennymi objaśniającymi**.
- ▶ Dodatkowa liczba zmiennych instrumentalnych **p** powinna być większa lub równa liczbie zmiennych objaśniających.

Estymator MZI

- ▶ Idea Metody Zmiennych Instrumentalnych polega na zastąpieniu **oryginalnych zmiennych objaśniających wartościami dopasowanymi tych zmiennych**, uzyskanymi z **regresji tych zmiennych na zmiennych instrumentalnych**.
- ▶ Oznaczmy przez **Z** macierz obserwacji dla zmiennych instrumentalnych.
- ▶ Wartości dopasowane zmiennych objaśniających **X**, uzyskane z regresji tych zmiennych na zmiennych instrumentalnych **Z**:

$$\hat{X} = Z\hat{B} = Z \underbrace{(Z'Z)^{-1}Z'X}_{\hat{B}} = P_Z X$$

gdzie: $P_Z = Z(Z'Z)^{-1}Z'$

Estymator MZI

- ▶ Estymator MZI uzyskujemy z regresji \mathbf{y} na wartościach dopasowanych \hat{X}

$$\begin{aligned}\tilde{b} &= (\hat{X}'\hat{X})^{-1}\hat{X}'\mathbf{y} = (X'P_ZX)^{-1}X'P_Z\mathbf{y} \\ &= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'\mathbf{y}\end{aligned}$$

Estymator MZI

- ▶ Estymator MZI określany jest jako **estymator dwustopniowej Metody Najmniejszych Kwadratów (2 MNK)**.
- ▶ Estymator ten można uzyskać w **dwóch krokach**:
 1. przeprowadzamy **regresję pomocniczą** endogenicznych zmiennych objaśniających na zmiennych instrumentalnych i uzyskujemy **wartości dopasowane endogenicznych zmiennych objaśniających**;
 2. zastępujemy **oryginalne wartości endogenicznych zmiennych objaśniających** ich **wartościami dopasowanymi** uzyskanymi w kroku 1. Następnie za pomocą **MNK** znajdujemy **oszacowania parametrów**.

Estymator MZI

- ▶ W przypadku gdy żadna ze zmiennych objaśniających nie jest endogeniczna a spełnione są założenia KMRL to estymator MZI jest estymatorem nieefektywnym.
- ▶ Wynika to bezpośrednio z **twierdzenia Gaussa-Markowa** - estymator MZI jest liniowy i nieobciążony ale różni się od estymatora MNK.
- ▶ Wniosek: **estymatora MZI nie powinno się używać** w przypadku gdy **nie występuje problem endogeniczności**.

Estymator MZI

Przykład I:

```
reg lnplaca eduk
```

Source	SS	df	MS	Number of obs	=	428
-----+-----				F(1, 426)	=	56.93
Model	26.3264193	1	26.3264193	Prob > F	=	0.0000
Residual	197.001022	426	.462443713	R-squared	=	0.1179
-----+-----				Adj R-squared	=	0.1158
Total	223.327441	427	.523015084	Root MSE	=	.68003

lnplaca	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
eduk	.1086487	.0143998	7.55	0.000	.0803451	.1369523
_cons	-.1851968	.1852259	-1.00	0.318	-.5492673	.1788736

Estymator MZI

```
reg eduk edukojca
```

Source	SS	df	MS	Number of obs	=	428
-----+-----				F(1, 426)	=	88.84
Model	384.841983	1	384.841983	Prob > F	=	0.0000
Residual	1845.35428	426	4.33181756	R-squared	=	0.1726
-----+-----				Adj R-squared	=	0.1706
Total	2230.19626	427	5.22294206	Root MSE	=	2.0813

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
edukojca	.2694416	.0285863	9.43	0.000	.2132538	.3256295
_cons	10.23705	.2759363	37.10	0.000	9.694685	10.77942

Estymator MZI

```
ivregress 2sls lnplaca (educ = edukojca)
```

Instrumental variables (2SLS) regression

```
Number of obs   =      428  
Wald chi2(1)    =      2.85  
Prob > chi2     =      0.0914  
R-squared       =      0.0934  
Root MSE       =      .68778
```

lnplaca	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
educ	.0591735	.0350596	1.69	0.091	-.009542	.127889
_cons	.4411034	.4450583	0.99	0.322	-.4311947	1.313402

Instrumented: educ

Instruments: edukojca

Estymator MZI

Przykład II:

```
reg lnplaca eduk
```

Source	SS	df	MS	Number of obs	=	935
-----+-----				F(1, 933)	=	100.70
Model	16.1377042	1	16.1377042	Prob > F	=	0.0000
Residual	149.518579	933	.160255712	R-squared	=	0.0974
-----+-----				Adj R-squared	=	0.0964
Total	165.656283	934	.177362188	Root MSE	=	.40032

lnplaca	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
eduk	.0598392	.0059631	10.03	0.000	.0481366	.0715418
_cons	5.973063	.0813737	73.40	0.000	5.813366	6.132759
-----+-----						

