



AKADEMIA GÓRNICZO-HUTNICZA  
im. Stanisława Staszica w Krakowie

WYDZIAŁ ZARZĄDZANIA



# Ceny za 1m<sup>2</sup> mieszkań w województwach w Polsce

---

Autor: Łukasz Chuchra  
Przedmiot: Ekonometria wielowymiarowa

Kraków, 16.01.2022 r.

1. Spis treści	
2. Cel projektu .....	2
3. Opis danych.....	2
4. Analiza danych .....	3
4.1. Statystyki opisowe.....	3
4.2. Wykresy zmiennych według województw .....	4
4.3. Wykresy rozrzutu .....	14
4.4. Analiza korelacji.....	16
5. Estymacja modelu .....	17
5.1. Model regresji łącznej .....	17
5.2. Model efektów ustalonych (FE) .....	18
5.3. Model efektów losowych (RE).....	19
6. Dobór zmiennych .....	20
7. Model dwukierunkowy .....	21
8. Weryfikacja modelu .....	24
9. Model dynamiczny.....	28
9.1. Model pierwszych różnic.....	28
9.2. Metoda zmiennych instrumentalnych .....	31
10. Podsumowanie .....	33

## 2. Cel projektu

Celem projektu jest wyestymowanie modelu najlepiej wyjaśniającego ceny za 1m<sup>2</sup> mieszkań w województwach w Polsce w latach 2010 – 2020. Estymację modeli przeprowadzano z użyciem programu GRETl.

## 3. Opis danych

Dane wykorzystane w projekcie w całości pochodzą ze strony Banku Danych Lokalnych<sup>1</sup>. Dane dotyczą wartości rocznych zmiennych w latach 2010 – 2020 dla wszystkich województw w Polsce

Zmienną objaśnianą w badaniu była *miesz*, czyli średnia cena 1m<sup>2</sup> lokali mieszkalnych. Po analizie wpływu zmiennych na ceny mieszkań wybrano 11 zmiennych objaśniających, które potencjalnie mogą wpływać na zmienną objaśnianą. Poniżej przedstawiono wybrany zbiór danych wraz z opisem.

- *miesz* - średnia cena za 1 m<sup>2</sup> lokali mieszkalnych
- *wynag* – średnie miesięczne wynagrodzenie brutto
- *komun* - przewozy pasażerskie przypadający na 1 osobę w ciągu roku
- *drogi* - drogi publiczne ogółem o nawierzchni twardej i gruntowej (w tys. km)
- *lud* – liczba ludności (w mln)
- *turys* - noclegi udzielone turystom zagranicznym (nierezydentom) w turystycznych obiektach noclegowych (w tys.)
- *urban* - ludność w miastach w % ogółu ludności
- *inf* – wskaźnik inflacji
- *bezrob* - stopa bezrobocia rejestrowanego (liczba bezrobotnych zarejestrowanych w stosunku do ogólnej liczby ludności w danym województwie w danym roku)
- *lek* - lekarze (personel pracujący ogółem) na 100 tys. ludności
- *skol* - współczynnik skolaryzacji netto w szkołach podstawowych (liczba uczniów szkół podstawowych w wieku 7-15 lat na początku danego roku szkolnego przez liczbę ludności w wieku 7-15 lat według stanu w dniu 31 XII tego samego roku; wynik podany w ujęciu procentowym)
- *kina* - liczba ludności na 1 miejsce w kinach stałych

Zmienne takie jak inflacja, stopa bezrobocia i średnie zarobki bardzo dobrze oddają sytuację ekonomiczną województwa w danym roku. Jako potencjalnych kandydatów wybrano również stopień urbanizacji i liczba ludności, które miały oddawać tendencje do osiedlania w miejscach o większym nagromadzeniu ludzi. Kolejnymi wskaźnikami, które wzięto pod uwagę był dostęp do komunikacji oraz liczba dróg (w kilometrach). Zmienne te miały oddawać stopień skomunikowania. Na warunki życia zdecydowanie wpływa również dostępność do służby

---

<sup>1</sup> <https://bdl.stat.gov.pl/BDL/start>

zdrowia, edukacji oraz szeroko pojętej kultury. Wpływy uwzględniono poprzez dodanie w badaniu zmiennych *lek*, *skol* oraz *kina* (jako wskaźnik dostępu do kultury). Dzięki temu sprawdzono, czy ceny mieszkań można wyjaśnić za pomocą turystyki w danym województwie.

Z uwagi na prowadzenie całej analizy w GRETl oznaczenia województw przekształcono na unikatowe identyfikatory:

- 1 = 'DOLNOŚLĄSKIE'
- 2 = 'KUJAWSKO-POMORSKIE'
- 3 = 'LUBELSKIE'
- 4 = 'LUBUSKIE'
- 5 = 'ŁÓDZKIE'
- 6 = 'MAŁOPOLSKIE'
- 7 = 'MAZOWIECKIE'
- 8 = 'OPOLSKIE'
- 9 = 'PODKARPACKIE'
- 10 = 'PODLASKIE'
- 11 = 'POMORSKIE'
- 12 = 'ŚLĄSKIE'
- 13 = 'ŚWIĘTOKRZYSKIE'
- 14 = 'WARMIŃSKO-MAZURSKIE'
- 15 = 'WIELKOPOLSKIE'
- 16 = 'ZACHODNIOPOMORSKIE'

## 4. Analiza danych

### 4.1. Statystyki opisowe

	zliczenia	średnia	odch. stand	min	25%	median	75%	max	wariancja	skośność	kurtoza	wsp. zmienności
<b>miesz</b>	176.00	3.97	1.18	2.27	3.15	3.72	4.52	7.61	1.39	1.08	0.71	30%
<b>wynag</b>	176.00	3.99	0.73	2.88	3.42	3.84	4.48	6.58	0.53	0.82	0.40	18%
<b>komun</b>	176.00	79.77	42.07	15.01	41.88	82.97	103.02	212.50	1770.19	0.71	0.40	53%
<b>drogi</b>	176.00	26.16	10.21	10.36	19.84	24.65	28.51	55.48	104.20	1.16	1.39	39%
<b>lud</b>	176.00	2.40	1.23	0.98	1.38	2.13	3.02	5.43	1.52	0.96	0.14	51%
<b>turys</b>	176.00	340.49	383.08	14.25	105.26	169.74	430.40	1684.23	146746.93	1.86	2.78	113%
<b>urban</b>	176.00	58.67	9.61	41.09	51.19	60.32	64.53	77.83	92.41	-0.14	-0.73	16%
<b>inf</b>	176.00	101.76	1.73	98.30	100.15	102.05	103.13	104.90	2.98	-0.17	-1.09	2%
<b>bezrob</b>	176.00	10.49	4.15	2.80	7.28	10.15	13.33	21.60	17.21	0.36	-0.51	40%
<b>lek</b>	176.00	4.94	1.10	2.86	4.08	4.89	5.55	8.05	1.21	0.42	-0.28	22%
<b>skol</b>	176.00	93.09	2.38	86.22	91.73	93.13	94.58	98.61	5.69	-0.21	0.09	3%
<b>kina</b>	176.00	164.85	41.38	88.00	132.00	165.50	196.00	285.00	1712.13	0.34	-0.58	25%

Można zauważyć, że zmienne *skol* oraz *inf* charakteryzowały się niskim współczynnikiem zmienności. Z uwagi na to postanowiono je wyłączyć z dalszej analizy.

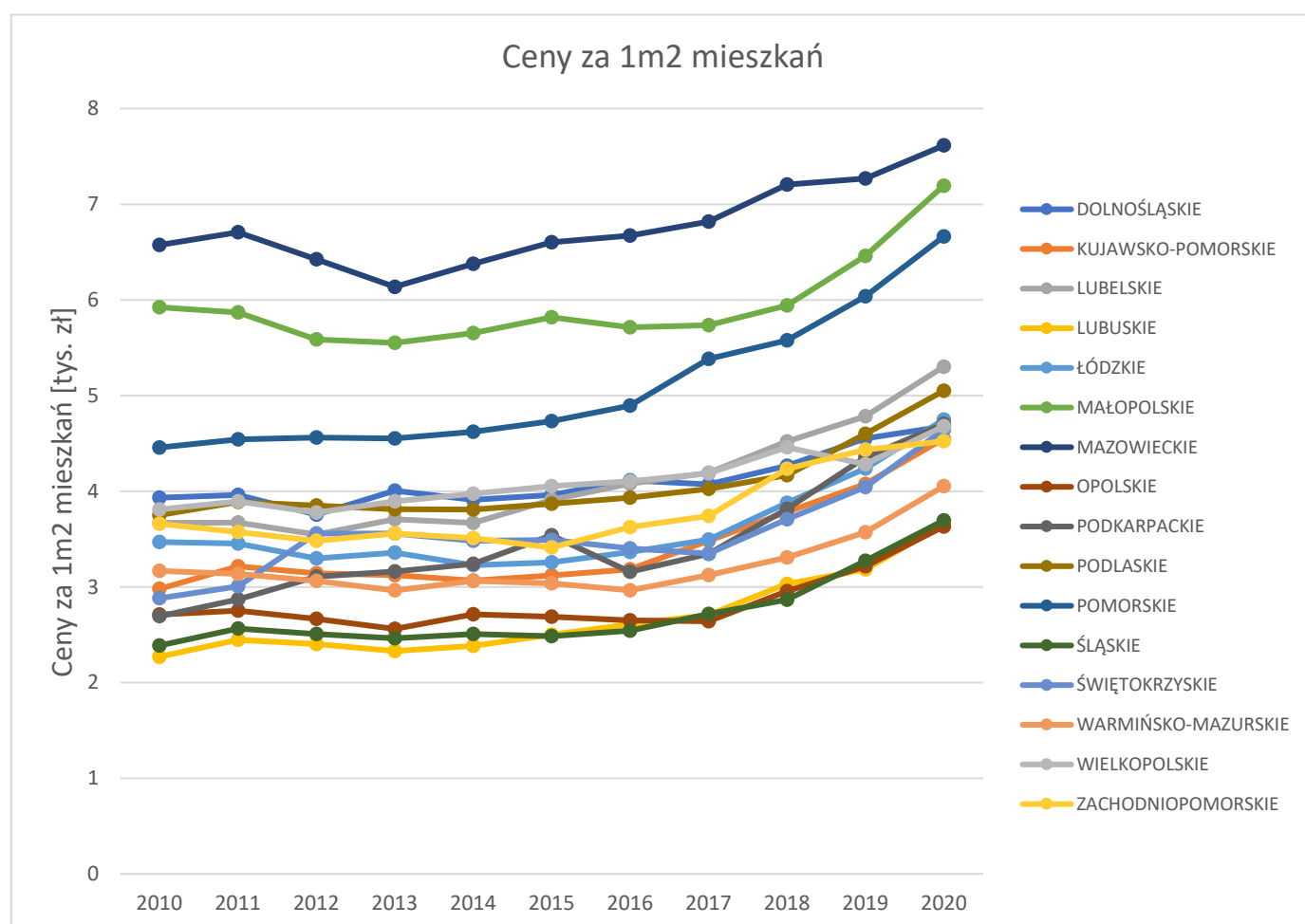
Można zauważyć, że zmienna *turys* ma znacznie wyższe wartości od pozostałych zmiennych (o 2 rzędy wielkości średnia od zmiennej objaśnianej). Przy okazji osiąga również wyższy współczynnik zmienności. Może być to spowodowane znacznym spadkiem turystyki w roku 2020.

Przy analizie skośności można zauważyć asymetrię rozkładu zmiennej objaśnianej *miesz* oraz objaśniających *drogi* oraz *turys*. Może to być spowodowane mocniejszymi tendencjami wzrostowymi zmiennych w przeciągu ostatnich kilku lat w porównaniu do całego okresu 2010 - 2020.

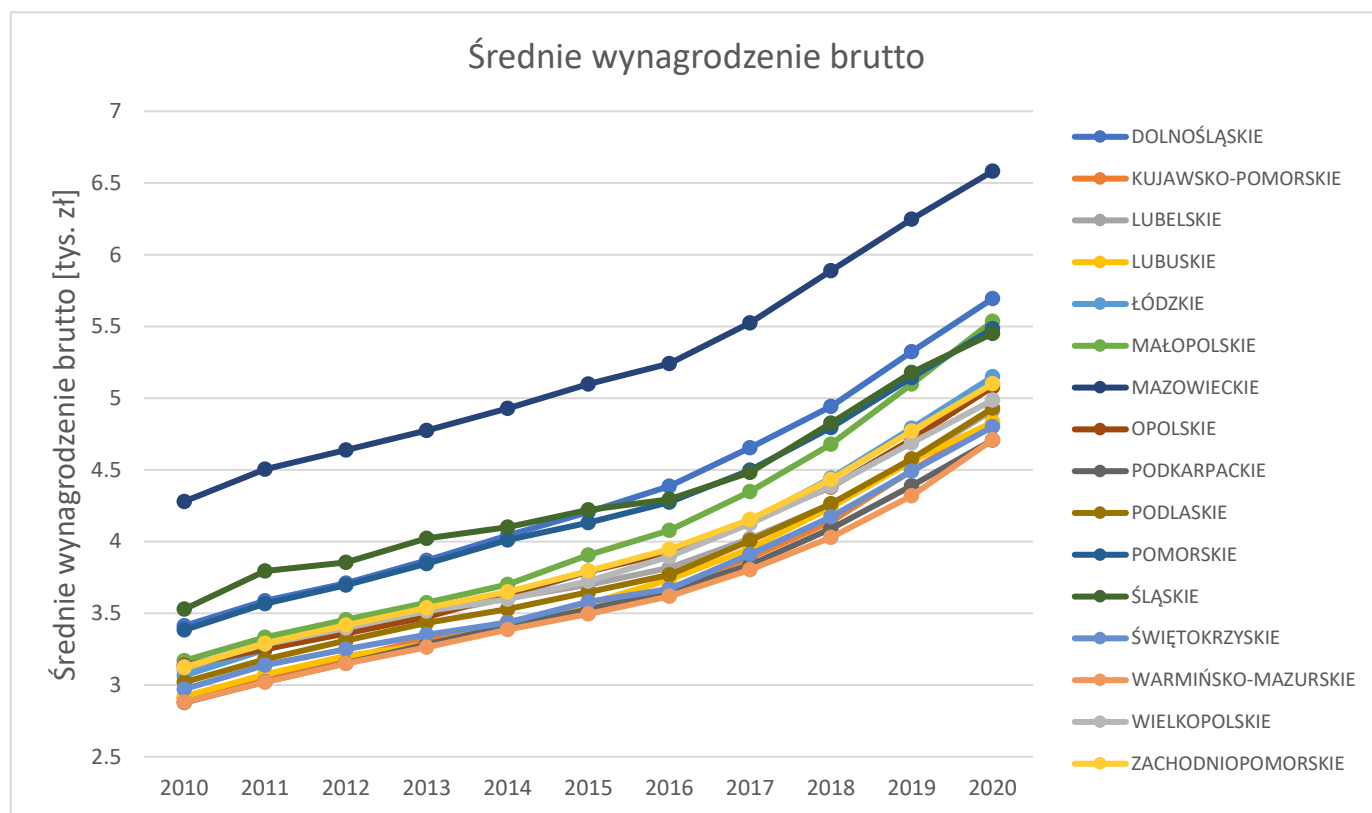
Jeśli chodzi o kurtozę to dla zmiennych *drogi* oraz *turys* przyjmuje ona mocno dodatnie wartości co świadczy o dużej koncentracji wyników wokół średniej. Odwrotnie dla zmiennej *inf*, która ma ujemną wartość, co świadczy o dużym rozrzucie wartości.

