

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

Aneta Gałuszewska
Marta Kwiatosz

**Determinanty podaży pracy zamężnych kobiet w Stanach
Zjednoczonych w roku 2002**

Model ekonometryczny przygotowany
pod kierunkiem mgr Dariusza Szymańskiego

Warszawa, styczeń 2008

Spis treści

1. Wstęp.....	2
2. Problem ekonometryczny.....	2
3. Hipotezy badawcze oraz fundamenty teoretyczne.....	3
4. Przegląd literatury pod kątem hipotez badawczych.....	4
5. Opis zbioru danych oraz zmiennych wykorzystanych w modelu.....	5
6. Analiza danych.....	7
✖ Wstępna analiza danych.....	7
✖ Analiza zmiennej zależnej.....	7
✖ Analiza zmiennej binarnej.....	9
✖ Analiza zmiennych ciągłych.....	9
7. Przewidywana forma funkcyjna.....	13
8. Estymacja parametrów.....	15
9. Problemy ze zbiorem danych.....	17
10. Diagnostyka modelu.....	18
11. Interpretacja wyników estymacji.....	21
12. Podsumowanie.....	22
13. Bibliografia.....	24
 Załączniki.....	25
✖ Praca ze zmienną miasto.....	25
✖ Praca ze zmienną wiek.....	32
✖ Porównanie modeli z zastosowania zmiennej niezlogarytmowanej oraz zlogarytmowanej dochm.....	33

1. Wstęp

Rynek pracy w USA znajduje się w nierównowadze. Występuje niedopasowanie w wielkości popytu na pracę oraz jej podaży, o czym świadczy wskaźnik bezrobocia na poziomie około 5.5% w roku badanym. Łatwo można zaobserwować różnice w sile roboczej mężczyzn i sile roboczej kobiet, która przez wiele lat w przeszłości utrzymywała się na niskim poziomie. Wskaźnik aktywności kobiet na rynku pracy (angielski skrót LFP) od ponad 100 lat rośnie, z 18,9% w 1890 roku do 57,5% w roku 1990. Od roku 1990, kobiety LFP ciągle rośnie, ale już w mniejszym stopniu. W 2002 kształtował się w okolicach 60 procent, podczas gdy ten sam wskaźnik, tylko obliczony dla mężczyzn wskazywał, że ponad 72% z nich pracowało w roku 2002. Dlaczego kobiety stanowią mniejszość na rynku pracy? Dlaczego prawie 40 % z nich podejmuje decyzje o nie-pracowaniu? Co wpływa na decyzję o podjęciu pracy i przepracowaniu określonej liczby godzin przez zamężne kobiety?

Przedstawiony poniżej model próbuje odpowiedzieć na wyżej postawione pytania czyli na pytanie o determinanty podaży pracy zamężnych kobiet w Stanach Zjednoczonych w roku 2002.

2. Problem ekonometryczny

Tak jak już zostało zaznaczone we wstępie, rola kobiety na rynku pracy nie jest tak znacząca jak rola mężczyzn, kobiety w mniejszym stopniu angażują się w aktywne uczestniczenie na tym rynku.

Rola kobiety w społeczeństwie na przełomie wieków miała różne aspekty. Dyskusje podjęte w latach 70 i 80 doprowadziły do wzrostu świadomości, że fundamentalną kwestią nie są zmiany w postrzeganiu ról kobiet i mężczyzn, lecz relacje między pracą a życiem rodzinnym. Kobieta bowiem od zawsze zajmowała się przede wszystkim domem, a mężczyzna zapewniał byt rodzinie. Obserwowany trend ciągłego wzrostu aktywności zawodowej jest związany niejako z „przymusem ekonomicznym” (praca jako źródło utrzymania), wzrostem wykształcenia kobiet i ich aspiracji zawodowych oraz ze zwykle małą dziećmi. Kobiety, w związku ich pozycją społeczną (opieka nad dziećmi, prowadzenie gospodarstwa domowego) podejmują często pracę w niepełnym wymiarze godzin. Często dobrze wykształcona kobieta, z ogromnymi ambicjami związanymi z pracą, musi decydować pomiędzy spędzeniem czasu w domu (z dziećmi) a rozwijaniem swoich umiejętności związanych z pracą. Dla większości kobiet jest to decyzja bardzo trudna, jest to bowiem swoista wymiana „trade – off”. Z jednej strony ogromnie ważną rolę dla kobiet spełnia potomstwo, rodzina z drugiej jednak strony dla coraz większej części kobiet kariera zawodowa jest równie ważna.

Większość ludzi spędza dużą część swojego życia w pracy. Zatrudnienie jest głównym źródłem utrzymania, daje często poczucie spełnienia, podczas gdy rodzina jest strukturą tworzącą i wychowującą następne pokolenia. Obydwie dziedziny są częstą głównym źródłem satysfakcji społecznej i psychicznej członków rodziny. Praca zawodowa może jednak ograniczać życie rodziny przez limitowanie ilości czasu na funkcje rodzinne. Podobnie rodzina ze swoimi powinnościami i obowiązkami może ograniczyć aktywność zawodową, wykonywanie pracy zawodowej.

Rozpatrywanie relacji między pracą a życiem rodzinnym koncentrowało się bardzo często na negatywnych konsekwencjach pracy zawodowej dla rodziny, powodujących separowanie pracy i domu, wymagających nieobecności ojca przez większość dnia. Jeden z rodziców, najczęściej kobieta, pozostawała w domu zajmując się rodziną, pracami domowymi. Jak wynika z badań (Eurobarometr 1990)¹ kobiety wolałyby nie pracować w pełnym wymiarze godzin, gdy ich dzieci były w wieku przedszkolnym.

¹ http://ec.europa.eu/public_opinion/archives/ebs/ebs_48_en.pdf

Można by zatem zadać sobie pytanie, co w rzeczywistości wpływa na liczbę godzin przepracowanych przez zamężne kobiety. Problem ten wydaje się być szczególnie ciekawy ze względu na coraz większą popularność kobiet na rynku pracy. Nie są one na razie tak aktywne jak mężczyźni, ale coraz częściej kobiety wolą poświęcić życie rodzinne na zdobywanie kariery. Do tematu podeszliśmy od strony decyzji kobiety a nie pracodawcy. Oczywiście, wiemy, że tak na prawdę to dostępność pracy kształtuje podaż, jednak w tym przypadku chcieliśmy zbadać determinanty indywidualnych decyzji kobiety.

Celem naszej pracy będzie zatem określenie istotności wybranych czynników, jakie wpływają na podaż pracy kobiet. Zmienne zastosowane w modelu odwołują się zarówno do kategorii czysto ekonomicznych, jak i społeczno-demograficznych.

3. Hipotezy badawcze oraz fundamenty teoretyczne

Zgodnie z teorią makroekonomiczną podaż jakiegoś dobra zależy od jego ceny. Tak samo dzieje się z pracą. Podaż pracy jest więc dodatnio skorelowana ze stawką płacy. Jednak dotyczy to zagregowanej funkcji podaży pracy. Jeśli chodzi o przedstawiony w modelu problem badawczy, to wszystko zależy od indywidualnych preferencji kobiety. Mogłoby się wydawać, że wzrost płacy zawsze będzie powodował wzrost podaży pracy, jednak jak okaże się później, możemy mieć również do czynienia z efektem substytucyjnym (czyli wzrost płacy spowoduje wzrost dochodu, co wpłynie na zmniejszenie przepracowanych przez kobietę godzin). Mamy więc do czynienia w przypadku płacy z dwoma efektami – dochodowym i substytucyjnym, które działają w różnych kierunkach. Istotny jest również fakt rozdzielenia wzrostu płacy, którą otrzymuje mąż.

Analizując teorie mikroekonomiczne można dojść do wniosku, że dochód którym już dysponuje rodzina również wpływa na decyzje podażowe kobiety. Oczywiście, znak przy tej zależności jest ujemny – im więcej rodzina posiada zasobów materialnych (w naszej pracy są to zarobki męża) tym mniej kobieta musi pracować.

Naszym zdaniem, bardzo istotnym wpływem odznacza się wiek kobiety na liczbę przepracowanych godzin. I ma on wpływ dwojaki – młodsze kobiety będą mniej pracować ponieważ większość z nich będzie się uczyć czy przeznaczać czas na opiekę nad dziećmi, dopiero później liczba godzin przez nie przepracowanych będzie wzrastać. Analitycy rynku pracy wskazują na ustabilizowaną i potem ujemną zależność wieku i podaży pracy po przekroczeniu pewnego progu wieku (około 45 lat) a tłumaczą to zjawisko występowaniem osiągnięcia stabilności zawodowej (młodzi ludzie muszą się bardziej starać, aby dostać lepszą pracę, aby być zauważonym przez pracodawcę) oraz zbliżeniem się do progu emerytury (im bliżej do emerytury, tym mniej pracy wykonują, są coraz mniej potrzebni na rynku pracy). Łatwo można zauważyć również elementy wypierania z rynku pracy starszych osób przez młode.

Edukacja ma również istotny wpływ na zachowanie się kobiet na rynku pracy. Kobiety, które poświęciły więcej czasu na zdobycie wykształcenia, wykorzystują zdobyte umiejętności i rozwinięte zdolności w pracy. Wydaje się być oczywistym faktem to, że kobiety o wyższym wykształceniu są bardziej ambitne od tych, które skończyły tylko obowiązkowe szkoły. Ambitna kobieta, idąc dalej, nie jest w stanie zrezygnować z dobrej pracy, ze swoich marzeń z nią związanych.

Doświadczenie zawodowe w podobny sposób wpływa na aktywność zawodową kobiet. Człowiek, który odznacza się znacznym doświadczeniem zawodowym, jest bardziej ceniony przez pracodawcę, co umożliwia mu rozwijanie swoich ambicji. Praca związana jest również z przebywaniem z innymi ludźmi, rozwijaniem znajomości, z czego tak łatwo się nie rezygnuje.

Tak jak już wcześniej zostało nakreślone, dzieci w bardzo istotny sposób wpływają na decyzje o aktywnym uczestniczeniu na rynku pracy kobiet. Jest to chyba oczywiste, że

kobiety, których najmłodsze dziecko jest bardzo małe, pracują mniej, niż te, które mają już dzieci, które potrafią o siebie zadbać. Dlatego w modelu zawarte zostały dwie zmienne dotyczące dzieci – liczba dzieci oraz wiek najmłodszego.

Wydaje się, że wpływ miejsca zamieszkania jest również zauważalny w decyzjach podaźowych dotyczących pracy kobiet. Kobiety, które mieszkają w dużych miastach będą pracowały więcej, niż te, które mieszkają w mniejszych miasteczkach a wiąże się to z tym, że w większych miastach są wyższe koszty utrzymania (dochód męża może nie wystarczyć na zapewnienie bytu rodzinie). Duże miasto wiąże się również z większą możliwością rozwijania umiejętności, co również kosztuje (bardzo często rodzice poświęcają dużą część zarobków na zapewnienie lepszej edukacji dzieciom – kursy językowe, zajęcia pozalekcyjne).

Niewątpliwie poziom bezrobocia również wpływa na aktywność kobiety na rynku pracy. Nawet wtedy, gdy kobieta chce pracować, często nie jest w stanie znaleźć pracy, co związane jest z niedopasowaniem popytu na prace i jej podaży. Faktem jest, iż mężczyźni są chętniej zatrudniani przez pracodawców, dlatego w czasach, gdy występuje silne bezrobocie, kobiety będą mniej pracowały.

4. Przegląd literatury pod kątem hipotez badawczych

Przed rozpoczęciem analizy interesującego nas zjawiska, dokonaliśmy krótkiego przeglądu literatury. Dotyczyła ona problematyki związanej z podażą pracy kobiet. Artykuły i badania, do których dotarliśmy, to wyłącznie pozycje napisane w języku angielskim (trudno się dziwić, przecież nasz problem ekonometryczny dotyczy Stanów Zjednoczonych). Poniżej zawarliśmy główne tezy dotyczące przedstawionych wcześniej hipotez badawczych.

Tak jak już wcześniej zaznaczyliśmy, teorie ekonomiczne, które wyjaśniają podaż pracy kobiet, kładą szczególny nacisk na efekty zmian w cenie pracy, czyli w wynagrodzeniu. W tym przypadku bardzo istotne jest rozróżnienie czy zmiana wynagrodzenia dotyczy kobiety, czy jej męża (jeśli następuje wzrost dochodu męża to prawdopodobne jest wystąpienie efektu substytucyjnego – wzrost dochodu rodziny spowoduje wycofanie się z rynku pracy kobiety). Jacob Mincer² (1962) zauważył, że wyższe wynagrodzenia były głównym powodem wzrostu siły roboczej kobiet do roku 1960. Zauważył on również bardzo istotną właściwość – liczba przepracowanych przez kobiety godzin rosła wraz ze wzrostem wynagrodzenia za ich prace, jednak malała wraz ze wzrostem wynagrodzeń mężczyzn (wy tłumaczył to jako zjawisko efektu substytucyjnego oraz wpływu dochodu rodziny – wzrost wynagrodzenia męża powoduje wzrost dochodu całej rodziny). Później przeprowadzane badania wskazują na mniejszy wpływ wynagrodzenia kobiet na podaż ich pracy (W latach 1970-tych podaż pracy zameężnych kobiet rosła nawet gdy wynagrodzenie pozostawało względnie stałe lub nawet zmniejszało się – badania przeprowadzane przez Serwina Rosena³). Leibowitz i Klerman⁴ wskazują na silniejszy wpływ dochodu mężczyzn na podaż pracy kobiet niż to wyjaśnił Mincer. Tak więc, płaca kobiet nie musi powodować wzrostu podaży pracy, tak jak by się to na pierwszy rzut oka wydawało.

Edukacja to kolejny bardzo istotny czynnik wpływający na liczbę przepracowanych godzin przez kobiety. Badania przeprowadzone np. przez Jamesa Smitha i Michaela Warda⁵ wskazują na związek (dodatni) pomiędzy wyższym poziomem wykształcenia a większą indywidualną podażą pracy wśród kobiet. Związek ten jest zazwyczaj wyjaśniany poprzez stwierdzenie, że kobiety, które będą chciały dużo pracować są skłonne do inwestowania w

² Mincer J. *Labour Force Participation of Married Women: a Study of Labour Supply*, [w:] *Aspects of Labour Economics* pod. Red. H. Gregg Lewis, Princeton University Press, 1962.

³ Rosen. S *Mincering Labor Economics*, [w:] *Journal of Economic Perspectives*, rok 1992 nr 6, s. 157-170.

⁴ Leibowitz A, Klerman J. A., *Explaining Changes in Married Mothers Employment Over time* [w:] *Demography*, rok 1995 nr 32, s. 365-370.

⁵ Smith, J. P., Ward M. P., *Women's Wages and Work in the Twentieth Century*, Rand Publications, 1984.

wyższe wykształcenie (z wyższym wykształceniem wiąże się oczywiście większa możliwość znalezienia pracy). Lepsze wykształcenie często powoduje chęć do rozwijania zawodowej kariery oraz odkrywania nowych talentów nie tylko jako matka, ale przede wszystkim jako pracownik.⁶

Teorie ekonomiczne pokazują również zależność pomiędzy wiekiem kobiety a ilością pracy przez nią wykonywaną. Często pokazuje się wpływ płodności⁷ (wieku, w którym kobieta może mieć dzieci) oraz jej zmian, na podaż pracy. Gdy kobieta jest bardzo młoda to wtedy może mniej pracować bo będzie się jeszcze uczyć. Później może ograniczyć pracę (gdy podejmie decyzję o urodzeniu dziecka), natomiast gdy dziecko osiągnie odpowiedni wiek to wtedy kobieta wraca na rynek pracy. Dlatego też wiek przedziałami wpływa na liczbę przepracowanych godzin.

