



AGH

**AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA
W KRAKOWIE**

Badanie istotności wpływu czynników na kształtowanie się przeciętnego wynagrodzenia brutto w polskich miastach

Przedmiot: Ekonometria

Informatyka i Ekonometria rok 2 semestr IV

Wiktoria Paluch

Kraków 2020r

Spis treści

1. Wprowadzenie	4
1.1 Wstęp	4
1.2 Cel projektu	5
1.3 Źródło danych	5
1.4 Program Gretl	5
2. Model	5
2.1 Postać analityczna modelu	5
2.2 Opis zmiennych i uzasadnienie ich wyboru	6
2.3 Statystyki opisowe danych	7
2.4 Interpretacja:	8
2.5 Outliery- wartości odstające	8
2.6 Statystyki opisowe po usunięciu wartości odstających	8
2.7 Interpretacja:	9
2.8 Współczynnik zmienności	9
2.9 Korelacja między zmiennymi	10
2.10 Podział próby	11
2.11 Poziom istotności	11
2.12 Estymacja modelu KMNK	11
2.13 Test normalności rozkładu reszt modelu ze wszystkimi zmiennymi	12
2.14 Centralne Twierdzenie Graniczne	12
2.15 Istotność zmiennych objaśniających	13
2.16 Współczynnik determinacji	13
1. Dobór zmiennych objaśniających do modelu	14
3.1 Metoda Hellwiga	14
3.2 Metoda Krokowa Wstecz	15
3.3 Porównanie modeli	16
3.4 Badanie efektu katalizy	17
3.6 Ostateczny wybór zmiennych do modelu	18
4. Weryfikacja statystyczna modelu	18
4.1 Koincydencja	19
4.2 Współliniowość	19
4.3 Badanie autokorelacji	20
4.4 Losowość rozkładu reszt modelu	20
4.5 Liniowość modelu	20
4.6 Test pominiętych/zbędnych zmiennych	21

4.7 Test stabilności parametrów modelu.....	21
4.8 Heteroskedastyczność.....	22
5. Prognoza.....	23
6. Interpretacja.....	24
7. Podsumowanie	24
8. Spis tabel i wykresów	25
9. Bibliografia.....	26

1.Wprowadzenie

1.1Wstęp

Wynagrodzenie jest istotnym elementem organizowania procesów gospodarczych. Proces produkcji, podziału, wymiany i konsumpcji nie może odbywać się bez udziału dobrze opracowanej strategii wynagrodzeń. Płaca jest jednym z wyznaczników świadczącym o kondycji gospodarki.

Przegląd teorii ekonomicznych pozwala na wyodrębnienie kilku grup czynników determinujących kształtowanie się wynagrodzeń.

1. Czynniki o charakterze mikroekonomicznym. Zaliczyć można do nich np. kwalifikacje pracowników, liczbę osób poszukujących pracę, intensywność i czas poszukiwań pracy, oczekiwania i ambicje pracowników, kontakty z pracodawcami. Większość są to czynniki jakościowe, trudno mierzalne.
2. Przyjęta przez przedsiębiorstwo strategia wynagrodzeń. Wielkości te również mają charakter mikroekonomiczny, jednak należy je interpretować z perspektywy pracodawcy. Można do nich zaliczyć: strategię przedsiębiorstwa, system szkoleń, strukturę organizacyjną przedsiębiorstwa, stopień trudności wykonywanej pracy. Wielkości te są w większości mierzalne i mogą posłużyć jako podstawa do relatywizmu.
3. Czynniki makroekonomiczne. Dotyczą one otoczenia wewnętrznego i zewnętrznego przedsiębiorstw oraz mierników świadczących o poziomie rozwoju gospodarczego. Można do nich zaliczyć: poziom produkcji globalnej, stopę bezrobocia, zasób liczby pracujących, koszty pracy.

Cena na rynku pracy zależy od popytu i podaży. Na rynku pracy popyt reprezentują przedsiębiorcy, którzy poszukują osób do pracy, a podaż, czyli gotowość do pracy zgłaszają pracownicy. To te dwie siły wpływają na poziom wynagrodzenia pojedynczego pracownika. Oczywiście czynniki jakościowe również w dużym stopniu wpływają na wysokość wynagrodzenia, jednak są one trudno mierzalne oraz podporządkowane popytowi i podaży.

Przewaga konkurencyjna gospodarki regionu jest zawsze powiązana z czynnikami wewnętrznymi, zasobami regionu oraz czynnikami zewnętrznymi wynikającymi z sąsiedztwa oraz polityki interregionalnej kraju. Wielkie miasta mogą stanowić o przewadze komparatywnej wynikającej z zasobów wewnętrznych, a w globalnej gospodarce, mogą też ułatwiać nawiązywanie współpracy międzynarodowej. W miastach wojewódzkich wynagrodzenia wynoszą od kilkuset do nawet tysiąca PLN więcej niż w pozostałych regionach danego województwa. Doskonałym przykładem jest Warszawa. Mediana wynagrodzeń w Warszawie wynosi 6 000 PLN, a w województwie mazowieckim (bez miasta stołecznego) tylko 3 800 PLN. Źródeł tak znaczącej różnicy należałoby się raczej doszukiwać w tym, iż w próbie pracowników z Warszawy znajdziemy stosunkowo wiele osób zatrudnionych w centralach dużych, międzynarodowych korporacji, operujących w branżach uznawanych za dobrze opłacane (finanse i bankowość, ubezpieczenia, IT etc.). Na pozostałym terenie województwa odsetek tego rodzaju zatrudnionych będzie znacząco mniejszy, co może tłumaczyć istotną część dysproporcji zarobków.

To prowadzi do kolejnego pytania, czy w takim razie ilość firm zagranicznych wpływa znacząco na wzrost wynagrodzenia brutto? Inwestycje zagraniczne są postrzegane jako jeden z ważniejszych czynników innowacyjności, wzrostu i modernizacji gospodarki. Znaczenie ich jest szczególnie duże, gdy niedobór środków finansowych uniemożliwia podejmowanie inwestycji.

1.2 Cel projektu

Celem badań jest sprawdzenie jakie czynniki istotnie mają wpływ na wysokość średniego wynagrodzenia brutto w polskich miastach. Czy wielkość miasta, ludność czy też ośrodki innowacyjne lub z kapitałem zagranicznym mają wpływ na zmianę cen na rynku pracy?

Wyniki mogą być sygnałem dla średnich i małych miast, aby zainwestować swoje środki w rozwój danej dziedziny, by ożywić gospodarczo swój rejon.

Model ekonometryczny może również być wskazówką jakiej wypłaty możemy oczekiwać w danym rejonie.

1.3 Źródło danych

W większości dane pochodzą z BANKU DANYCH LOKALNYCH opracowanego dla Głównego Urzędu Statystycznego. Skompletowano dane dla 66 miast na prawach powiatu w Polsce dla 2018r, gdyż baza nie była kompletna dla 2019r.

Informację na temat liczby firm z kapitałem zagranicznym znaleziono na stronie:

- Centralnego Ośrodka Informacji Gospodarczej

- www.polskawliczbach.pl

1.4 Program Gretl

Badania przeprowadzone zostały w pakiecie ekonometrycznym „Gretl”. Należy on do oprogramowani Powszechnej Licencji Publicznej (GNU), czyli o swobodnym dostępie dla wszystkich użytkowników.

2. Model

2.1 Postać analityczna modelu

W celu opisanie zjawiska wykorzystano model regresji wielorakiej postaci:

Wzór 1-Postać analityczna modelu

$$y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5 + \alpha_6 X_6 + \alpha_7 X_7 + \varepsilon$$

gdzie:

- zmienna objaśniana:
y- przeciętne miesięczne wynagrodzenia brutto;
- zmienne objaśniające ilościowe:
x₁- ludność
x₂-liczba nowoutworzonych miejsc pracy
x₃-bezrobotny na ofertę pracy
x₄-cena za 1m² na rynku wtórnym
x₅-liczba firm z kapitałem zagranicznym
x₆-liczba przedsiębiorstw innowacyjnych
- zmienna objaśniająca binarna:
x₇- miasto wojewódzkie

2.2 Opis zmiennych i uzasadnienie ich wyboru

1. Przeciętne miesięczne wynagrodzenia brutto [zł]

Dane o wynagrodzeniu podane są w ujęciu brutto, tj. łącznie z zaliczkami na poczet podatku dochodowego od osób fizycznych oraz ze składkami na obowiązkowe ubezpieczenia społeczne (emerytalne, rentowe i chorobowe) płaconymi przez ubezpieczonego pracownika.