#### 4.2. Wykresy zmiennych według województw

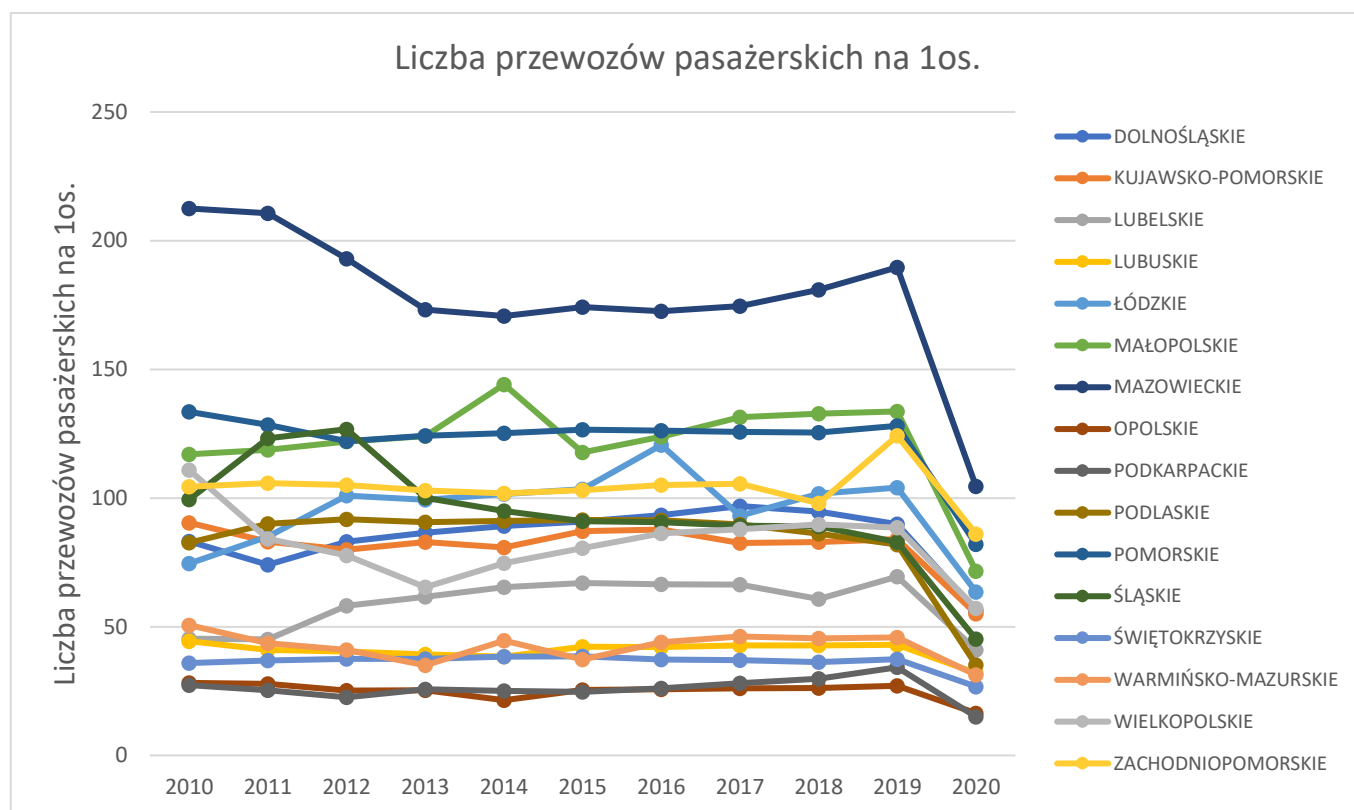
Poniżej przedstawiono wykresy analizowanych zmiennych w danych województwach



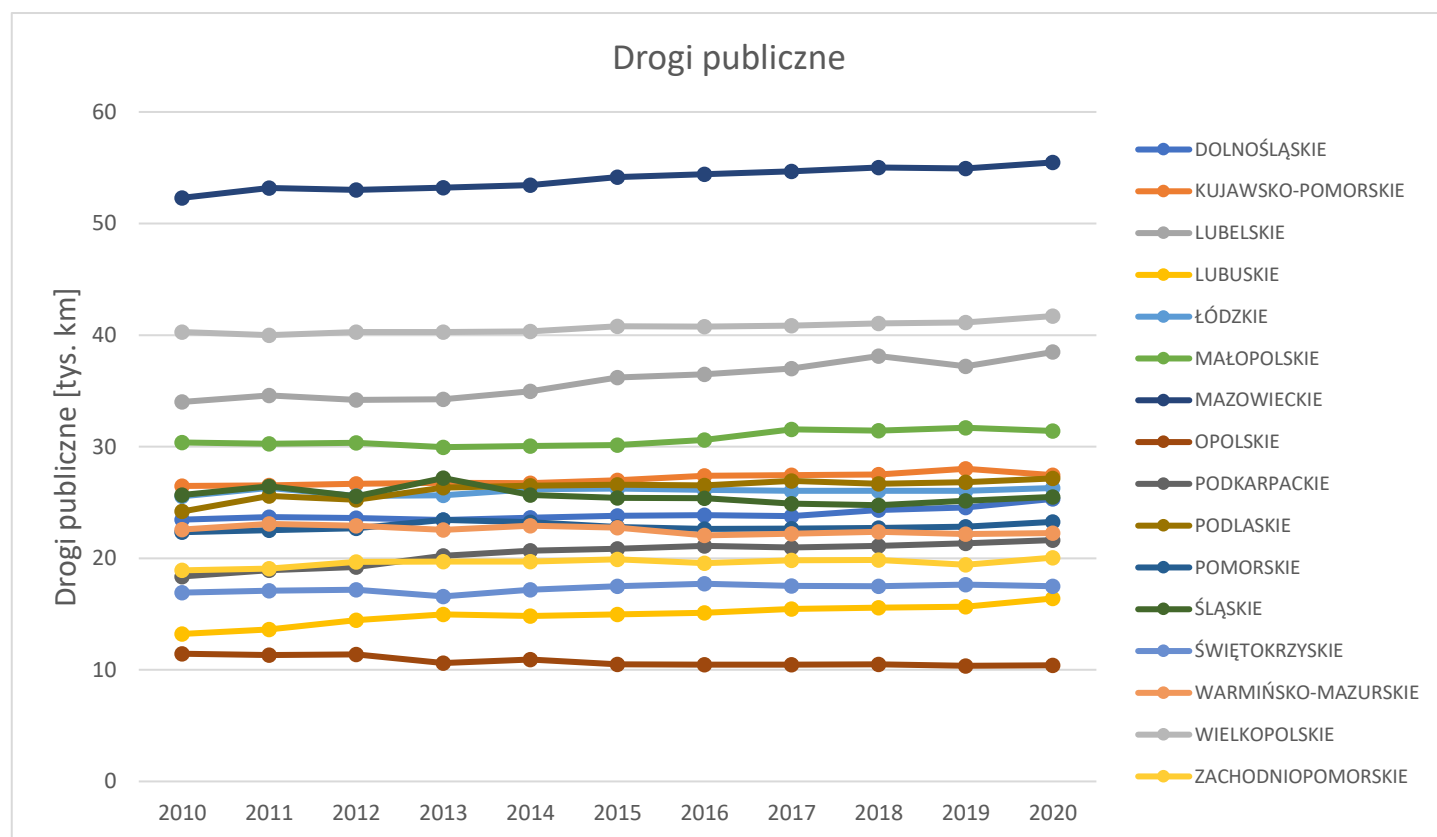
W każdym województwie można zauważyć tendencję wzrostową z biegiem lat. Szczególny wzrost można zauważyć zwłaszcza po 2015. Najwyższe ceny można zauważyć zwłaszcza w woj. Mazowieckim, Małopolskim oraz Pomorskim.



Tendencję wzrostową można zauważyć również w średnich wynagrodzeniach brutto. Dla każdego województwa można zaobserwować wzrost wynagrodzeń. Powyżej średnią dla województw znacznie wybijają się woj. Mazowieckie. Tam zarobki są wyższe o około 1000 zł w stosunku do reszty.

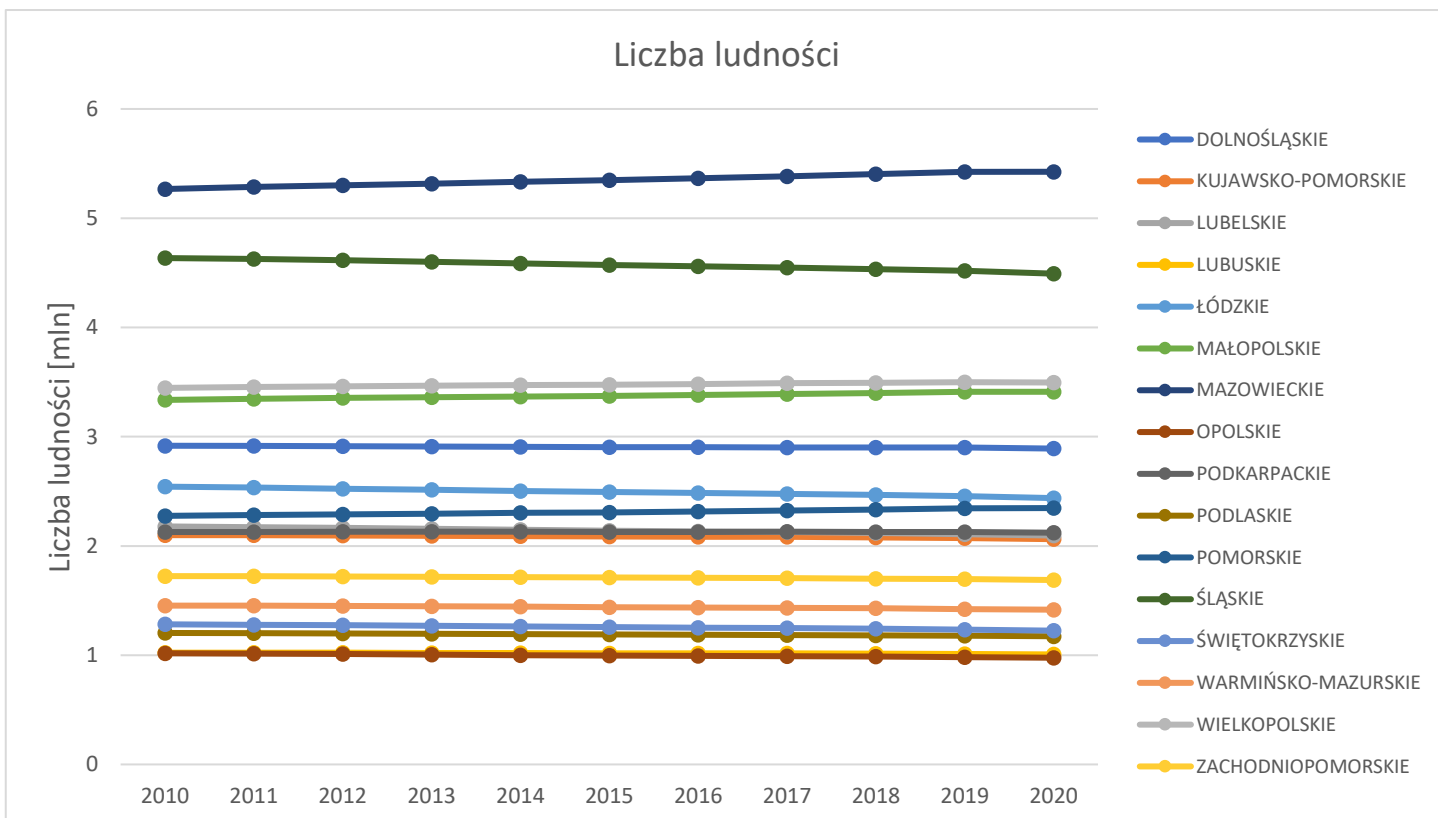


Najwięcej przewozów odnotowano dla woj. Mazowieckiego. Dużą liczbę przewozów odnotowano również w małopolskim i pomorskim. Warto zauważyć znaczny spadek liczby przejazdów w 2020 roku. Na skutek pandemii wiele linii miało ograniczenia na liczbę pasażerów, co jak widać poskutkowało znacznymi spadkami średniej liczby przewozów.

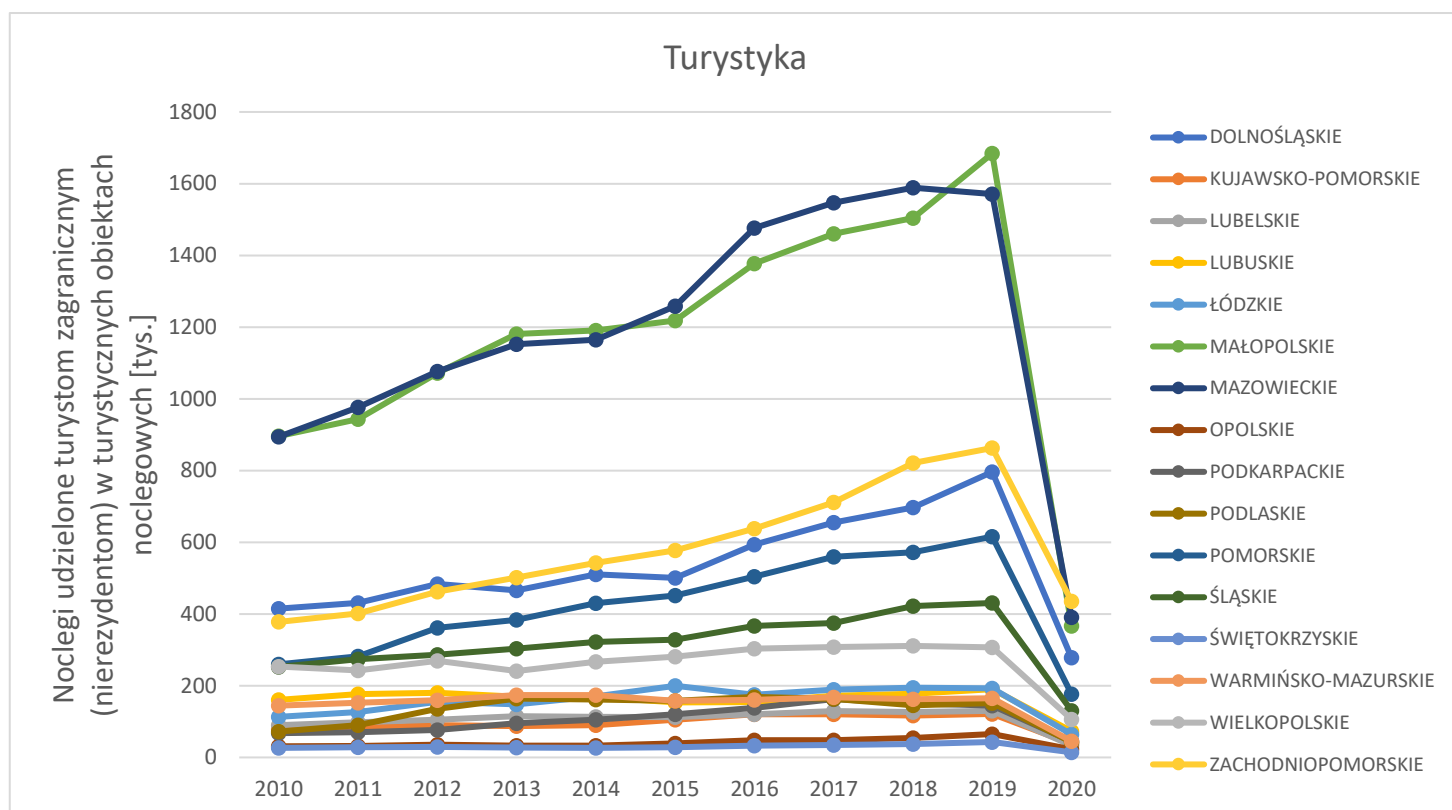


Najwięcej dróg znowu odnotowano w mazowieckim. Sporo dróg również posiadają woj. Lubelskie i Dolnośląskie.



