

Egzamin z Ekonometrii 18.06.2015

Pytania teoretyczne

1. Opisać procedurę od ogólnego do szczegółowego na przykładzie doboru liczby opóźnień w modelu.
2. Na czym polega najważniejsza różnica między testowaniem stacjonarności za pomocą testu *ADF* i *KPSS*?
3. Co modelowane jest w przypadku, kiedy zmienna zależna jest zmienną binarną (zerojedynkową)? Wyjaśnić, jaka jest relacja między zmienną obserwowalną a ukrytą w przypadku modeli dla zmiennych binarnych i jak tę relację można uzasadnić w oparciu o teorię ekonomii.
4. Wyjaśnić, dlaczego za pomocą estymatora efektów stałych nie można oszacować parametrów przy zmiennych nie zmieniających się w czasie.

ZADANIE 1 Mamy następujący model podaży i popytu na pracę

$$l_t = \alpha_0 + \alpha_1 w_t + \alpha_2 p_t + \varepsilon_{1t}$$
$$l_t = \beta_1 w_t + \beta_2 q_t + \varepsilon_{2t}$$

gdzie l_t oznacza logarytm zatrudnienia, w_t to logarytm płacy realnej, p_t logarytm wielkości siły roboczej, a q_t logarytm produkcji. Zmiennymi endogenicznymi są l_t i w_t a p_t i q_t są traktowane jako egzogeniczne.

1. Które z tych równań jest według Ciebie równaniem popytu, a które podaży? Odpowiedź uzasadnij.
2. Sprawdź identyfikację poszczególnych równań.
3. Wyjaśnij jak wyglądać będzie forma zredukowana tego modelu
4. Wyjaśnij dlaczego równania popytu nie można poprawnie wyestymować za pomocą *MNK*.
5. Jeśli zastosowalibyśmy do estymacji tego równania *MZI*, to jakie zmienne mogłyby być użyte jako instrumenty?

Rozwiązanie:

1. Podaż pracy zależy od wielkości siły roboczej oraz wysokości płacy. Popyt na pracę zależy od wysokości płacy i wielkości produkcji. Wynika z tego, że równanie pierwsze jest równanie podaży a równanie drugie równaniem popytu.
2. Podział na zmienne endo i egzogeniczne

zmienne egzogeniczne (wliczamy stałą)	$1, p_t, q_t$
zmienne endogeniczne	l_t, w_t

G_s liczba zmiennych endogenicznych w s -tym równaniu, K_s liczba zmiennych egzogenicznych w s -tym równaniu:

$$K = 3 \quad G_1 = 2 \quad K_1 = 2$$
$$G_2 = 2 \quad K_2 = 1$$

Identyfikacja:

$$3 = 2 + 2 - 1 \quad (+)$$
$$3 > 2 + 1 - 1 \quad (+)$$

Równanie popytu i podaży na pracę są zidentyfikowane.

3. Równania formy zredukowanej dla pracy i płacy:

$$l_t = \pi_{10} + \pi_{11} p_t + \pi_{12} q_t + \varepsilon_{1t}$$
$$w_t = \pi_{20} + \pi_{21} p_t + \pi_{22} q_t + \varepsilon_{2t}$$

- Równania popytu na pracę nie da się wyestymować standardowym MNK, ponieważ zmienna objaśniająca w_t (płaca) jest skorelowana z błędem losowym ε_{1t} z powodu występowania sprzężenia zwrotnego między popytem na pracę i wysokością płacy (wysokość płacy wpływa na popyt na pracę a równocześnie wielkość popytu na pracę wpływa na płacę). Taka korelacja oznacza występowanie problemu równoczesności, który powoduje, że estymator MNK nie jest zgodny.
- Instrumentami w przypadku estymacji równania popytu na pracę powinny być wszystkie zmienne egzogeniczne występujące w modelu a więc $1, p_t, q_t$.