2. Ludność [osoby/km²]

Zmienna ukazująca gęstość zaludnienia na 1 km² w danym mieście. Gęstość zaludnienia lepiej obrazuje stan ludności, gdyż dodatkowo daje informacje o wykorzystaniu obszaru miasta. Jest ona uzupełnieniem zmiennej x3, gdyż konkurencja na rynku pracy to nie tylko bezrobotni zarejestrowani, ale także np. studenci i osoby, które chcą zmienić pracę.

3. Liczba nowoutworzonych miejsc pracy [tys. msc.]

Nowoutworzone miejsca pracy to miejsca powstałe w wyniku zmian organizacyjnych, rozszerzenia lub zmiany profilu działalności oraz wszystkie miejsca pracy w jednostkach nowo powstałych. Dane nie uwzględniają pracujących w jednostkach budżetowych działających w zakresie obrony narodowej i bezpieczeństwa publicznego. Obrazuje ona popyt na rynku pracy.

4. Bezrobotny na ofertę pracy

Dane o liczbie bezrobotnych zarejestrowanych. Jest to podaż rynku pracy. Dane nie uwzględniają pracujących w jednostkach budżetowych działających w zakresie obrony narodowej i bezpieczeństwa publicznego.

5. Cena za 1 m² na rynku wtórnym [zł]

Prezentowane dane dotyczą sprzedaży o charakterze rynkowym tj. sprzedaży na wolnym rynku (z pewnymi włączeniami, np. sprzedaży między stronami spokrewnionymi) oraz sprzedaży przetargowej. W przypadku braku odpowiedniej informacji w zbiorach pochodzących z RCiWN, podział na rynek pierwotny i wtórny został przeprowadzony przyjmując założenie, że sprzedaż na rynku pierwotnym dotyczy transakcji dokonywanych na wolnym rynku, gdzie stroną sprzedającą jest osoba prawna i średnia cena za 1 m² powierzchni użytkowej lokalu wynosi co najmniej 2 000 zł. Do transakcji na rynku wtórnym zaliczono pozostałe transakcje rynkowe przeprowadzone na wolnym rynku oraz sprzedaż przetargową.

6. Liczba firm z kapitałem zagranicznym

Liczba firm z kapitałem zagranicznym w danym mieście. Zmienna w celu sprawdzenia czy kapitał zagraniczny istotnie ma wpływ na rozwój gospodarki lokalnej.

7. Liczba przedsiębiorstw innowacyjnych

Liczba przedsiębiorstw, która prowadzi działalność innowacyjną w danym województwie. Działalność innowacyjna to całokształt działań naukowych, technicznych, organizacyjnych, finansowych i komercyjnych, które rzeczywiście prowadzą lub mają w zamierzeniu prowadzić do wdrażania innowacji. Niektóre z tych działań same z siebie mają charakter innowacyjny, natomiast inne nie są nowością, lecz są konieczne do wdrażania innowacji. Działalność innowacyjna obejmuje także

działalność badawczo-rozwojową (B+R), która nie jest bezpośrednio związana z tworzeniem konkretnej innowacji. Innowacje powstają w celu modernizacji i rozwoju mającego na celu rozwój regionu i przedsiębiorstwa oraz zwiększania korzyści z prowadzenia działalności. Należy sprawdzić czy istotnie działalność innowacyjna ma wpływ na wynagrodzenie.

8. Miasto wojewódzkie

Zmienna binarna określająca czy dane miasto jest miastem wojewódzkim.

- 1- Jest to miasto wojewódzkie
- 0- Nie jest to miasto wojewódzkie

W celu sprawdzenia czy istotnie status miasta wojewódzkiego wiąże się z większymi zarobkami.

2.3 Statystyki opisowe danych

Zmienna	Średnia	Mediana	Minimalna	Maksymalna
y	4612,3	4392,2	3743,1	8121,1
X1	1645,8	1562,5	202	3757
X2	63,058	47,600	11,200	151,7
X3	12,424	8	1	71
X4	3736,3	3244,5	1926	10640
X5	943,39	125	24	31977
X6	21,311	20,6	14,3	27,3
X7	0,27273	0,00	0	1
Zmienna	Odch.stand.	Wsp. zmienności	Skośność	Kurtoza
y	738,47	0,16011	2,0388	6,2772
X1	686,26	0,41697	0,64790	0,85281
X2	39,993	0,63424	0,59513	-0,51926
X3	12,083	0,97254	2,4285	7,6446
X4	19,196	0,57303	0,00000	-1,2006
X5	3986,3	4,2255	7,3365	54,406
X6	3,0736	0,14423	0,068180	0,020962
X7	0,44877	1,6455	1,0206	-0,95833
Zmienna	Percentyl 5%	Percentyl 95%	Zakres Q3-Q1	Brakujące obs.
y	3815,3	5798,3	657,28	0
X1	568,15	3139,8	902,50	0
X2	11,600	151,70	65,950	0
X3	2,0000	40,500	11,500	0
X4	3,3500	63,650	33,500	0
X5	39,750	3904,3	311,75	0
X6	15,250	27,300	2,8750	0
X7	0,00000	1,0000	1,0000	0

Tabela 1 Statystyki opisowe dla 66 obserwacji

2.4 Interpretacja:

-y przeciętne wynagrodzenie w miastach jest na poziomie 4612,3 zł, mediana jest aż o 300zł mniejsza, zatem jest ona bardziej wiarygodna, gdyż nie uwzględnia odstającej wartości maksymalnej 8121,1 zł.

-x1 średnio w badanych miastach na 1km² mieszka ok. 1640 osób, wartość ta odchyła się średnio o 686. Maksymalna gęstość zaludnienia wynosi 3757, a najmniejsza 202. Duża dysproporcja,

-x2 średnio powstało ok 63 tys. nowych miejsc pracy, kurtoza jest ujemna, zatem więcej wartości są rozproszone wokół średniej,

-x3 przeciętnie na jedną ofertę pracy jest 12 osób bezrobotnych, najmniejsza wartość wynosi 1 a największa 71, odchylenie wynosi 12,

-x4 znów widać duże dysproporcje w cenie za 1 m² na rynku wtórnym pomiędzy maksymalną a minimalną wartością ok 8 tys zł różnicy, średnio 1m² kosztuje 3700zł, wartości są rozproszone wokół średniej,

-x5 średnia ilość firm z kapitałem zagranicznych w miastach to 940, wartość ta odchyła się przeciętnie o 3990, bardzo duże odchylenie standardowe, tak samo jak różnica między maksymalną a minimalną wartością- 31953

-x6 średnia ilość przedsiębiorstw innowacyjnych w danym województwie to 21 tys, wartość ta przeciętnie odchyła się o 3 tys

-x7 prawie 30% badanych miast to miasta wojewódzkie

2.5 Outliery- wartości odstające

Ze statystyki opisowej zmiennych widać, że w obserwacjach występują outliery. Jest to dlatego, że do badań wybrano 66 miast na prawach powiatu, a zatem takie miasta jak np. Warszawa, Kraków, Poznań, Wrocław istotnie odbiegają wartościami od reszty miast i zawyżają statystyki opisowe. W celu uzyskania lepszego obrazu zjawiska usunięto 6 obserwacji, pozostałe wartości odstające są normalnym zjawiskiem i dzięki nim próba jest kompletna.

2.6 Statystyki opisowe po usunięciu wartości odstających

Zmienna	Średnia	Mediana	Minimalna	Maksymalna
y	4468,2	4372,4	3743,1	5848,3
x1	1591,0	1506,0	202,00	3757,0
x2	61,243	45,550	11,200	151,70
x3	12,917	8,5000	1,0000	71,000
x4	3447,3	3168,0	1926,0	6887,0
x5	288,67	116,00	24,000	1645,0
x6	21,150	20,600	14,300	27,300
x7	0,23333	0,00000	0,00000	1,0000
Zmienna	Odch.stand.	Wsp. zmienności	Skośność	Kurtoza
y	516,50	0,11559	1,0165	0,50543
x1	660,05	0,41486	0,68598	1,1716
x2	39,893	0,65139	0,60335	-0,56725
x3	12,501	0,96784	2,3147	6,8625
x4	1017,2	0,29507	1,4927	2,5925
x5	393,84	1,3643	2,2281	4,1963

X6	3,0515	0,14428	0,064699	0,097868
X7	0,42652	1,8280	1,2610	-0,40994
Zmienna	Percentyl 5%	Percentyl 95%	Zakres Q3-Q1	Brakujące obs.
y	3809,2	5696,1	664,78	0
X1	553,45	2899,0	802,25	0
X2	11,600	151,70	67,000	0
X3	2,0000	43,500	12,750	0
X4	2256,0	6352,5	1220,8	0
X5	43,350	1437,4	255,25	0
X6	14,950	27,300	2,9750	0
X7	0,00000	1,0000	0,00000	0

Tabela 2 Statystyki opisowe dla 60 obserwacji

2.7 Interpretacja:

-y przeciętne wynagrodzenie w miastach jest na poziomie 4468,2 zł, nie ma już tak dużych dysproporcji pomiędzy wartością minimalną i maksymalną, odchylenie standardowe wynosi 516 zł

-x1 średnio w badanych miastach na 1km² mieszka ok. 1590 osób, wartość ta odchyła się średnio o 660. Maksymalna gęstość zaludnienia wynosi 3757, a najmniejsza 202.