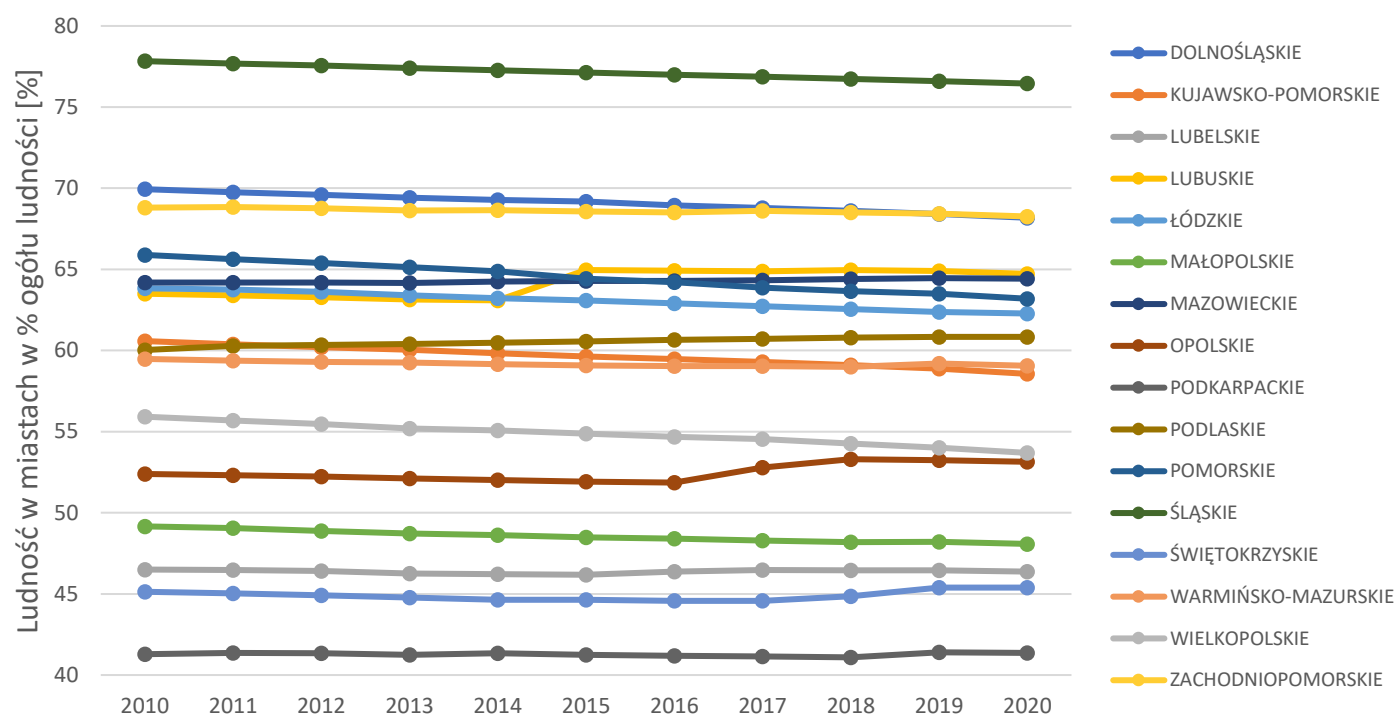


Liczba ludności w poszczególnych województwach jest niemal stała. Dla większości województw można zauważyć jednak spadku ludności. Wzrosty można za to odnotować w województwach, na terenach których znajdują się największe miasta w Polsce (Kraków, Warszawa, Poznań, Gdańsk).

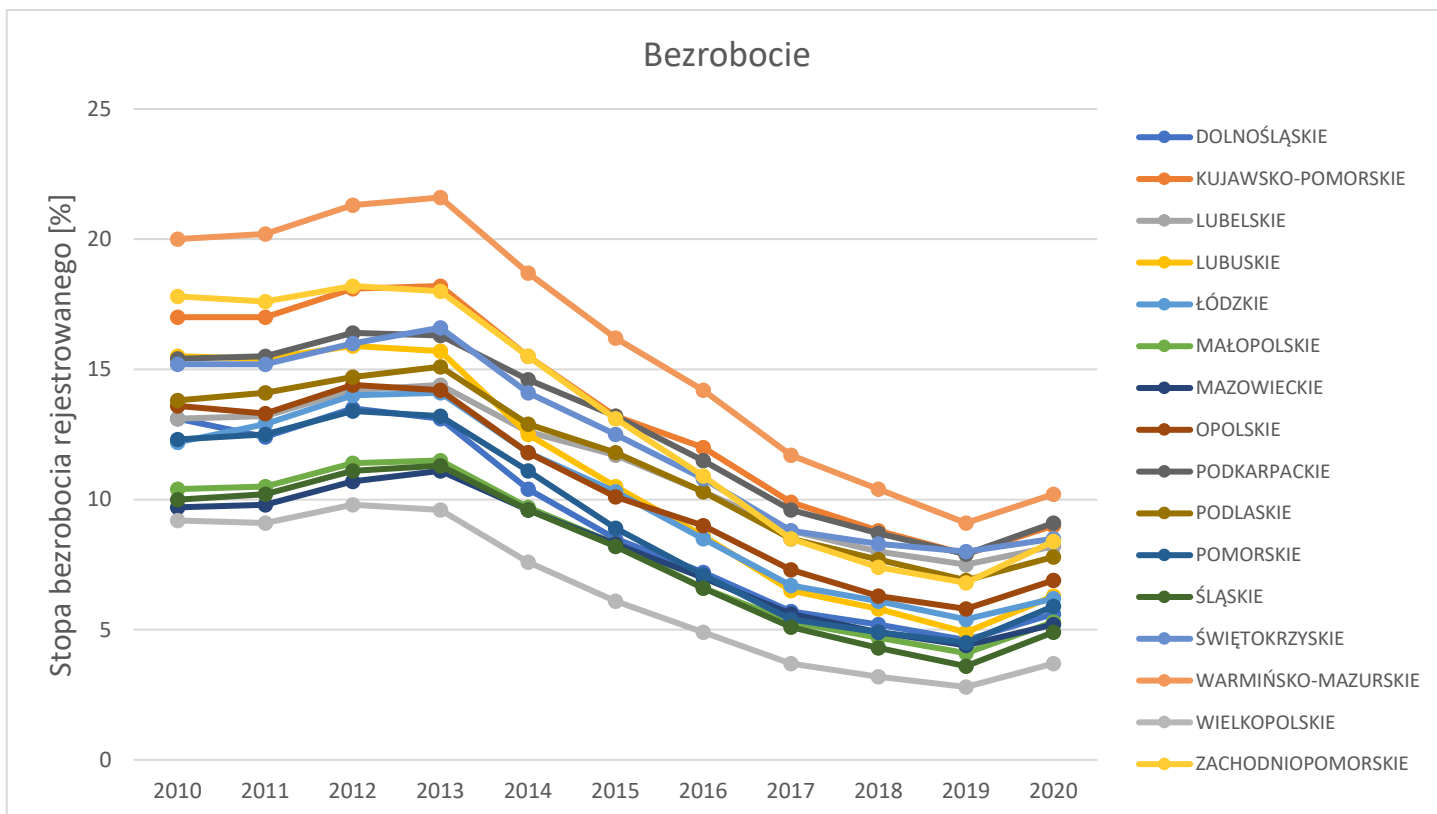


Jeśli chodzi o turystykę, to w statystykach przoduje małopolskie i mazowieckie. Dla najbardziej turystycznych województw widać tendencję wzrostową. Również tutaj można zauważyć spadki związane z pandemią w 2020 roku.

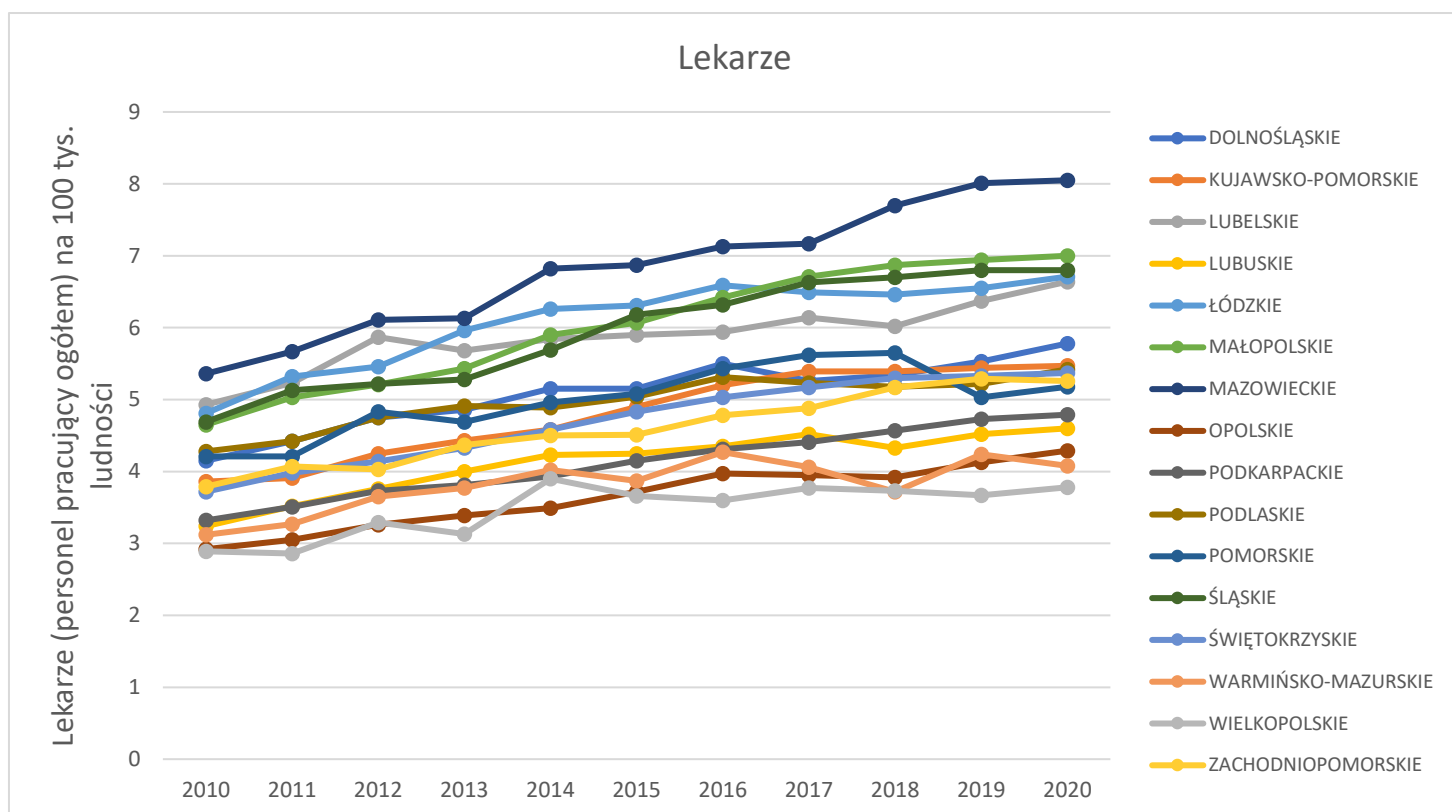
## Urbanizacja



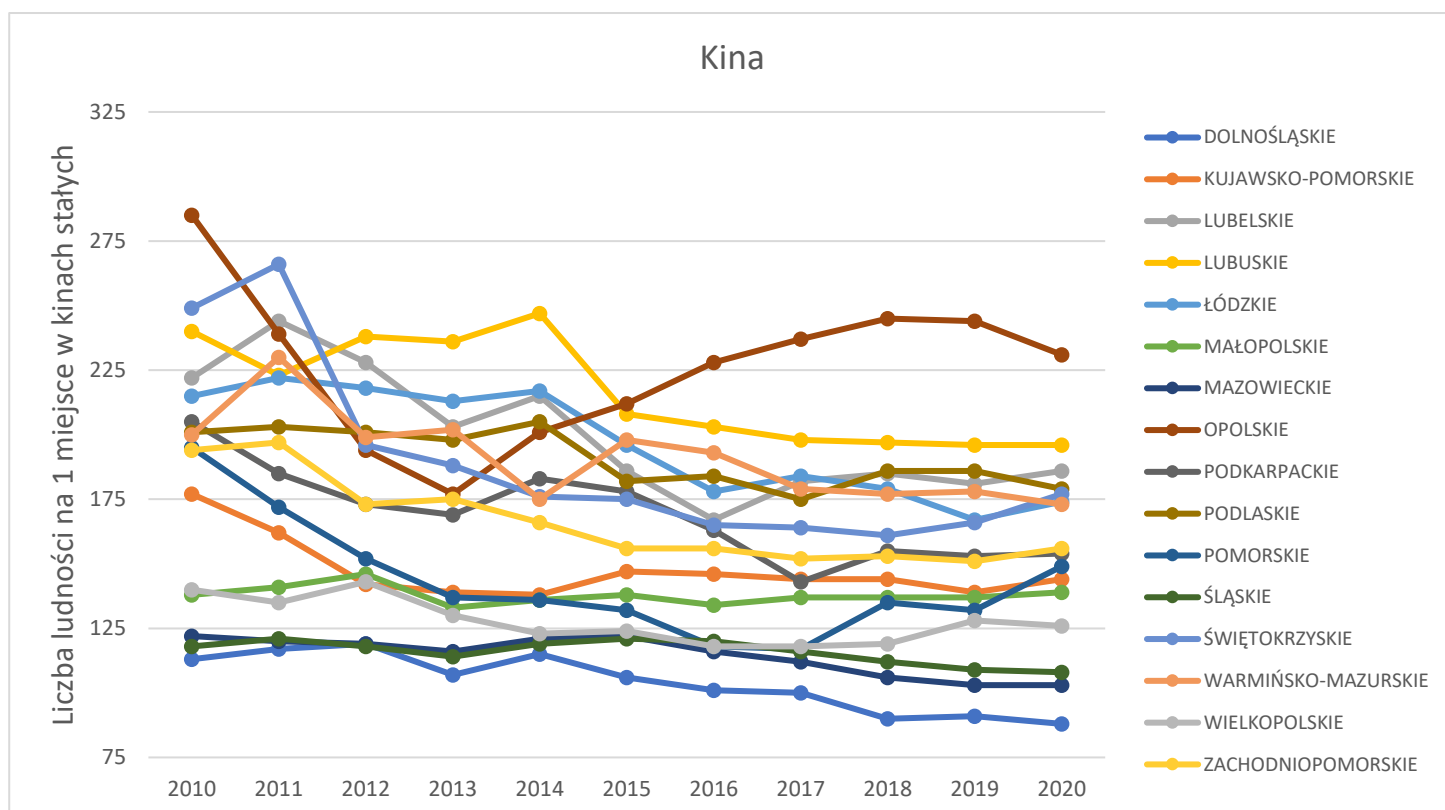
Dla większości województw można zauważyć spadki stopnia urbanizacji. Wyjątkami są podkarpackie, świętokrzyskie, lubuskie oraz mazowieckie



Można zauważyć znaczny spadek bezrobocia w poprzednich latach. Najwyższe bezrobocie rejestrowane odnotowano w 2013 roku, a najniższe w 2019. W 2020 (prawdopodobnie na skutek pandemii) obostrzeń związanych z pandemią odnotowano wzrost stopy bezrobocia.