Badania pokazują również dwojaki wpływ liczby dzieci (posiadania dzieci) na decyzje o podaży pracy. Uważa się, że posiadanie małych dzieci będzie skutkowało wzrostem odpowiedzialności za nie co przekłada się najczęściej na ograniczenie pracy kobiet (chęć poświęcać więcej czasu i wysiłku na opiekę nad dzieckiem). Szczególnie ważne w tym problemie jest rozgraniczenie na dzieci małe (do lat 6) które potrzebują większej uwagi rodziców i na dzieci starsze.

Wydaje się, że te kobiety, które mieszkają w dużych miastach, będą więcej pracowały niż te, które mieszkają na wsi i w miastach małych. W dużych miastach są większe koszty utrzymania, dlatego też kobiety mogą poświęcać więcej czasu na pracę.

5. Opis zbioru danych oraz zmiennych wykorzystanych w modelu

Baza danych pochodzi ze strony internetowej: <http://www.psidonline.org>. Są to dane z roku 2003 (dotyczące roku 2002) z badania panelowego przeprowadzanego regularnie w Stanach Zjednoczonych (Panel Study of Income Dynamics).

Dane te zostały przez nas uporządkowane oraz trochę zmienione. Ostatecznie, badana próbka to zamężne kobiety, które w roku badania mieszkały na terenie USA. Są to kobiety, które: pracowały w roku 2002 (ich mężowie również pracowali), mają przynajmniej 18 lat, naukę pobierały na terenie USA. Dane dotyczące bezrobocia pobrałyśmy ze strony Departamentu Pracy - <http://www.bls.gov/lau/lastrk02.htm> (odpowiednie numery stanów z PSID zamieniłyśmy na stopę bezrobocia).

Liczba obserwacji – 2067

Liczba zmiennych – 10

Zmienne modelu:

- a) Liczba przepracowanych godzin w roku 2002
- b) Ilość posiadanych dzieci
- c) Wiek najmłodszego dziecka
- d) Wiek kobiety
- e) Liczba lat poświęconych na edukację
- f) Zarobki badanej kobiety (za godzinę pracy)
- g) Dochód z pracy męża
- h) Wskaźnik bezrobocia
- i) Miejsce zamieszkania
- j) Doświadczenie zawodowe zdobyte w poprzednich miejscach pracy (w latach)

⁶ Heer D. M., Grossbard-Shechtman A., *The Impact of the Female Marriage Squeeze and the Contraceptive Revolution on Sex Roles and the Women's Liberation Movement in the United States, 1960 to 1975* [w:] *Journal of Marriage and the Family*, rok 1981 nr 43, s. 49-65.

⁷ Smith J. P., Ward M. P., Op. cit.

Zmienna objaśniana:

- ✖ Liczba godzin przepracowanych przez kobiety w roku 2002. W modelu próbka kobiet składa się tylko z kobiet, które w roku 2002 przepracowały chociaż jedną godzinę.

$$y = \textit{godzk}$$

Jest to zmienna ciągła przyjmująca wartości z przedziału (7;5712).

Zmienne objaśniające:

- ✖ Wiek badanej kobiety.

$$x_1 = \textit{wiek}$$

Jest to zmienna ciągła przyjmująca wartości z przedziału (18;76).

- ✖ Wiek badanej kobiety podniesiony do kwadratu.

$$x_2 = \textit{wiek_2}$$

Jest to zmienna ciągła zbudowana na podstawie zmiennej *wiekk*

- ✖ Ilość posiadanych dzieci w wieku do 18 lat

$$x_3 = \textit{dzieci}$$

Jest to zmienna dyskretna przyjmująca wartości od 0 do 6. 0 oznacza że kobieta nie posiada dzieci.

- ✖ Wiek najmłodszego dziecka:

$$x_4 = \textit{najml}$$

Jest to zmienna dyskretna przyjmująca wartości od 0 do 17. 0 oznacza, że kobieta nie posiada dzieci.

- ✖ Liczba lat poświęconych na edukację

$$x_5 = \textit{edu}$$

Jest to zmienna dyskretna przyjmująca wartości od 0 do 17, gdzie wartości od 0 do 16 to liczba lat spędzonych w szkole podstawowej, średniej, bądź wyższej, a 17 oznacza dokończanie na kursach podyplomowych.

- ✖ Średnie zarobki kobiety (za godzinę) w roku 2002

$$x_6 = \textit{plac}$$

Jest to zmienna ciągła. Są to wartości zadeklarowane przez badanych.

- ✖ Dochód

$$x_7 = \textit{doch}$$

męża

Jest to zmienna ciągła. W wyjściowej bazie danych znajdowały się zmienne *plac* i *godzm*, jednak stwierdziłyśmy, że istotniejszy w estymacji będzie dochód uzyskiwany przez męża. Tak więc zmienna ta otrzymałaśmy poprzez pomnożenie zmiennej *plac* i *godzm*.

- ✖ Wskaźnik bezrobocia

$$x_8 = \textit{bezrob}$$

Zmienna ciągła powstała na podstawie zmiennej opisującej stan w USA, w którym mieszkała respondentka i informacji odnośnie wskaźnika bezrobocia w poszczególnych stanach w 2002 roku w USA. W wyjściowej bazie danych znajdowała się informacja odnośnie stanu, zakodowana za pomocą State FIPS Code⁸. Przekodowałyśmy ją następnie, dzięki informacjom ze strony departamentu pracy Stanów Zjednoczonych⁹, na odpowiednie wskaźniki bezrobocia w roku 2002.

- ✖ Miejsce zamieszkania

$$x_9 = \textit{miasto}$$

Jest to zmienna binarna. W kwestionariuszu pytanie odnoszące się do miejsca zamieszkania dotyczyło wielkości największego miasta w hrabstwie, co uznałyśmy za dobrą informację. I tak zmienna przyjmuje 6 poziomów:

⁸ <http://www.bls.gov/lau/lausfips.htm>

⁹ <http://www.bls.gov/lau/lastrk02.htm>

- 1 - Okręg metropolii¹⁰: 500 000 lub więcej mieszkańców
- 2 - Okręg metropolii: 100 000-499 999 mieszkańców
- 3 - Okręg metropolii: 50 000-99 999 mieszkańców
- 4 - Okręg poza metropolią: 25 000-49 999 mieszkańców
- 5 - Okręg poza metropolią: 10 000-24 999 mieszkańców
- 6 - Okręg poza metropolią: poniżej 10 000 mieszkańców

✖ Doświadczenie zawodowe kobiety (w latach)

$$x_{10} = \text{dosw}$$

Jest to zmienna ciągła. Jest to zadeklarowana ilość lat przepracowanych przez kobietę.

6. Analiza danych

a) Wstępna analiza danych

Statystyki opisowe

Poniższa tabela zawiera podstawowe statystyki (średnia, odchylenie standardowe, minimalna oraz maksymalna wartość) zmiennych używanych w naszym modelu.

```
. sum godzkk wiekk placak dosw edu dzieci najml miasto bezrob dochm
```

Variable	Obs	Mean	Std. Dev.	Min	Max
godzkk	2067	1777.135	676.0656	7	5712
wiekk	2067	41.02274	10.23012	18	76
placak	2067	17.09207	11.68854	1	115.38
dosw	2067	10.82583	7.488723	0	55
edu	2067	13.53604	2.370771	0	17
dzieci	2067	1.047412	1.095921	0	6
najml	2067	4.545235	5.50264	0	17
miasto	2067	3.494436	1.756014	1	6
bezrob	2067	5.707789	.8860335	3.3	7.6
dochm	2067	50435.09	75550.43	300.04	1999999

Macierz korelacji

```
. corr godzkk wiekk placak dosw edu dzieci najml miasto bezrob dochm
(obs=2067)
```

	godzkk	wiekk	placak	dosw	edu	dzieci	najml	miasto	bezrob	dochm
godzkk	1.0000									
wiekk	0.0139	1.0000								
placak	-0.0513	0.0977	1.0000							
dosw	0.0922	0.5591	0.1377	1.0000						
edu	0.0538	-0.0222	0.3536	0.0354	1.0000					
dzieci	-0.1405	-0.2980	-0.0403	-0.2536	-0.0502	1.0000				
najml	-0.0025	0.0320	-0.0410	-0.0588	-0.0313	0.4763	1.0000			
miasto	-0.0230	-0.0003	-0.1038	-0.0560	-0.0795	-0.0124	-0.0241	1.0000		
bezrob	-0.0526	-0.0048	0.0308	0.0110	-0.0216	0.0462	0.0232	-0.2423	1.0000	
dochm	-0.0576	0.1044	0.1713	0.0292	0.1725	0.0173	0.0850	-0.0500	0.0044	1.0000

Powyższe informacje wskazują tylko na istotną korelację pomiędzy wiekiem a doświadczeniem, pomiędzy wiekiem najmłodszego a liczbą dzieci oraz pomiędzy liczbą lat przeznaczonych na edukację a otrzymywaną płacą. Wydaje nam się, że poszczególne korelacje wynikają z tego, iż

1. Wraz ze wzrostem wieku kobiety, liczba lat, które już przepracowała także rośnie (stąd korelacja dodatnia na poziomie 0.5591)
2. Wynika to tylko z tego, że w naszej próbce 872 kobiety nie miały dzieci (czyli wiek najmłodszego równał się zero).

¹⁰ Wyjściowo SMSA: Central city of a Standard Metropolitan Statistical Area (SMSA) with (a) less than 400,000 population or (b) density of less than 6,000 per square mile

3. Wraz ze wzrostem liczby lat przeznaczonych na edukację rośnie płaca otrzymywana za wykonywaną pracę, czyli kobiety o wyższym wykształceniu otrzymują wyższe wynagrodzenie.

Przyjrzyjmy się dokładniej jak poszczególne zmienne wpływają na kształtowanie się zmiennej zależnej.

b) Analiza zmiennej zależnej

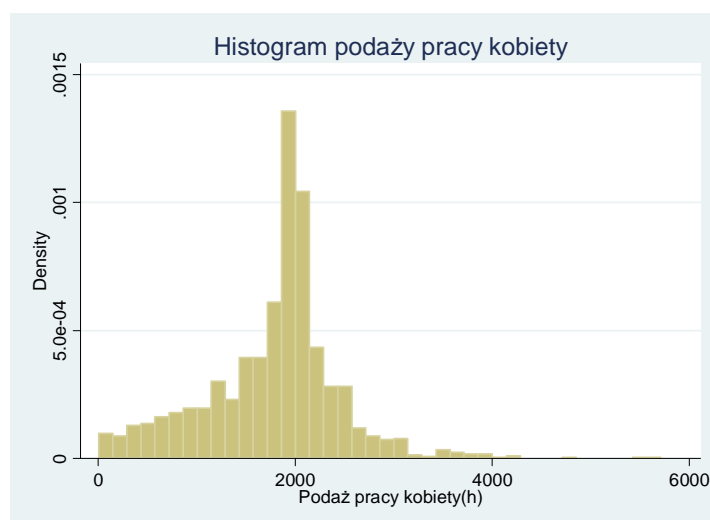
. summarize godz, detail

Podaż pracy kobiety(h)				
	Percentiles	Smallest		
1%	104	7		
5%	480	7		
10%	780	8	Obs	2067
25%	1444	20	Sum of wgt.	2067
50%	1920		Mean	1777.135
		Largest	Std. Dev.	676.0656
75%	2080	4160		
90%	2460	4845	Variance	457064.7
95%	2750	5500	Skewness	-.0379147
99%	3630	5712	Kurtosis	4.786122

Mediana jest wyższa od średniej, czyli w próbie przeważają kobiety o liczbie przepracowanych godzinach wyższych niż średnia (asymetria lewostronna, ujemna, słaba).

Współczynnik skupienia – kurtosis, która jest miarą skupienia poszczególnych obserwacji wokół średniej wynosi 4,78 a im wyższa wartość współczynnika (powyżej 3) tym bardziej wysmukła krzywa liczebności. To samo widać na histogramie:

Rys1. Histogram zmiennej godz



Średnia liczba godzin przepracowanych przez kobiety zaczyna się od poziomu 7 godzin w roku a kończą na poziomie ponad 5700. Dla wartości ok. 2000 przyjmuje swoje maksimum. Mediana kształtująca się na poziomie ok. 1900 potwierdza, że rozkład cechuje się słabą asymetrią.

c) Analiza zmiennej binarnej

Poniżej przedstawiamy wartości średnie oraz odchylenia zmiennej zależnej godz w zależności od poszczególnych poziomów zmiennej binarnej miasto.