-x2 średnio powstało ok 61 tys. nowych miejsc pracy, kurtoza jest ujemna, zatem więcej wartości są rozproszone wokół średniej,

-x3 przeciętnie na jedną ofertę pracy jest 12 osób bezrobotnych, najmniejsza wartość wynosi 1 a największa 71, odchylenie wynosi 12,

-x4 zmniejszyły się dysproporcje w cenie za 1 m² na rynku wtórnym pomiędzy maksymalną a minimalną wartością, średnio 1m² kosztuje 3400zł, wartości są skumulowane wokół średniej,

-x5 średnia ilość firm z kapitałem zagranicznych w miastach to 288(duża zmiana w porównaniu do poprzednich statystyk), wartość ta odchyła się przeciętnie o 393,zmniejszyła się też różnica między maks i min.

-x6 średnia ilość przedsiębiorstw innowacyjnych w danym województwie to 21 tys, wartość ta przeciętnie odchyła się o 3 tys

-x7 20% badanych miast to miasta wojewódzkie

2.8 Współczynnik zmienności

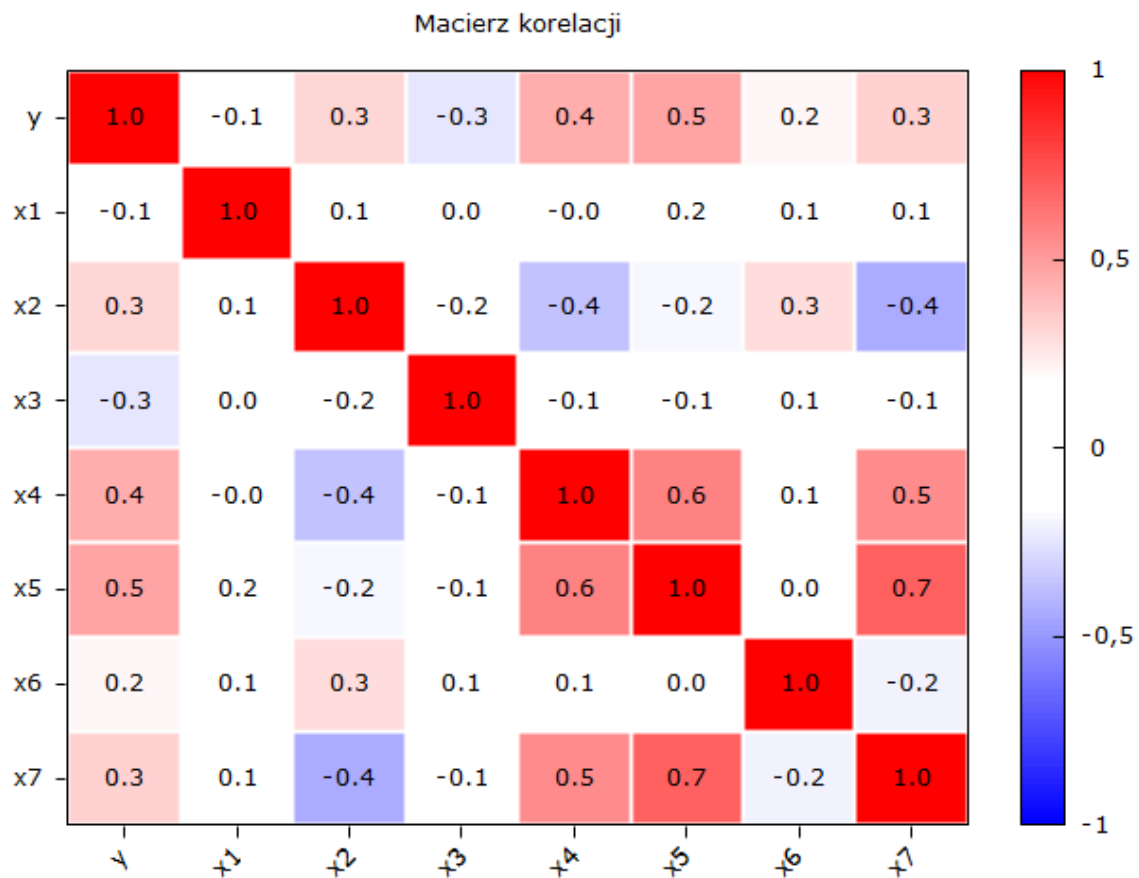
Miara zróżnicowania cechy. Gdy współczynnik zmienności jest mniejszy niż 10% oznacza to, że zmienna jest quasi-stała i nie powinna być uwzględniona w modelu.

Wszystkie badane zmienne mają współczynnik większy niż 10%, zatem nie ma podstaw do odrzucenia któreś z nich.

Wsp. zmienności	Zmienna
0,11559	y
0,41486	X1
0,65139	X2
0,96784	X3
0,29507	X4
1,3643	X5
0,14428	X6
1,8280	X7

Tabela 3 współczynnik zmienności

2.9 Korelacja między zmiennymi



Rysunek 1 Macierz korelacji

Aby zmienne jak najlepiej opisywały zjawisko pożądane są wysokie wartości korelacji zmiennych objaśniających ze zmienną objaśnianą- y oraz jak najmniejsze wartości bezwzględne korelacji między zmiennymi objaśniającymi.

- zapewne w modelu powinny się znaleźć zmienne x5(liczba firm z kapitałem zagranicznym) i x4(cena za 1m² na rynku wtórnym), gdyż są one najsilniej skorelowane ze zmienną y na poziomie 0.5 i 0.4, jednakże zmienne te są silnie między sobą skorelowane(na poziomie 0.6) dlatego zapewne tylko jedna z nich wejdzie do modelu.

-zmienne x2(liczba nowoutworzonych miejsc pracy), x3(bezrobotny na ofertę pracy) i x7(miasto wojewódzkie) są skorelowane ze zmienną y na poziomie 0.3 wartości bezwzględnej, zatem też powinny być brane pod uwagę podczas tworzenia modelu

-najsłabiej skorelowana ze zmienną y jest zmienna x1(ludność), co może świadczyć o tym, że zmienna nie jest istotna

-silne korelacje pomiędzy zmiennymi objaśniającymi występują w parach: x7 i x5, x4 i x7, x4 i x5. Można od razu założyć, że któraś ze zmiennych x4(cena za 1m² na rynku wtórnym), x5(liczba firm z kapitałem zagranicznym), x7(miasto wojewódzkie) nie powinna się znaleźć w modelu.

-dodatkowo widać, że niektóre wartości współczynnika korelacji między zmiennymi objaśniającymi są ujemne, zatem należy się spodziewać katalizatorów.

2.10 Podział próby

Obserwacje zostały podzielone na dwie próby:

- ucząca 54 obserwacji
- testowa 10 obserwacji

2.11 Poziom istotności

Przyjęty poziom istotności $\alpha = 0,05$. Jest to dopuszczalne ryzyko popełnienia błędu I rodzaju, czyli uznania prawdziwej hipotezy zerowej za fałszywą. Jeżeli w przeprowadzonych testach poziom p-value będzie mniejszy niż 5% oznacza to odrzucenie H_0 .

2.12 Estymacja modelu KMNK

Idea tej metody polega na znalezieniu takich wartości nieznanego wektora parametrów β , który minimalizują sumę kwadratów reszt, czyli różnic pomiędzy wartościami obserwowanymi a teoretycznymi.