ZADANIE 2 Badane są czynniki skłaniające respondentów do emigracji. W badaniu wykorzystywane są dane z badania "Diagnoza społeczna" z lat 2000, 2003, 2005, 2007. Poniżej znajdują się oszacowania parametrów i efektów cząstkowych policzonych dla średnich w próbie dla modelu probitowego oszacowanego dla tego problemu, w którym zmienną zależną jest odpowiedź na pytanie, czy dany członek gospodarstwa w momencie badania pracuje zagranicą (1 pracuje, 0 nie pracuje). Zmiennymi objaśniającymi w modelu są: logarytm dochodu na głowę w gospodarstwie, liczba lat nauki, zmienna zerojedynkowa związana z płcią respondenta (1 kobieta, 0 mężczyzna) oraz zmienne zerojedynkowe związane z rokiem przeprowadzenia badania.

	parametr	efekt cząstkowy	z	p
ln(dochodu)	-0.292	-0.004	-8.387	0.000
lata nauki	0.064	0.001	8.332	0.000
kobieta	-0.334	-0.005	-7.417	0.000
2003	-0.118	-0.002	-1.240	0.215
2005	0.388	0.008	4.885	0.000
2007	0.598	0.012	7.958	0.000
Constant	-1.550		-7.544	0.000
N	38916	38916		
$R_{\text{McKelvey'a i Zavoiny}}^2$	0.134			
$R_{\text{liczebnościowe}}^2$	0.992			
$\bar{R}_{\text{liczebnościowe}}^2$	0.000			

Testy przeprowadzamy na poziomie istotności $\alpha = 0.05$.

- Wypisz założenia modelu probitowego.
- Które założenie $KMRL$ będzie najprawdopodobniej fałszywe dla modelu liniowego opisującego zależność p-stwa emigracji od charakterystyk respondenta?
- Na jakie problemy natkniemy się, jeśli model ten oszacujemy za pomocą MNK ?
- Jakie wnioski można wyciągnąć na podstawie oszacowań parametrów w tym modelu?
- Podaj interpretację $R_{\text{McKelvey'a i Zavoiny}}^2$ i $R_{\text{liczebnościowe}}^2$. Wyjaśnij dlaczego występuje tak duża rozbieżność między $R_{\text{liczebnościowe}}^2$ i miarą skorygowaną $\bar{R}_{\text{liczebnościowe}}^2$.
Podpowiedź: W badanej próbie znajduje się jedynie 317 obserwacji, dla których respondenci w momencie badania pracowali zagranicą.
- Podaj interpretację efektu cząstkowego dla płci (kodowanie 0 - mężczyzna, 1 kobieta).
- Zinterpretuj efekt cząstkowy dla logarytmu dochodu na głowę w gospodarstwie.
- Jaki model można by zastosować, by zbadać hipotezę, że wykształcenie wpływa w różny sposób na skłonność do emigracji kobiet i mężczyzn?
- Przetestowano hipotezę, że rok badania nie ma wpływu na skłonność do emigracji respondentów uzyskano i statystykę $LR = 154.95$. Zweryfikuj powyższą hipotezę.
Podpowiedź: $\chi_{0.95}^2(1) = 3.84$, $\chi_{0.95}^2(2) = 5.99$, $\chi_{0.95}^2(3) = 7.81$.

Rozwiązanie:

1. Zmienna ukryta

$$y^* = \mathbf{x}\boldsymbol{\beta} + \varepsilon$$
$$\varepsilon \sim N(0, 1)$$

i poszczególne obserwacje są niezależne. Obserwujemy

$$y = 0 \quad \text{dla} \quad y^* \leq 0$$
$$y = 1 \quad \text{dla} \quad y^* > 0$$

Alternatywna odpowiedź: prawdopodobieństwa zajścia pojedynczego zdarzenia:

$$\Pr(y_i) = \begin{cases} 1 - \Phi(\mathbf{x}\boldsymbol{\beta}) & \text{dla } y = 0 \\ \Phi(\mathbf{x}\boldsymbol{\beta}) & \text{dla } y = 1 \end{cases}$$

i poszczególne obserwacje są niezależne.