. by miasto: summarize godzkk

-> miasto = Okręg metropolii: 500 000 lub więcej

Variable	Obs	Mean	Std. Dev.	Min	Max
godzk	295	1831.563	623.3241	120	3840

-> miasto = Okręg metropolii: 100 000-499 999

Variable	Obs	Mean	Std. Dev.	Min	Max
godzk	522	1788.41	756.6147	39	5712

-> miasto = Okręg metropolii: 50 000-99 999

Variable	Obs	Mean	Std. Dev.	Min	Max
godzk	245	1724.367	665.877	50	4000

-> miasto = Okręg poza metropolią: 25 000-49 999

Variable	Obs	Mean	Std. Dev.	Min	Max
godzk	267	1764.06	718.8806	7	4160

-> miasto = Okręg poza metropolią: 10 000-24 999

Variable	Obs	Mean	Std. Dev.	Min	Max
godzk	347	1772.654	607.4206	7	3430

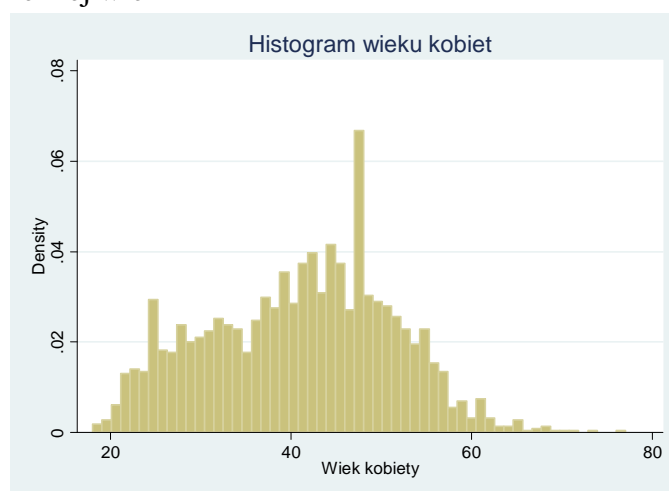
-> miasto = Okręg poza metropolią: poniżej 10 000

Variable	Obs	Mean	Std. Dev.	Min	Max
godzk	391	1766.987	634.0413	34	3968

Jak widzimy, średnia liczby godzin przepracowanych przez kobietę nie różni się znacząco od siebie w poszczególnych poziomach zmiennej miasto. Wydaje się, że będzie ona miała nieistotny wpływ na zmienną zależną, ale to stwierdzenie potwierdzimy/obalimy dopiero później, przeprowadzając formalne testy.

c) Analiza zmiennych ciągłych

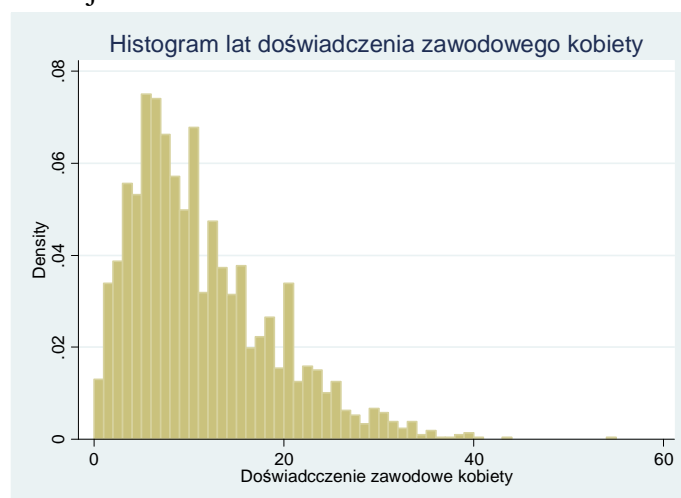
Rys 2. Histogram zmiennej wiekk



W badaniu udział wzięły kobiety z przedziału 18 – 76 lat, co wskazuje na to że w badaniu udział wzięły kobiety pełnoletnie, jak też te które są już na emeryturze (ale ciągle pracują). Rozkład zmiennej jest symetryczny i nieco spłaszczony. Najwięcej przebadanych zostało kobiet w wieku 40 – 50 lat co potwierdza wartość mediany wynosząca 42 lata. Najwięcej było

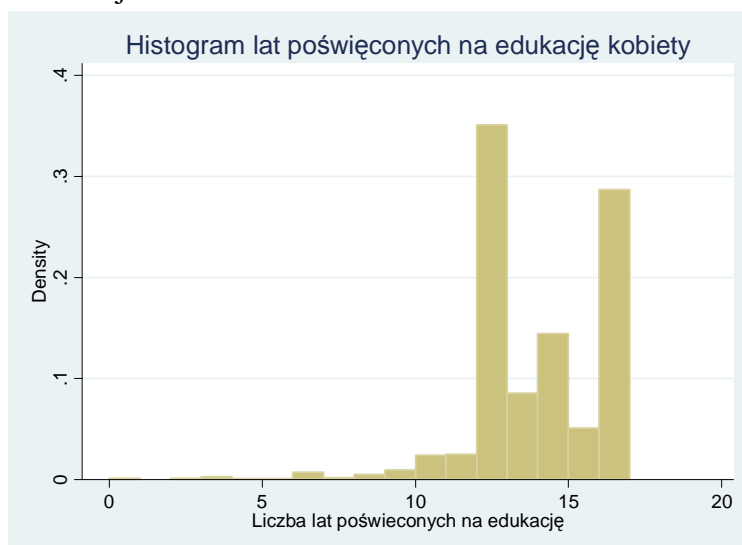
kobiet w wieku 47 lata. W modelu uwzględniamy również zmienną wiekk_2 jednak łatwiejsza w interpretacji jest zmienna wiek.

Rys 3. Histogram zmiennej dosw



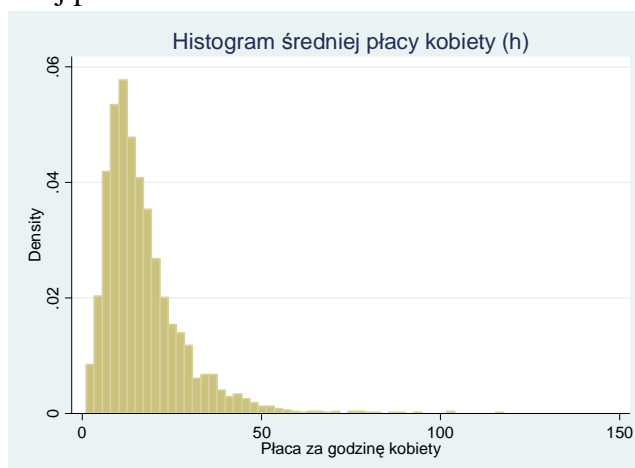
Z histogramu odczytujemy, że ilość przepracowanych lat zawiera się w przedziale (0;55). Można zauważyć, że wykres nie jest symetryczny, co wynika z faktu, iż większość przebadanych kobiet pracowała od 5 do 15 lat. Fakt ten potwierdza mediana wynosząca 9 lat.

Rys 4. Histogram zmiennej edu.



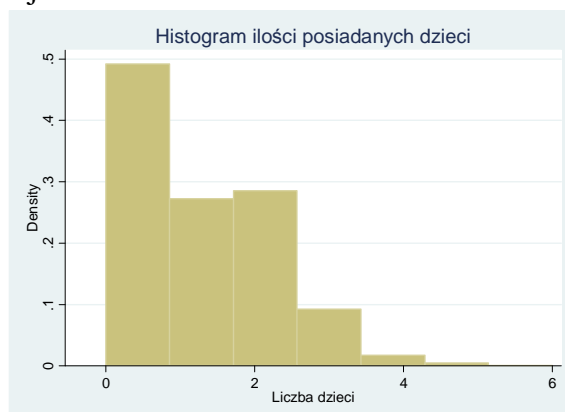
Jak widać na wykresie rozkład zmiennej nie jest symetryczny. Wykazuje zdecydowaną lewoskośność, ponieważ większość respondentek poświęciła na naukę 12 lub więcej lat. Potwierdza to wartość mediany równa 13 lat.

Rys 5. histogram zmiennej placak



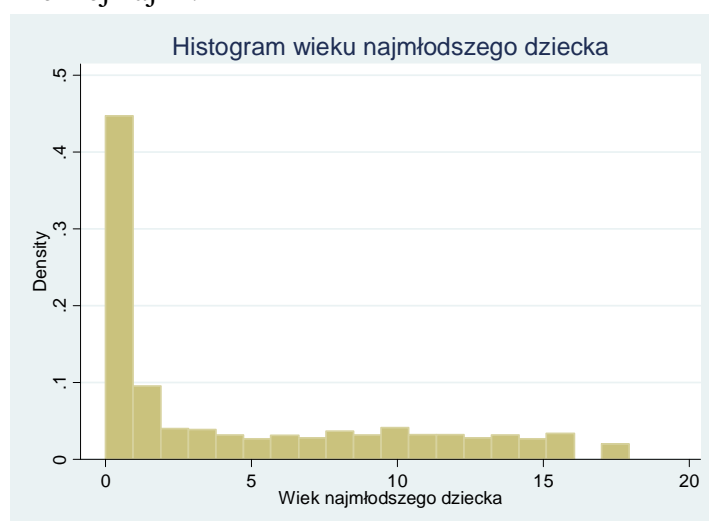
Jak widać większość kobiet zarabia za godzinę od 9 do 25 dolarów. Wykres również nie jest symetryczny co potwierdza mediana równa 14 dolarów. Płaca powyżej 50 dolarów jest raczej rzadko spotykana.

Rys 6. Histogram zmiennej dzieci.



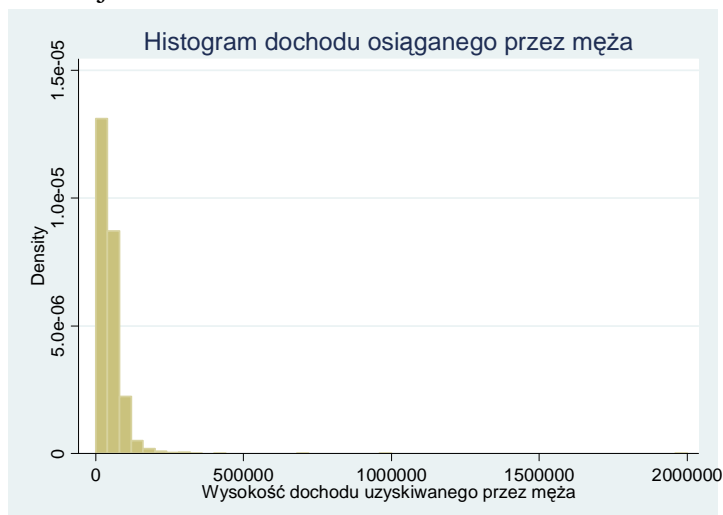
Jak możemy odczytać z histogramu zdecydowana większość kobiet nie posiada dzieci. Jeśli już ma dzieci to albo 1 albo 2. Rzadko kobiety decydują się na więcej niż 3 dzieci, a w naszej próbie żadna kobieta nie ma więcej niż 6 dzieci.

Rys 7. Histogram zmiennej najml.



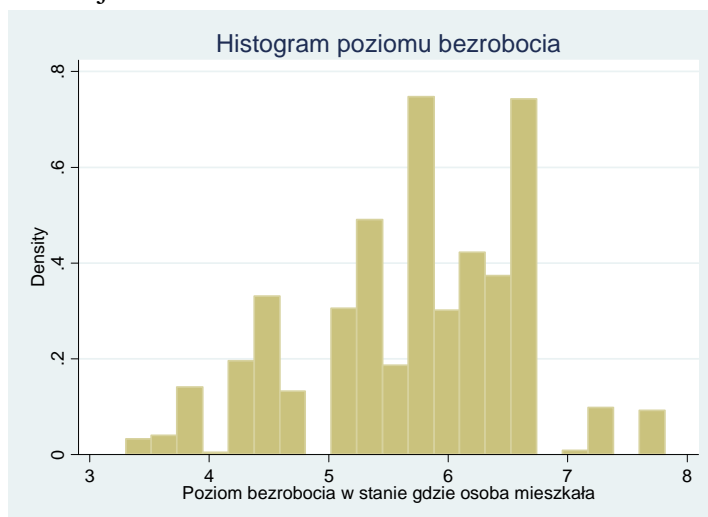
Histogram pokazuje, iż najczęściej występującą wartością jest 0 co wynika z tego, iż większość kobiet nie posiada dzieci. Jednak drugą najczęściej występującą wartością jest 1 co oznacza, że większość kobiet ma najmłodsze dziecko w wieku 1 roku. Rozkład pozostałych wartości jest mniej więcej taki sam.

Rys 8. Histogram zmiennej dochm



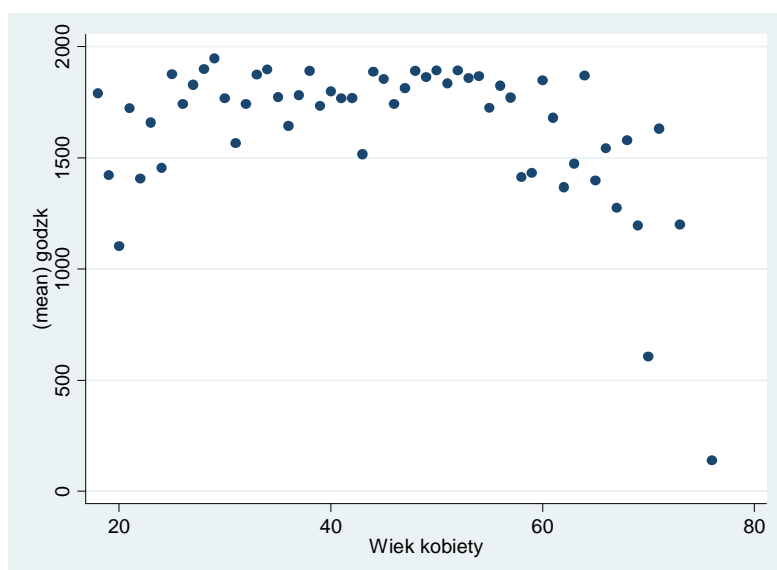
Minimalny dochód uzyskany przez męża wyniósł 300.04 zaś maksymalny, aż 1999998.5. Histogram dochodu uzyskiwanego przez męża jest mało pomocny w analizie tej zmiennej, ponieważ w tej próbie występuje kilka obserwacji, które powodują rozszerzenie zakresu zmiennej.

Rys 9. Histogram zmiennej bezrob



Zmienna bezrobocie przyjmowała wartości z przedziału (3.3,7.6). Rozkład jest w miarę symetryczny, co potwierdza porównanie mediany i średniej. Mediana równa 5.7, podczas gdy średnia 5.70779.

Rys 10. Wykres średniego poziomu przepracowanych godzin przez kobietę w zależności od jej wieku.



Powyższy wykres sugeruje, iż zależność między średnim poziomem podaży pracy a wiekiem jest raczej kwadratowa niż liniowa. Jedynie prawa strona wykresu odbiega od wyglądu funkcji kwadratowej co może być przyczyną tego, iż osoby powyżej 65 roku życia stanowiły mały odsetek badanych osób i ich podaż pracy była silnie zróżnicowana. Skłoniło nas to do uwzględnienia w modelu zmiennej wiek podniesionej do kwadratu.

7. Przewidywana forma funkcyjna

boxcox godz. k dochm wiekk wiekk_2 placak dosw edu dzieci najml bezrob, model(lhsonly)

Oszacowanie parametru w przekształceniu Boxa-Coxa wynosi 1,0168 – wartość bliska 1, co wskazuje na model ze niezlogarytmowana zmienna dochód. Hipotezy zerowe $H_0 : \theta = 0$ oraz $H_0 : \theta = -1$ zostają odrzucone (p -value = 0,000, czyli mniejsze od przyjętego poziomu istotności na poziomie 0,05), natomiast brak jest podstaw do odrzucenia $H_0 : \theta = 1$ (p -value = 0,596).

Log likelihood = **-16331.84** Number of obs = **2067**
 LR chi2(9) = **138.30**
 Prob > chi2 = **0.000**

godzk	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
/theta	1.016833	.0318122	31.96	0.000	.9544824 1.079184

Estimates of scale-variant parameters

	Coef.
Notrans	
dochm	-.0006151
wiekk	65.19556
wiekk_2	-.8860401
placak	-5.565531
dosw	12.20733
edu	24.59425
dzieci	-142.5449
najml	7.67656
bezrob	-39.69601
_cons	895.4973
/sigma	739.6263

Test H0:	Restricted log likelihood	LR statistic chi2	P-value Prob > chi2
theta = -1	-22583.114	12502.55	0.000
theta = 0	-17109.667	1555.65	0.000
theta = 1	-16331.981	0.28	0.596

Formalny test Boxa-coxa przeprowadzony dla zmiennej zależnej wskazał na to, iż zmienna godzk powinna zostać niezlogarytmowana.