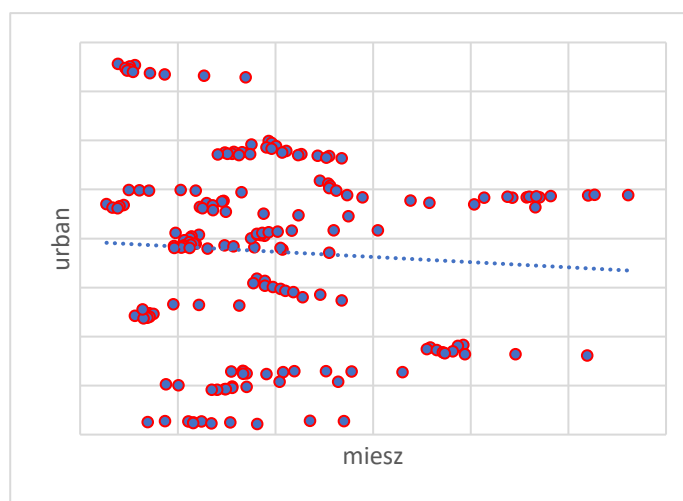
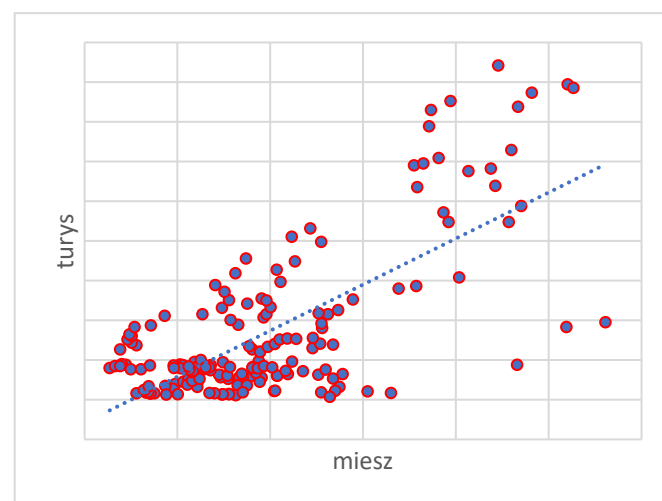
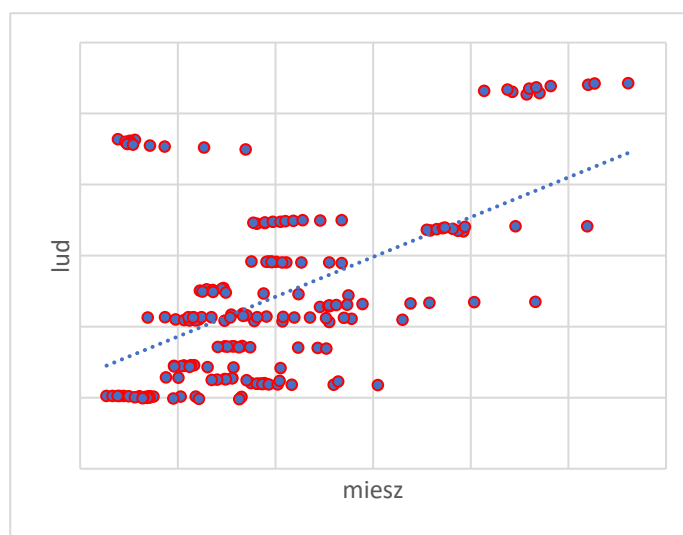
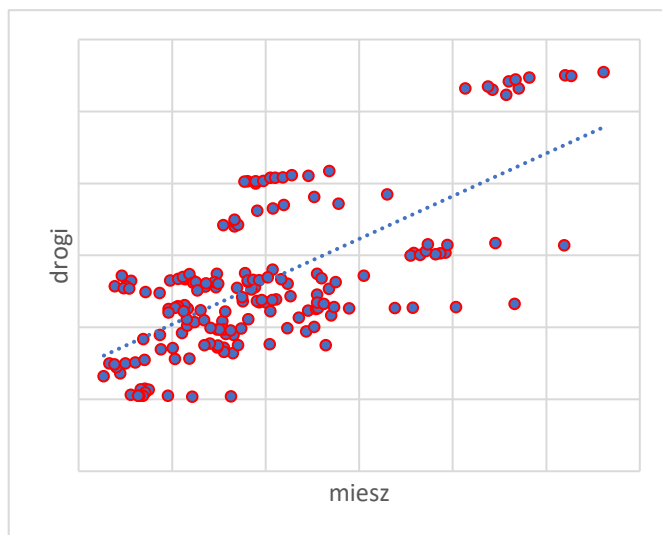
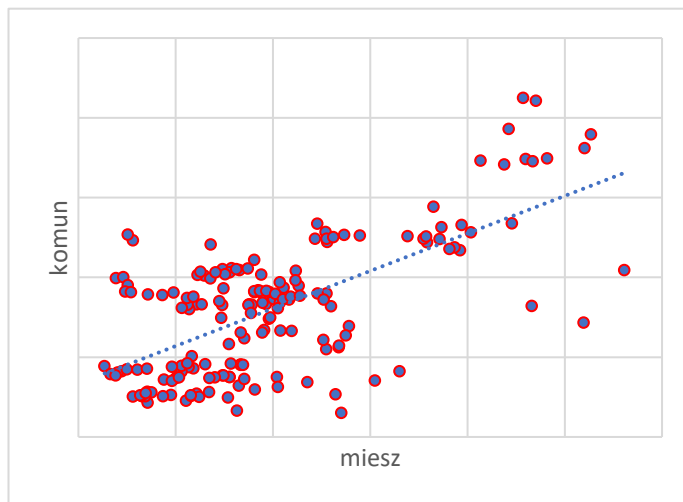
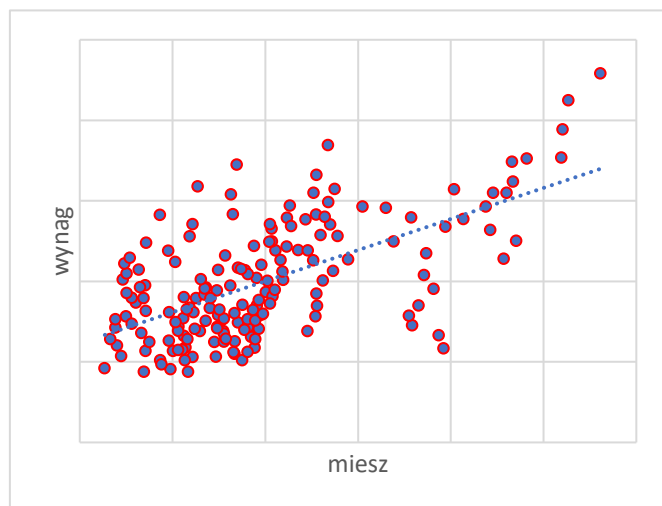
Najwięcej lekarzy w przeliczeniu na liczbę mieszkańców odnotowano na Mazowszu, mniej w małopolsce, śląsku, łódzkim i lubelskim. Można, więc zauważyć, że są to województwa o wysoko rozwiniętych szkołach wyższych z kierunkami medycznymi. Tendencja w każdym województwie jest wzrostowa.

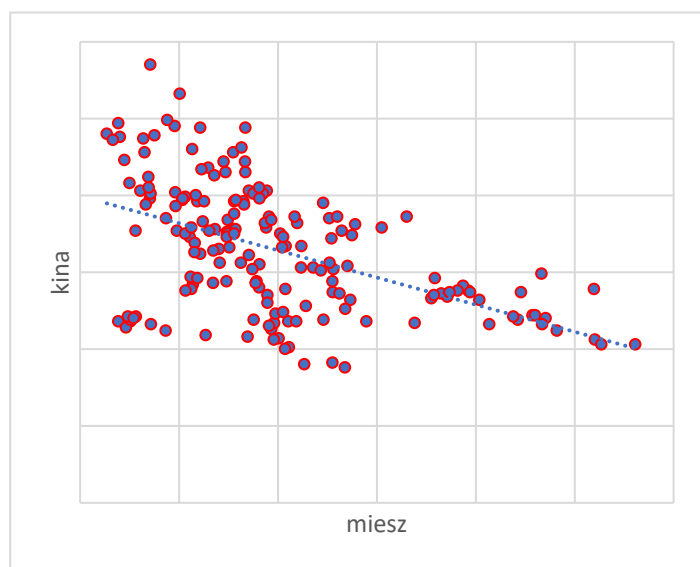
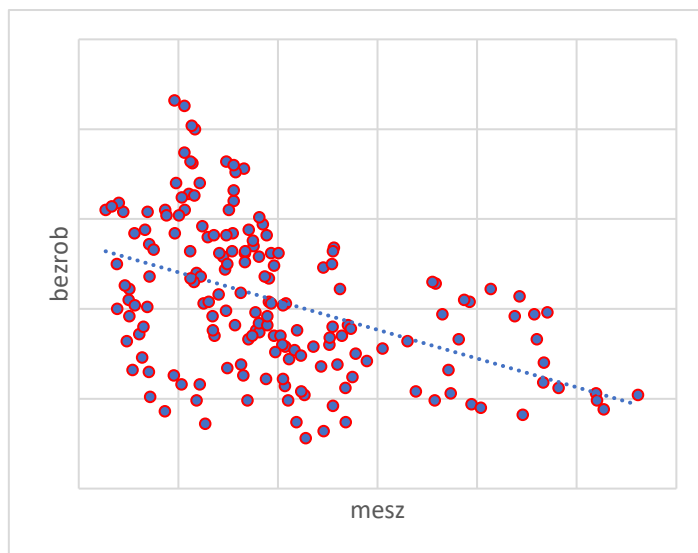
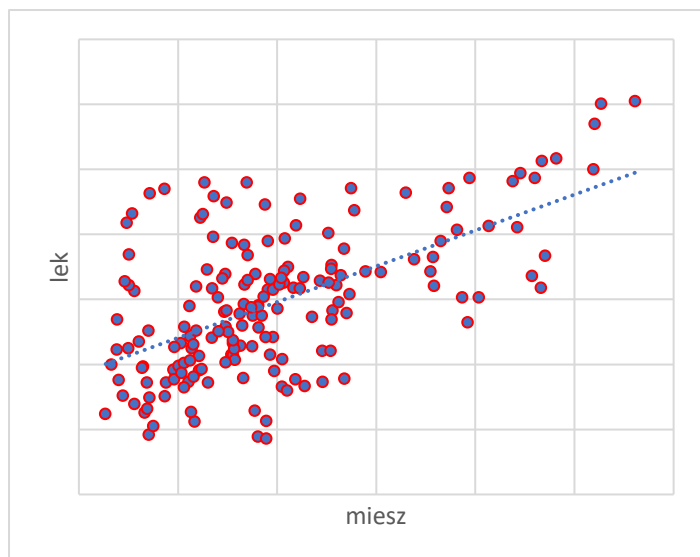


Ostatnią zmienną jest liczba ludności przypadająca na 1 miejsce w kinach. Do roku 2015 liczba ta znacznie się zmieniała w zależności od województwa. W ostatnich 5 latach nastąpiła lekka stabilizacja i można zauważyć, że od 5 lat najwięcej osób przepada na 1 miejsce w opolskim i lubuskim. Najmniej w dolnośląskim, śląskim i mazowieckim.

### 4.3. Wykresy rozrzutu

Na poniższych wykresach przedstawiono wykresy zmiennych objaśniających od zmiennej objaśnianej.





Na wykresy naniesiono również linię trendu. Dzięki temu łatwiej jest dostrzec zależności pomiędzy poszczególnymi zmiennymi. Wyraźną korelację dodatnią można zauważyć dla zmiennych *wynag*, *komun*, *drogi*, *lud*, *lek* oraz *turys*. Dla zmiennej *bezrob* oraz *kina* odnotowano niezerową korelację ujemną. Zmienne *skol*, *urban*, *inf* wydają się słabo skorelowane ze zmienną objaśniającą. Szczegółowe współczynniki korelacji przedstawiono w następnym rozdziale.



#### 4.4. Analiza korelacji

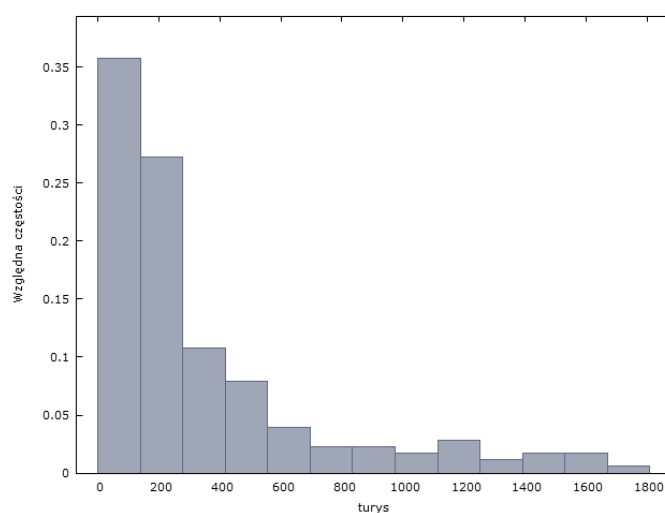
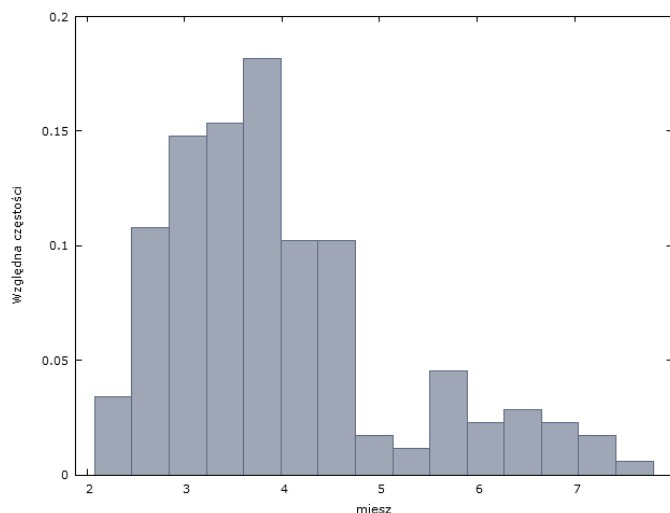
Poniżej przedstawiono tabelę korelacji pomiędzy poszczególnymi zmiennymi.