Założenia MNK

- Zmienne objaśniające są nielosowe i nieskorelowane ze składnikiem losowym
- Ilość obserwacji jest większa od $k+1$, gdzie k to ilość zmiennych objaśniających
- Wartość oczekiwana składnika losowego wynosi 0
- $D^2(\epsilon) = E(\epsilon^T \epsilon) = \delta^2 I$
- Składnik losowy ma rozkład normalny

Estymator $\hat{\beta}$ uzyskany Klasyczną Metodą Najmniejszych Kwadratów jest estymatorem BLUE [best linear unbiased estimator], tj. zgodnym, nieobciążonym i najefektywniejszy w klasie liniowych estymatorów wektora β .¹

Do oszacowania modelu KMNK wykorzystano wszystkie zmienne

Model 1: Estymacja KMNK, wykorzystane obserwacje 1-54
Zmienna zależna (Y): y

	Współczynnik	Błąd stand.	t-Studenta	wartość p	
const	2974,73	489,637	6,075	<0,0001	***
x1	-0,224848	0,0754260	-2,981	0,0046	***
x2	9,57318	2,15729	4,438	<0,0001	***
x3	-1,11290	4,84118	-0,2299	0,8192	
x4	0,174049	0,0619817	2,808	0,0073	***
x5	0,251233	0,183970	1,366	0,1787	
x6	25,0762	19,8329	1,264	0,2125	
x7	446,462	192,402	2,320	0,0248	**

Tabela 4 Model 1

¹ <http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja1.pdf>

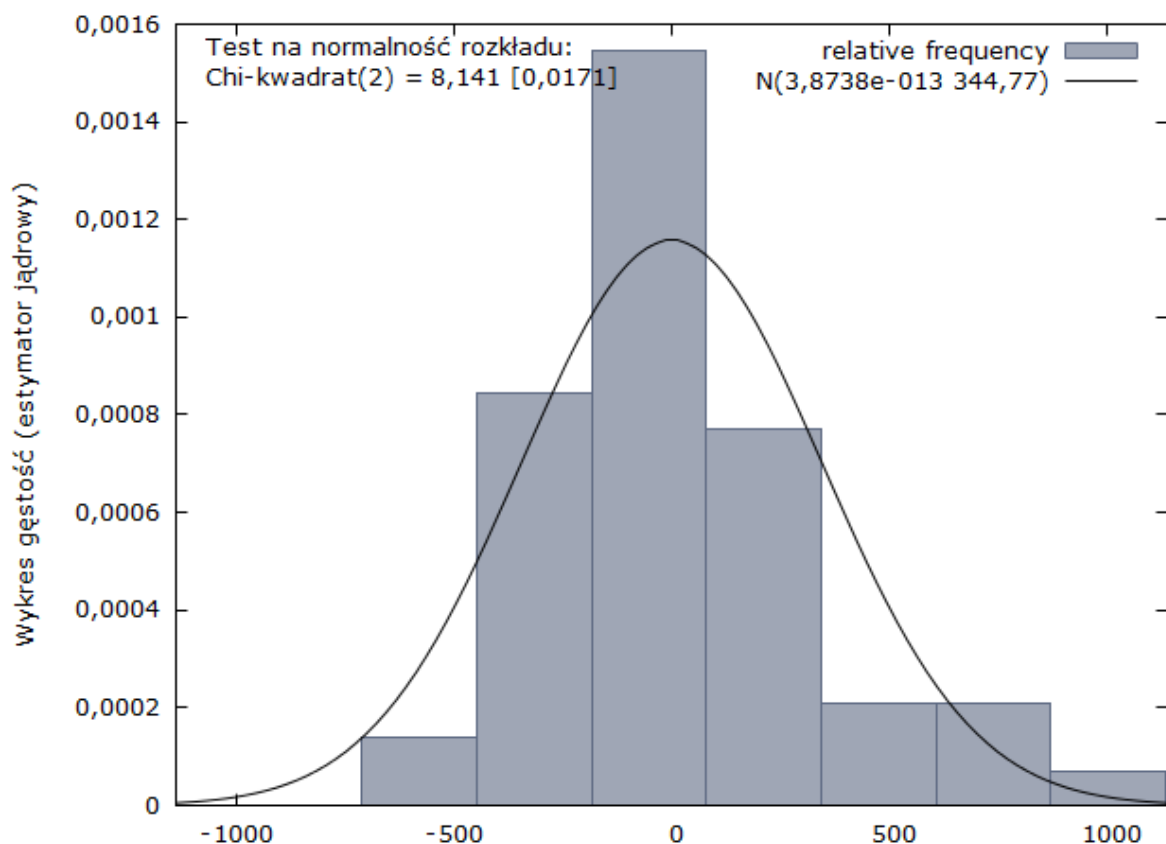
Średn.aryt.zm.zależnej	4458,144	Odch.stand.zm.zależnej	507,8728
Suma kwadratów reszt	5467962	Błąd standardowy reszt	344,7735
Wsp. determ. R-kwadrat	0,600019	Skorygowany R-kwadrat	0,539152
F(7, 46)	9,857909	Wartość p dla testu F	1,93e-07
Logarytm wiarygodności	-387,8094	Kryt. inform. Akaike'a	791,6187
Kryt. bayes. Schwarza	807,5306	Kryt. Hannana-Quinna	797,7553

Tabela 5 współczynniki dla model 1

Aby móc testować istotność zmiennych objaśniających należy najpierw zbadać normalność rozkładu reszt modelu. Jest to warunek konieczny i jedno z założeń testu t-Studenta na weryfikację istotności zmiennych.

2.13 Test normalności rozkładu reszt modelu ze wszystkimi zmiennymi

Do badania normalności rozkładu reszt można się posłużyć np. testem Doornika-Hansena



Rysunek 2 Reszty modelu ze wszystkimi zmiennymi

Wartość p-value testu Doornika-Hansena jest równa 0,0171. Oznacza to, że należy odrzucić H_0 o normalności rozkładu reszt modelu, co uniemożliwia sprawdzenie istotności poszczególnych zmiennych.

2.14 Centralne Twierdzenie Graniczne

Centralne twierdzenie graniczne to teoria uważana za najważniejszą dla wnioskowania statystycznego i przeprowadzania parametrycznych metod analiz. Wskazuje ona na to, że jeśli będziemy losować z populacji wiele próbek to rozkład średnich z tych próbek będzie zbliżony do rozkładu normalnego (nawet jeśli rozkład danej właściwości w populacji nie ma rozkładu normalnego), a średnia z tych średnich będzie taka sama (lub bardzo zbliżona) do średniej w

populacji. Po drugie wartość błędu standardowego z tego rozkładu będzie tym mniejsza, im większe będą losowane próbki.

W praktyce Centralne Twierdzenie Graniczne oznacza, że im większa będzie liczebność naszej badanej próby w pojedynczym badaniu, tym większa jest szansa, że pochodzi ona z rozkładu średnich z cechy o rozkładzie normalnym. Założeniem testów parametrycznych nie jest to, aby rozkład z próby był normalny, ale aby normalny był rozkład z populacji, z której pochodzi próba. Dodatkowo, im większa jest badana próba, tym większa jest szansa na bardziej precyzyjne odzwierciedlenie średniej z populacji (z racji tego że maleje błąd standardowy średnich z cechy).²

Wielkość badanej próby wynosi 54, a zatem na mocy Centralnego Twierdzenia Granicznego można przyjąć, że **reszty pochodzą z populacji o rozkładzie normalnym**.

2.15 Istotność zmiennych objaśniających

Skoro już normalność reszt została zweryfikowana można przejść do weryfikacji istotności zmiennych losowych. Posłużymy się testem t-Studenta istotności parametru

H0: zmienna jest nieistotna

H1: zmienna jest istotna

Zatem zależy na wartościach p-value < 5%

	Współczynnik	Błąd stand.	t-Studenta	wartość p	
const	2974,73	489,637	6,075	<0,0001	***
x1	-0,224848	0,0754260	-2,981	0,0046	***
x2	9,57318	2,15729	4,438	<0,0001	***
x3	-1,11290	4,84118	-0,2299	0,8192	
x4	0,174049	0,0619817	2,808	0,0073	***
x5	0,251233	0,183970	1,366	0,1787	
x6	25,0762	19,8329	1,264	0,2125	
x7	446,462	192,402	2,320	0,0248	**

Tabela 6 Model 1- badanie istotności zmiennych

Według wartości p-value testu t-Studenta zmienne x3, x5 i x6 (liczba przedsiębiorstw innowacyjnych) są nieistotne. Największa wartość p-value jest dla zmiennej x3 (bezrobotny na ofertę pracy), co nie wynikało z macierzy korelacji. Zmienna x5 (liczba firm z kapitałem zagranicznym) jest istotna zapewne ze względu na dużą korelację z pozostałymi zmiennymi.

Zmienna x1 (ludność) jest istotna mimo iż miała najmniejszą wartość korelacji ze zmienną y (na poziomie 0.1). Również istotnymi zmiennymi są zmienne x2 (liczba nowoutworzonych miejsc pracy), x4 (cena za 1m² na rynku wtórnym) oraz x7 (miasto wojewódzkie).