Rozpatrzmy jeszcze model, w którym przekształceniu poddamy dochm:

```
boxcox godzk dochm, notrans(wiekk wiekk_2 placak dosw edu dzieci najml
bezrob) model(rhsonly)
```

Log likelihood = **-16322.27**

Number of obs = **2067**
 LR chi2(10) = **158.69**
 Prob > chi2 = **0.000**

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
godzk						
/lambda	.2636564	.1570355	1.68	0.093	-.0441275	.5714403

Estimates of scale-variant parameters

	Coef.
Notrans	
wiekk	64.81334
wiekk_2	-.8614755
placak	-4.312732
dosw	10.19509
edu	26.73237
dzieci	-129.1171
najml	7.101206
bezrob	-35.04259
_cons	934.8943
Trans	
dochm	-6.387792
/sigma	650.4475

Test H0:	Restricted log likelihood	LR statistic chi2	P-value Prob > chi2
lambda = -1	-16332.446	20.35	0.000
lambda = 0	-16323.392	2.24	0.134
lambda = 1	-16331.981	19.42	0.000

Oszacowanie parametru wynosi 0,26365 a więc zbliżone jest do zera, co wskazuje na model ze logarytmowaną zmienną dochm (p-value przy teście mówiącym o tym, że zmienna powinna zostać logarytmowana wynosi 0,134, dlatego nie mamy podstaw do odrzucenia tej hipotezy).

Poniżej przedstawiamy zakładany model końcowy:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i} + \beta_8 x_{8i} + \beta_9 x_{9i} + \beta_{10} x_{10i} + \beta_{11} x_{11i} + \beta_{12} x_{12i} + \beta_{13} x_{13i} + \beta_{14} x_{14i} + \varepsilon_i$$

Postać analityczna szacowanego równania regresji (wraz z przewidywanymi znakami współczynników):

$$\begin{aligned} \text{godzk} = & \beta_0 + \overset{(+)}{\beta_1} \overset{(-)}{\text{wiek}} + \overset{(+)}{\beta_2} \overset{(+)}{\text{wiek_2}} + \overset{(-)}{\beta_3} \overset{(-)}{\text{dosw}} + \overset{(-)}{\beta_4} \overset{(-)}{\text{edu}} + \overset{(+)}{\beta_5} \overset{(-)}{\text{placak}} + \overset{(-)}{\beta_6} \overset{(-)}{\text{dzieci}} + \overset{(+)}{\beta_7} \overset{(-)}{\text{najml}} + \\ & \overset{(-)}{\beta_8} \overset{(-)}{\text{bezrob}} + \overset{(-)}{\beta_9} \ln_ \overset{(-)}{\text{dochm}} + \overset{(-)}{\beta_{10}} \overset{(-)}{\text{miasto2}} + \overset{(-)}{\beta_{11}} \overset{(-)}{\text{miasto3}} + \overset{(-)}{\beta_{12}} \overset{(-)}{\text{miasto4}} + \overset{(-)}{\beta_{13}} \overset{(-)}{\text{miasto5}} + \overset{(-)}{\beta_{14}} \overset{(-)}{\text{miasto6}} + \varepsilon_i \end{aligned}$$

8. Estymacja parametrów modelu

Regresja przeprowadzona została na wszystkich zmiennych w modelu (dochm zlogarytmowany) oraz wraz ze zmienną `wiekk_2`, która odpowiada zmiennej `wiek` podniesionej do kwadratu (załącznik nr 2 oraz nr 3).

```
. regress godzk wiekk wiekk_2 placak dosw edu dzieci bezrob ln_dochm najml miasto3 miasto2 miasto4 miasto5 miasto6
```

Source	SS	df	MS	Number of obs = 2067		
Model	71205931.5	14	5086137.97	F(14, 2052) = 11.95		
Residual	873089820	2052	425482.368	Prob > F = 0.0000		
				R-squared = 0.0754		
				Adj R-squared = 0.0691		
				Root MSE = 652.29		
Total	944295751	2066	457064.739			

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	65.98868	10.71225	6.16	0.000	44.98066	86.9967
wiekk_2	-.877621	.1295592	-6.77	0.000	-1.131702	-.6235398
placak	-4.656101	1.348854	-3.45	0.001	-7.301366	-2.010836
dosw	10.03373	2.353142	4.26	0.000	5.418931	14.64852
edu	25.41549	6.638749	3.83	0.000	12.3961	38.43488
dzieci	-128.9651	16.21538	-7.95	0.000	-160.7654	-97.16475
bezrob	-43.82491	16.79024	-2.61	0.009	-76.7526	-10.89723
ln_dochm	-96.62067	19.21937	-5.03	0.000	-134.3122	-58.92917
najml	6.562027	3.171645	2.07	0.039	.3420486	12.782
miasto3	-104.1392	56.77388	-1.83	0.067	-215.4796	7.201222
miasto2	-42.56881	47.59746	-0.89	0.371	-135.9132	50.77557
miasto4	-76.77097	55.4419	-1.38	0.166	-185.4992	31.95729
miasto5	-68.89118	52.22642	-1.32	0.187	-171.3135	33.53113
miasto6	-104.1753	52.01117	-2.00	0.045	-206.1755	-2.175126
_cons	1701.996	285.413	5.96	0.000	1142.267	2261.726

Test na łączną istotność równania regresji wskazuje, że nie odrzucamy hipotezy zerowej o nieistotności równania regresji (ponieważ p -value = 0,0000 jest mniejsze od przyjętego poziomu istotności 0,05).

Pomimo łącznej istotności zmiennych występują w modelu zmienne nieistotne: `miasto2` (p -value=0,371), `miasto3` (p -value = 0,067), `miasto4` (p -value = 0,166), `miasto5` (p -value = 0,187). Tylko `miasto6` z pozostałych zmiennych binarnych dotyczących rozkodowanego miasta jest istotne na poziomie istotności 0,05 (p -value = 0,045).

Pozostałe zmienne w modelu są istotne na poziomie istotności 0,05.

Zaczynamy od przetestowania łącznej istotności zmiennych zerojedynkowych dotyczących miejsca zamieszkania.

```
. test miasto2 miasto3 miasto4 miasto5 miasto6
```

```
( 1) miasto2 = 0
( 2) miasto3 = 0
( 3) miasto4 = 0
( 4) miasto5 = 0
( 5) miasto6 = 0
```

```
F( 5, 2052) = 1.11
Prob > F = 0.3506
```

Nie odrzucamy hipotezy o łącznej nieistotności tych zmiennych, ponieważ p -value wyniósł 0,3506 (czyli są łącznie nieistotne).

Próbowałyśmy tak przekształcać zmienne binarne `miasto1/6` (łączyć w różne poziomy, dodając interakcje) aby odnaleźć jakieś zmienne istotne, jednak zawsze wychodziły nieistotne w modelu (odpowiednie przekształcenia znajdują się w załączniku nr 1).

Source	SS	df	MS	Number of obs = 2067		
Model	69486432.5	10	6948643.25	F(10, 2056) = 16.33		
Residual	874809319	2056	425490.914	Prob > F = 0.0000		
				R-squared = 0.0736		
				Adj R-squared = 0.0691		
				Root MSE = 652.3		
Total	944295751	2066	457064.739			

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	65.542	10.70757	6.12	0.000	44.54318	86.54082
wiekk_2	-.873999	.1295145	-6.75	0.000	-1.127992	-.6200057
placak	-4.610038	1.347382	-3.42	0.001	-7.252413	-1.967663
dosw	10.21818	2.348192	4.35	0.000	5.613097	14.82326
edu	25.29691	6.630156	3.82	0.000	12.29439	38.29943
dzieci	-130.0424	16.19361	-8.03	0.000	-161.8	-98.28478
bezrob	-39.29146	16.60557	-2.37	0.018	-71.85695	-6.725964
ln_dochm	-96.76438	19.19334	-5.04	0.000	-134.4048	-59.12397
najml	6.808884	3.167374	2.15	0.032	.597287	13.02048
miasto6	-46.83501	37.86676	-1.24	0.216	-121.0962	27.42618
_cons	1632.887	280.8945	5.81	0.000	1082.019	2183.754

Po wyrzuceniu usunięciu zmiennych związanych z miastem oprócz miasta , które wyszło na początku istotne, w końcowej estymacji również wyszło nieistotne (p-value = 0,079). Dlatego postanowiliśmy usunąć zmienne dotyczące miasta z modelu. Pomimo tego, że usunięcie zmiennych pogorszy jakość dopasowania modelu, to usunięcie zmiennych nieistotnych poprawia dokładność oszacowań parametrów przy zmiennych istotnych.

Wyniki estymacji po usunięciu zmiennych nieistotnych przedstawiają się następująco:

. regress godzk wiekk wiekk_2 placak dosw edu dzieci bezrob ln_dochm najml

Source	SS	df	MS	Number of obs = 2067		
Model	68835530.9	9	7648392.32	F(9, 2057) = 17.97		
Residual	875460220	2057	425600.496	Prob > F = 0.0000		
				R-squared = 0.0729		
				Adj R-squared = 0.0688		
				Root MSE = 652.38		
Total	944295751	2066	457064.739			

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	65.03274	10.70103	6.08	0.000	44.04675	86.01872
wiekk_2	-.8681008	.1294434	-6.71	0.000	-1.121954	-.6142471
placak	-4.474517	1.343092	-3.33	0.001	-7.108479	-1.840556
dosw	10.2755	2.348037	4.38	0.000	5.670727	14.88028
edu	25.70963	6.622606	3.88	0.000	12.72192	38.69735
dzieci	-130.0112	16.19568	-8.03	0.000	-161.7728	-98.24954
bezrob	-34.99069	16.23951	-2.15	0.031	-66.83829	-3.143089
ln_dochm	-95.68548	19.17597	-4.99	0.000	-133.2918	-58.07913
najml	6.937126	3.166085	2.19	0.029	.7280611	13.14619
_cons	1589.337	278.7149	5.70	0.000	1042.745	2135.93

9. Problemy ze zbiorem danych

Współliniowość
`. estat vif`

Variable	VIF	1/VIF
wiekk_2	58.43	0.017115
wiekk	58.18	0.017189
dzieci	1.53	0.653906
dosw	1.50	0.666264
najml	1.47	0.678710
edu	1.20	0.835670
placak	1.20	0.835872
ln_dochm	1.14	0.879611
bezrob	1.01	0.995005
Mean VIF	13.96	

Widzimy, że zmienne `wiek` oraz `wiekk_2` cechują się współliniowością, co nie powinno dziwić, bowiem `wiekk_2` jest wszakże wiekiem podniesionym do kwadratu. Inne wartości `vif`ów są małe, dlatego między nimi nie występuje współliniowość.

Standardowym sposobem rozwiązania problemu współliniowości jest usunięcie zmiennej o najwyższym VIF (w tym przypadku jest to `wiekk_2`). Jednak jest to zmienna istotna, tak więc może spowodować obciążenie estymatora zmiennej `wiek`, dlatego nie usunęliśmy ani zmiennej `wiek`, ani zmiennej `wiekk_2`.

Obserwacje nietypowe

Kolejnym naszym krokiem było zbadanie obserwacji nietypowych, których w modelu występuje jedenaście. Jednak po dokładnej analizie każdego przypadku doszliśmy do wniosku, że nie możemy wyrzucić żadnej z tych obserwacji, ponieważ pogorszyłoby to dopasowanie modelu.

```
. list godzk wiekk wiekk_2 placak dosw edu dzieci bezrob ln_dochm najml dzwignia
> a reszty_st cook_dist if dzwignia > 2*e(df_m)/e(N)
> & abs(reszty_st)>2;
```

2.	godzk 56	wiekk 37	wiekk_2 1369	placak 4.46	dosw 18	edu 3	dzieci 0	bezrob 6.7
	ln_dochm 10.08581	najml 0	dzwignia .0139997	reszty_st -2.768992	cook_d~t .0108864			

Spójrzmy na obserwację drugą, w której płaca kobiety kształtuje się na poziomie 4,46, czyli poniżej płacy minimalnej (5,15 w roku 2002). Jest to jednak możliwe otrzymywanie płacy poniżej płacy minimalnej (choć nielegalne), dlatego nie możemy usunąć tej obserwacji.

77.	godzk 575	wiekk 45	wiekk_2 2025	placak 8.35	dosw 25	edu 6	dzieci 1	bezrob 6.7
	ln_dochm 9.431883	najml 2	dzwignia .0092268	reszty_st -2.007535	cook_d~t .0037532			

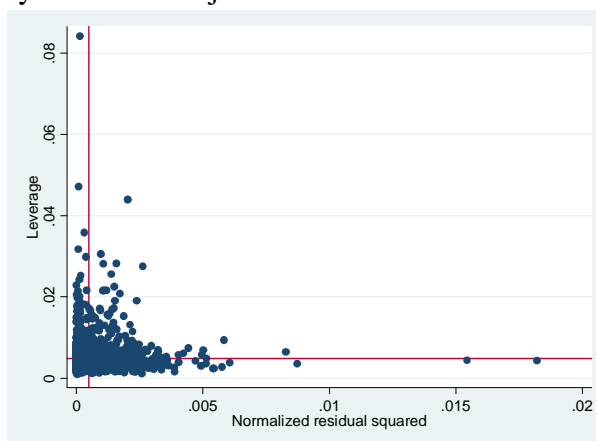
Kolejną, ciekawą obserwacją jest obserwacja nr 77, w której to kobieta mająca 45 lat ma dwuletnie dziecko. Jest to możliwe, dlatego również nie możemy usunąć tej obserwacji z próbki.

2062.	godzk 3528	wiekk 28	wiekk_2 784	placak 7.5	dosw 6	edu 11	dzieci 5	bezrob 5.7
	ln_dochm 10.04447	najml 6	dzwignia .009371	reszty_st 3.475078	cook_d~t .0114237			

Z odwrotną sytuacją mamy do czynienia w obserwacji 2062, gdzie młoda, 28 letnia kobieta ma pięcioro dzieci, z których najmłodsze ma 6 lat. Jest to jednak możliwe, dlatego nie powinno wyrzucać się tej obserwacji z próbki. Jest ona nietypowa, ale nie błędna z punktu widzenia tego modelu (model nie potrafi wytłumaczyć dlaczego jest nietypowa).

. lvr2plot

Rys 11. Dźwignie i reszty dla obserwacji w modelu



Tak jak już wcześniej wspomniałam, żadnej z wymienionych powyżej obserwacji nie powinno się usunąć, co zobaczyć można na powyższym wykresie. Widzimy, że żadna z obserwacji nie znalazła się w prawej-górnej części wykresu. W związku z tym upewniliśmy się, że nie mamy obserwacji, które znacząco wpływałyby na precyzję oszacowań parametrów.

10. Diagnostyka modelu

Test RESET na poprawność formy funkcyjnej

ovtest

Ramsey RESET test using powers of the fitted values of godzk

Ho: model has no omitted variables

$F(3, 2054) = 1.29$

$\text{Prob} > F = 0.2754$

$\text{Prob} > F = 0.2754 > 0.05$ zatem nie ma podstaw do odrzucenia hipotezy zerowej zakładającej poprawność formy funkcyjnej. Stąd wnioskujemy, że forma funkcyjna zastosowana w modelu jest poprawna, w związku z tym estymator MNK w zbudowanym modelu ma porządane własności.