	miesz	wynag	komun	drogi	lud	turys	urban	inf	bezrob	lek	skol	kina
miesz	1.000	0.625	0.659	0.685	0.536	0.714	-0.065	0.087	-0.454	0.589	0.386	-0.503
wynag	0.625	1.000	0.318	0.380	0.442	0.411	0.233	0.001	-0.791	0.671	0.172	-0.512
komun	0.659	0.318	1.000	0.660	0.701	0.777	0.454	-0.077	-0.255	0.542	0.399	-0.569
drogi	0.685	0.380	0.660	1.000	0.771	0.533	0.045	-0.023	-0.304	0.499	0.547	-0.516
lud	0.536	0.442	0.701	0.771	1.000	0.635	0.333	-0.018	-0.398	0.551	0.600	-0.733
turys	0.714	0.411	0.777	0.533	0.635	1.000	0.195	-0.096	-0.333	0.517	0.265	-0.541
urban	-0.065	0.233	0.454	0.045	0.333	0.195	1.000	-0.016	-0.122	0.206	0.076	-0.354
inf	0.087	0.001	-0.077	-0.023	-0.018	-0.096	-0.016	1.000	0.086	-0.165	0.487	0.135
bezrob	-0.454	-0.791	-0.255	-0.304	-0.398	-0.333	-0.122	0.086	1.000	-0.511	-0.173	0.476
lek	0.589	0.671	0.542	0.499	0.551	0.517	0.206	-0.165	-0.511	1.000	0.206	-0.420
skol	0.386	0.172	0.399	0.547	0.600	0.265	0.076	0.487	-0.173	0.206	1.000	-0.226
kina	-0.503	-0.512	-0.569	-0.516	-0.733	-0.541	-0.354	0.135	0.476	-0.420	-0.226	1.000

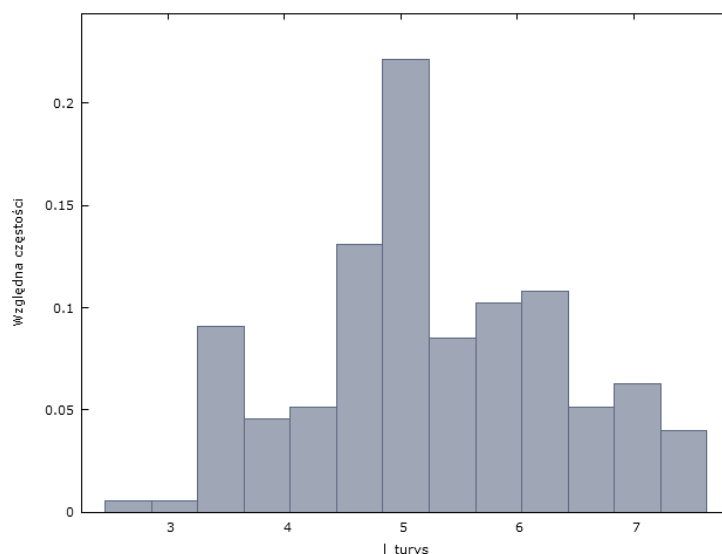
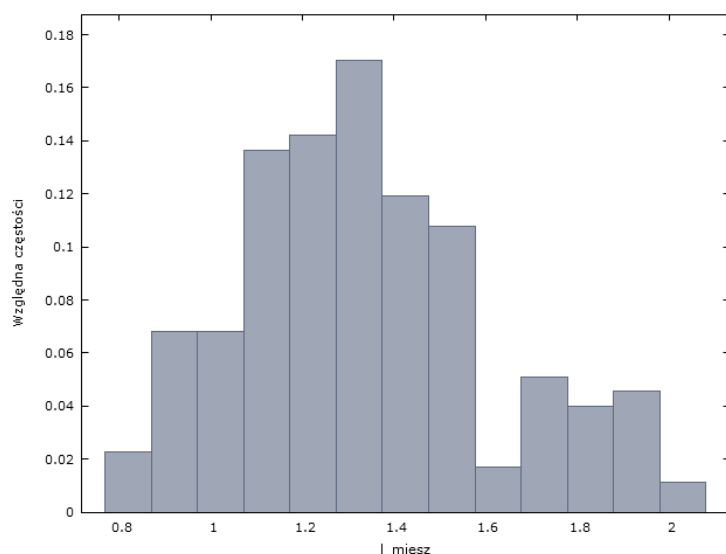
Można zauważyć silną korelację pomiędzy zmienną objaśnianą i zmiennymi *wynag*, *komun*, *drogi*, *lud*, *turys*, *lek*. Niestety zanotowano również niepożądane korelacje. Zmienna *kino* silnie korelowała ze zmiennymi *wynag*, *komun*, *drogi* oraz *lud*. Z tego powodów postanowiono wyłączyć ją z dalszej analizy. Podobnie zaobserwowano korelację między zmienną *bezrob*, a zmienną *wynagrodzenie*. Zmienną *bezrob* również wyłączono z dalszej analizy.

Z uwagi na to, że zmienne *miesz* oraz *turys* charakteryzowały się dużą asymetrią rozkładu, a dodatkowo zmienna *turys* przyjmowała duże wartości w analizie postanowiono użyć logarytmów zmiennych

Wykresy zmiennych przed zlogarytmowaniem



## Wykresy zmiennych po zlogarytmowaniu:



Podsumowując w estymacji modelu nie brano pod uwagę zmiennych *inf* oraz *skol* ze względu na mały stopień zmienności danych oraz zmiennych *bezrob* i *kina* ze względu na duży współczynnik korelacji z pozostałymi zmiennymi objaśniającymi. W dalszej analizie użyto ostatecznie 7 zmiennych objaśniających.

## 5. Estymacja modelu

W rozdziale zostaną przedstawione wyniki estymacji 3 modeli: regresji łącznej, efektów ustalonych (FE) oraz efektów losowych (RE).

### 5.1. Model regresji łącznej

Poniżej przedstawiono wyniki uzyskane w programie GRET

```
# regresja laczna
? panel l_miesz const wynag komun drogi lud l_turys urban lek --pooled
```

Model 19: Estymacja Panelowe MNK, z wykorzystaniem 176 obserwacji  
Włączono 16 jednostek danych przekrojowych  
Szereg czasowy długości = 11  
Zmienna zależna (Y): l\_miesz

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	0.821520	0.0913552	8.993	4.92e-016 ***
wynag	0.208114	0.0166497	12.50	9.07e-026 ***
komun	0.00402062	0.000444287	9.050	3.47e-016 ***
drogi	0.00689213	0.00158246	4.355	2.30e-05 ***
lud	-0.0699485	0.0132228	-5.290	3.77e-07 ***
l_turys	0.0491200	0.0141397	3.474	0.0007 ***
urban	-0.0138161	0.00116208	-11.89	4.83e-024 ***
lek	-0.0191527	0.0120931	-1.584	0.1151

Średn.aryt.zm.zależnej 1.338883    Odch.stand.zm.zależnej 0.276736  
Suma kwadratów reszt 2.127303    Błąd standardowy reszt 0.112528

Wsp. determ. R-kwadrat	0.841270	Skorygowany R-kwadrat	0.834656
F(7, 168)	127.2002	Wartość p dla testu F	9.58e-64
Logarytm wiarygodności	138.8422	Kryt. inform. Akaike'a	-261.6844
Kryt. bayes. Schwarza	-236.3205	Kryt. Hannana-Quinna	-251.3969
Autokorel.reszt - rho1	0.681634	Stat. Durbina-Watsona	0.567552

Wyłączając stałą, największa wartość p jest dla zmiennej 12 (lek)

Wartość  $p$  dla testu  $F$  świadczy o istotności chociaż jednej zmiennej w wyestymowanym modelu.

Po sprawdzeniu wartości  $p$  dla poszczególnych zmiennych wszystkie okazały się istotne statystycznie na poziomie istotności 0.10 z wyjątkiem zmiennej *lek*.

Dla modelu uzyskano współczynnik determinacji na poziomie  $R^2 = 0.85$ , co świadczy o tym, że model wyjaśnia 84% zmienności.

## 5.2. Model efektów ustalonych (FE)

Drugim testowanym modelem był model efektów ustalonych. Poniżej zaprezentowano wynik estymacji

```
# model FE
? panel l_miesz const wynag komun drogi lud l_turys urban lek

Model 20: Estymacja Ustalone efekty, z wykorzystaniem 176 obserwacji
Włączono 16 jednostek danych przekrojowych
Szereg czasowy długości = 11
Zmienna zależna (Y): l_miesz
```

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	1.98596	0.658004	3.018	0.0030	***
wynag	0.194761	0.0142786	13.64	3.17e-028	***
komun	0.000130103	0.000420814	0.3092	0.7576	
drogi	0.0118455	0.00668381	1.772	0.0783	*
lud	-0.720764	0.169174	-4.260	3.55e-05	***
l_turys	-0.0401268	0.0167883	-2.390	0.0181	**
urban	0.00876684	0.00876018	1.001	0.3185	
lek	-0.0637538	0.0154273	-4.133	5.89e-05	***

Średn.aryt.zm.zależnej	1.338883	Odch.stand.zm.zależnej	0.276736
Suma kwadratów reszt	0.373028	Błąd standardowy reszt	0.049377
LSDV R-kwadrat	0.972166	Within R-kwadrat	0.825111
LSDV F(22, 153)	242.9063	Wartość p dla testu F	2.9e-107
Logarytm wiarygodności	292.0465	Kryt. inform. Akaike'a	-538.0929
Kryt. bayes. Schwarza	-465.1718	Kryt. Hannana-Quinna	-508.5165
Autokorel.reszt - rho1	0.597624	Stat. Durbina-Watsona	0.651895

Joint test on named regressors -  
 Statystyka testu:  $F(7, 153) = 103.12$   
 z wartością  $p = P(F(7, 153) > 103.12) = 1.19606e-054$

Test na zróżnicowanie wyrazu wolnego w grupach -  
 Hipoteza zerowa: grupy posiadają wspólny wyraz wolny  
 Statystyka testu:  $F(15, 153) = 47.9686$   
 z wartością  $p = P(F(15, 153) > 47.9686) = 5.35527e-050$

W modelu FE istotne statystycznie okazały się zmienne *wynag*, *drogi*, *l\_turys*, *lud* oraz *lek*.

Wartość współczynnika determinacji wzrosła w modelu do 97%. Wewnątrzgrupowy  $R^2$  jest znacznie niższy (83%) co może świadczyć o przeparametryzowaniu modelu.

Analizując reszty modelu i błąd standardowy reszt można zauważyć, że dla obydwu model FE przyjmuje niższe wartości.

Zarówno kryterium Schwartz'a jak i Akaike wskazują na wybór modelu FE, jednak ostateczną weryfikację dostarcza test na zróżnicowanie wyrazu wolnego w grupach. W teście hipoteza  $H_0$  wskazuje na przyjęcie modelu regresji łącznej i brak zróżnicowania wyrazu wolnego. Uzyskana wartość  $p$  wskazuje na przyjęcie modelu FE kosztem regresji łącznej.

### 5.3. Model efektów losowych (RE)

Kolejnym modelem był model efektów losowych. Poniżej zaprezentowano wynik estymacji

```
# model RE
? panel l_miesz const wynag komun drogi lud l_turys urban lek --random-effects
```

Model 21: Estymacja Losowe efekty (GLS), z wykorzystaniem 176 obserwacji  
Włączono 16 jednostek danych przekrojowych  
Szereg czasowy długości = 11  
Zmienna zależna (Y): l\_miesz

	współczynnik	błąd standardowy	z	wartość p	
const	0.847406	0.173974	4.871	1.11e-06	***
wynag	0.195375	0.0153198	12.75	3.00e-037	***
komun	0.00119793	0.000469923	2.549	0.0108	**
drogi	0.0150399	0.00361041	4.166	3.10e-05	***
lud	-0.0121607	0.0319429	-0.3807	0.7034	
l_turys	-0.0400981	0.0179366	-2.236	0.0254	**
urban	-0.00465784	0.00273683	-1.702	0.0888	*
lek	-0.0532395	0.0163506	-3.256	0.0011	***

Średn.arytm.zm.zależnej	1.338883	Odch.stand.zm.zależnej	0.276736
Suma kwadratów reszt	5.368601	Błąd standardowy reszt	0.178233
Logarytm wiarygodności	57.37949	Kryt. inform. Akaike'a	-98.75898
Kryt. bayes. Schwarza	-73.39511	Kryt. Hannana-Quinna	-88.47152
Autokorel.reszt - rho1	0.597624	Stat. Durbina-Watsona	0.651895

```
'Between' wariancji = 0.00550179
'Within' wariancji = 0.00243809
theta wykorzystuje quasi-demeaning = 0.803211
corr(y,yhat)^2 = 0.600266
```

```
Joint test on named regressors -
Asymptotyczna statystyka testu: Chi-kwadrat(7) = 501.235
z wartością p = 4.35029e-104
```

```
Test Breuscha-Pagana na -
Hipoteza zerowa: Wariancja błędu w jednostce = 0
Asymptotyczna statystyka testu: Chi-kwadrat(1) = 72.9221
z wartością p = 1.34867e-017
```

Test Hausmana -

Hipoteza zerowa: Estymator UMNK (GLS) jest zgodny

Asymptotyczna statystyka testu: Chi-kwadrat(7) = 113.194

z wartością  $p = 1.99478e-021$

W modelu RE istotne statystycznie są wszystkie zmienne z wyjątkiem *lud*. Suma kwadratów reszt i błędu standardowego również znacznie wzrosła w stosunku do modelu FE. Kryteria Schwartz oraz Akaike również faworyzują model FE (a nawet regresji łącznej).

Program GRETl zapewnia również automatyczną generację testu Breuscha – Pagana oraz Hausmana. Hipotezę zerową testu BP można połączyć z wyborem regresji łącznej ponad modelem RE – odwrotnie dla hipotezy alternatywnej. W teście BP otrzymano wartość  $p=0$ , a więc wybieramy model RE. Test Hausmana z kolei wskazuje, czy lepszym modelem będzie RE (H0), czy FE (H1). Wartość  $p$  jest niemal równa 0, a więc wybrano model FE.

## 6. Dobór zmiennych

We wcześniej wygenerowanym modelu część zmiennych była nieistotna. W programie wykonano, więc sekwencyjną eliminację zmiennych nieistotnych. Poniżej zaprezentowano wyniki

Sekwencyjna eliminacja nieistotnych zmiennych przy dwustronnym obszarze krytycznym,  $\alpha = 0.10$

Wyeliminowano nieistotną zmienną: komun (wartość  $p = 0.758$ )

Wyeliminowano nieistotną zmienną: urban (wartość  $p = 0.306$ )

Test porównawczy z Modelem 23

Hipoteza zerowa: parametry regresji dla wskazanych zmiennych są równe zero  
komun, urban

Statystyka testu:  $F(2, 153) = 0.572058$ , wartość  $p = 0.565565$

Pominięcie zmiennych poprawiło 3 z 3 kryteriów informacyjnych (AIC, BIC, HQC).