2.16 Współczynnik determinacji

Współczynnik R-kwadrat jest miarą jakości dopasowania modelu. Mówi on o tym, jaki procent jednej zmiennej wyjaśnia zmienność drugiej zmiennej. Przyjmuje on wartości od 0 do 1. Dzięki współczynnikowi determinacji R-kwadrat można się dowiedzieć jak bardzo zmiany jakiejś wartości są zdeterminowane zmianami w zakresie innej cechy.³

² <https://pogotowiestatystyczne.pl/slowniczek/centralne-twierdzenie-graniczne/>

³ <https://pogotowiestatystyczne.pl/slowniczek/wspolczynnik-r-kwadrat/>

Wsp. determ. R-kwadrat	0,600019
------------------------	----------

Tabela Współczynnik determinacji

Dla modelu 1 współczynnik determinacji R-kwadrat wynosi 0,6, zatem zmienne obrazują zjawisko w 60%.

1. Dobór zmiennych objaśniających do modelu

Wykorzystane metody doboru zmiennych to: Metoda Hellwiga i Metoda Krokowa Wstecz

3.1 Metoda Hellwiga

Metoda ta służy do wyodrębnienia kandydatów na zmienne objaśniające modelu spośród pewnego ich zbioru. Należy rozważyć wszystkie możliwe niepuste kombinacje kandydatek na zmienne objaśniające. Dla każdej kombinacji należy obliczyć pojemność integralną nośników informacji.

$$H_s = \sum_{j \in C_s} h_{sj}$$

gdzie:

C_s – podzbiór zmiennych objaśniających o numerze 's'

$$h_{sj} = \frac{r_j^2}{1 + \sum_{i \in C_s} |r_{ij}|}$$

r_j - współczynnik korelacji (zależności) między Y oraz X_j

r_{ij} - współczynnik korelacji (zależności) między X_i oraz X_j

Spośród wszystkich kombinacji należy wybrać tą o największej integralnej pojemności informacyjnej.⁴

Metodę Hellwiga przeprowadzono dla zbioru wszystkich zmiennych objaśnianych w celu dobrania optymalnego podzbioru. Skrypt następnie wskazał następujące zmienne:

```
0,014507106
x1
? H_max
0,38743754
? najlepszalista
x3 x4 x5
```

Rysunek 3wynik skryptu Metoda Hellwiga

Zatem według Metody Hellwiga optymalna postać modelu zawiera trzy zmienne objaśniające

x_3 -bezrobotny na ofertę pracy

x_4 -cena za 1m2 na rynku wtórnym

x_5 -liczba firm z kapitałem zagranicznym

Zatem sprawdzili się przewidywania z analizy macierzy korelacji. Zmienna x_5 miała największą wartość współczynnika korelacji ze zmienną y, zmienna x_4 niewiele mniejszą, a następnie była zmienna x_3 .

⁴ Metoda Hellwiga, <http://www.ekonometria.4me.pl/metoda-hellwiga/>

Model 2: Estymacja KMNK, wykorzystane obserwacje 1-54
Zmienna zależna (Y): y

	Współczynnik	Błąd stand.	t-Studenta	wartość p	
const	4058,40	216,446	18,75	<0,0001	***
x3	-12,0466	4,90467	-2,456	0,0176	**
x4	0,114436	0,0647228	1,768	0,0831	*
x5	0,458897	0,168645	2,721	0,0089	***

Tabela 8 Model otrzymany metodą Hellwiga

Średn. arytm. zm. zależnej	4458,144	Odch. stand. zm. zależnej	507,8728
Suma kwadratów reszt	8337576	Błąd standardowy reszt	408,3522
Wsp. determ. R-kwadrat	0,390106	Skorygowany R-kwadrat	0,353513
F(3, 50)	10,66051	Wartość p dla testu F	0,000016
Logarytm wiarygodności	-399,1998	Kryt. inform. Akaike'a	806,3995
Kryt. bayes. Schwarza	814,3554	Kryt. Hannana-Quinna	809,4678

Tabela 9 Wartości wskaźników dla modelu 2

Po wyestymowaniu modelu należy sprawdzić normalność rozkładu reszt. Wykonując test Dooornika-Hansena otrzymano p-value równe $p = 1,64082e-005$. Również i w tym przypadku p-value jest mniejsze niż 5%, jednak powołując się na Centralne Twierdzenie Graniczne i wielkość próby $n=54$ można przyjąć, że reszty pochodzą z populacji o rozkładzie normalnym, a zatem ich rozkład też jest normalny.

Gdy założenie o normalności rozkładu reszt zostało spełnione można przystąpić do sprawdzenia istotności zmiennych. Analizując wartości p-value testu t-Studenta można stwierdzić, że wszystkie zmienne są istotne, jednakże zmienna x4 ma największą wartość p-value i być może trzeba będzie ją usunąć z modelu

3.2 Metoda Krokowa Wstecz

Należy zacząć od modelu zawierającego wszystkie zmienne objaśniające i w każdym kolejnym kroku usuwać najmniej istotną z nich (dla której p-value testu t-Studenta jest najwyższe), aż do momentu kiedy wszystkie zmienne będą istotne. Konieczne jest sprawdzanie czy normalność rozkładu reszt modelu została zachowana. Zakończenie stosowania metody następuje po tym, gdy wszystkie zmienne w modelu są statystycznie istotne.⁵

Postępując zgodnie z Metodą Krokową Wstecz zostały usunięte zmienne w kolejności:

1. x_3 -bezrobotny na ofertę pracy
2. x_6 -liczba przedsiębiorstw innowacyjnych
3. x_5 -liczba firm z kapitałem zagranicznym

⁵ Regresja krokowa, https://www.naukowiec.org/wiedza/statystyka/regresja-krokowa_913.html

Model otrzymany Metodą Krokową Wstecz

Model 3: Estymacja KMNK, wykorzystane obserwacje 1-54
Zmienna zależna (Y): y

	Współczynnik	Błąd stand.	t-Studenta	wartość p	
const	3233,28	252,674	12,80	<0,0001	***
x1	-0,196041	0,0742190	-2,641	0,0110	**
x2	10,5215	1,80415	5,832	<0,0001	***
x4	0,232724	0,0551429	4,220	0,0001	***
x7	532,478	144,172	3,693	0,0006	***

Tabela 10 Model otrzymany MKW

Średn. aryt. zm. zależnej	4458,144	Odch. stand. zm. zależnej	507,8728
Suma kwadratów reszt	5989463	Błąd standardowy reszt	349,6197
Wsp. determ. R-kwadrat	0,561871	Skorygowany R-kwadrat	0,526105
F(4, 49)	15,70980	Wartość p dla testu F	2,45e-08
Logarytm wiarygodności	-390,2689	Kryt. inform. Akaike'a	790,5379
Kryt. bayes. Schwarza	800,4828	Kryt. Hannana-Quinna	794,3732

Tabela 11 Wskaźniki dla modelu MKW

Pierwszym krokiem po estymacji modelu jest sprawdzenie normalności rozkładu reszt. P-value dla testu Doornika-Hansena wynosi 0,02057 < 5%. Powołując się na Centralne Twierdzenie Graniczne i wielkość próby n=54 można przyjąć, że reszty pochodzą z populacji o rozkładzie normalnym, a zatem ich rozkład też jest normalny.

Dopiero po tym teście można zbadać istotność zmiennych. Dla każdej zmiennej p-value testu t-Studenta jest mniejsze niż 5%, zatem wszystkie zmienne są istotne.

3.3 Porównanie modeli

W celu porównania modeli, należy spojrzeć na mierniki dopasowania modelu:

- **Współczynnik determinacji R-kwadrat** pozwala zmierzyć jaką część zmienności (wariancji) zmiennej objaśnianej jest wyjaśniana zmiennością wartości teoretycznych wynikających z modelu. Współczynnik determinacji R-kwadrat zawsze będzie faworyzował modele z większą liczbą zmiennych. Zatem do porównywania modeli z różną liczbą zmiennych objaśniających wykorzystuje się skorygowany R-kwadrat. Oczywiście pożądana jest jak największa wartość tego współczynnika.
- **Kryteria Informacyjne** maleją wraz ze wzrostem funkcji wiarygodności oraz rosną wraz ze wzrostem liczby parametrów. W gretlu dostępne są kryteria informacyjne, jak np. bayesowskie kryterium Schwarza (BIC), Hannana-Quinna (HQ), Kryterium Akaike'a (AIC). Kryteria te różnią się od siebie ponieważ uwzględniają różną liczbę stopni swobody. Pożądana jest jak najmniejsza wartość tych kryteriów.⁶

⁶Ekonometria Weryfikacja liniowego modelu jednorównaniowego,
http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja2017L_2.pdf

Dla lepszej przejrzystości mierniki, o których mowa wyżej przedstawiono w tabeli:

	Model Hellwiga	Model MKW
Skorygowany R-kwadrat	0,353513	0,526105
Kryt. inform. Akaike'a	806,3995	790,5379
Kryt. Hannana-Quinna	809,4678	794,3732
Kryt. bayes. Schwarza	814,3554	800,4828

Tabela 12 Porównanie modeli

Widać, że pod każdym względem Model MKW jest lepszy od modelu otrzymanego za pomocą Metody Hellwiga. Jednakże dysproporcja między skorygowanymi R-kwadrat jest dość duża i również z macierzy korelacji można było odczytać, że występują katalizatory wśród zmiennych. Zatem przed ostatecznym wyborem postaci modelu wypada sprawdzić czy w danych modelach występują katalizatory, które sztucznie zawyżają R-kwadrat.