Test BREUSCHA-PAGANA na heteroskedastyczność

. hettest, rhs

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: wiekk wiekk_2 placak dosw edu dzieci najml bezrob ln_dochm

chi2(9) = 30.89
 Prob > chi2 = 0.0003

Prob>chi2 = 0.0003<0.05 zatem odrzucamy hipotezę zerową zakładającą homoskedastyczności. Stąd wnioskujemy, że składnik losowy jest heteroskedastyczny. Potwierdzeniem tego testu jest kolejny test: White'a:

Test WHITE'A

White's test for Ho: homoskedasticity
 against Ha: unrestricted heteroskedasticity

chi2(53) = 113.14
 Prob > chi2 = 0.0000

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	113.14	53	0.0000
Skewness	16.43	9	0.0584
Kurtosis	7.14	1	0.0076
Total	136.71	63	0.0000

Potwierdza on poprzednie spostrzeżenia wyciągnięte na podstawie testu BREUSCHA-PAGANA, iż w naszym modelu występuje heteroskedastyczność. P-value = 0.000 < 0.05 czyli odrzucamy hipotezę zerową o homoskedastyczności. Aby poprawić złą postać macierzy wariancji – kowariancji estymatora b, wynikającą z występowania heteroskedastyczności zastosowałyśmy "odporną" macierz wariancji-kowariancji White'a (co dało poprawę wyników oszacowań parametru b):

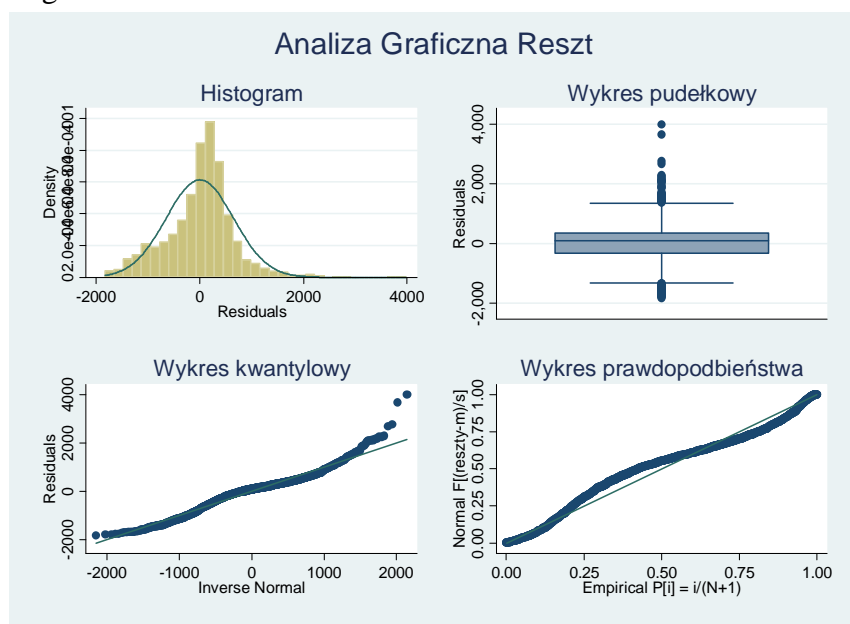
Linear regression

Number of obs = 2067
 F(9, 2057) = 16.98
 Prob > F = 0.0000
 R-squared = 0.0729
 Root MSE = 652.38

godzk	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	65.03274	10.39698	6.25	0.000	44.64302	85.42245
wiekk_2	-.8681008	.1252601	-6.93	0.000	-1.113751	-.6224511
placak	-4.474517	1.773332	-2.52	0.012	-7.95223	-.9968046
dosw	10.2755	2.324182	4.42	0.000	5.717508	14.8335
edu	25.70963	7.561821	3.40	0.001	10.88001	40.53926
dzieci	-130.0112	17.30263	-7.51	0.000	-163.9437	-96.07866
najml	6.937126	3.051561	2.27	0.023	.9526558	12.9216
bezrob	-34.99069	16.54234	-2.12	0.035	-67.43216	-2.549219
ln_dochm	-95.68548	20.20801	-4.74	0.000	-135.3158	-56.05519
_cons	1589.337	295.5078	5.38	0.000	1009.812	2168.863

Oszacowania parametrów pozostają takie same, jednak błędy standardowe uległy nieznacznej poprawie.

Rys 12. Analiza graficzna reszt

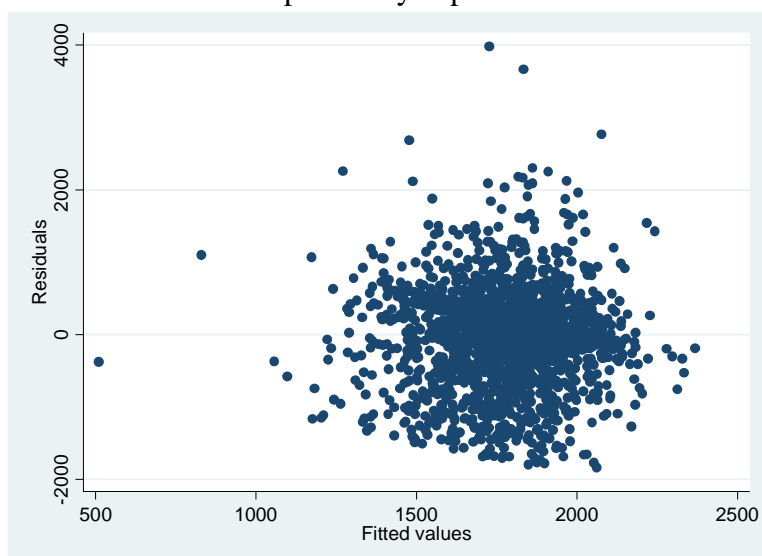


Histogram rozkładu reszt na tle rozkładu normalnego, pokazują, że reszty nie mają rozkładu normalnego (wskazuje również na silną asymetrię prawostronną). Wykres pudełkowy wskazuje na dość dużą liczbę obserwacji nietypowych (jeśli chodzi o reszty). Wykres kwantylowy wskazuje natomiast że pojawiają się odchylenia od wartości z rozkładu normalnego, co świadczy o tym, że ogony nie są dopasowane do rozkładu normalnego. Ostatni, czwarty wykres również wskazuje, że pojawiają się odchylenia od wartości pochodzących z rozkładu normalnego (świadczy o tym środek rozkładu niedopasowany rozkładu reszt).

Przeprowadzony formalny test Jarque-Berra potwierdził to, co widać na rysunkach – zaburzenie losowe nie ma rozkładu normalnego ($p\text{-value} = 0,000$, dlatego odrzucamy hipotezę zerową o normalności zaburzenia losowego).

Rozkłady testów t i F zostały wyprowadzone przy założeniu normalności zaburzenia losowego. Brak normalności rozkładu błędów losowych nie jest w przypadku dużych prób zbyt istotny, ponieważ rozkłady statystyk w takich próbach są bliskie standardowym rozkładom. Czyli duża liczebność próby gwarantuje nam, że przeprowadzone wnioskowanie statystyczne jest poprawne.

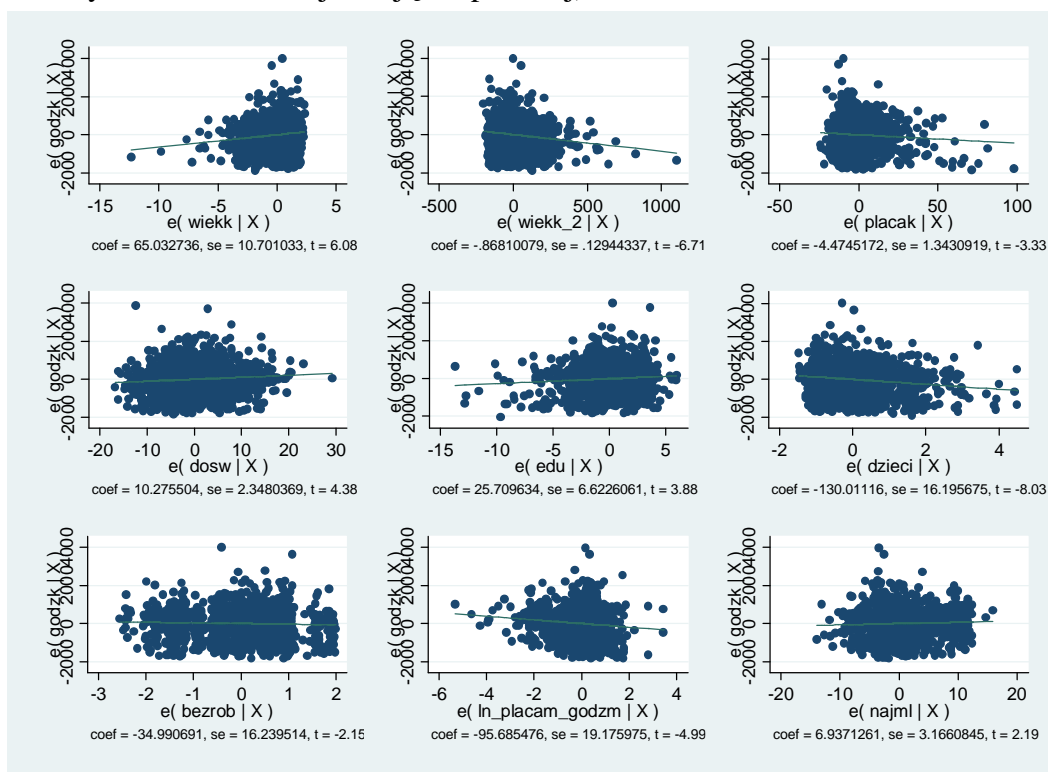
Rys 13. Zależność reszt od wartości dopasowanych przez model



Z wykresu wynika, iż w wyestymowanym przez nas modelu nie występuje zależność pomiędzy resztami a wartością dopasowaną przez model.

Istotność zmiennych w modelu

Rys 14. Wykres reszt z regresji liniowej y na X vs reszty z regresji liniowej x_j na X (X zawiera wszystkie zmienne objaśniające oprócz x_j)



W przypadku wszystkich przedstawionych tutaj zmiennych, nachylenie jest różne od zera, co świadczy o istotności każdej zmiennej.

11. Interpretacja wyników estymacji

Tak jak już wcześniej zostało pokazane, zmiennymi istotnymi w modelu są: wiek, wiek₂, placak, dosw, edu, dzieci, najml, ln_dochm, bezrob, natomiast nieistotną zmienną okazało się miasto. Są to zmienne istotne łącznie, a model zbudowany wyjaśnia zmienność zmiennej zależnej w 7,29%.

Wyestymowane parametry można zinterpretować jako:

Wiek - Wzrost wieku kobiety o rok powoduje (średnio rzecz biorąc) wzrost liczby przepracowanych godzin w ciągu roku o 65, *ceteris paribus*.

Płaca - Wzrost stawki godzinowej o 1 dolara powoduje (średnio rzecz biorąc) spadek liczby przepracowanych godzin w ciągu roku o 4,5, *ceteris paribus*.

Doświadczenie - Wzrost doświadczenia kobiety o 1 rok powoduje (średnio rzecz biorąc) wzrost liczby przepracowanych godzin w ciągu roku o 10, *ceteris paribus*.

Liczba lat przeznaczonych na edukację - Wzrost liczby lat poświęconych na edukację o 1 rok powoduje (średnio rzecz biorąc) wzrost liczby przepracowanych godzin w ciągu roku o 25,7, *ceteris paribus*.

Dzieci – Wzrost liczby dzieci o jeden, powoduje oczekiwany (średni) spadek liczby przepracowanych przez kobietę godzin o około 130 w ciągu roku (przy pozostałych zmiennych niezmiennych).

Bezrob – Wzrost stopy bezrobocia o jeden punkt procentowy powoduje spadek oczekiwanej (średniej) liczby godzin przepracowanych przez kobietę o 35 w ciągu roku, założenie *ceteris paribus*.

Ln_dochm – tego parametru nie można tak wprost zinterpretować, ponieważ zmienna zależna nie jest logarytmowana. Parametr ten o wartości -95.68548 wskazuje nam tylko kierunek zmian, czyli gdy mężczyzna zarabia więcej, kobieta mniej pracuje.

Najmł – Wzrost wieku najmłodszego dziecka o jeden, powoduje wzrost podaży pracy kobiety o, średnio rzecz biorąc, 6.937126 godziny w ciągu roku, *ceteris paribus*.

Stała – nie jest interpretowalna.

12. Interpretacja ekonomiczna zbudowanego modelu - podsumowanie

W naszym modelu udało się dowiedzieć, że istnieją pewne determinanty, za pomocą których możemy wyjaśnić ilość przepracowanych godzin przez kobiety. Niestety nie wszystkie postawione przez nas we wstępie hipotezy okazały się być prawdziwe.

Tak jak zakładaliśmy na samym początku, okazało się, że płaca, którą kobieta otrzymuje, wpływa na ilość godzin przez nią przepracowanych. W tym przypadku w większym stopniu zadziałał efekt substytucyjny, bowiem jak się okazało, wraz ze wzrostem płacy, kobiety coraz mniej pracują. Jest to swoistym potwierdzeniem hipotezy, że większość kobiet nie pracuje po to, żeby spełniać swoje ambicje. Pracuje po to, żeby mieć określony dochód. Jeśli określony, pożądany dochód może być otrzymany poprzez wyższą płacę, to nie pracują więcej. Po otrzymaniu tego dochodu, cenniejszy staje się czas spędzony z rodziną.

Chcieliśmy pokazać również wpływ dochodu uzyskanego przez męża na aktywność kobiety na rynku pracy. Konstrukcja naszego modelu nie pozwoliła nam ocenić siły tej zależności, wiemy jednak że jest to zależność istotna, oraz że ma charakter ujemny. Gdy mężczyzna więcej zarabia, kobieta mniej pracuje. To kolejny dowód na to, że tak większy dochód rodziny ma charakter substytucyjny, czyli pozwala na ograniczenie pracy na rzecz czasu wolnego spędzonego w domu.

Wyniki, które otrzymaliśmy za pomocą regresji liniowej potwierdzają hipotezę, iż zmienna wiek wpływa dodatnio na podaż pracy kobiety. Potwierdziły się więc w naszym badaniu tezy o wpływie wieku na decyzje kobiet o aktywności na rynku pracy stawiane przez wielu autorów artykułów (np. Smith, Ward 1984). Postawiona przez nas we wstępie hipoteza zakładała, iż kobiety na początku swojej kariery pracują mało ze względu posiadania małych dzieci (przeznaczają dużo czasu nad opiekę nad nimi) kolejno ilość przepracowanych godzin zaczyna wzrastać, aby w wieku ok. 45 lat ustabilizować się a następnie spada. Zyskało to potwierdzenie w naszym modelu poprzez wprowadzenie zmiennej wiek podniesionej do kwadratu (jak się okazało istotnej). Zgodnie z uzyskanymi przez nas oszacowaniami zależność między przepracowanymi godzinami a wiekiem ma postać funkcji kwadratowej (na początku rośnie, następnie stabilizuje się aby potem spaść).

Wyniki badań przeprowadzony między innymi przez Heer i Grossbard-Shechtmana (1981) dotyczące wpływu edukacji na liczbę przepracowanych godzin przez zamężne kobiety, również znalazły potwierdzenie w naszym modelu ekonometrycznym. Kobieta, która więcej lat poświęciła na zdobywanie wykształcenia, czyli *de facto* odznacza się wyższym poziomem wykształcenia, więcej pracuje. Wydaje nam się, że ma na to wpływ

ambicja oraz determinacja w dążeniu do celu (te kobiety, które były w stanie poświęcić 16 lat swojego życia na zdobycie wykształcenia, są bardziej zdeterminowane i ambitne od tych, które zadowolili się wykształceniem podstawowym). Zdobywanie wykształcenia jest również swoistą inwestycją, dlatego, ta kobieta, która zainwestowała w edukację, nie będzie marnowała swoich umiejętności pozostając nieaktywna na rynku pracy. To właśnie te czynniki kształtują według nas dodatni wpływ edukacji na liczbę przepracowanych godzin przez zamężne kobiety.