Model 24: Estymacja Ustalonych efektów, z wykorzystaniem 176 obserwacji

Włączono 16 jednostek danych przekrojowych

Szereg czasowy długości = 11

Zmienna zależna (Y): *l\_miesz*

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	2.53140	0.396941	6.377	1.97e-09	***
wynag	0.192042	0.0140115	13.71	1.64e-028	***
drogi	0.0135329	0.00647541	2.090	0.0383	**
lud	-0.743971	0.164408	-4.525	1.19e-05	***
l_turys	-0.0373162	0.0126631	-2.947	0.0037	***
lek	-0.0664044	0.0149081	-4.454	1.60e-05	***

Średn. aryt. zm. zależnej	1.338883	Odch. stand. zm. zależnej	0.276736
Suma kwadratów reszt	0.375817	Błąd standardowy reszt	0.049240
LSDV R-kwadrat	0.971958	Within R-kwadrat	0.823803
LSDV F(20, 155)	268.6230	Wartość p dla testu F	1.9e-109
Logarytm wiarygodności	291.3909	Kryt. inform. Akaike'a	-540.7817
Kryt. bayes. Schwarza	-474.2016	Kryt. Hannana-Quinna	-513.7771
Autokorel. reszt - rho1	0.593545	Stat. Durbina-Watsona	0.664416

Joint test on named regressors -

Statystyka testu:  $F(5, 155) = 144.939$

z wartością  $p = P(F(5, 155) > 144.939) = 1.44932e-056$

Test na zróżnicowanie wyrazu wolnego w grupach -  
 Hipoteza zerowa: grupy posiadają wspólny wyraz wolny  
 Statystyka testu:  $F(15, 155) = 106.593$   
 z wartością  $p = P(F(15, 155) > 106.593) = 1.6736e-073$

W efekcie przeprowadzonej eliminacji wyeliminowano zmienne *urban* oraz *komun*. Nowy model posiada tylko zmienne istotne na poziomie 0.1. Nowy model posiada również niższe kryteria AIC, BIC oraz HQC co faworyzuje wyestymowany model.

## 7. Model dwukierunkowy

Dla modelu wyestymowanego w poprzednim rozdziale sprawdzono, czy dołożenie efektów czasowych do modelu jest zasadne. W tym celu wyestymowano model FE dwukierunkowy.

```
# model FE dwukierunkowy
? panel l_miesz const wynag drogi lud l_turys lek --time-dummies
```

Model 25: Estymacja Ustalone efekty, z wykorzystaniem 176 obserwacji  
 Włączono 16 jednostek danych przekrojowych  
 Szereg czasowy długości = 11  
 Zmienna zależna (Y): l\_miesz

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	2.13271	0.378027	5.642	8.55e-08	***
wynag	-0.173899	0.0694284	-2.505	0.0134	**
drogi	0.0147374	0.00596486	2.471	0.0146	**
lud	-0.318493	0.181858	-1.751	0.0820	*
l_turys	0.0155098	0.0270519	0.5733	0.5673	
lek	-0.00368768	0.0174606	-0.2112	0.8330	
dt_2	0.0488201	0.0192359	2.538	0.0122	**
dt_3	0.0592381	0.0247795	2.391	0.0181	**
dt_4	0.0776234	0.0317153	2.448	0.0156	**
dt_5	0.104816	0.0391349	2.678	0.0083	***
dt_6	0.144439	0.0477107	3.027	0.0029	***
dt_7	0.175421	0.0561798	3.122	0.0022	***
dt_8	0.245088	0.0701759	3.492	0.0006	***
dt_9	0.367716	0.0883607	4.162	5.40e-05	***
dt_10	0.495037	0.110508	4.480	1.51e-05	***
dt_11	0.661054	0.138936	4.758	4.68e-06	***

Średn.aryt.zm.zależnej	1.338883	Odch.stand.zm.zależnej	0.276736
Suma kwadratów reszt	0.281562	Błąd standardowy reszt	0.044066
LSDV R-kwadrat	0.978991	Within R-kwadrat	0.867993
LSDV F(30, 145)	225.2276	Wartość p dla testu F	8.8e-107
Logarytm wiarygodności	316.8008	Kryt. inform. Akaike'a	-571.6016
Kryt. bayes. Schwarza	-473.3166	Kryt. Hannana-Quinna	-531.7377
Autokorel.reszt - rho1	0.535063	Stat. Durbina-Watsona	0.734023

Joint test on named regressors -  
 Statystyka testu:  $F(5, 145) = 5.19856$   
 z wartością  $p = P(F(5, 145) > 5.19856) = 0.000203016$

Test na zróżnicowanie wyrazu wolnego w grupach -  
 Hipoteza zerowa: grupy posiadają wspólny wyraz wolny  
 Statystyka testu:  $F(15, 145) = 110.563$   
 z wartością  $p = P(F(15, 145) > 110.563) = 2.29092e-071$

Test Walda na łączną istotność zmiennych 0-1 jednostek czasu -  
 Hipoteza zerowa: No time effects  
 Asymptotyczna statystyka testu: Chi-kwadrat(10) = 48.5398  
 z wartością  $p = 4.94541e-007$

Na końcu wyników modelu można zauważyć wyniki testu Walda na łączną istotność efektów czasowych. Wartość  $p$  jest niemal zerowa, co świadczy o tym, że w modelu powinny zostać zróżnicowane efekty czasowe. W wyestymowanym modelu wzrosła również wartość współczynnika  $R^2$  oraz spadło kryterium AIC co również faworyzuje model FE dwukierunkowy. Również test na zróżnicowanie wyrazu wolnego w grupach wskazuje na wybranie dwukierunkowego modelu FE kosztem regresji łącznej.

Porównano również model dwukierunkowy FE oraz RE.

```
# model RE dwukierunkowy
? panel l_miesz const wynag drogi lud l_turys lek --time-dummies \par --random-effects
```

Model 27: Estymacja Losowe efekty (GLS), z wykorzystaniem 176 obserwacji  
Włączono 16 jednostek danych przekrojowych  
Szereg czasowy długości = 11  
Zmienna zależna (Y): l\_miesz

	współczynnik	błąd standardowy	z	wartość p	
const	1.20455	0.186526	6.458	1.06e-010	***
wynag	-0.198953	0.0602742	-3.301	0.0010	***
drogi	0.0165616	0.00452079	3.663	0.0002	***
lud	0.0114446	0.0500114	0.2288	0.8190	
l_turys	0.0461715	0.0250935	1.840	0.0658	*
lek	-0.000760131	0.0173102	-0.04391	0.9650	
dt_2	0.0493968	0.0188214	2.624	0.0087	***
dt_3	0.0584584	0.0232750	2.512	0.0120	**
dt_4	0.0790585	0.0286869	2.756	0.0059	***
dt_5	0.107335	0.0348185	3.083	0.0021	***
dt_6	0.149276	0.0417572	3.575	0.0004	***
dt_7	0.180121	0.0489901	3.677	0.0002	***
dt_8	0.253121	0.0607681	4.165	3.11e-05	***
dt_9	0.382109	0.0760809	5.022	5.10e-07	***
dt_10	0.516490	0.0949260	5.441	5.30e-08	***
dt_11	0.728260	0.118196	6.161	7.21e-010	***

Średn.aryt.zm.zależnej	1.338883	Odch.stand.zm.zależnej	0.276736
Suma kwadratów reszt	6.496589	Błąd standardowy reszt	0.200877
Logarytm wiarygodności	40.59701	Kryt. inform. Akaike'a	-49.19403
Kryt. bayes. Schwarza	1.533715	Kryt. Hannana-Quinna	-28.61911
Autokorel.reszt - rho1	0.535063	Stat. Durbina-Watsona	0.734023

'Between' wariancji = 0.0272213  
'Within' wariancji = 0.00194181  
theta wykorzystuje quasi-demeaning = 0.919731  
corr(y,yhat)^2 = 0.515791

Joint test on named regressors -  
Asymptotyczna statystyka testu: Chi-kwadrat(5) = 34.1271  
z wartością p = 2.24623e-006

Test Walda na łączną istotność zmiennych 0-1 jednostek czasu -  
Hipoteza zerowa: No time effects  
Asymptotyczna statystyka testu: Chi-kwadrat(10) = 66.7017  
z wartością p = 1.91308e-010

Test Breuscha-Pagana na -  
Hipoteza zerowa: Wariancja błędu w jednostce = 0  
Asymptotyczna statystyka testu: Chi-kwadrat(1) = 636.131  
z wartością p = 2.3192e-140

Test Hausmana -  
Hipoteza zerowa: Estymator UMNK (GLS) jest zgodny

Asymptotyczna statystyka testu: Chi-kwadrat(5) = 19.0209  
z wartością  $p = 0.00190502$

W teście Hausmana znowu otrzymano wartość  $p = 0.001$ , co oznacza, że lepszym wyborem będzie dwukierunkowy model efektów ustalonych.

Ostatecznie wybrano, więc dwukierunkowy model efektów ustalonych.

Dla modelu dwukierunkowego FE również przeprowadzono sekwencyjną eliminację zmiennych nieistotnych. W efekcie uzyskano końcową wersję modelu

Sekwencyjna eliminacja nieistotnych zmiennych przy dwustronnym obszarze krytycznym,  $\alpha = 0.10$

Wyeliminowano nieistotną zmienną: lek (wartość  $p = 0.833$ )  
Wyeliminowano nieistotną zmienną: l\_turys (wartość  $p = 0.576$ )

Test porównawczy z Modelem 30

Hipoteza zerowa: parametry regresji dla wskazanych zmiennych są równe zero  
l\_turys, lek

Statystyka testu:  $F(2, 145) = 0.178057$ , wartość  $p = 0.837077$

Pominięcie zmiennych poprawiło 3 z 3 kryteriów informacyjnych (AIC, BIC, HQC).

Model 31: Estymacja Ustalonych efekty, z wykorzystaniem 176 obserwacji

Włączono 16 jednostek danych przekrojowych

Szereg czasowy długości = 11

Zmienna zależna (Y): l\_miesz

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	2.20548	0.352905	6.250	4.22e-09	***
wynag	-0.172780	0.0643657	-2.684	0.0081	***
drogi	0.0148299	0.00592410	2.503	0.0134	**
lud	-0.324657	0.180431	-1.799	0.0740	*
dt_2	0.0488239	0.0191277	2.553	0.0117	**
dt_3	0.0598707	0.0244616	2.448	0.0156	**
dt_4	0.0781625	0.0314721	2.484	0.0141	**
dt_5	0.104803	0.0387360	2.706	0.0076	***
dt_6	0.144532	0.0473805	3.050	0.0027	***
dt_7	0.175979	0.0557491	3.157	0.0019	***
dt_8	0.246006	0.0697647	3.526	0.0006	***
dt_9	0.368520	0.0876168	4.206	4.50e-05	***
dt_10	0.495858	0.109145	4.543	1.15e-05	***
dt_11	0.643044	0.131842	4.877	2.76e-06	***

Średn.aryt.zm.zależnej	1.338883	Odch.stand.zm.zależnej	0.276736
Suma kwadratów reszt	0.282254	Błąd standardowy reszt	0.043819
LSDV R-kwadrat	0.978939	Within R-kwadrat	0.867669
LSDV F(28, 147)	244.0316	Wartość p dla testu F	4.3e-109
Logarytm wiarygodności	316.5850	Kryt. inform. Akaike'a	-575.1699
Kryt. bayes. Schwarza	-483.2259	Kryt. Hannana-Quinna	-537.8779
Autokorel.reszt - rho1	0.540840	Stat. Durbina-Watsona	0.728923

Joint test on named regressors -

Statystyka testu:  $F(13, 147) = 74.1423$

z wartością  $p = P(F(13, 147) > 74.1423) = 1.03807e-057$

Test na zróżnicowanie wyrazu wolnego w grupach -

Hipoteza zerowa: grupy posiadają wspólny wyraz wolny

Statystyka testu:  $F(15, 147) = 157.803$

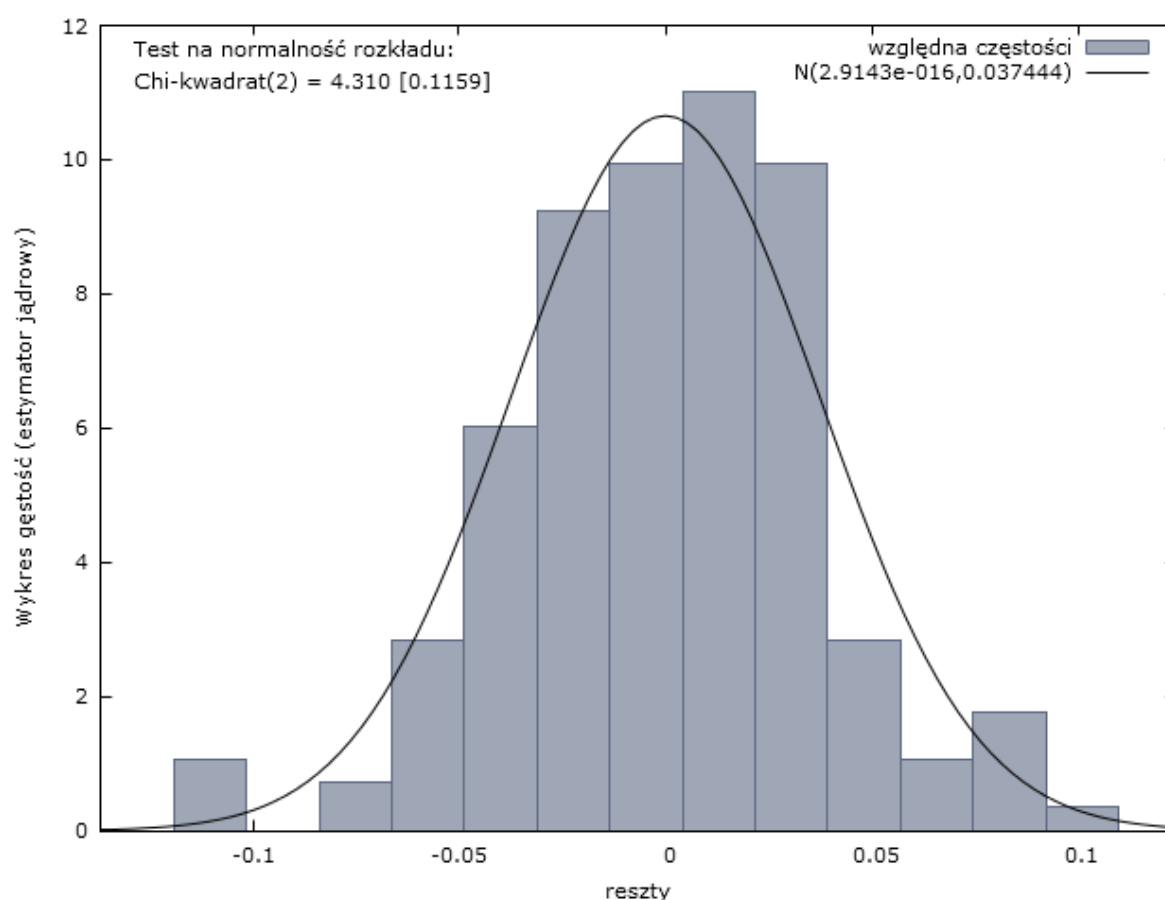
z wartością  $p = P(F(15, 147) > 157.803) = 1.58915e-082$



Jak widać po wyeliminowaniu zmiennych nieistotnych model jest wyjaśniany za pomocą 10 zmiennych różnicujących czas oraz 3 oryginalnych zmiennych objaśniających *wynag*, *drogi*, *lud*. To oznacza, że wpływ czasu (a raczej inna zmienna, która się pod nim kryje) w bardzo dużym stopniu wyjaśnia ceny mieszkań i tak naprawdę nie znamy wszystkich czynników, który różnicują ceny. Widać, że różnicowanie postępuje z kolejnymi latami. Im rok bliższy aktualnemu tym współczynnika przy zmiennej wyższy.