2. W przypadku modelu liniowego (Liniowego Modelu Prawdopodobieństwa - LPM) fałszywe jest założenie o homoskedastyczności, ponieważ $\text{Var}(y) = \text{Var}(\varepsilon) = p(\mathbf{x})[1 - p(\mathbf{x})] = \mathbf{x}\boldsymbol{\beta}(1 - \mathbf{x}\boldsymbol{\beta})$ i jest zależne od \mathbf{x} .
3. Szacując model LPM natknijemy się na dwa problemy: po pierwsze na wspomnianą wyżej heteroskedastyczność a po drugie możemy uzyskać nieinterpretowalne dopasowane wartości prawdopodobieństw (spoza przedziału $[0, 1]$)
4. Na podstawie oszacowań uzyskanych z modelu możemy zbadać istotność oraz kierunek wpływu poszczególnych zmiennych. Istotne są zmienne: lata nauki, kobieta, 2005, 2007, oraz stała. Dodatni wpływ na p-stwo posiadania emigracji ma zmienna lata nauki, 2005, 2007. Zmienna kobieta ma ujemny wpływ na p-stwo emigracji. Zmienna 2003 ma nieistotny wpływ na p-stwo emigracji co oznacza, że nie było istotnej różnicy między rokiem 2000 i 2003 jeśli chodzi o p-stwo emigracji osób o identycznych charakterystykach.
5. Interpretacja $R_{\text{McKelvey'a i Zavoiny}}^2$: 13,4% zmienności nieobserwowalnej zmiennej ukrytej została wyjaśniona przez zmienne niezależne. Interpretacja $R_{\text{liczebnościowe}}^2$: w przypadku 99,2% obserwacji model prawidłowo przewidział sukces bądź porażkę. $\bar{R}_{\text{liczebnościowe}}^2$ jest skorygowane o liczebność najbardziej liczebnej kategorii. W analizowanym przypadku najliczniejsza kategoria (porażka), to $\frac{38916-317}{38916} = 99.2\%$ obserwacji. Model, który zawsze wskazywałby na tą kategorię przewidywałby prawidłowo 99.2% obserwacji. Po zastosowaniu korekty okazuje się, że zmienne mają bardzo mały wpływ na liczbę prawidłowo przewidzianych zdarzeń.
6. Wielkość efektu cząstkowego dla zmiennej kobieta oznacza, że kobiety o charakterystykach na poziomie średnich w próbie mają o 0.5 punkta procentowego mniejsze prawdopodobieństwo emigracji w stosunku do mężczyzn o charakterystykach na poziomie średnich w próbie.
7. Wielkość efektu cząstkowego dla zmiennej $\ln(\text{dochód})$ oznacza, że wzrost dochodu na głowę w gospodarstwie o 1% związany jest ze spadkiem p-stwa emigracji o 0.4 punktu procentowego.
8. Należy do modelu wstawić interakcję między płcią a latami nauki (kobieta \times lata nauki) oszacować tak rozbudowany model i przetestować istotność współczynnika przy dodanej zmiennej
9. Testując hipotezę łączną o nieistotności roku badania badamy hipotezę łączną o nieistotności parametrów przy trzech zmiennych: 2003, 2005, 2007. Statystyka testu LR ma więc rozkład $\chi^2(3)$. Ponieważ $LR = 154.95 > 7.81$ więc H_0 o nieistotności roku badania odrzucamy.

ZADANIE 3 Badacz analizował stopę inflacji płac (inaczej mówiąc dynamikę płac w wyrażeniu nominalnym) w Polsce w okresie od 3 kwartału 1998 do 4 kwartału 2013 roku. Do badania użył danych kwartalnych. Założył, że dynamika płac zależy od inflacji cen towarów konsumpcyjnych, od stopy wzrostu produktywności pracy oraz od wysokości bezrobocia. Zależności te wynikają z teoretycznych rozważań dotyczących negocjacji płacowych, z których wynika, że zatrudnieni mają lepszą pozycję negocjacyjną jeśli ich produktywność rośnie i kiedy rośnie ogólny poziom cen. Jednak w przypadku, gdy poziom bezrobocia jest wysoki pozycja zatrudnionych w negocjacjach płacowych jest słabsza, ponieważ łatwiej ich zastąpić innymi pracownikami, którzy