3.4 Badanie efektu katalizy

Informacja jaką niesie o modelu współczynnik R-kwadrat może być fałszywa, jeżeli w modelu występują zmienne, które są katalizatorami.

Założenie: (R, R_0) - regularna para korelacyjna

Zmienna X_i z pary (X_i, X_j) , $i < j$, jest katalizatorem jeżeli $r_{ij} < 0$ lub $r_{ij} > \frac{r_i}{r_j}$

Ponieważ efekt katalizy może występować w modelu z różnym natężeniem, dlatego został określony miernik : $\eta = R^2 - H$, gdzie H to integralna pojemność informacyjna zmiennych objaśniających.

Natężenie efektu katalizy, określa o jaką wartość jest zakłamanie R-kwadrat.

- **Model otrzymany metodą Hellwiga**

W tym modelu nie występują katalizatory, zatem wartość R-kwadrat nie jest zafałszowana i wynosi 0,35

- **Model otrzymany metodą krokową wstecz**

```
KATALIZATOR:
x1
W PARZE:
x1 x2
KATALIZATOR:
x2
W PARZE:
x2 x4
KATALIZATOR:
x1
W PARZE:
x1 x7
KATALIZATOR:
x2
W PARZE:
x2 x7
#nat[ ]enie efektu katalizy
? ols Y const xlist --quiet
? H=hellwig(Y,xlist)
Wygenerowano skalar H = 0,202119
? scalar natezenie_efektu_katalizy=$rsq-H
Wygenerowano skalar natezenie_efektu_katalizy = 0,359752
? natezenie_efektu_katalizy
0,35975201
```

W modelu otrzymanym Metodą Krokową Wstecz występują katalizatory. Widać, że na pewno katalizatorem jest zmienna x_1 i zmienna x_2 , po usunięciu tych zmiennych model nie powinien zawierać katalizatorów.

Natężenie katalizy jest duże, aż 0,35. Oznacza to, że współczynnik R-kwadrat jest zakłamanym o 0,35.

Rysunek 4 Wynik skryptu na badanie katalizy model MKW

Model MKW po usunięciu katalizatorów

Model 4: Estymacja KMNK, wykorzystane obserwacje 1-54
Zmienna zależna (Y): y

	Współczynnik	Błąd stand.	t-Studenta	wartość p	
const	3763,05	226,873	16,59	<0,0001	***
x4	0,189035	0,0693348	2,726	0,0088	***
x7	165,618	167,436	0,9891	0,3273	

Tabela 13 Model MKW bez katalizatorów

Średn. aryt. zm. zależnej	4458,144	Odch. stand. zm. zależnej	507,8728
Suma kwadratów reszt	10364713	Błąd standardowy reszt	450,8100
Wsp. determ. R-kwadrat	0,241821	Skorygowany R-kwadrat	0,212089
F(2, 51)	8,133238	Wartość p dla testu F	0,000859
Logarytm wiarygodności	-405,0759	Kryt. inform. Akaike'a	816,1518
Kryt. bayes. Schwarza	822,1187	Kryt. Hannana-Quinna	818,4530

Tabela 14 Współczynniki dla Model 4

P-value testu Doornika-Hansena jest równe 0, jednak na mocy Centralnego Twierdzenia Granicznego można przyjąć, że reszty te mają rozkład zbliżony do normalnego.

Po analizie wyników testu t-Studenta dla zmiennych można zauważyć, że x7 jest nieistotna.

3.6 Ostateczny wybór zmiennych do modelu

Po analizie mierników dopasowania oraz po uwzględnieniu efektu katalizy do dalszych badań wybrano model otrzymany metodą Hellwiga. Postać modelu:

$$y = 4058,40 - 12,0466x_3 + 0,114436x_4 + 0,458897x_5 + \varepsilon$$

gdzie:

- zmienna objaśniana:
y- przeciętne miesięczne wynagrodzenia brutto;
- zmienne objaśniające ilościowe:
x₃-bezrobotny na ofertę pracy
x₄-cena za 1m² na rynku wtórnym
x₅-liczba firm z kapitałem zagranicznym

4. Weryfikacja statystyczna modelu

Aby otrzymane metodą najmniejszych kwadratów estymatory aj współczynników α_j (j = 0, 1, 2, ..., k) były efektywne, muszą być spełnione założenia Gaussa–Markowa, a mianowicie:

- Związek między zmienną objaśnianą y a zmiennymi objaśniającymi x₁, x₂, ..., x_k ma charakter liniowy.
- Wartości zmiennych objaśniających są ustalone (nie są losowe) – losowość wartości zmiennej objaśnianej y wynika z losowości składnika ε.
- Składniki losowe ε dla poszczególnych wartości zmiennych objaśniających mają rozkład normalny (lub bardzo silnie zbliżony do normalnego) o wartości oczekiwanej zero i stałej wariancji: N(0, δε).

- Składowiki losowe nie są ze sobą skorelowane.⁷

4.1 Koincydencja

Model jest koincydentny, gdy wszystkie znaki stojące w wektorze R_0 przy zmiennych objaśniających X_i , czyli znaki współczynników r_i są zgodne ze znakami stojącymi przy tych zmiennych w oszacowanych parametrach a_i , tzn.:

$$\forall_{i \leq k} \operatorname{sgn}(r_i) = \operatorname{sgn}(a_i)$$

Brak koincydencji często świadczy o współliniowości zmiennych objaśniających.

zmienna	r_i	a_i
x_3 -bezrobotny na ofertę pracy	-0,3051	-12,0466
x_4 -cena za 1m ² na rynku wtórnym	0,4767	0,114436
x_5 -liczba firm z kapitałem zagranicznym	0,5204	0,458897

Tabela 15 badanie koincydencji

Można zauważyć, że wszystkie pary mają ten sam znak, zatem model jest koincydentny.

4.2 Współliniowość

Zgodnie z założeniami zmienne objaśniające w modelu powinny być skorelowane ze zmienną objaśnianą i nieskorelowane między sobą. Współliniowość nie wpływa na cechy estymatora MNK, ale sztucznie zawyża współczynnik determinacji, gdyż wprowadza do modelu tę samą informację.

Do oszacowania współliniowości w modelu można się posłużyć statystyką zwaną Variance Inflation Factor (VIF). Jest to prosty test oparty na statystyce R^2

$$VIF_j = \frac{1}{1 - r_j^2}$$

r_j^2 – współczynnik korelacji wielorakiej pomiędzy zmienną x_j , a pozostałymi zmiennymi

Mierzy on jaką część wariancji estymatora jest powodowana przez to, że zmienna j nie jest ortogonalna względem pozostałych zmiennych objaśniających w modelu regresji.⁸

Minimalna wartość jaką może przyjąć statystyka VIF to 1, co oznacza nieskorelowanie zmiennej z pozostałymi. Jeżeli wartość statystyki jest większa od 10 to wskazuje ona na współliniowość.

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

```
x3    1,006
x4    1,519
x5    1,513
```

Rysunek 5 Wartości statystyki VIF

⁷ Modelowanie ekonometryczne,

https://dbc.wroc.pl/Content/2182/PDF/Gladysz_modelowanie_ekonometryczne.pdf

⁸ <http://coin.wne.uw.edu.pl/pstrawinski/notatki/wspol.pdf>

W modelu nie ma problemu ze współliniowością zmiennych, gdyż dla każdej zmiennej wartość statystyki VIF jest mniejsza niż 10.

4.3 Badanie autokorelacji

Autokorelacja składnika losowego jest problemem najczęściej występującym w przypadku szeregów czasowych i polega na zależności (skorelowaniu) bieżących wartości składnika losowego od wartości przeszłych.⁹ Zjawisko autokorelacji pierwszego rzędu składników losowych można weryfikować między innymi za pomocą testów Durбина–Watsona, von Neumanna, Durбина, a występowanie autokorelacji dowolnego rzędu testem istotności współczynników autokorelacji.¹⁰

Dane są rozpatrywane w jednym okresie, zatem nie ma możliwości aby wartości były skorelowane z wartościami przeszłymi.

