

Egzamin z ekonometrii IiE 22.06.2011

Pytania teoretyczne

1. Jakie trzy testy stosuje się do testowania hipotez parametrycznych postaci $h(\theta) = 0$ w kontekście estymacji MNW? Porównać wady i zalety tych testów.
2. Wyjaśnij, czym jest specyficzność i wrażliwość modelu dla zmiennej binarnej i jak one zależą od progowego prawdopodobieństwa p^* .
3. Opisz dwustopniową procedurę estymacji modelu Heckmana.
4. Kiedy mówimy, że w modelu występuje równoczesność? Za pomocą jakiej statystyki można przetestować występowanie równoczesności? Jaką metodę estymacji powinno się użyć w przypadku występowania równoczesności?

ZADANIE 1 Przy użyciu danych dotyczących 133 krajów świata z lat 1970-2009 badacz zbudował panel niezbilansowany. Przy użyciu tego panelu badacz oszacował model, którego celem jest wyjaśnienie zróżnicowania średniej długości trwania życia w poszczególnych krajach. Zmienną zależną jest zmienna *life* oznaczająca oczekiwaną długość życia w latach a zmiennymi objaśniającymi poziom alfabetyzacji (*lit* - wyrażony w procentach udział ludności umiejącej czytać), PKB per capita (*GDP_pc* wyrażone w dolarach), trend czasowy (*year* - rok z którego pochodzą dane) umieszczony, by uwzględnić postęp medycyny. Dodatkowo do regresji włączono zmienną zerojedynkową *post_comm* oznaczającą, że danych kraj był w pewnym okresie lub dalej jest krajem komunistycznym po to, by zbadać, czy rzeczywiście w krajach komunistycznych opieka medyczna była na względnie dobrym poziomie. Model został przez badacza oszacowany za pomocą MNK, estymatora efektów losowych i estymatora efektów stałych. Wyniki estymacji znajdują się na następnej stronie.

Założony poziom istotności przy testowaniu hipotez statystycznych $\alpha = 0.05$. Uzyskane wyniki testów należy uzasadnić wielkościami odpowiednich statystyk bądź wartościami p .

1. Wyjaśnij, dlaczego badacz użył w regresji MNK odpornego warstwowego estymatora macierzy wariancji i kowariancji.
2. Weźmy pod uwagę jedynie wyniki dla MNK i estymatora efektów losowych. Który z nich jest estymatorem efektywnym w przypadku rozpatrywanego problemu? Odpowiedź uzasadnij odpowiednią statystyką testową.
3. Na podstawie znajdujących się na wydruku statystyk testowych wybierz spośród estymatorów POLS, RE i FE estymator, który powinno się użyć w kontekście analizowanego problemu. Wyjaśnij jakie hipotezy zerowe testujemy za pomocą użytych testów.
4. Dlaczego w przypadku estymatora efektów stałych nie udało się oszacować współczynnika dla zmiennej *post_comm*?
5. Zinterpretuj wielkości wszystkich trzech statystyk R^2 uzyskanych dla estymatora efektów stałych.
6. Zinterpretuj wielkość istotnych współczynników w modelu efektów stałych.
7. Czy w modelu efektów stałych poprawne byłoby pominięcie efektów indywidualnych dla krajów? Odpowiedź uzasadnij wielkością odpowiedniej statystyki.
8. Dlaczego zarówno w modelu efektów stałych jak i w modelu efektów losowych uzyskujemy dwa oszacowania błędów standardowych czynników losowych (*sigma_u*, *sigma_e*)?

Linear regression

		Robust				
life	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lit	.3069304	.0160516	19.12	0.000	.2751788	.338682
GDP_pc	.0002563	.0000399	6.43	0.000	.0001775	.0003352
post_comm	-1.302394	.8825781	-1.48	0.142	-3.048221	.4434327
year	-.0098472	.0260935	-0.38	0.706	-.0614626	.0417683

_cons		60.10586	51.21889	1.17	0.243	-41.21018	161.4219
-------	--	----------	----------	------	-------	-----------	----------

Random-effects GLS regression

	life		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
	lit		.1054896	.007226	14.60	0.000	.0913269 .1196522
	GDP_pc		.0000376	7.16e-06	5.24	0.000	.0000235 .0000516
	post_comm		1.562903	1.079774	1.45	0.148	-.5534152 3.679221
	year		.1900848	.0068268	27.84	0.000	.1767046 .2034649
	_cons		-321.9032	13.15153	-24.48	0.000	-347.6797 -296.1266

	sigma_u		4.4966449				
	sigma_e		2.5231168				
	rho		.76054561	(fraction of variance due to u_i)			

Breusch and Pagan Lagrangian multiplier test for random effects

chi2(1) = 43672.49
 Prob > chi2 = 0.0000

Fixed-effects (within) regression

R-sq: within = 0.5793
 between = 0.4524
 overall = 0.2970

	life		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	lit		.0468977	.0078084	6.01	0.000	.0315894 .062206
	GDP_pc		-2.12e-07	7.28e-06	-0.03	0.977	-.0000145 .0000141
	post_comm		(omitted)				
	year		.2384537	.0071899	33.17	0.000	.2243581 .2525493
	_cons		-413.1269	13.79678	-29.94	0.000	-440.1753 -386.0785

	sigma_u		8.8759641				
	sigma_e		2.5231168				
	rho		.92523542	(fraction of variance due to u_i)			