Kolejną postawioną przez nas hipotezą był wpływ doświadczenia zawodowego na ilość przepracowanych godzin. W odniesieniu do wyników naszej estymacji udowodniłyśmy, iż liczba lat przepracowanych w karierze zawodowej ma znaczenie. Ma to swoje uzasadnienie w oparciu o fakt, iż człowiek, który odznacza się wysokim doświadczeniem zawodowym, jest bardziej ceniony przez pracodawcę, ma rozwinięte zdolności niezbędne do efektywnej pracy, a także zna rynek pracy i zasady panujące na nim. Są to zalety, które niejako procentują na korzyść pracownika i jego wartość na rynku pracy. Ma tu również znaczenie fakt, że człowiek, który pracował już jakiś czas, czyli otrzymywał własne dochody, spotykał się na co dzień z ludźmi, nie jest w stanie tak łatwo zrezygnować z pracy, czy zredukować liczbę przepracowanych godzin.

Kolejną przedstawioną przez nas hipotezą była hipoteza odnośnie istotności wpływu ilości posiadanych dzieci. Dzięki oszacowaniom naszego modelu udało nam się potwierdzić zależność, iż kobiety posiadające większą ilość dzieci poświęcają im więcej czasu redukując jednocześnie ilość godzin poświęconych na pracę zawodową. Wnioski te znajdują potwierdzenie w niektórych z przytoczonych przez nas artykułów.

Niejako z poprzednią zależnością jest związana zależność pomiędzy wiekiem najmłodszego dziecka a podażą pracy kobiety. Udało nam się potwierdzić wstępną hipotezę, iż kobiety mające małe dzieci pracują mniej. Wiąże się to zapewne z faktem, że dzieci w swoim dzieciństwie (mamy tu na myśli wiek do 5 lat) potrzebują więcej czasu, uwagi. Absorbują one więcej czasu niż dzieci „odchowane”, które można zapisać do przedszkola lub te które chodzą już do szkoły i stają się już w pewnym stopniu samodzielne. Problem ten jest rzadko omawiany w literaturze dlatego też opieramy się głównie na naszych przekonaniach i wynikach estymacji w naszym modelu.

Kolejnym czynnikiem ekonomicznym, który w istotny sposób wpływa na badaną zależność jest też stopa bezrobocia. Potwierdziły się więc nasze hipotezy odnośnie istotności tej zmiennej, a co więcej również kierunek tej zmiennej został potwierdzony naszym badaniem. Wzrost stopy bezrobocia powoduje bowiem spadek przepracowanej przez kobiety liczby godzin. Związane jest to pewnie z tym, że kobietom w czasie wysokiego bezrobocia, trudno znaleźć pracę (trudniej niż mężczyznom). Jest to jeden z przejawów dyskryminacji płciowej na rynku pracy.

Jedyną zmienną, którą usunęłyśmy z modelu była zmienna określająca miejsce zamieszkania badanych kobiet. Okazało się, że wielkość miasta w którym mieszka kobieta, nie wpływa na jej decyzje o liczbie przepracowanych godzin.

Podsumowując nasze wnioski, możemy stwierdzić iż model dowodzi, że dokonywanie decyzji o tym czy przepracować 10 czy 200 godzin w roku przez zamężne kobiety w USA nie jest procesem całkowicie losowym, a liczba godzin przepracowanych przez nie w sposób istotny zależy od zmiennych ekonomicznych, demograficznych a przede wszystkim rodzinnych.

13. Bibliografia

Dane:

1. PSID: <http://www.psidonline.org>
2. Dane odnośnie bezrobocia: <http://www.bls.gov/lau/lastrk02.htm>

Literatura:

1. Heer D. M., Grossbard-Shechtman A., *The Impact of the Female Marriage Squeeze and the Contraceptive Revolution on Sex Roles and the Women's Liberation Movement in the United States, 1960 to 1975* [w:] *Journal of Marriage and the Family*, rok 1981 nr 43, s. 49-65
2. Leibowitz A., Klerman J. A., *Explaining Changes in Married Mothers Employment Over time* [w:] *Demography*, rok 1995 nr 32, s. 365-370.
3. Mincer J. *Labour Force Participation of Married Women: a Study of Labour Supply*, [w:] *Aspects of Labour Economics* pod. Red. H. Gregg Lewis, Princeton University Press, 1962.
4. Rosen. S *Mincering Labor Economics*, [w:] *Journal of Economic Perspectives*, rok 1992 nr 6, s. 157-170.
5. Smith, J. P., Ward M. P., *Women's Wages and Work in the Twentieth Century*, Rand Publications, 1984

Załączniki

a) Analiza zmiennej miasto

Oprócz przeprowadzonej powyżej analizy zmiennej miasto rozkodowanej na zmienna zerojedynkowa przeprowadziłyśmy jeszcze inne analizy, które przedstawiamy poniżej.

Kolejnym naszym krokiem było zakodowanie jej na dwa poziomy miasto1 i miasto2, gdzie są to zmienne zerojedynkowe. Miasto1 to obszar SMSA (Standard Metropolitan Statistical Area, czyli obszar metropolii), a miasto2 to obszar non-SMSA.

```
. regress godzk wiekk wiekk_2 placak dosw edu dzieci najml bezrob ln_dochm miasto_2
```

Source	SS	df	MS	Number of obs = 2067		
Model	69540828	10	6954082.8	F(10, 2056) =	16.34	
Residual	874754923	2056	425464.457	Prob > F =	0.0000	
				R-squared =	0.0736	
				Adj R-squared =	0.0691	
Total	944295751	2066	457064.739	Root MSE =	652.28	

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	65.43297	10.70384	6.11	0.000	44.44148	86.42447
wiekk_2	-.8720076	.1294582	-6.74	0.000	-1.125891	-.6181247
placak	-4.54467	1.343982	-3.38	0.001	-7.180378	-1.908962
dosw	10.10689	2.351311	4.30	0.000	5.495692	14.71809
edu	25.44757	6.624675	3.84	0.000	12.4558	38.43935
dzieci	-130.3293	16.19497	-8.05	0.000	-162.0896	-98.56905
najml	6.86347	3.166095	2.17	0.030	.6543816	13.07256
bezrob	-39.06701	16.54271	-2.36	0.018	-71.50922	-6.6248
ln_dochm	-96.84098	19.1939	-5.05	0.000	-134.4825	-59.19946
miasto_2	-37.89703	29.4341	-1.29	0.198	-95.62078	19.82673
_cons	1640.992	281.5434	5.83	0.000	1088.852	2193.132

```
. end of do-file
```

```
. ovtest
```

```
Ramsey RESET test using powers of the fitted values of godzk
Ho: model has no omitted variables
F(3, 2053) = 1.36
Prob > F = 0.2542
```

Przeprowadzając regresje z takimi zmiennymi nowo utworzona zmienna jest jednak nieistotna (p-value= 0.198). Sprawdziłyśmy również jak zachowuje się test RESET na poprawność formy funkcyjnej. Statystyka testowa nam wzrosła, więc sugeruje to nam, iż nie jest to najlepszy kierunek zmian.

Postanowiłyśmy również sprawdzić czy wpływ niezależnych zmiennych ciągłych na zmienna zależna zależy od poziomów przyjmowanych przez dyskretna zmienna, którą w naszym modelu jest zmienna miasto, poprzez wprowadzeniu do modelu interakcji.

Pierwszą interakcją, którą postanowiłyśmy przebadac jest wpływ miejsca zamieszkania na zależność między podażą pracy a stopą bezrobocia. Postanowiłyśmy wprowadzić taką interakcję, ponieważ sądziłyśmy, iż w obszarze metropolii duże bezrobocie nie musi mieć tak istotnego wpływu na ilość przepracowanych godzin, jak to może mieć miejsce w mniej uprzemysłowionych obszarach:

```
. xi: regress godzk wiekk wiekk_2 placak dosw edu dzieci najml ln_dochm i.miasto*bezrob
i.miasto _Imiasto_1-6 (naturally coded; _Imiasto_1 omitted)
i.miasto*bezrob _ImiaXbezro_# (coded as above)
```

Source	SS	df	MS	Number of obs =	2067
Model	72366823.2	19	3808780.17	F(19, 2047) =	8.94
Residual	871928928	2047	425954.533	Prob > F =	0.0000
Total	944295751	2066	457064.739	R-squared =	0.0766
				Adj R-squared =	0.0681
				Root MSE =	652.65

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	66.29382	10.73894	6.17	0.000	45.23344	87.35421
wiekk_2	-.8800362	.1299051	-6.77	0.000	-1.134796	-.6252762
placak	-4.688306	1.351572	-3.47	0.001	-7.338906	-2.037706
dosw	10.06976	2.3549	4.28	0.000	5.451515	14.68801
edu	25.4626	6.655522	3.83	0.000	12.4103	38.5149
dzieci	-127.7925	16.27598	-7.85	0.000	-159.7117	-95.87333
najml	6.628701	3.175098	2.09	0.037	.4019413	12.85546
ln_dochm	-97.16343	19.24046	-5.05	0.000	-134.8963	-59.43051
_Imiasto_2	290.3248	411.4856	0.71	0.481	-516.6492	1097.299
_Imiasto_3	14.47255	459.4609	0.03	0.975	-886.587	915.5321
_Imiasto_4	-196.8097	438.8218	-0.45	0.654	-1057.393	663.7741
_Imiasto_5	-102.3313	432.4428	-0.24	0.813	-950.405	745.7424
_Imiasto_6	-73.57209	407.0885	-0.18	0.857	-871.923	724.7788
bezrob	-30.4966	59.59429	-0.51	0.609	-147.3684	86.37516
_ImiaXbezr~2	-56.26363	68.21892	-0.82	0.410	-190.0494	77.5221
_ImiaXbezr~3	-20.10487	77.52728	-0.26	0.795	-172.1455	131.9357
_ImiaXbezr~4	21.82434	73.61461	0.30	0.767	-122.543	166.1917
_ImiaXbezr~5	6.835682	72.76901	0.09	0.925	-135.8733	149.5447
_ImiaXbezr~6	-4.013574	69.01807	-0.06	0.954	-139.3665	131.3394
_cons	1617.251	458.5011	3.53	0.000	718.0734	2516.428

```
. end of do-file
```

```
. ovtest
```

```
Ramsey RESET test using powers of the fitted values of godzk
```

```
Ho: model has no omitted variables
```

```
F(3, 2044) = 1.64
Prob > F = 0.1790
```

Otrzymane wyniki nadal mówią o nieistotności zmiennej miasto (p-value we wszystkich poziomach miasta nadal >0.05), a także wszystkich wprowadzonych interakcji (p-value >0.05).

Statystyka testowa znów wzrosła, więc to skłoniło nas do wniosku ze wprowadzone interakcje należy usunąć.

W ramach formalności przetestowałyśmy istotność wprowadzonych do modelu interakcji za pomocą funkcji test.

```
. test _ImiaXbezro_2 _ImiaXbezro_3 _ImiaXbezro_4 _ImiaXbezro_5 _ImiaXbezro_6
```

```
( 1) _ImiaXbezro_2 = 0
```

```
( 2) _ImiaXbezro_3 = 0
```

```
( 3) _ImiaXbezro_4 = 0
```

```
( 4) _ImiaXbezro_5 = 0
```

```
( 5) _ImiaXbezro_6 = 0
```

```
F( 5, 2047) = 0.55
```

```
Prob > F = 0.7422
```

Interakcje okazały się nieistotne, czyli miejsce zamieszkania o dziwo nie wpływa na zależność między podażą pracy kobiety a bezrobociem.

Kolejna interakcja to wpływ miejsca zamieszkania na zależność między podażą pracy a ilością posiadanych dzieci. Wydawało nam się, iż to że kobieta mieszka w zurbanizowanym okręgu ma wpływ na to czy ilość posiadanych dzieci wpływa na ilość przepracowanych godzin. Bowiem na obszarze metropolii łatwiej jest zorganizować opiekę dla swoich dzieci

np. znaleźć opiekunkę do dziecka. W bardziej rozwiniętych regionach sieć żłobków, przedszkoli czy świetlic dla dzieci jest bardziej rozwinięta, przez co kobieta spokojnie może poświęcić więcej czasu na pracę zawodową.

```
. xi: regress godzk wiekk wiekk_2 placak dosw edu najml bezrob ln_dochm i.miasto*dzieci
i.miasto _Imiasto_1-6 (naturally coded; _Imiasto_1 omitted)
i.miasto*dzieci _ImiaXdziec_# (coded as above)
```

Source	SS	df	MS	Number of obs =	2067
Model	73371514.9	19	3861658.68	F(19, 2047) =	9.08
Residual	870924236	2047	425463.721	Prob > F =	0.0000
				R-squared =	0.0777
				Adj R-squared =	0.0691
Total	944295751	2066	457064.739	Root MSE =	652.28

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	66.85479	10.74215	6.22	0.000	45.7881	87.92147
wiekk_2	-8858079	1298724	-6.82	0.000	-1.140504	-6.6311121
placak	-4.690667	1.351569	-3.47	0.001	-7.341261	-2.040074
dosw	10.03409	2.353872	4.26	0.000	5.417857	14.65032
edu	24.98721	6.65244	3.76	0.000	11.94095	38.03346
najml	6.423267	3.173539	2.02	0.043	1.1995653	12.64697
bezrob	-45.5155	16.84545	-2.70	0.007	-78.5515	-12.4795
ln_dochm	-97.88001	19.22736	-5.09	0.000	-135.5872	-60.17279
_Imiasto_2	-80.42377	66.41387	-1.21	0.226	-210.6696	49.82203
_Imiasto_3	-128.6971	80.64054	-1.60	0.111	-286.8432	29.44895
_Imiasto_4	-74.48053	76.46368	-0.97	0.330	-224.4352	75.47419
_Imiasto_5	-134.3883	73.88049	-1.82	0.069	-279.2771	10.50044
_Imiasto_6	-189.247	71.75328	-2.64	0.008	-329.9641	-48.52997
dzieci	-164.4527	36.52705	-4.50	0.000	-236.0868	-92.81864
_ImiaXdziec_2	35.11062	43.85643	0.80	0.423	-50.89726	121.1185
_ImiaXdziec_3	23.45425	50.89899	0.46	0.645	-76.36496	123.2735
_ImiaXdziec_4	-4.992838	50.75525	-0.10	0.922	-104.5301	94.54447
_ImiaXdziec_5	61.25467	49.23536	1.24	0.214	-35.30196	157.8113
_ImiaXdziec_6	79.663	46.3838	1.72	0.086	-11.30136	170.6274
_cons	1749.677	288.2576	6.07	0.000	1184.368	2314.986

end of do-file

. ovtest

```
Ramsey RESET test using powers of the fitted values of godzk
Ho: model has no omitted variables
F(3, 2044) = 1.07
Prob > F = 0.3597
```

Otrzymane wyniki znów przemawiają za nieistotnością prawie wszystkich poziomów zmiennej miasto (oprócz zmiennej miasto6: p-value= 0.008<0.05), jak i wszystkich wprowadzonych interakcji.