4.4 Losowość rozkładu reszt modelu

Do weryfikacji losowości rozkładu reszt modelu względem równania regresji \hat{y} można zastosować między innymi testy serii (test liczby serii, test maksymalnej długości serii).¹¹

Do badanego przykładu wykorzystano test liczby serii, o hipotezach:

H_0 : próba losowa,

$H_1 \sim H_0$

Dla reszt przeprowadzono nieparametryczny test serii dla dodatnich i ujemnych serii

Test serii

```
Liczba serii (R) dla zmiennej 'uhat2' = 23
Test niezależności oparty na liczbie dodatnich i ujemnych serii.
Hipoteza zerowa: próba jest losowa, dla R odpowiednio N(28, 3,64005),
test z-score = -1,37361, przy dwustronnym obszarze krytycznym p = 0,169564
```

Rysunek 6 test serii

p-value dla testu liczby serii wynosi 0,169564 > 5%, nie ma podstaw do odrzucenia H_0 , dobór jednostek do próby jest losowy.

4.5 Liniowość modelu

Liniowość modelu można sprawdzić za pomocą testu Ramseya RESET.

Test poprawnej specyfikacji modelu Ramseya RESET (ang. Regression Equation Specification Error Test) jest ogólnym testem, który pozwala zidentyfikować niepoprawną postać funkcyjną modelu ekonometrycznego. Hipoteza zerowa oznacza poprawną postać funkcyjną modelu.

H_0 : dobrana liniowa postać modelu jest poprawna,

$H_1 \sim H_0$ ¹²

Test ten porównuje postać liniową z dodatkowym modelem poszerzonym o:

- kwadraty zmiennej objaśnianej,
- sześciany zmiennej objaśnianej,
- kwadraty i sześciany zmiennej objaśnianej.

⁹ <http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja3.pdf>

¹⁰ https://dbc.wroc.pl/Content/2182/PDF/Gladysz_modelowanie_ekonometryczne.pdf

¹¹ https://dbc.wroc.pl/Content/2182/PDF/Gladysz_modelowanie_ekonometryczne.pdf

¹² http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja2017L_2.pdf

```

Test RESET na specyfikację (kwadrat i sześćcian zmiennej)
Statystyka testu: F = 0,130002,
z wartością p = P(F(2,48) > 0,130002) = 0,878

Test RESET na specyfikację (tylko kwadrat zmiennej)
Statystyka testu: F = 0,225487,
z wartością p = P(F(1,49) > 0,225487) = 0,637

Test RESET na specyfikację (tylko sześćcian zmiennej)
Statystyka testu: F = 0,216828,
z wartością p = P(F(1,49) > 0,216828) = 0,644

```

Rysunek 7 Test Ramsey RESET

W każdym przypadku p-value >5%, nie ma podstaw do odrzucenia hipotezy H_0 , dobrana postać liniowa modelu jest poprawna.

4.6 Test pominiętych/zbędnych zmiennych

Ponieważ dla zmiennej x_4 wartość p-value testu t-Studenta była na granicy, to żeby przekonać się faktycznie o istotności tej zmiennej należy przeprowadzić test pominiętych zmiennych. Wariant testu Walda!

$H_0 : \beta_{k-h}, \dots, \beta_k = 0$ w modelu słusznie nie uwzględniono zmiennych x_{k-h}, \dots, x_k

H_1 : ww. zmienne powinny zostać uwzględnione w modelu¹³

Test porównawczy z Modelem 1

```

Hipoteza zerowa: parametr regresji jest równy zero dla  $x_4$ 
Statystyka testu: F(1, 56) = 3,13259, wartość p 0,0821858
Pominięcie zmiennych poprawiło 1 z 3 kryteriów informacyjnych (AIC, BIC, HQC) .

```

Rysunek 8 Test pominiętych zmiennych

p-value = 0,0822 >5%, nie ma podstaw do odrzucenia hipotezy zerowej, zmienna x_4 powinna być usunięta z modelu. Jednakże usunięcie zmiennej x_4 powoduje, że zmienna x_3 przestaje być istotna, a także współczynnik R-kwadrat w dużym stopniu maleje.

4.7 Test stabilności parametrów modelu

Stabilność oszacowań parametrów modeli ekonometrycznych jest zasadna zarówno w przypadku prognozowania jak i analizy strukturalnej. Test Chowa pozwala na statystyczną identyfikację zmiany strukturalnej parametrów.¹⁴ Służy do weryfikacji czy parametry modelu będą takie same dla kilku różnych podpróbek. Punktem wyjściowym w teście Chowa jest wybór punktu załamania strukturalnego, a więc momentu w czasie, po którym nastąpiła zmiana strukturalna.

H_0 : parametry są takie same w podpróbach

H_1 : parametry różnią się w podpróbach¹⁵

¹³ https://web.sgh.waw.pl/~atoroj/ekonometria/cwiczenia_02.pdf

¹⁴ <http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja4.pdf>

¹⁵ <http://www.ekonometria.wne.uw.edu.pl/uploads/Main/Wyklad%2010.01.2014.pdf>

Test Chowa na zmiany strukturalne przy podziale próby w obserwacji 27 -
Hipoteza zerowa: brak zmian strukturalnych
Statystyka testu: $F(4, 46) = 1,93417$
z wartością $p = P(F(4, 46) > 1,93417) = 0,120731$

Test Chowa na zmiany strukturalne przy podziale próby w obserwacji 17 -
Hipoteza zerowa: brak zmian strukturalnych
Statystyka testu: $F(4, 46) = 1,14156$
z wartością $p = P(F(4, 46) > 1,14156) = 0,348945$

Test Chowa na zmiany strukturalne przy podziale próby w obserwacji 3 -
Hipoteza zerowa: brak zmian strukturalnych
Statystyka testu: $F(2, 48) = 0,724413$
z wartością $p = P(F(2, 48) > 0,724413) = 0,489831$

Rysunek 9 Test Chowa

W każdym przypadku p-value >5%, nie ma podstaw do odrzucenia H_0 , parametry są takie same w każdej podpróbie.

4.8 Heteroskedastyczność

Heteroskedastyczność oznacza, że jedno z założeń MNK (poniżej) zostało naruszone:

$$\text{var}(e_i) = \sigma^2$$

Konsekwencje heteroskedastyczności:

1. Estymator MNK jest nadal liniowym i nieobciążonym estymatorem, ale nie jest najbardziej efektywny. Istnieje inny estymator o mniejszej wariancji (błędzie).
2. Błędy standardowe zwykle obliczane dla estymatora MNK są niepoprawne.
3. Przedziały ufności i testy hipotez, które obliczają błędy standardowe są niepoprawne i mogą prowadzić do niewłaściwych wniosków.¹⁶

Do wykrywania heteroskedastyczności służą różne testy m.in.

- Breuscha-Pagana,
- White'a,
- Koenkera.

Każdy z nich ma ten sam zestaw hipotez:

H_0 : reszty modelu są homoskedastyczne,

H_1 : reszty modelu są heteroskedastyczne.

Przeprowadzono wszystkie trzy testy dla modelu:

¹⁶ coig.com.pl/firmy-z-kapitałem-zagranicznym.php

Test White'a na heteroskedastyczność reszt (zmiennosc wariacji resztowej) -
Hipoteza zerowa: heteroskedastyczność reszt nie występuje
Statystyka testu: LM = 2,06133
z wartością $p = P(\text{Chi-kwadrat}(9) > 2,06133) = 0,99046$

Test Breuscha-Pagana na heteroskedastyczność -
Hipoteza zerowa: heteroskedastyczność reszt nie występuje
Statystyka testu: LM = 4,34955
z wartością $p = P(\text{Chi-kwadrat}(3) > 4,34955) = 0,226109$

Test Breuscha-Pagana na heteroskedastyczność (robust variant) -
Hipoteza zerowa: heteroskedastyczność reszt nie występuje
Statystyka testu: LM = 1,79955
z wartością $p = P(\text{Chi-kwadrat}(3) > 1,79955) = 0,615032$

Rysunek 10 Testy na heteroskedastyczność

Dla każdego testu wartości $p\text{-value} > 5\%$, nie ma podstaw do odrzucenia H_0 , zmienna „y” oraz składnik losowy są homoskedastyczne.