Estymator MZI

```
reg eduk liczbarodzen
```

Source	SS	df	MS	Number of obs	=	935
-----+-----				F(1, 933)	=	56.67
Model	258.055048	1	258.055048	Prob > F	=	0.0000
Residual	4248.7642	933	4.55387374	R-squared	=	0.0573
-----+-----				Adj R-squared	=	0.0562
Total	4506.81925	934	4.82528828	Root MSE	=	2.134

	educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----						
liczbarodzen		-.2279164	.0302768	-7.53	0.000	-.287335 - .1684979
_cons		14.13879	.1131382	124.97	0.000	13.91676 14.36083

Estymator MZI

```
ivregress 2sls lnplaca (educ = liczbarodzen)
```

Instrumental variables (2SLS) regression

```
Number of obs   =      935  
Wald chi2(1)    =      21.63  
Prob > chi2     =      0.0000  
R-squared       =          .  
Root MSE       =     .42284
```

lnplaca	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
educ	.1224326	.0263224	4.65	0.000	.0708417	.1740235
_cons	5.130026	.3547911	14.46	0.000	4.434648	5.825404

Instrumented: educ

Instruments: liczbarodzen

Estymator MZI

Przykład III:

```
reg lnwaga paczki
```

Source	SS	df	MS	Number of obs	=	1,388
-----+-----				F(1, 1386)	=	27.98
Model	.997781141	1	.997781141	Prob > F	=	0.0000
Residual	49.4225525	1,386	.035658407	R-squared	=	0.0198
-----+-----				Adj R-squared	=	0.0191
Total	50.4203336	1,387	.036352079	Root MSE	=	.18883

lnwaga	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
paczki	-.0898131	.0169786	-5.29	0.000	-.1231197	-.0565065
_cons	4.769404	.0053694	888.26	0.000	4.758871	4.779937

Estymator MZI

```
reg paczki cenapapier
```

Source	SS	df	MS	Number of obs	=	1,388
-----+-----						
Model	.011648626	1	.011648626	F(1, 1386)	=	0.13
Residual	123.684481	1,386	.089238442	Prob > F	=	0.7179
-----+-----						
Total	123.696129	1,387	.089182501	R-squared	=	0.0001
-----+-----						
				Adj R-squared	=	-0.0006
				Root MSE	=	.29873

paczki	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
cenapapier	.0002829	.000783	0.36	0.718	-.0012531	.0018188
_cons	.0674257	.1025384	0.66	0.511	-.1337215	.2685728
-----+-----						

Estymator MZI

```
ivregress 2sls lnwaga (paczki = cenapapier)
```

Instrumental variables (2SLS) regression

```
Number of obs   =    1,388  
Wald chi2(1)    =      0.12  
Prob > chi2     =    0.7310  
R-squared       =          .  
Root MSE       =    .93818
```

lnwaga	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
paczki	2.988676	8.692619	0.34	0.731	-14.04854	20.0259
_cons	4.448136	.9075006	4.90	0.000	2.669468	6.226805

Instrumented: paczki

Instruments: cenapapier

Estymator MZI

Przykład IV:

```
reg lnplaca eduk doswiad doswiad2
```

Source	SS	df	MS	Number of obs	=	428
-----+-----				F(3, 424)	=	26.29
Model	35.0222967	3	11.6740989	Prob > F	=	0.0000
Residual	188.305144	424	.444115906	R-squared	=	0.1568
-----+-----				Adj R-squared	=	0.1509
Total	223.327441	427	.523015084	Root MSE	=	.66642

lnplaca	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
eduk	.1074896	.0141465	7.60	0.000	.0796837	.1352956
doswiad	.0415665	.0131752	3.15	0.002	.0156697	.0674633
doswiad2	-.0008112	.0003932	-2.06	0.040	-.0015841	-.0000382
_cons	-.5220406	.1986321	-2.63	0.009	-.9124667	-.1316144
-----+-----						

Estymator MZI

```
reg eduk doswiad doswiad2 edukojsca edukmat
```

Source	SS	df	MS	Number of obs	=	428
-----+-----				F(4, 423)	=	28.36
Model	471.620998	4	117.90525	Prob > F	=	0.0000
Residual	1758.57526	423	4.15738833	R-squared	=	0.2115
-----+-----				Adj R-squared	=	0.2040
Total	2230.19626	427	5.22294206	Root MSE	=	2.039

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
doswiad	.0452254	.0402507	1.12	0.262	-.0338909	.1243417
doswiad2	-.0010091	.0012033	-0.84	0.402	-.0033744	.0013562
edukojsca	.1895484	.0337565	5.62	0.000	.1231971	.2558997
edukmat	.157597	.0358941	4.39	0.000	.087044	.2281501
_cons	9.10264	.4265614	21.34	0.000	8.264196	9.941084

-----+-----

Estymator MZI

test edukojca edukmat

(1) edukojca = 0

(2) edukmat = 0

F(2, 423) = 55.40

Prob > F = 0.0000

Estymator MZI

```
ivregress 2sls lnplaca doswiad doswiad2 (educ = edukojca edukmat)
```