aktualnie są bezrobotni. Niektórzy autorzy wskazują wszakże, że słabszy wpływ na negocjacje płacowe liczba bezrobotnych długotrwale, ponieważ część tego rodzaju bezrobotnych jest z różnych przyczyn niezatrudnialna. W związku z powyższym w swej analizie badacz wykorzystał następujące zmienne: dynamikę płac nominalnych $d_wn_t = \ln\left(\frac{w_t}{w_{t-1}}\right)$ (w_t jest średnim poziomem płacy brutto), indeks inflacji $cpi_t = \ln\left(\frac{p_t}{p_{t-1}}\right)$ (p_t jest poziomem cen artykułów konsumpcyjnych), poziom bezrobocia $u_t = \ln(U_t)$ (U_t udział bezrobotnych wśród aktywnych zawodowo), poziom bezrobocia długotrwałego $ult_t = \ln(ULT_t)$ (ULT_t udział bezrobotnych długotrwale wśród aktywnych zawodowo), zmianę produktywności pracy $d_gprod = \ln\left(\frac{g_t}{g_{t-1}}\right)$ ($g_t = \frac{PKB_t}{L_t}$, gdzie PKB_t to produkt krajowy brutto a L_t to liczba pracujących). Badacz przeprowadził zbadął najpierw poziom integracji zmiennych, następnie oszacował równanie długookresowe i zbadął występowanie kointegracji a na koniec oszacował Model Korekty Błędem. Wyniki regresji oraz wielkość statystyk testowych podane są poniżej. Testy przeprowadzamy na poziomie istotności $\alpha = 0.05$. Odpowiedzi należy uzasadnić np. wielkościami statystyk testowych.

1. Zbadaj, czy między zmiennymi może wystąpić kointegracja.
2. Zweryfikuj, czy między zmiennymi występuje kointegracja. Reszty uzyskane z oszacowanej zależności długookresowej oznaczone są jako zmienna *resid*.
3. Zinterpretuj długookresowe zależności między inflacją płac a zmiennymi objaśniającymi. Czy uzyskane znaki przy parametrach zgadzają się z teorią opisaną w treści zadania?
4. Zinterpretuj współczynnik przy zmiennej *resid* w oszacowanym Modelu Korekty Błędem.
5. O czym świadczy wielkość statystyki Breuscha-Godfrey'a policzonej dla Modelu Korekty Błędem? Jakie konsekwencje miałyby autokorelacja błędu losowego, gdyby wystąpiła w tym równaniu?

Testy Dickey-Fullera

Dickey-Fuller test for unit root			Number of obs	=	61
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value	
d_wn	-2.824	-3.565	-2.921	-2.596	
cpi	-2.504	-3.565	-2.921	-2.596	
u	-0.819	-3.565	-2.921	-2.596	
ult	-0.712	-3.565	-2.921	-2.596	
d_gprod	-2.127	-3.565	-2.921	-2.596	

Dickey-Fuller test for unit root			Number of obs	=	61
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value	
D.d_wn	-8.298	-3.565	-2.921	-2.596	
D.cpi	-5.037	-3.565	-2.921	-2.596	
D.u	-6.242	-3.565	-2.921	-2.596	
D.ult	-5.610	-3.565	-2.921	-2.596	
D.d_gprod	-4.831	-3.565	-2.921	-2.596	

Równanie długookresowe

Source	SS	df	MS	Number of obs = 62		
Model	.050805539	4	.012701385	F(4, 57)	=	62.82
Residual	.011525399	57	.0002022	Prob > F	=	0.0000
-----				R-squared	=	0.8151
Total	.062330938	61	.001021819	Adj R-squared	=	0.8021
-----				Root MSE	=	.01422
d_wn	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
cpi	1.010264	.0804399	12.56	0.000	.8491855	1.171342
u	-.1275192	.0208994	-6.10	0.000	-.1693696	-.0856688
ult	.067068	.0124911	5.37	0.000	.042055	.0920811
d_gprod	.4173495	.1026229	4.07	0.000	.2118507	.6228483
_cons	-.05035	.0147197	-3.42	0.001	-.0798257	-.0208744