5. Prognoza

Zbudowane na podstawie modelu ekonometrycznego prognozy można podzielić na dwie zasadnicze grupy: prognozy ex post oraz prognozy ex ante. Prognozy ex post oparte są na znanych wartościach zmiennych objaśniających. W momencie, gdy poznajemy zrealizowaną wartość zmiennej prognozowanej – prognoza ex post staje się prognozą wygasłą. Prognozy ex ante są natomiast oparte na nieznanymi wartościach zmiennej objaśnianej.¹⁷

Przeprowadzono prognozowanie punktowe ex-post. Punktem, jaki wybrano do prognozy jest ten wyznaczony przez średnie wartości zmiennych objaśniających z modelu, czyli:

x_3 -średnia ilość bezrobotnych na ofertę pracy

x_4 - średnia cena za 1m² na rynku wtórnym

x_5 -średnia liczba firm z kapitałem zagranicznym

```
? scalar me_x3=mean(x3)
Zamieniono skalar me_x3 = 11,4259
? scalar me_x4=median(x4)
Zamieniono skalar me_x4 = 3145
? scalar me_x5=median(x5)
Zamieniono skalar me_x5 = 125
? scalar y_pred=stala +a1*me_x3+a2*me_x4+a3*me_x5
Zamieniono skalar y_pred = 4338,02
```

Rysunek 11 Prognoza punktowa

Prognozowana średnia wartość wynagrodzenia brutto dla takich obserwacji to 4338,02

Błąd prognozy: 412,906

95% przedział ufności: (3508,68 ; 5167,37)

¹⁷ <http://www.prognozowanie.info/ex-ante-post/>

W celu zbadania dokładności prognozy wybrano 2 rodzaje błędów, które niosą takie informacje:

- MAE(mean absolute error)- średni błąd absolutny, podaje w jednostkach bezwzględnych o ile średnio prognoza różni się od wartości rzeczywistej

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - y_t^p|$$

Wartość dla modelu wynosi MAE = 451,23 zł, co oznacza, że prognoza średnio różni się od rzeczywistej wartości o 451,23 zł

- MAPE(mean absolute percentage error)- średni bezwzględny błąd procentowy, podaje o ile procent średnio prognoza różni się od wartości rzeczywistej.

$$MAPE = \frac{1}{S} \sum_{t=1}^S \left| \frac{y_t - y_t^p}{y_t} \right| \cdot 100$$

MAPE = 0,0907415, zatem średni bezwzględny błąd procentowy wynosi zaledwie 9%.

6. Interpretacja

Ostatecznie model ma postać:

$$y = 4058,40 - 12,0466x_3 + 0,114436x_4 + 0,458897x_5 + \varepsilon$$

gdzie:

- zmienna objaśniana:
y- przeciętne miesięczne wynagrodzenia brutto;
- zmienne objaśniające ilościowe:
x₃-bezrobotny na ofertę pracy
x₄-cena za 1m² na rynku wtórnym
x₅-liczba firm z kapitałem zagranicznym

Oznacza to, że wraz ze wzrostem ilości osób bezrobotnych na ofertę pracy, przeciętne wynagrodzenie brutto maleje o wartość równą iloczynowi ilości tych osób razy 12zł. Wraz ze wzrostem ceny za 1 m² mieszkania na rynku wtórnym o złotówkę przeciętne wynagrodzenie wzrośnie w danym mieście o 0,11 zł. Gdy liczba firm z kapitałem zagranicznym wzrośnie o 1 to przeciętne wynagrodzenie brutto wzrośnie o 0,46 zł.

7. Podsumowanie

W badanym przypadku analizowano wpływ różnych czynników na wysokość przeciętnego wynagrodzenia brutto w danym mieście. Badania potwierdziły wstęp teoretyczny do zagadnienia. Na wysokość przeciętnego wynagrodzenia wpływa przede wszystkim:

- Podaż i popyt, podaż obrazuje zmienna bezrobotni na ofertę pracy, zgodnie z teorią jeżeli podaż wzrasta to cena (wynagrodzenie) maleje.
- Drugim ważnym czynnikiem wpływającym na wysokość wynagrodzenia są ceny mieszkań, które zapewne są ceną zaporową i wpływają na wymagania pracowników wobec pracodawców w kwestiach wynagrodzenia.

- Trzecim czynnikiem jest liczba firm z kapitałem zagranicznym, wzrost tego czynnika przyczynia się do wzrostu wynagrodzeń. Zatem prawdziwe jest stwierdzenie, że inwestycje zagraniczne są postrzegane jako jeden z ważniejszych czynników innowacyjności, wzrostu i modernizacji gospodarki.

R-kwadrat modelu jest na poziomie 39%, zatem teoretycznie model nie obrazuje zjawiska nawet w połowie. Wpływ na to zapewne jest spowodowany tym, że nie wzięto pod uwagę czynników jakościowych, trudno mierzalnych, a mających na pewno istotny wpływ na kształtowanie się cen na rynku pracy. Jednakże wyżej opisane czynniki są istotną wskazówką dla inwestorów bądź też władz miast.

8. Spis tabel i wykresów

Rysunek 1 Macierz korelacji	10
Rysunek 2 Reszty modelu ze wszystkimi zmiennymi	12
Rysunek 3 Wynik skryptu Metoda Hellwiga	14
Rysunek 4 Wynik skryptu na badanie katalizy model MKW	17
Rysunek 5 Wartości statystyki VIF	19
Rysunek 6 test serii.....	20
Rysunek 7 Test Ramsey RESET.....	21
Rysunek 8 Test pominiętych zmiennych	21
Rysunek 9 Test Chowa.....	22
Rysunek 10 Testy na heteroskedastyczność	23
Rysunek 11 Prognoza punktowa	23
Tabela 1 Statystyki opisowe dla 66 obserwacji	7
Tabela 2 Statystyki opisowe dla 60 obserwacji	9
Tabela 3 współczynnik zmienności.....	9
Tabela 4 Model 1	11
Tabela 5 współczynniki dla model 1	12
Tabela 6 Model 1- badanie istotności zmiennych	13
Tabela 7 współczynnik determinacji	14
Tabela 8 Model otrzymany metodą Hellwiga	15
Tabela 9 Wartości wskaźników dla modelu 2	15
Tabela 10 Model otrzymany MKW.....	16
Tabela 11 Wskaźniki dla modelu MKW	16
Tabela 12 Porównanie modeli.....	17
Tabela 13 Model MKW bez katalizatorów	18
Tabela 14 Współczynniki dla Model 4	18
Tabela 15 badanie koincydencji	19

9. Bibliografia

- P. Carlsson, Kształtowanie wynagrodzeń, BL Info, Gdańsk 2008, s. 13.
- M. Nyk, Wynagrodzenie a pracujący w gospodarce polskiej, SWSPiZ, s.113
- Wanda Maria Gaczek, METROPOLIA JAKO ŹRÓDŁO PRZEWAGI KONKURENCYJNEJ GOSPODARKI REGIONU,
- Katarzyna Puchalska, Bezpośrednie inwestycje zagraniczne jako czynnik modernizacji polskiej gospodarki
- <http://www.praca.egospodarka.pl/108439,Co-wplywa-na-wysokosc-wynagrodzenia,1,55,1.html>
- <https://www.edufinanse.pl/moje-finanse/news-od-czego-zalezy-poziom-wynagrodzenia,nld,2405274>
- <http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja1.pdf>
- <https://pogotowiestatystyczne.pl/slowniczek/centralne-twierdzenie-graniczne/>
- <https://pogotowiestatystyczne.pl/slowniczek/wspolczynnik-r-kwadrat/>
- Metoda Hellwiga,<http://www.ekonometria.4me.pl/metoda-hellwiga/>
- Regresja krokowa,https://www.naukowiec.org/wiedza/statystyka/regresja-krokowa_913.html
- Ekonometria Weryfikacja liniowego modelu jednorównaniowego,
- http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja2017L_2.pdf
- Modelowanie ekonometryczne,
- https://dbc.wroc.pl/Content/2182/PDF/Gladysz_modelowanie_ekonometryczne.pdf
- <http://coin.wne.uw.edu.pl/pstrawinski/notatki/wspol.pdf>
- <http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja3.pdf>
- https://dbc.wroc.pl/Content/2182/PDF/Gladysz_modelowanie_ekonometryczne.pdf
- https://dbc.wroc.pl/Content/2182/PDF/Gladysz_modelowanie_ekonometryczne.pdf
- http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja2017L_2.pdf
- https://web.sgh.waw.pl/~atoroj/ekonometria/cwiczenia_02.pdf
- <http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja4.pdf>
- <http://www.ekonometria.wne.uw.edu.pl/uploads/Main/Wyklad%2010.01.2014.pdf>
- coig.com.pl/firmy-z-kapitalem-zagranicznym.php
- <http://www.prognozowanie.info/ex-ante-post/>