## 8. Weryfikacja modelu

Weryfikację modelu uzyskanego w rozdziale 7 rozpoczęto od analizy rozkładu reszt modelu



Dodatkowo przeprowadzono 4 testy sprawdzające normalność rozkładu reszt

Test na normalność rozkładu reszty:

Test Doornika-Hansena = 4.31039, z wartością p 0.115881

Test Shapiro-Wilka = 0.986774, z wartością p 0.134246

Test Lillieforsa = 0.0568893, z wartością p ~ 0.23

Test Jarque'a-Bera = 3.0307, z wartością p 0.219731

We wszystkich testach otrzymano wynik, że reszty posiadają charakteru rozkładu normalnego.

Kolejnym krokiem było zbadanie autokorelacji, w tym celu zbudowano model MNK z resztami jako zmienną objaśnianą i resztami opóźnionymi jako objaśniającymi.

```
? ols reszty reszty(-1) reszty(-2) reszty(-3)
```

Model 41: Estymacja Panelowe MNK, z wykorzystaniem 112 obserwacji  
Włączono 16 jednostek danych przekrojowych  
Szereg czasowy długości = 7  
Zmienna zależna (Y): reszty

	współczynnik	błąd standardowy	t-Studenta	wartość p	
reszty_1	0.574194	0.0989082	5.805	6.41e-08	***
reszty_2_aaa	-0.121121	0.108878	-1.112	0.2684	
reszty_3	-0.108063	0.0908213	-1.190	0.2367	

Średn.aryt.zm.zależnej	0.000000	Odch.stand.zm.zależnej	0.035256
Suma kwadratów reszt	0.099673	Błąd standardowy reszt	0.030240
Wsp. determ. R-kwadrat	0.277560	Skorygowany R-kwadrat	0.264305
F(2, 109)	20.93882	Wartość p dla testu F	2.02e-08
Logarytm wiarygodności	234.4428	Kryt. inform. Akaike'a	-462.8856
Kryt. bayes. Schwarza	-454.7301	Kryt. Hannana-Quinna	-459.5767
Autokorel.reszt - rho1	-0.053505	Stat. Durbina-Watsona	1.675763

Największa wartość p jest dla zmiennej 32 (reszty\_2\_aaa)

Jak widać występuje autokorelacja 1 rzędu. Wskazuje na to również statystyka testu F, która dowodzi, że co najmniej jedna ze zmiennych objaśniających jest istotna w modelu.

W celu wyeliminowania autokorelacji zastosowano przekształcenie Prais-Winstena. Dla przekształconych zmiennych wyestymowano znowu model i sprawdzono, czy nadal występuje autokorelacja.

Model 45: Estymacja Panelowe MNK, z wykorzystaniem 111 obserwacji  
Włączono 16 jednostek danych przekrojowych  
Szereg czasowy długości: minimum 6, maximum 7  
Zmienna zależna (Y): reszty

	współczynnik	błąd standardowy	t-Studenta	wartość p	
reszty_1	0.246248	0.0951878	2.587	0.0110	**
reszty_2_aaa	-0.0533945	0.0990667	-0.5390	0.5910	
reszty_3	-0.155306	0.0945698	-1.642	0.1035	

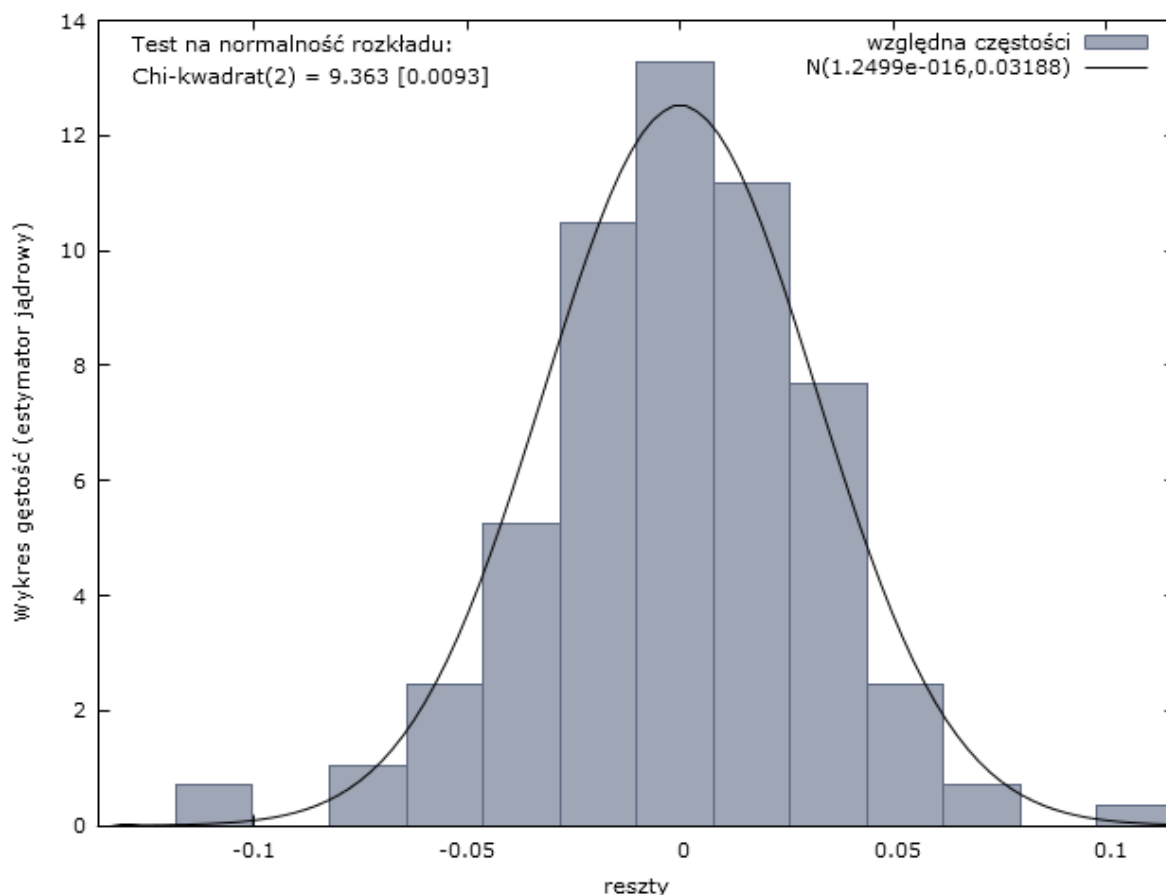
  

Średn.aryt.zm.zależnej	0.000000	Odch.stand.zm.zależnej	0.031913
Suma kwadratów reszt	0.102775	Błąd standardowy reszt	0.030848
Niecentrowany R-kwadr.	0.082598	Centrowany R-kwadrat	0.082598
F(3, 108)	3.241261	Wartość p dla testu F	0.024944
Logarytm wiarygodności	230.1513	Kryt. inform. Akaike'a	-454.3026
Kryt. bayes. Schwarza	-446.1740	Kryt. Hannana-Quinna	-451.0051
Autokorel.reszt - rho1	-0.026848	Stat. Durbina-Watsona	1.795212

Największa wartość p jest dla zmiennej 32 (reszty\_2\_aaa)

Niestety nie udało się usunąć autokorelacji reszt

Sprawdzono również, czy w nowym modelu reszty mają rozkład normalny



Test na normalność rozkładu reszty:

Test Doornika-Hansena = 9.36326, z wartością p 0.0092639

Test Shapiro-Wilka = 0.985116, z wartością p 0.0865085

Test Lillieforsa = 0.0463856, z wartością p ≈ 0.54

Test Jarque'a-Bera = 10.5591, z wartością p 0.00509473

W dwóch testach z 4 reszty nadal mają rozkład normalny.

Z tego powodu przeprowadzono przekształcenie Prais-Winstena drugi raz. Dla przekształconych zmiennych wyestymowano znowu model i sprawdzono, czy nadal występuje autokorelacja.

Model 48: Estymacja Panelowe MNK, z wykorzystaniem 95 obserwacji

Włączono 16 jednostek danych przekrojowych

Szereg czasowy długości: minimum 5, maximum 6

Zmienna zależna (Y): reszty

	współczynnik	błąd standardowy	t-Studenta	wartość p	
reszty_1	0.0776998	0.101953	0.7621	0.4479	
reszty_2_aaa	-0.0738874	0.109486	-0.6749	0.5015	
reszty_3	-0.188103	0.103882	-1.811	0.0734	*
Średn.aryt.zm.zależnej	0.000000	Odch.stand.zm.zależnej	0.030941		
Suma kwadratów reszt	0.085356	Błąd standardowy reszt	0.030460		

Niecentrowany R-kwadr.	0.051474	Centrowany R-kwadrat	0.051474
F(3, 92)	1.664194	Wartość p dla testu F	0.180184
Logarytm wiarygodności	198.4040	Kryt. inform. Akaike'a	-390.8079
Kryt. bayes. Schwarza	-383.1463	Kryt. Hannana-Quinna	-387.7120
Autokorel.reszt - rho1	-0.035116	Stat. Durbina-Watsona	1.832196

Największa wartość p jest dla zmiennej 32 (reszty\_2\_aaa)

Autokorelacja zniknęła. Sprawdzono, czy składnik losowy ma nadal rozkład normalny.

Test na normalność rozkładu reszty:

Test Doornika-Hansena = 16.8786, z wartością p 0.000216203

Test Shapiro-Wilka = 0.976824, z wartością p 0.0157451

Test Lillieforsa = 0.0655342, z wartością p ~ 0.13

Test Jarque'a-Bera = 17.1086, z wartością p 0.000192718

Tylko w jednym teście reszty mają rozkład normalny. Przy tak jednoznacznie odrzuconych wynikach w 3 testach przyjęto, że reszty nie mają rozkładu normalnego.

Dla nowego modelu sprawdzono, czy występuje heteroskedastyczność. W tym celu przeprowadzono test Breuscha – Pagana oraz White'a.

Test BP:

```
#heteroskedastyczność
? sq_reszty = reszty * reszty
? yh = $yhat
? yh2 = yh*yh
#BP
? ols sq_reszty const new_new_wynag new_new_drogi new_new_lud
```

Model 50: Estymacja Panelowe MNK, z wykorzystaniem 143 obserwacji  
Włączono 16 jednostek danych przekrojowych  
Szereg czasowy długości: minimum 8, maximum 9  
Zmienna zależna (Y): sq\_reszty

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	0.00113140	0.000304872	3.711	0.0003	***
new_new_wynag	7.50568e-05	0.000260600	0.2880	0.7738	
new_new_drogi	-2.56441e-05	3.32401e-05	-0.7715	0.4417	
new_new_lud	-3.05858e-05	0.000299156	-0.1022	0.9187	
Średn.aryt.zm.zależnej	0.000942	Odch.stand.zm.zależnej	0.001815		
Suma kwadratów reszt	0.000462	Błąd standardowy reszt	0.001824		
Wsp. determ. R-kwadrat	0.011411	Skorygowany R-kwadrat	-0.009925		
F(3, 139)	0.534820	Wartość p dla testu F	0.659166		
Logarytm wiarygodności	700.9863	Kryt. inform. Akaike'a	-1393.973		
Kryt. bayes. Schwarza	-1382.121	Kryt. Hannana-Quinna	-1389.157		
Autokorel.reszt - rho1	0.000106	Stat. Durbina-Watsona	1.679831		

Wyłączając stałą, największa wartość p jest dla zmiennej 45 (new\_new\_lud)

White:

Model 51: Estymacja Panelowe MNK, z wykorzystaniem 143 obserwacji  
Włączono 16 jednostek danych przekrojowych  
Szereg czasowy długości: minimum 8, maximum 9  
Zmienna zależna (Y): sq\_reszty

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	0.000877009	0.000447198	1.961	0.0518	*
yh	0.000640824	0.00126282	0.5075	0.6126	
yh2	-0.000587846	0.000740734	-0.7936	0.4288	
Średn.aryt.zm.zależnej	0.000942	Odch.stand.zm.zależnej	0.001815		
Suma kwadratów reszt	0.000463	Błąd standardowy reszt	0.001819		
Wsp. determ. R-kwadrat	0.009858	Skorygowany R-kwadrat	-0.004286		
F(2, 140)	0.696958	Wartość p dla testu F	0.499818		
Logarytm wiarygodności	700.8741	Kryt. inform. Akaike'a	-1395.748		
Kryt. bayes. Schwarza	-1386.860	Kryt. Hannana-Quinna	-1392.136		
Autokorel.reszt - rho1	0.000113	Stat. Durbina-Watsona	1.678190		

W obydwu testach wartość  $p > 0.05$  dla testu  $F$ , co oznacza, że nie występuje heteroskedastyczność.

Podsumowując zweryfikowano model otrzymany w poprzednim rozdziale i usunięto autokorelację. Ostateczny model nie ma autokorelacji reszt i heteroskedastyczności, ale za to reszty nie mają również rozkładu normalnego.

## 9. Model dynamiczny

Postanowiono poszukać modelu dynamicznego. Poszukiwania modelu dynamicznego rozpoczęto od modelu pierwszych różnic

### 9.1. Model pierwszych różnic

```
#model pierwszych różnic
? panel l_miesz(-1) const wynag(-1) komun(-1) drogi(-1) lud(-1) l_turys(-1) \par urban(-1)
lek(-1)
```

Model 3: Estymacja Ustalone efekty, z wykorzystaniem 160 obserwacji  
Włączono 16 jednostek danych przekrojowych  
Szereg czasowy długości = 10  
Zmienna zależna (Y): l\_miesz\_1

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	1.76521	0.742440	2.378	0.0188	**
wynag_1	0.191167	0.0168358	11.35	1.80e-021	***
komun_1	-0.000369800	0.000488388	-0.7572	0.4502	
drogi_1	0.0147804	0.00701971	2.106	0.0371	**
lud_1	-0.550387	0.182983	-3.008	0.0031	***
l_turys_1	-0.0180073	0.0356789	-0.5047	0.6146	
urban_1	0.00374992	0.00948167	0.3955	0.6931	
lek_1	-0.0708592	0.0155121	-4.568	1.09e-05	***
Średn.aryt.zm.zależnej	1.315080	Odch.stand.zm.zależnej	0.270690		
Suma kwadratów reszt	0.297209	Błąd standardowy reszt	0.046577		
LSDV R-kwadrat	0.974489	Within R-kwadrat	0.719431		

LSDV F(22, 137)	237.8770	Wartość p dla testu F	7.57e-98
Logarytm wiarygodności	276.0492	Kryt. inform. Akaike'a	-506.0984
Kryt. bayes. Schwarza	-435.3694	Kryt. Hannana-Quinna	-477.3778
Autokorel.reszt - rho1	0.497440	Stat. Durbina-Watsona	0.801821

Joint test on named regressors -

Statystyka testu:  $F(7, 137) = 50.1848$

z wartością  $p = P(F(7, 137) > 50.1848) = 8.59594e-035$

Test na zróżnicowanie wyrazu wolnego w grupach -

Hipoteza zerowa: grupy posiadają wspólny wyraz wolny

Statystyka testu:  $F(15, 137) = 27.068$

z wartością  $p = P(F(15, 137) > 27.068) = 1.07892e-033$

Jak widać model ma wysoki współczynnik  $R^2$ , ale za to niższy  $R^2$  wewnątrzgrupowy. Dokonano eliminacji zmiennych nieistotnych i otrzymano model postaci

Model 4: Estymacja Ustalonych efektów, z wykorzystaniem 160 obserwacji

Włączono 16 jednostek danych przekrojowych

Szereg czasowy długości = 10

Zmienna zależna (Y): l\_miesz\_1

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	1.86154	0.428146	4.348	2.63e-05	***
wynag_1	0.186065	0.0155141	11.99	3.07e-023	***
drogi_1	0.0143730	0.00663252	2.167	0.0319	**
lud_1	-0.534344	0.179325	-2.980	0.0034	***
lek_1	-0.0732151	0.0140681	-5.204	6.79e-07	***

Średn.aryt.zm.zależnej	1.315080	Odch.stand.zm.zależnej	0.270690
Suma kwadratów reszt	0.299723	Błąd standardowy reszt	0.046270
LSDV R-kwadrat	0.974274	Within R-kwadrat	0.717058
LSDV F(19, 140)	279.0459	Wartość p dla testu F	3.0e-101
Logarytm wiarygodności	275.3755	Kryt. inform. Akaike'a	-510.7510
Kryt. bayes. Schwarza	-449.2476	Kryt. Hannana-Quinna	-485.7766
Autokorel.reszt - rho1	0.500745	Stat. Durbina-Watsona	0.795708

Joint test on named regressors -

Statystyka testu:  $F(4, 140) = 88.7004$

z wartością  $p = P(F(4, 140) > 88.7004) = 2.12819e-037$

Test na zróżnicowanie wyrazu wolnego w grupach -

Hipoteza zerowa: grupy posiadają wspólny wyraz wolny

Statystyka testu:  $F(15, 140) = 139.675$

z wartością  $p = P(F(15, 140) > 139.675) = 2.91954e-076$

Wszystkie zmienne w nowym modelu są istotne statystycznie. Postanowiono zweryfikować uzyskany model.