Wartość statystyki testowej spadła, jednak wprowadzone interakcje są nieistotne.

Test na łączną nieistotność wprowadzonych interakcji potwierdza te spostrzeżenia:

```
. test _ImiaXdziec_2 _ImiaXdziec_3 _ImiaXdziec_4 _ImiaXdziec_5 _ImiaXdziec_6
```

(1) _ImiaXdziec_2 = 0

(2) _ImiaXdziec_3 = 0

(3) _ImiaXdziec_4 = 0

(4) _ImiaXdziec_5 = 0

(5) _ImiaXdziec_6 = 0

F(5, 2047) = 1.02

Prob > F = 0.4053

Interakcje okazały się nieistotne, czyli miejsce zamieszkania nie wpływa na zależność między podażą pracy kobiety a ilością posiadanych dzieci.

Następna interakcja, która wiąże się niejako z poprzednią, to wpływ miejsca zamieszkania na zależność między podażą pracy a wiekiem najmłodszego dziecka. Zdecydowaliśmy się

przeprowadzić taką interakcję albowiem możliwe jest, iż kobiety w dużych miastach decydują się na mniejszą ilość dzieci, a co z tym się wiąże wiek najmłodszego dziecka w obszarach metropolii może być wyższy.

```
. xi: regress godzk wiekk wiekk_2 placak dosw edu dzieci bezrob ln_dochm i.miasto*najml
i.miasto _Imiasto_1-6 (naturally coded; _Imiasto_1 omitted)
i.miasto*najml _ImiaXnajml_# (coded as above)
```

Source	SS	df	MS	Number of obs =	2067
Model	72444783.3	19	3812883.33	F(19, 2047) =	8.95
Residual	871850968	2047	425916.448	Prob > F =	0.0000
				R-squared =	0.0767
				Adj R-squared =	0.0681
Total	944295751	2066	457064.739	Root MSE =	652.62

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	66.22149	10.72322	6.18	0.000	45.19193	87.25105
wiekk_2	-.8803322	.1297141	-6.79	0.000	-1.134718	-.6259468
placak	-4.621319	1.350604	-3.42	0.001	-7.27002	-1.972618
dosw	10.05764	2.355593	4.27	0.000	5.438035	14.67725
edu	25.30527	6.648359	3.81	0.000	12.26702	38.34352
dzieci	-129.0045	16.23756	-7.94	0.000	-160.8484	-97.16066
bezrob	-43.41367	16.84267	-2.58	0.010	-76.44423	-10.38311
ln_dochm	-96.57481	19.2705	-5.01	0.000	-134.3667	-58.78297
_Imiasto_2	-46.1122	63.13281	-0.73	0.465	-169.9235	77.69905
_Imiasto_3	-140.7081	74.79857	-1.88	0.060	-287.3974	5.981083
_Imiasto_4	-84.05672	72.18453	-1.16	0.244	-225.6195	57.50607
_Imiasto_5	-84.22716	68.98507	-1.22	0.222	-219.5154	51.06108
_Imiasto_6	-157.1931	67.38678	-2.33	0.020	-289.347	-25.03934
najml	2.547321	7.194723	0.35	0.723	-11.56242	16.65706
_ImiaXnajml_2	.6184564	8.69203	0.07	0.943	-16.42769	17.6646
_ImiaXnajml_3	7.885825	10.49396	0.75	0.452	-12.69412	28.46577
_ImiaXnajml_4	1.21683	10.22075	0.12	0.905	-18.82731	21.26097
_ImiaXnajml_5	3.217912	9.300092	0.35	0.729	-15.02072	21.45654
_ImiaXnajml_6	12.02294	9.251452	1.30	0.194	-6.1203	30.16618
_cons	1714.324	290.0544	5.91	0.000	1145.492	2283.156

end of do-file

. ovtest

```
Ramsey RESET test using powers of the fitted values of godzk
Ho: model has no omitted variables
F(3, 2044) = 0.94
Prob > F = 0.4200
```

Otrzymane wyniki nadal przemawiają za nieistotnością prawie wszystkich poziomów zmiennej miasto (oprócz zmiennej miasto6: p-value = 0.020 < 0.05), jak i wszystkich wprowadzonych interakcji (p-value>0.05).

Test na łączną nieistotność wprowadzonych interakcji potwierdza te spostrzeżenia:

```
. test _ImiaXnajml_2 _ImiaXnajml_3 _ImiaXnajml_4 _ImiaXnajml_5 _ImiaXnajml_6
```

(1) _ImiaXnajml_2 = 0

(2) _ImiaXnajml_3 = 0

(3) _ImiaXnajml_4 = 0

(4) _ImiaXnajml_5 = 0

(5) _ImiaXnajml_6 = 0

F(5, 2047) = 0.58

Prob > F = 0.7141

Interakcje okazały się nieistotne, czyli miejsce zamieszkania nie wpływa na zależność między podażą pracy kobiety a wiekiem najmłodszego dziecka.

Kolejną analizowaną przez nas interakcją był wpływ miejsca zamieszkania na zależność między podażą pracy a ilością lat poświęconych na naukę. Wydawać by się mogło, że kobiety które zamieszkują zurbanizowane obszary posiadając wyższe wykształcenie pracują krócej, ponieważ nie muszą poświęcać aż tyle czasu na pracę ze względu na poziom swojego wykształcenia (za czym, idzie również wysokość wynagrodzenia). Na obszarze mniej

zurbanizowanym mogą występować pewne trudności ze znalezieniem pracy która odpowiadałaby wykształceniu:

```
. xi: regress godzk wiekk wiekk_2 placak dosw dzieci najml bezrob ln_dochm i.miasto*edu
i.miasto      _Imiasto_1-6      (naturally coded; _Imiasto_1 omitted)
i.miasto*edu   _ImiaXedu_#      (coded as above)
```

Source	SS	df	MS	Number of obs =	2067
Model	73913492.2	19	3890183.8	F(19, 2047) =	9.15
Residual	870382259	2047	425198.954	Prob > F =	0.0000
				R-squared =	0.0783
				Adj R-squared =	0.0697
				Root MSE =	652.07
Total	944295751	2066	457064.739		

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	66.20606	10.71594	6.18	0.000	45.19077	87.22134
wiekk_2	-.8793163	.1295705	-6.79	0.000	-1.13342	-.6252125
placak	-4.849168	1.351524	-3.59	0.000	-7.499674	-2.198663
dosw	9.881449	2.360425	4.19	0.000	5.252364	14.51053
dzieci	-128.8541	16.25812	-7.93	0.000	-160.7383	-96.96995
najml	6.49645	3.179539	2.04	0.041	.260982	12.73192
bezrob	-41.97076	16.81391	-2.50	0.013	-74.9449	-8.996608
ln_dochm	-96.58362	19.23777	-5.02	0.000	-134.3113	-58.85597
_Imiasto_2	69.66343	256.7255	0.27	0.786	-433.807	573.1338
_Imiasto_3	290.6687	331.2369	0.88	0.380	-358.9278	940.2652
_Imiasto_4	-71.10231	334.3723	-0.21	0.832	-726.8476	584.643
_Imiasto_5	-169.7731	287.8838	-0.59	0.555	-734.3489	394.8027
_Imiasto_6	382.8803	281.374	1.36	0.174	-168.9288	934.6894
edu	36.11518	14.69569	2.46	0.014	7.295118	64.93524
_ImiaXedu_2	-8.173288	18.4759	-0.44	0.658	-44.40682	28.06024
_ImiaXedu_3	-28.7552	23.77288	-1.21	0.227	-75.37675	17.86636
_ImiaXedu_4	-.5497716	23.83916	-0.02	0.982	-47.30131	46.20177
_ImiaXedu_5	7.859095	20.97636	0.37	0.708	-33.27814	48.99633
_ImiaXedu_6	-36.68952	20.66249	-1.78	0.076	-77.21121	3.832177
_cons	1543.617	333.4068	4.63	0.000	889.7647	2197.469

end of do-file

. ovtest

```
Ramsey RESET test using powers of the fitted values of godzk
Ho: model has no omitted variables
F(3, 2044) = 2.54
Prob > F = 0.0552
```

Tak jak poprzednio wprowadzone interakcje jak i poziom zmiennej miasto nadal jest nieistotny.

Statystyka testowa znacznie wzrosła więc wprowadzone interakcje nie prowadziły nas w dobrym kierunku.

test _ImiaXedu_2 _ImiaXedu_3 _ImiaXedu_4 _ImiaXedu_5 _ImiaXedu_6

(1) _ImiaXedu_2 = 0

(2) _ImiaXedu_3 = 0

(3) _ImiaXedu_4 = 0

(4) _ImiaXedu_5 = 0

(5) _ImiaXedu_6 = 0

F(5, 2047) = 1.27

Prob > F = 0.2725

Co potwierdza poprzednie wnioski.

Kolejną analizowaną przez nas interakcją był wpływ miejsca zamieszkania na zależność między podażą pracy a średnią stawką godzinową.

```
. xi: regress godzk wiekk wiekk_2 dosw edu dzieci najml bezrob ln_dochm i.miasto*placak
i.miasto _Imiasto_1-6 (naturally coded; _Imiasto_1 omitted)
i.miasto*placak _Imiaxplaca_# (coded as above)
```

Source	SS	df	MS	Number of obs =	2067
Model	73608535.2	19	3874133.43	F(19, 2047) =	9.11
Residual	870687216	2047	425347.932	Prob > F =	0.0000
Total	944295751	2066	457064.739	R-squared =	0.0780
				Adj R-squared =	0.0694
				Root MSE =	652.19

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	66.88529	10.7481	6.22	0.000	45.80693	87.96365
wiekk_2	-.8886424	.1300991	-6.83	0.000	-1.143783	-.6335019
dosw	10.22638	2.35434	4.34	0.000	5.60923	14.84353
edu	24.61361	6.651354	3.70	0.000	11.56948	37.65774
dzieci	-128.3269	16.22492	-7.91	0.000	-160.1459	-96.50781
najml	6.419696	3.177437	2.02	0.043	.1883502	12.65104
bezrob	-44.69868	16.80581	-2.66	0.008	-77.65695	-11.74041
ln_dochm	-96.12834	19.25921	-4.99	0.000	-133.898	-58.35865
_Imiasto_2	-10.07459	87.22149	-0.12	0.908	-181.1267	160.9775
_Imiasto_3	-21.76114	106.7143	-0.20	0.838	-231.0411	187.5188
_Imiasto_4	11.14766	98.44914	0.11	0.910	-181.9233	204.2186
_Imiasto_5	2.633256	92.43457	0.03	0.977	-178.6424	183.9089
_Imiasto_6	52.03947	91.24131	0.57	0.569	-126.896	230.975
placak	-.6873607	3.384973	-0.20	0.839	-7.325711	5.95099
_Imiaxplac~2	-1.941983	4.078107	-0.48	0.634	-9.939656	6.055689
_Imiaxplac~3	-4.707755	5.218001	-0.90	0.367	-14.9409	5.525391
_Imiaxplac~4	-4.92787	4.47917	-1.10	0.271	-13.71208	3.856336
_Imiaxplac~5	-4.073606	4.458281	-0.91	0.361	-12.81685	4.669635
_Imiaxplac~6	-10.10435	4.724204	-2.14	0.033	-19.3691	-.839604
_cons	1623.721	292.3292	5.55	0.000	1050.427	2197.014

end of do-file

```
. ovtest
```

Ramsey RESET test using powers of the fitted values of godzk

Ho: model has no omitted variables

F(3, 2044) = **2.09**
Prob > F = **0.1001**

Tu wszystkie poziomy zmiennej miasto jak i prawie wszystkie interakcje wyszły nieistotne jednak jedna interakcja wyszła istotna. Postanowiliśmy wyrzucić pozostałe (po przeprowadzeniu testu na łączną nieistotność) i wprowadzić do modelu nową zmienną:

```
generate miasto_6_placak = (placak & miasto == 6)
```

po oszacowaniu regresji wyszło jednak, że nowo utworzona zmienna jest nieistotna więc również ją usunęliśmy.

```
. regress godzk wiekk wiekk_2 placak dosw edu dzieci najml bezrob ln_dochm miasto_6_placak
```

Source	SS	df	MS	Number of obs =	2067
Model	69486432.5	10	6948643.25	F(10, 2056) =	16.33
Residual	874809319	2056	425490.914	Prob > F =	0.0000
Total	944295751	2066	457064.739	R-squared =	0.0736
				Adj R-squared =	0.0691
				Root MSE =	652.3

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	65.542	10.70757	6.12	0.000	44.54318	86.54082
wiekk_2	-.873999	.1295145	-6.75	0.000	-1.127992	-.6200057
placak	-4.610038	1.347382	-3.42	0.001	-7.252413	-1.967663
dosw	10.21818	2.348192	4.35	0.000	5.613097	14.82326
edu	25.29691	6.630156	3.82	0.000	12.29439	38.29943
dzieci	-130.0424	16.19361	-8.03	0.000	-161.8	-98.28478
najml	6.808884	3.167374	2.15	0.032	.597287	13.02048
bezrob	-39.29146	16.60557	-2.37	0.018	-71.85695	-6.725964
ln_dochm	-96.76438	19.19334	-5.04	0.000	-134.4048	-59.12397
miasto_6_p~k	-46.83501	37.86676	-1.24	0.216	-121.0962	27.42618
_cons	1632.887	280.8945	5.81	0.000	1082.019	2183.754

Próbowaliśmy przeprowadzić następujące interakcje, jednak po ich wprowadzeniu były one nieistotne:

```
. xi: regress godzk wiekk wiekk_2 placak dzieci edu najml ln_dochm bezrob i.miasto*dosw
i.miasto      _Imiasto_1-6      (naturally coded; _Imiasto_1 omitted)
i.miasto*dosw  _Imiaxdosw_#      (coded as above)
```

Source	SS	df	MS	Number of obs =	2067
Model	72537911.1	19	3817784.8	F(19, 2047) =	8.96
Residual	871757840	2047	425870.953	Prob > F =	0.0000
				R-squared =	0.0768
				Adj R-squared =	0.0682
Total	944295751	2066	457064.739	Root MSE =	652.59