F test that all u_i=0: F(132, 4533) = 127.85 Prob > F = 0.0000

Hausman test

chi2(2) = (b-B)' [(V_b-V_B)^(-1)] (b-B) = 301.07
 Prob>chi2 = 0.0000

Rozwiązanie:

1. Badacz użył w regresji MNK warstwowego estymatora odpornego, ponieważ w przypadku regresji na panelu można spodziewać się niediagonalnej macierzy wariancji i kowariancji ze względu na występowanie efektów indywidualnych
2. Estymatorem efektywnym jest w tym przypadku estymator efektów losowych, ponieważ z wielkości statystyki Breuscha-Pagana 43672.49 [0.0000] wynika, że efekty losowe są istotne w modelu, macierz wariancji kowariancji łącznego błędu losowego jest niediagonalna, a w takim przypadku estymator SUMNK, którego szczególnym przypadkiem jest estymator efektów losowych, jest efektywniejszy od estymatora MNK.
3. Podstawą wyboru odpowiedniego estymatora w rozpatrywanym przypadku powinien być wynik testu Hausmana. Hipotezą zerową w tym teście jest warunek konieczny dla zgodności estymatora efektów losowych to jest brak korelacji między efektem indywidualnym a zmiennymi objaśniającymi $Cov(u_i, X_i) = 0$. W naszym przypadku hipoteza ta jest odrzucana 301.07 [0.0000] a tym samym jedynym zgodnym estymatorem jest estymator efektów stałych.

4. Za pomocą estymatora efektów stałych nie jest możliwe oszacowanie wpływu zmiennych, które nie zmieniają się w czasie. Zmienna zerojedynkowa oznaczająca, że dany kraj był bądź jest krajem komunistycznym nie zmienia się w czasie.
5. Wielkość R_{within}^2 oznacza, że 58% zróżnicowania wewnątrz obiektowego (to jest zmian długości oczekiwanego życia dla danego kraju) udało się wyjaśnić za pomocą zróżnicowania zmiennych objaśniających dla tego kraju. Wielkość $R_{between}^2$ oznacza, że 45% zróżnicowania długości trwania życia między krajami udało się wyjaśnić za pomocą różnic w wielkościach zmiennych objaśniających pomiędzy krajami, $R_{overall}^2$ oznacza, że 30% całkowitej zmienności zmiennej zależnej udało się wyjaśnić zmiennością zmiennych niezależnych.
6. Wzrost poziomu alfabetyzacji o 1 punkt procentowy wydłuża średnią długość życia o 0.04 roku, średnia długość życia wydłuża się o 0.23 roku w każdym kolejnym okresie badanym.
7. W modelu efektów stałych nie można pominąć charakterystyk indywidualnych krajów ponieważ są one istotne co wnioskujemy z wyniku testu, że wszystkie $u_i = 0$ (statystyka 127.85 [0.0000]) a zarazem wiemy z wyniku testu Hausmana, że efekty indywidualne są skorelowane ze zmiennymi objaśniającymi. Pominięcie efektów indywidualnych wywołałoby w tym przypadku pojawienie się problemu endogeniczności.
8. Ponieważ w przypadku liniowego modelu efektów nieobserwowalnych mamy dwa składniki losowe: efekt indywidualny u_i oraz błąd czystolosowy ε_i . Oszacowane odchylenia standardowe odpowiadają odchyleniom standardowym u_i oraz ε_i .

ZADANIE 2 Dany jest model logitowy, ze zmiennymi niezależnymi x_1, x_2, x_3 . Hipotezy testujemy na poziomie istotności $\alpha = 0.05$.

1. Należy przetestować łączną nieistotność x_1 i x_2 . Dla modelu z pełną liczbą zmiennych wielkość logarytmu funkcji wiarygodności wyniosła -279.5 . Dla modelu bez zmiennych x_1 i x_2 wartość logarytmu funkcji wiarygodności wyniosła -281 . Policzyc odpowiednią statystykę testową i podać wynik testu.
2. Policzono wielkość statystyki Walda dla hipotezy $H_0 : \beta_1^2 = \beta_2$ oraz $H_0 : \frac{\beta_1^2}{\beta_2} - 1 = 0$ i otrzymano wielkość statystyk testowych 5 i 2. Przeprowadzić testy dla tych dwóch hipotez i skomentować uzyskane wyniki.
3. Mamy następujący problem: chcemy przetestować hipotezę $\gamma = 1$, w modelu nieliniowym postaci:

$$y_i = (\beta_1 x_{1i} + \beta_2 x_{2i})^\gamma + \varepsilon_i$$

przy czym $\varepsilon \sim N(0, I)$. Dysponujemy programem, który umożliwia jedynie szacowanie zwykłej regresji liniowej. Jaką rodzajem testu powinniśmy się posłużyć w tym przypadku?