Test na normalność rozkładu reszty:

Test Doornika-Hansena = 1.33605, z wartością  $p$  0.512721

Test Shapiro-Wilka = 0.991352, z wartością  $p$  0.444228

Test Lillieforsa = 0.0516183, z wartością  $p$   $\approx$  0.36

Test Jarque'a-Bera = 0.80731, z wartością  $p$  0.667874

Okazało się, że reszty tego modelu mają rozkład normalny. Sprawdzono, więc czy występuje autokorelacja

gretl wersja 2021d  
Bieżąca sesja: 2022-01-17 11:29

? ols reszty reszty(-1)

Model 7: Estymacja Panelowe MNK, z wykorzystaniem 144 obserwacji  
Włączono 16 jednostek danych przekrojowych  
Szereg czasowy długości = 9  
Zmienna zależna (Y): reszty

	współczynnik	błąd standardowy	t-Studenta	wartość p
reszty_1	0.505637	0.0721474	7.008	8.77e-011 ***

Średn.aryt.zm.zależnej	-0.001269	Odch.stand.zm.zależnej	0.041052
Suma kwadratów reszt	0.179552	Błąd standardowy reszt	0.035435
Niecentrowany R-kwadr.	0.255663	Centrowany R-kwadrat	0.254947
F(1, 143)	49.11740	Wartość p dla testu F	8.77e-11
Logarytm wiarygodności	277.1442	Kryt. inform. Akaike'a	-552.2883
Kryt. bayes. Schwarza	-549.3185	Kryt. Hannana-Quinna	-551.0816
Autokorel.reszt - rho1	0.080818	Stat. Durbina-Watsona	1.593454

W modelu stwierdzono występowanie autokorelacji składnika losowego. Zastosowano przekształcenie Prais – Winstena.

Model 9: Estymacja Panelowe MNK, z wykorzystaniem 128 obserwacji  
Włączono 16 jednostek danych przekrojowych  
Szereg czasowy długości = 8  
Zmienna zależna (Y): reszty

	współczynnik	błąd standardowy	t-Studenta	wartość p
reszty_1	-0.00671815	0.0890482	-0.07544	0.9400

Średn.aryt.zm.zależnej	-0.002902	Odch.stand.zm.zależnej	0.031169
Suma kwadratów reszt	0.124451	Błąd standardowy reszt	0.031304
Niecentrowany R-kwadr.	0.000045	Centrowany R-kwadrat	-0.008690
F(1, 127)	0.005692	Wartość p dla testu F	0.939980
Logarytm wiarygodności	262.2718	Kryt. inform. Akaike'a	-522.5436
Kryt. bayes. Schwarza	-519.6916	Kryt. Hannana-Quinna	-521.3848
Autokorel.reszt - rho1	0.065215	Stat. Durbina-Watsona	1.574090

Autokorelacja zniknęła, jednak również reszty straciły charakter rozkładu normalnego

Test na normalność rozkładu reszty:

Test Doornika-Hansena = 16.5153, z wartością p 0.000259267

Test Shapiro-Wilka = 0.978228, z wartością p 0.0214999

Test Lillieforsa = 0.0635681, z wartością p ~ 0.16

Test Jarque'a-Bera = 16.4947, z wartością p 0.000261949

Postanowiono porzucić model pierwszych różnic

## 9.2. Metoda zmiennych instrumentalnych

W drugim podejściu postanowiono wykorzystać metodę zmiennych instrumentalnych do estymacji modelu. Podejrzewano, że wpływ na ceny za 1m<sup>2</sup> mieszkań będą zależały od cen z poprzedniego roku. Wyestymowano, więc model dynamiczny z dodaniem opóźnionej zmiennej *l\_miesz*. Do estymacji wykorzystano 1-krokowy model pierwszych różnic. Poniżej zaprezentowano wyniki

```
? list X = wynag komun drogi lud turys urban lek
```

```
Wygenerowano listę X
```

```
? dpanel 1 ; l_miesz X
```

Model 40: Estymacja dynamiczny panel 1-step, z wykorzystaniem 144 obserwacji

Włączono 16 jednostek danych przekrojowych

Zmienna zależna (Y): *l\_miesz*

	współczynnik	błąd standardowy	z	wartość p	
l_miesz(-1)	0.643604	0.152958	4.208	2.58e-05	***
wynag	0.129659	0.0249560	5.195	2.04e-07	***
komun	-0.000229937	0.000334896	-0.6866	0.4923	
drogi	-0.000529243	0.00667375	-0.07930	0.9368	
lud	-0.483542	0.159854	-3.025	0.0025	***
turys	1.37908e-05	2.45840e-05	0.5610	0.5748	
urban	0.0286121	0.0111282	2.571	0.0101	**
lek	-0.0470685	0.0165000	-2.853	0.0043	***

Suma kwadratów reszt 0.303942 Błąd standardowy reszt 0.045942

Liczba instrumentów = 52

Test AR(1) dla błędu: z = -3.09612 [0.0020]

Test AR(2) dla błędu: z = -0.432828 [0.6651]

Test Sargana - nadmiernej identyfikacji: Chi-kwadrat(44) = 55.2406 [0.1192]

Test Walda (joint): Chi-kwadrat(8) = 1375.95 [0.0000]

Wartość *p* dla testu Sargana wskazuje, że wybrany model jest poprawny. Można zauważyć jednak występowanie ujemnej autokorelacji 1 rzędu (test AR(1)) oraz brak autokorelacji rzędu drugiego AR(2)). Świadczy to zgodności i efektywności estymatora UMM. Wyestymowano również model uwzględniając jedynie istotne statystycznie parametry.

```
? dpanel 1 ; l_miesz const wynag lud urban lek
```

Model 4: Estymacja dynamiczny panel 1-step, z wykorzystaniem 144 obserwacji

Włączono 16 jednostek danych przekrojowych

Zmienna zależna (Y): *l\_miesz*

	współczynnik	błąd standardowy	z	wartość p	
l_miesz(-1)	0.669284	0.149410	4.480	7.48e-06	***
wynag	0.128207	0.0252945	5.069	4.01e-07	***
lud	-0.436505	0.148660	-2.936	0.0033	***
urban	0.0279175	0.0111693	2.499	0.0124	**
lek	-0.0463339	0.0181390	-2.554	0.0106	**

Suma kwadratów reszt 0.311565 Błąd standardowy reszt 0.046515

Liczba instrumentów = 49

Test AR(1) dla błędu: z = -3.10303 [0.0019]

Test AR(2) dla błędu: z = -0.324263 [0.7457]

Test Sargana - nadmiernej identyfikacji: Chi-kwadrat(44) = 51.9125 [0.1928]

Test Walda (joint): Chi-kwadrat(5) = 949.041 [0.0000]



Jak widać test Sargana nadal potwierdza poprawność modelu. Również wyniki testów AR(1) i AR(2) się nie zmieniły. Jeśli chodzi o weryfikację modelu to w 2 testach z 4 otrzymano wyniki wskazujące na brak rozkładu normalnego reszt. Przeprowadzono jednak testy BP i White'a w celu sprawdzenia heteroskedastyczności.

Model 5: Estymacja Panelowe MNK, z wykorzystaniem 144 obserwacji  
Włączono 16 jednostek danych przekrojowych  
Szereg czasowy długości = 9  
Zmienna zależna (Y): sq\_reszty

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	0.0109336	0.00297055	3.681	0.0003	***
l_miesz_1	-0.00198738	0.00167217	-1.189	0.2367	
wynag	0.000445363	0.000632864	0.7037	0.4828	
lud	0.000564777	0.000347769	1.624	0.1067	
urban	-9.02841e-05	3.83106e-05	-2.357	0.0198	**
lek	-0.000791315	0.000416099	-1.902	0.0593	*
Średn.aryt.zm.zależnej	0.002164	Odch.stand.zm.zależnej	0.003813		
Suma kwadratów reszt	0.001914	Błąd standardowy reszt	0.003724		
Wsp. determ. R-kwadrat	0.079257	Skorygowany R-kwadrat	0.045897		
F(5, 138)	2.375794	Wartość p dla testu F	0.042028		
Logarytm wiarygodności	604.1174	Kryt. inform. Akaike'a	-1196.235		
Kryt. bayes. Schwarza	-1178.416	Kryt. Hannana-Quinna	-1188.994		
Autokorel.reszt - rho1	0.335025	Stat. Durbina-Watsona	1.172910		

Wyłączając stałą, największa wartość p jest dla zmiennej 4 (wynag)

Model 6: Estymacja Panelowe MNK, z wykorzystaniem 144 obserwacji  
Włączono 16 jednostek danych przekrojowych  
Szereg czasowy długości = 9  
Zmienna zależna (Y): sq\_reszty

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	-0.00378842	0.00687198	-0.5513	0.5823	
yh	0.0100320	0.00987247	1.016	0.3113	
yh2	-0.00399800	0.00345618	-1.157	0.2493	
Średn.aryt.zm.zależnej	0.002164	Odch.stand.zm.zależnej	0.003813		
Suma kwadratów reszt	0.002041	Błąd standardowy reszt	0.003805		
Wsp. determ. R-kwadrat	0.018191	Skorygowany R-kwadrat	0.004265		
F(2, 141)	1.306218	Wartość p dla testu F	0.274099		
Logarytm wiarygodności	599.4939	Kryt. inform. Akaike'a	-1192.988		
Kryt. bayes. Schwarza	-1184.078	Kryt. Hannana-Quinna	-1189.367		
Autokorel.reszt - rho1	0.379053	Stat. Durbina-Watsona	1.091244		

Co prawda wynik testu White'a nie wskazuje na występowanie heteroskedastyczności, ale test BP pokazuje odmienne wyniki. W modelu występuje heteroskedastyczność. Ze względu na brak rozkładu normalnego oraz występowanie heteroskedastyczności porzucono model. Sprawdzone inne kombinacje modelu z różnymi opóźnieniami zmiennych, ale nie udało się wyestymować odpowiedniego modelu.

## 10. Podsumowanie

Ostatecznie wybrano dwukierunkowy model efektów ustalonych postaci

```
? panel new_new_l_miesz const new_new_wynag new_new_drogi new_new_lud \par --time-dummies
```

Model 34: Estymacja Ustalonych efekty, z wykorzystaniem 128 obserwacji

Włączono 16 jednostek danych przekrojowych

Szereg czasowy długości = 8

Zmienna zależna (Y): new\_new\_l\_miesz

Z powodu ścisłej współliniowości pominięto zmienną: dt\_10

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	1.90948	0.445943	4.282	4.20e-05	***
new_new_wynag	0.0768222	0.0310217	2.476	0.0149	**
new_new_drogi	0.0116117	0.00633673	1.832	0.0698	*
new_new_lud	-1.08687	0.299107	-3.634	0.0004	***
dt_3	-0.0754806	0.0260150	-2.901	0.0046	***
dt_4	-0.0803809	0.0245350	-3.276	0.0014	***
dt_5	-0.0755428	0.0228897	-3.300	0.0013	***
dt_6	-0.0724183	0.0211869	-3.418	0.0009	***
dt_7	-0.0781904	0.0194404	-4.022	0.0001	***
dt_8	-0.0617649	0.0159297	-3.877	0.0002	***
dt_9	-0.0178809	0.0130673	-1.368	0.1742	
Średn.aryt.zm.zależnej	0.701042	Odch.stand.zm.zależnej	0.384841		
Suma kwadratów reszt	0.109870	Błąd standardowy reszt	0.032820		
LSDV R-kwadrat	0.994159	Within R-kwadrat	0.788423		
LSDV F(25, 102)	694.3895	Wartość p dla testu F	1.4e-102		
Logarytm wiarygodności	270.2469	Kryt. inform. Akaike'a	-488.4939		
Kryt. bayes. Schwarza	-414.3411	Kryt. Hannana-Quinna	-458.3652		
Autokorel.reszt - rho1	0.091738	Stat. Durbina-Watsona	1.461362		

Joint test on named regressors -

Statystyka testu: F(3, 102) = 13.3097

z wartością p = P(F(3, 102) > 13.3097) = 2.12466e-007

Test na zróżnicowanie wyrazu wolnego w grupach -

Hipoteza zerowa: grupy posiadają wspólny wyraz wolny

Statystyka testu: F(15, 102) = 34.0441

z wartością p = P(F(15, 102) > 34.0441) = 6.50103e-033

Test Walda na łączną istotność zmiennych 0-1 jednostek czasu -

Hipoteza zerowa: No time effects

Asymptotyczna statystyka testu: Chi-kwadrat(7) = 24.0529

z wartością p = 0.00111523

Model charakteryzuje się współczynnikiem determinacji  $R^2 = 0.99$  oraz wewnątrzgrupowym  $R^2 = 0.79$ . Niestety w trakcie usuwania autokorelacji z modelu wyjściowego reszty utraciły charakter rozkładu normalnego, co powoduje, że nie ma możliwości oszacowania istotności zmiennych modelu.

Analiza innych modeli (jednokierunkowych oraz dynamicznych) również nie przyniosła lepszych rezultatów.

W modelu nie występuje autokorelacja reszt modelu, a także heteroskedastyczność.

Testowanie modeli wskazało, że model FE był lepszy od modeli regresji łącznej oraz efektów losowych.

Ze względu na brak normalnego rozkładu reszt nie przeprowadzono interpretacji współczynników przy zmiennych.