Dickey-Fuller test for unit root

Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
resid	-4.810	-5.15	-4.17

Mechanizm Korekty Błędem

Source	SS	df	MS	Number of obs = 60		
Model	.004178242	6	.000696374	F(6, 53)	=	4.91
Residual	.007520207	53	.000141891	Prob > F	=	0.0005
-----				R-squared	=	0.3572
Total	.011698449	59	.000198279	Adj R-squared	=	0.2844
-----				Root MSE	=	.01191
D.d_wn	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
resid						
L1.	-.3559416	.1411676	-2.52	0.015	-.6390879	-.0727953
d_wn						
LD.	-.1656946	.1403192	-1.18	0.243	-.4471393	.1157501
cpi						
LD.	.3746666	.2113921	1.77	0.082	-.0493323	.7986656
u						
LD.	.0197791	.0272029	0.73	0.470	-.0347831	.0743413
ult						
LD.	-.0511408	.0214131	-2.39	0.021	-.0940901	-.0081915
d_gprod						
LD.	.3786317	.1315273	2.88	0.006	.1148213	.6424421
_cons	-.0011098	.0015683	-0.71	0.482	-.0042553	.0020357

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	0.488	1	0.4846

H0: no serial correlation

Rozwiązanie:

1. Między zmiennymi może wystąpić kointegracja, ponieważ wszystkie rozważane zmienne są, zgodnie z wynikami testów ADF, zmiennymi $I(1)$: d_wn_t ($-2.824 > -2.921$, $-8.298 < -2.921$), cpi_t ($-2.504 > -2.921$; $-5.037 < -2.921$), u_t ($-0.819 > -2.921$; $-6.242 < -2.921$), ult_t ($-0.712 > -2.921$; $-5.610 < -2.921$), d_gprod ($2.127 > -2.921$; $-4.831 < -2.921$)
2. Między zmiennymi występuje kointegracja, ponieważ zgodnie z wynikiem testu ADF reszty uzyskane z równania długookresowego są stacjonarne ($-4.810 > -4.59$)
3. Wzrost inflacji o 1% powoduje wzrost płac nominalnych o 1.01%, wzrost bezrobocia o 1% powoduje wzrost spadek płac nominalnych o -0.128%, wzrost bezrobocia długoterminowego o 1% powoduje wzrost płac nominalnych o 0.067%, wzrost produktywności pracy o 1% powoduje wzrost płac nominalnych o 0.417%. Uzyskane znaki parametrów wydają się być zgodne z teorią. W szczególności płace nominalne zgodnie z oszacowanym modelem rosną wraz z inflacją i wzrostem produktywności. Wzrost bezrobocia ogółem *ceateris paribus* powoduje spadek dynamiki płac jednak wzrost bezrobocia długotrwałego *ceateris paribus* (w szczególności przy niezmiennym bezrobociu ogółem) powoduje wzrost dynamiki płac, co jest zgodne z teorią opisaną w treści zadania.
4. Oszacowanie współczynnika przy zmiennej *resid* interpretujemy jako wielkość współczynnika szybkości dostosowań. Jego oszacowana wielkość oznacza, że 36.6% odchylenia od stanu równowagi korygowana jest w ciągu jednego kwartału.
5. Wielkość statystyki Breuscha-Godfrey'a oznacza, że w oszacowanym Modelu Korekty Błędem nie występuje problem autokorelacji błędu losowego. Gdyby problem ten wystąpił mógłby spowodować brak zgodności estymatora MNK, ponieważ w modelach z opóźnioną zmienną zależną występowanie autokorelacji powoduje wystąpienie problemu równoczesności.