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	66.13452	10.75871	6.15	0.000	45.03536	87.23369
wiekk_2	- .8796056	.130105	-6.76	0.000	-1.134758	-.6244536
placak	-4.615871	1.351153	-3.42	0.001	-7.265649	-1.966093
dzieci	-128.7572	16.22885	-7.93	0.000	-160.584	-96.93046
edu	24.56847	6.668692	3.68	0.000	11.49034	37.6466
najml	6.492803	3.175145	2.04	0.041	.2659519	12.71965
ln_dochm	-95.74586	19.25812	-4.97	0.000	-133.5134	-57.97831
bezrob	-43.18513	16.81355	-2.57	0.010	-76.15859	-10.21168
_Imiasto_2	-98.74886	84.65966	-1.17	0.244	-264.7769	67.27919
_Imiasto_3	-123.6002	101.5082	-1.22	0.224	-322.6703	75.46982
_Imiasto_4	-204.045	100.7812	-2.02	0.043	-401.6894	-6.400665
_Imiasto_5	-158.583	90.73499	-1.75	0.081	-336.5255	19.35953
_Imiasto_6	-152.8915	88.67587	-1.72	0.085	-326.7958	21.01284
dosw	5.008221	5.190952	0.96	0.335	-5.171877	15.18832
_Imiaxdosw_2	5.019886	6.208046	0.81	0.419	-7.15486	17.19463
_Imiaxdosw_3	1.605197	7.72605	0.21	0.835	-13.54654	16.75694
_Imiaxdosw_4	11.45906	7.579168	1.51	0.131	-3.404627	26.32274
_Imiaxdosw_5	8.341772	6.936131	1.20	0.229	-5.260838	21.94438
_Imiaxdosw_6	4.313252	6.625138	0.65	0.515	-8.679462	17.30597
_cons	1753.726	292.3111	6.00	0.000	1180.468	2326.985

```
. xi: regress godzk wiekk_2 placak dosw dzieci edu najml ln_dochm bezrob i.miasto*wiekk
i.miasto      _Imiasto_1-6      (naturally coded; _Imiasto_1 omitted)
i.miasto*wiekk _Imiaxwiek_#      (coded as above)
```

Source	SS	df	MS	Number of obs =	2067
Model	72210712.8	19	3800563.83	F(19, 2047) =	8.92
Residual	872085039	2047	426030.796	Prob > F =	0.0000
				R-squared =	0.0765
				Adj R-squared =	0.0679
Total	944295751	2066	457064.739	Root MSE =	652.71

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk_2	- .8693806	.130837	-6.64	0.000	-1.125968	-.6127931
placak	-4.522029	1.353979	-3.34	0.001	-7.177349	-1.866709
dosw	10.09653	2.355595	4.29	0.000	5.476913	14.71614
dzieci	-128.0053	16.28006	-7.86	0.000	-159.9325	-96.07812
edu	24.9946	6.658068	3.75	0.000	11.93731	38.0519
najml	6.547438	3.173935	2.06	0.039	.322959	12.77192
ln_dochm	-96.69396	19.26465	-5.02	0.000	-134.4743	-58.91359
bezrob	-43.31711	16.83813	-2.57	0.010	-76.33876	-10.29547
_Imiasto_2	-282.6002	196.0605	-1.44	0.150	-667.0991	101.8987
_Imiasto_3	-414.7977	237.3325	-1.75	0.081	-880.236	50.64061
_Imiasto_4	-314.5874	231.3859	-1.36	0.174	-768.3638	139.1889
_Imiasto_5	-301.1555	211.6429	-1.42	0.155	-716.2134	113.9024
_Imiasto_6	-312.0595	208.818	-1.49	0.135	-721.5774	97.45845
wiekk	60.21542	11.44726	5.26	0.000	37.76592	82.66491
_Imiaxwiek~2	5.900373	4.658243	1.27	0.205	-3.235016	15.03576
_Imiaxwiek~3	7.639578	5.655235	1.35	0.177	-3.451037	18.73019
_Imiaxwiek~4	5.848931	5.478177	1.07	0.286	-4.894451	16.59231
_Imiaxwiek~5	5.723601	5.0334	1.14	0.256	-4.147519	15.59472
_Imiaxwiek~6	5.139182	4.976699	1.03	0.302	-4.62074	14.8991
_cons	1921.283	320.9976	5.99	0.000	1291.767	2550.798


```
. xi: regress godzkk wiekk wiekk_2 placcak dosw edu dzieci najm1 bezrob i.miasto*dochm
i.miasto      _Imiasto_1-6      (naturally coded; _Imiasto_1 omitted)
i.miasto*dochm      _Imiaxdochm_#      (coded as above)
```

Source	SS	df	MS	Number of obs =	2067
Model	70017852.8	19	3685150.15	F(19, 2047) =	8.63
Residual	874277899	2047	427102.051	Prob > F =	0.0000
Total	944295751	2066	457064.739	R-squared =	0.0741
				Adj R-squared =	0.0656
				Root MSE =	653.53

godzkk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	61.94976	10.64985	5.82	0.000	41.06409	82.83544
wiekk_2	- .8278853	.1288985	-6.42	0.000	-1.080671	-.5750994
placcak	-5.059066	1.348423	-3.75	0.000	-7.70349	-2.414642
dosw	10.45696	2.355995	4.44	0.000	5.836565	15.07736
edu	24.46525	6.676517	3.66	0.000	11.37178	37.55873
dzieci	-125.959	16.24613	-7.75	0.000	-157.8197	-94.09836
najm1	6.741766	3.186396	2.12	0.034	.492849	12.99068
bezrob	-43.08256	16.8255	-2.56	0.011	-76.07945	-10.08567
_Imiasto_2	-72.26077	71.32576	-1.01	0.311	-212.1394	67.61785
_Imiasto_3	23.14244	97.01373	0.24	0.811	-167.1135	213.3984
_Imiasto_4	-88.75752	77.31543	-1.15	0.251	-240.3826	62.86759
_Imiasto_5	8.594658	89.80368	0.10	0.924	-167.5215	184.7108
_Imiasto_6	-3.55193	84.76975	-0.04	0.967	-169.7959	162.692
dochm	- .00081	.0010307	-0.79	0.432	-.0028314	.0012113
_Imiaxdoch~2	.0006397	.0010693	0.60	0.550	-.0014572	.0027367
_Imiaxdoch~3	-.0025316	.0015587	-1.62	0.104	-.0055885	.0005252
_Imiaxdoch~4	.0003122	.0010649	0.29	0.769	-.0017762	.0024007
_Imiaxdoch~5	-.0016522	.0015968	-1.03	0.301	-.0047836	.0014792
_Imiaxdoch~6	-.0022146	.0014791	-1.50	0.134	-.0051153	.0006861
_cons	805.1903	249.5291	3.23	0.001	315.8329	1294.548

b) Praca ze zmienną dotyczącą wieku

Source	SS	df	MS	Number of obs =	2067
Model	47523669	13	3655666.85	F(13, 2053) =	8.37
Residual	896772082	2053	436810.561	Prob > F =	0.0000
Total	944295751	2066	457064.739	R-squared =	0.0503
				Adj R-squared =	0.0443
				Root MSE =	660.92

godzkk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	-5.915626	1.80157	-3.28	0.001	-9.448721	-2.382531
placcak	-4.640951	1.360942	-3.41	0.001	-7.309921	-1.97198
dosw	10.00411	2.378585	4.21	0.000	5.339419	14.6688
edu	22.01481	6.66407	3.30	0.001	8.945773	35.08386
dzieci	-111.4168	16.26263	-6.85	0.000	-143.3098	-79.52384
najm1	11.89812	3.089847	3.85	0.000	5.838563	17.95769
bezrob	-40.79879	17.00679	-2.40	0.017	-74.15114	-7.44644
dochm	-.0005092	.0001994	-2.55	0.011	-.0009003	-.0001181
miasto2	-41.01395	48.23778	-0.85	0.395	-135.614	53.58614
miasto3	-102.7808	57.51714	-1.79	0.074	-215.5789	10.01718
miasto4	-67.09125	56.22254	-1.19	0.233	-177.3504	43.16791
miasto5	-65.28915	52.89072	-1.23	0.217	-169.0142	38.43592
miasto6	-88.33483	52.64836	-1.68	0.094	-191.5846	14.91492
_cons	2072.886	158.8906	13.05	0.000	1761.282	2384.489

```
. ovtest
```

Ramsey RESET test using powers of the fitted values of godzkk

Ho: model has no omitted variables

F(3, 2050) = **6.71**

Prob > F = **0.0002**

P-value w tym teście wyszedł na poziomie 0,0000, dlatego odrzucamy hipotezę o właściwej formie funkcyjnej. W związku z tym, że test RESET we wstępnej regresji mówił nam o niepoprawnej formie funkcyjnej postanowiliśmy się przyjrzeć zmiennej wiek.

Po wstępnej analizie wykresu średniej ilości godzin w zależności od wieku (przedstawionego wyżej w pracy) doszliśmy do wniosku, że przedstawia on bardziej zależność kwadratową niż liniową. Może mieć to swoje uzasadnienie w tym, iż na początku swojej kariery zawodowej kobiety mogą się jeszcze uczyć, bądź też dokształcać na przeróżnych kursach, które

zwiększają ich zdolności, a także czynią je bardziej wartościowe na rynku pracy. Innym czynnikiem powodującym mniejszą ilość godzin przepracowanych u kobiet do 30 – 35 roku życia może mieć fakt, że właśnie w tym okresie kobiety decydują się na dzieci. Kiedy ich pociechy są jeszcze małe potrzebują więcej uwagi przez co matki często skracają ilość czasu, w którym przebywają w pracy.

Po wstępnych obliczeniach, a także patrząc na wykres możemy wywnioskować, iż kobiety najwięcej czasu na pracę poświęcają w wieku 42 lat. Następnie ilość przepracowanych godzin spada, co może być spowodowane faktem, iż ludzie starsi są w pewien sposób wypierani z rynku pracy przez młodych ludzi, którzy dla pracodawców często są bardziej efektywni niż osoby zbliżające się do wieku emerytalnego.

Wprowadzenie do modelu zmiennej `wiekk_2` znacznie poprawiło wartość statystyki w teście RESET co upewniło nas o słuszności wprowadzenia tej zmiennej do modelu, tym bardziej, że obie zmienne `wiekk` i `wiekk_2` są istotne ($p\text{-value} = 0.000$).

```
. regress godzk wiekk wiekk_2 placak dosw edu dzieci najml bezrob dochm miasto2
> miasto3 miasto4 miasto5 miasto6
```

Source	SS	df	MS	Number of obs = 2067		
Model	63772246.7	14	4555160.48	F(14, 2052) = 10.62		
Residual	880523505	2052	429105.022	Prob > F = 0.0000		
Total	944295751	2066	457064.739	R-squared = 0.0675		
				Adj R-squared = 0.0612		
				Root MSE = 655.06		

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	58.51018	10.62087	5.51	0.000	37.68137	79.33899
wiekk_2	-.792043	.1287131	-6.15	0.000	-1.044465	-.539621
placak	-5.12466	1.351173	-3.79	0.000	-7.774473	-2.474846
dosw	10.59046	2.359437	4.49	0.000	5.96332	15.2176
edu	21.57881	6.605409	3.27	0.001	8.624804	34.53281
dzieci	-124.9843	16.26865	-7.68	0.000	-156.8891	-93.07951
najml	6.445343	3.188094	2.02	0.043	.1931052	12.69758
bezrob	-43.33831	16.86117	-2.57	0.010	-76.40509	-10.27153
dochm	-.00055	.0001978	-2.78	0.005	-.0009379	-.0001622
miasto2	-38.39521	47.81231	-0.80	0.422	-132.1609	55.37051
miasto3	-106.853	57.01141	-1.87	0.061	-218.6593	4.953267
miasto4	-71.51491	55.72908	-1.28	0.200	-180.8064	37.77654
miasto5	-62.28467	52.42441	-1.19	0.235	-165.0953	40.52593
miasto6	-95.55633	52.19512	-1.83	0.067	-197.9173	6.804593
_cons	910.3421	245.9523	3.70	0.000	428	1392.684

```
. ovtest
```

Ramsey RESET test using powers of the fitted values of `godzk`

Ho: model has no omitted variables

F(3, 2049) = 2.57

Prob > F = 0.0529

c) Porównanie modeli z `ln_dochm` i `dochm`

Aby dalej poprawić formę funkcyjną naszego modelu skupiliśmy się również na zmiennej `dochm`.

Wyniki oszacowania regresji ze zmienna niezlogarytmowaną:

```
. regress godzk wiekk wiekk_2 placak dosw edu dzieci najml bezrob dochm miasto2
> miasto3 miasto4 miasto5 miasto6
```

Source	SS	df	MS	Number of obs = 2067		
Model	63772246.7	14	4555160.48	F(14, 2052) = 10.62		
Residual	880523505	2052	429105.022	Prob > F = 0.0000		
				R-squared = 0.0675		
				Adj R-squared = 0.0612		
				Root MSE = 655.06		
Total	944295751	2066	457064.739			

godzk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiekk	58.51018	10.62087	5.51	0.000	37.68137	79.33899
wiekk_2	-.792043	.1287131	-6.15	0.000	-1.044465	-.539621
placak	-5.12466	1.351173	-3.79	0.000	-7.774473	-2.474846
dosw	10.59046	2.359437	4.49	0.000	5.96332	15.2176
edu	21.57881	6.605409	3.27	0.001	8.624804	34.53281
dzieci	-124.9843	16.26865	-7.68	0.000	-156.8891	-93.07951
najml	6.445343	3.188094	2.02	0.043	.1931052	12.69758
bezrob	-43.33831	16.86117	-2.57	0.010	-76.40509	-10.27153
dochm	-.00055	.0001978	-2.78	0.005	-.0009379	-.0001622
miasto2	-38.39521	47.81231	-0.80	0.422	-132.1609	55.37051
miasto3	-106.853	57.01141	-1.87	0.061	-218.6593	4.953267
miasto4	-71.51491	55.72908	-1.28	0.200	-180.8064	37.77654
miasto5	-62.28467	52.42441	-1.19	0.235	-165.0953	40.52593
miasto6	-95.55633	52.19512	-1.83	0.067	-197.9173	6.804593
_cons	910.3421	245.9523	3.70	0.000	428	1392.684

```
. ovtest
```

Ramsey RESET test using powers of the fitted values of godzk

Ho: model has no omitted variables

F(3, 2049) = 2.57
Prob > F = 0.0529

mówią o tym, że owszem zmienna dochm jest istotna (p-value= 0.005) ale test RESET mówi nam o niepoprawności formy funkcyjnej na poziomie istotności 0.01 (p-value = 0,0529).

Przeprowadzając test Boxacoxa stata zwróciła, że nasz dochód powinien zostać zlogarytmowany:

```
boxcox godzk dochm, notrans(wiekk wiekk_2 placak dosw edu dzieci najml
bezrob) model(rhsonly)
```

Log likelihood = -16322.27

Number of obs = 2067
LR chi2(10) = 158.69
Prob > chi2 = 0.000

godzk	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
/lambda	.2636564	.1570355	1.68	0.093	-.0441275	.5714403

Estimates of scale-variant parameters

	Coef.
Notrans	
wiekk	64.81334
wiekk_2	-.8614755
placak	-4.312732
dosw	10.19509
edu	26.73237
dzieci	-129.1171
najml	7.101206
bezrob	-35.04259
_cons	934.8943
Trans	
dochm	-6.387792
/sigma	650.4475

Test H0:	Restricted log likelihood	LR statistic chi2	P-value Prob > chi2
lambda = -1	-16332.446	20.35	0.000
lambda = 0	-16323.392	2.24	0.134
lambda = 1	-16331.981	19.42	0.000

Kolejno sprawdziliśmy jak zmienia się statystyka testowa w teście RESET po umieszczeniu w modelu logarytmowanego dochodu:

Ramsey RESET test using powers of the fitted values of godzk

Ho: model has no omitted variables

$$F(3, 2054) = 1.29$$

$$\text{Prob} > F = 0.2754$$

Statystyka testowa znacznie się poprawia, co utwierdziło nas w przekonaniu o słuszności umieszczenia zmiennej \ln_dochm .