Instrumental variables (2SLS) regression

```
Number of obs   =      428  
Wald chi2(3)    =      24.65  
Prob > chi2     =      0.0000  
R-squared       =      0.1357  
Root MSE       =      .67155
```

```
-----  
      lnplaca |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]  
-----+-----  
      educ    |   .0613966   .0312895     1.96   0.050     .0000704     .1227228  
      doswiad |   .0441704   .0133696     3.30   0.001     .0179665     .0703742  
      doswiad2|  -.000899    .0003998    -2.25   0.025    -.0016826    -.0001154  
      _cons   |   .0481003   .398453     0.12   0.904    -.7328532     .8290538  
-----
```

Instrumented: educ

Instruments: doswiad doswiad2 edukojca edukmat

Plan wykładu

- ▶ 1. Założenia KMRL
- ▶ 2. MNK dla losowych zmiennych niezależnych
- ▶ 3. Równoczesność
- ▶ 4. Estymator MZI
- ▶ 5. Testy diagnostyczne w MZI

Testy diagnostyczne w MZI

- ▶ Wszystkie testy diagnostyczne, które można przeprowadzić w kontekście MNK można także przeprowadzić w przypadku MZI. Dotyczy to m.in. testów na poprawność formy funkcyjnej, testów na normalność składników losowych, testów na heteroskedastyczność i testów na autokorelację.
- ▶ W **MZI** można jednak przeprowadzić **kilka dodatkowych testów**:
 - a) test Hausmanna-Wu (test na egzogeniczność);
 - b) test Sargana (test na poprawność instrumentów)

Testy diagnostyczne w MZI

- ▶ **Test Hausmana-Wu:**

Hipoteza zerowa: zmienne objaśniające są egzogeniczne

Hipoteza alternatywna: przynajmniej jedna zmienna objaśniająca nie jest egzogeniczna

Testy diagnostyczne w MZI

```
ivregress 2sls lnplaca doswiad doswiad2 (educ = edukojca edukmat)
```

```
Instrumental variables (2SLS) regression          Number of obs   =          428
                                                Wald chi2(3)    =          24.65
                                                Prob > chi2     =          0.0000
                                                R-squared       =          0.1357
                                                Root MSE       =          .67155
```

lnplaca	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
educ	.0613966	.0312895	1.96	0.050	.0000704	.1227228
doswiad	.0441704	.0133696	3.30	0.001	.0179665	.0703742
doswiad2	-.000899	.0003998	-2.25	0.025	-.0016826	-.0001154
_cons	.0481003	.398453	0.12	0.904	-.7328532	.8290538

```
Instrumented:  educ
```

```
Instruments:  doswiad doswiad2 edukojca edukmat
```

Testy diagnostyczne w MZI

```
estat endog
```

```
Tests of endogeneity
```

```
Ho: variables are exogenous
```

```
Durbin (score) chi2(1) = 2.80707 (p = 0.0938)
```

```
Wu-Hausman F(1,423) = 2.79259 (p = 0.0954)
```

Testy diagnostyczne w MZI

```
ivregress 2sls lnplaca (educ = liczbarodzen)
```

Instrumental variables (2SLS) regression

```
Number of obs   =      935  
Wald chi2(1)    =      21.63  
Prob > chi2     =      0.0000  
R-squared       =          .  
Root MSE       =     .42284
```

lnplaca	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
educ	.1224326	.0263224	4.65	0.000	.0708417	.1740235
_cons	5.130026	.3547911	14.46	0.000	4.434648	5.825404

Instrumented: educ

Instruments: liczbarodzen

Testy diagnostyczne w MZI

```
estat endog
```

```
Tests of endogeneity
```

```
Ho: variables are exogenous
```

```
Durbin (score) chi2(1) = 6.70647 (p = 0.0096)
```

```
Wu-Hausman F(1,932) = 6.73325 (p = 0.0096)
```

Testy diagnostyczne w MZI

- ▶ **Test Sargana:**

Hipoteza zerowa: instrumenty zostały dobrane prawidłowo

Hipoteza alternatywna: instrumenty nie zostały dobrane prawidłowo

Testy diagnostyczne w MZI

```
ivregress 2sls lnplaca doswiad doswiad2 (educ = edukojca edukmat)
```

```
Instrumental variables (2SLS) regression          Number of obs   =           428
                                                Wald chi2(3)    =           24.65
                                                Prob > chi2     =           0.0000
                                                R-squared       =           0.1357
                                                Root MSE       =           .67155
```

lnplaca	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
educ	.0613966	.0312895	1.96	0.050	.0000704	.1227228
doswiad	.0441704	.0133696	3.30	0.001	.0179665	.0703742
doswiad2	-.000899	.0003998	-2.25	0.025	-.0016826	-.0001154
_cons	.0481003	.398453	0.12	0.904	-.7328532	.8290538

```
Instrumented:   educ
```

```
Instruments:   doswiad doswiad2 edukojca edukmat
```


Testy diagnostyczne w MZI

estat overid

Tests of overidentifying restrictions:

Sargan (score) $\chi^2(1) = .378071$ (p = 0.5386)

Basman $\chi^2(1) = .373985$ (p = 0.5408)

Dziękuję za uwagę