Rozwiązanie:

1. Statystyka testowa $LR = -2(L_0 - L_1)$ i ma rozkład $\chi^2(k)$, gdzie k to liczba zmiennych o których zakładamy że są nieistotne, L_0 wartość logarytmu wiarygodności dla modelu bez ograniczeń, L_1 wartość logarytmu wiarygodności dla modelu z ograniczeniami. W tym przypadku:

$$LR = -2(-281 - (-279.5)) = 3 < \chi^2(2) = 5.99$$

Wobec tego należy uznać, że te modele nie różnią się w sposób statystycznie istotny. Wobec tego zmienne są łącznie nieistotne.

2. W pierwszym przypadku odrzucamy $H_0 : \beta_1^2 = \beta_2$ [$5 > 3.84$], w drugim przypadku hipotezę $H_0 : \frac{\beta_1^2}{\beta_2} - 1 = 0$ nie ma podstaw do odrzucenia H_0 [$2 < 3.84$]. Analizowane hipotezy są równoważne algebraicznie a więc wynik dwóch testów daje sprzeczne wyniki. W przypadku testu Walda jest to możliwe, ponieważ w małych próbach i dla hipotez nieliniowych wynik testu może zależeć od sposobu sformułowania H_0 .
3. Statystyką mnożników Lagrange'a - umożliwia on zweryfikowanie H_0 przy znajomości oszacowań modelu z ograniczeniami. W tym przypadku model z ograniczeniami jest bardzo zwykłym modelem liniowym $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ spełniającym założenia *KMRL*.

ZADANIE 3 Mamy następujący model

$$y_i = \beta x_i + \varepsilon_i, \text{ dla } i = 1, \dots, N$$

Wiemy, że próba jest losowa i że błędy losowe są sferyczne. Wiemy jednak też, że $\text{Cov}(\varepsilon_i, x_i) = \gamma \neq 0$. Załóżmy dla uproszczenia $E(\varepsilon_i) = E(x_i) = 0$ i oznaczmy $\text{Var}(\varepsilon_i) = \sigma_\varepsilon^2$, $\text{Var}(x_i) = \sigma_x^2$.

1. Policz asymptotyczne obciążenie estymatora MNK to jest $\text{plim}(b - \beta)$

Założmy, że mamy zmienną z_i , o której wiemy, że $E(z_i) = 0$, $\text{Var}(z_i) = \sigma_z^2$, $\text{Cov}(z_i, \varepsilon_i) = 0$, oraz $\text{Cov}(z_i, x_i) = \delta$.

2. Podaj postać estymatora MZI , który można zastosować w tym przypadku, pokaż, że jest on zgodny i policz jego asymptotyczną wariację (to jest wariację $\sqrt{N}(\beta_{MZI} - \beta)$)

Uwaga: Załóż, że spełnione są założenia Prawa Wielkich Liczb tam, gdzie będziesz go używał.

Rozwiązanie:

1. Estymator MNK

$$b = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{y} = \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\varepsilon}$$

$$b - \beta = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\varepsilon} = \frac{\sum_{i=1}^N x_i \varepsilon_i}{\sum_{i=1}^N x_i^2}$$

$$\text{plim} \left(\frac{\sum_{i=1}^N x_i \varepsilon_i}{\sum_{i=1}^N x_i^2} \right) = \frac{\text{plim} \left(N^{-1} \sum_{i=1}^N x_i \varepsilon_i \right)}{\text{plim} \left(N^{-1} \sum_{i=1}^N x_i^2 \right)} = \frac{\text{Cov}(\varepsilon_i, x_i)}{\text{Var}(x_i)} = \frac{\gamma}{\sigma_x^2} \neq 0$$

2. Estymator MZI

$$b_{MZI} = (\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}' \mathbf{y} = \beta + (\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}' \boldsymbol{\varepsilon}$$

$$b_{MZI} - \beta = (\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}' \boldsymbol{\varepsilon} = \frac{\sum_{i=1}^N z_i \varepsilon_i}{\sum_{i=1}^N z_i x_i} = \frac{\text{Cov}(\varepsilon_i, z_i)}{\text{Cov}(x_i, z_i)} = \frac{0}{\delta} = 0$$

Wariancja

$$\text{Var}(b_{MZI}) = \sigma_\varepsilon^2 (\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}' \mathbf{Z} (\mathbf{X}'\mathbf{Z})^{-1} = \sigma_\varepsilon^2 \frac{\sum_{i=1}^N z_i^2}{\left(\sum_{i=1}^N z_i x_i \right)^2}$$

$$\text{plim} \left(\sqrt{N} (b_{MZI} - \beta) \right) = \sigma_\varepsilon^2 \frac{\text{plim} \left(N^{-1} \sum_{i=1}^N z_i^2 \right)}{\left[\text{plim} \left(N^{-1} \sum_{i=1}^N z_i x_i \right) \right]^2} = \frac{\sigma_\varepsilon^2 \sigma_z^2}{\delta^2}$$