

Scriptie bespreking

Anne de Hond

27 juni

1 Agenda bespreking 27 juni

- Ik heb gekeken naar de eigenschappen van de individuen die in aanmerking komen voor het sample op basis van hun gerapporteerde ADL/IADL. De resultaten staan onder **Sample**. Ik zou graag bespreken of hier nog verder op gefilterd moet worden op basis van deze resultaten.
- Ik heb mijn model aangepast van probit naar logit, vanwege de schattingsproblemen in R met de uiteindelijke dataset. Een korte uitleg staat onder **Logit model**.
- Ik wil graag nog een financiële variabele toevoegen die een indicatie geeft van het inkomen, maar welke financiële variabelen kan ik daarvoor het best gebruiken? (Ofwel, wat is gebruikelijk?) Keus uit: earnings from employment (EP205), earnings from self-employment (EP207), pensions (EP078), total household net income (gvimputations).
- Ik moet bepalen hoe ik duratie ga meten, sinds we niet het precieze moment van de verandering weten, alleen sinds welke wave personen een verhoging in ADL/IADL aangeven. Ik kan duratie dus definiëren als “sinds de vorige wave” / “sinds 2 waves” etc.. De periodes tussen de waves zijn alleen niet allemaal even lang, soms 2 en soms 3 jaar. Een andere optie is het gemiddelde van de tijd tussen de twee periodes te nemen. (Dus als iemand in wave 5 een verhoging in ADL/IADL aangeeft en de er zit twee jaar tussen wave 5 en 4, dan wordt de duratie gelijk gesteld aan 1.)

2 Sample

Het sample dat nu geselecteerd is bestaat uit individuen die pas na de eerste wave (in wave 2, 4, 5 of 6) een verhoging in ADL (of IADL) ondervinden van 0 naar > 0 . In de opeenvolgende waves kan de ADL (IADL) vervolgens nog hoger worden, gelijk blijven, of verminderen. Bovendien bestaat het sample uit individuen die vanaf hun eerste wave iedere wave zijn waargenomen tot aan hun laatste wave.

Table 1: Aantal observaties in sample

Selectie op basis van ADL	
Totaal aantal observaties	22010
Totaal aantal individuen	6776
Totaal met constante ADL*	4201
Individueen constante verhoogde ADL in > 2 waves	431
Individueen met (uitsluitend) stijgende ADL	469
Selectie op basis van IADL	
Totaal aantal observaties	29872
Totaal aantal individuen	9214
Totaal met constante IADL*	5698
Individueen constante verhoogde IADL in > 2 waves	630
Individueen met (uitsluitend) stijgende IADL	931
Selectie op basis van ADL + IADL	
Totaal aantal observaties	32764
Totaal aantal individuen	10072
Totaal met constante ADL + IADL*	5972
Individueen constante verhoogde ADL + IADL in > 2 waves	593
Individueen met (uitsluitend) stijgende ADL + IADL	1010

* Dit zijn de individuen die een verhoging in ADL/IADL ondervinden en vervolgens geen verandering aangeven in de resterende periodes. Dit omvat ook de individuen die alleen in hun laatste periode een verhoogde ADL/IADL aangeven.

3 Logit model

Hier volgt een korte beschrijving van het nieuwe logit model dat ik voorstel en mijn redenen om van het probit model af te zien.

Stel we hebben een latente variabele (bijvoorbeeld de werkelijke gezondheidstoestand):

$$y_{it}^* = x'_{it}\beta + \alpha_i + \varepsilon_{it},$$

Maar je observeert

$$y_{it} = k \text{ als } \tau_k < y_{it}^* \leq \tau_{k+1}, k = 1, \dots, K,$$

waar K het aantal categoriën is die waargenomen worden. Het (enige) verschil met mijn eerder voorgestelde probit model is de veronderstelde “logistic” distributie functie van ε_{it} :

$$F(\varepsilon_{it}|x_{it}, \alpha_i) = F(\varepsilon_{it}) = \frac{1}{1 + \exp(-\varepsilon_{it})} \equiv \Lambda(\varepsilon_{it}),$$

in plaats van de normale verdeling die verondersteld wordt in een probit model. Deze functie wordt vervolgens gebruikt in het opstellen van het logit model:

$$P(y_{it} = k|x_{it}, \alpha_i) = \Lambda(\tau_{k+1} - x'_{it}\beta - \alpha_i) - \Lambda(\tau_k - x'_{it}\beta - \alpha_i).$$

Het voordeel van het logit model is dat ik doormiddel van een soort van “concentrated likelihood” procedure de parameters α_i en τ_k niet meer hoeft te schatten. De

reden hiervoor is vergelijkbaar met de lineaire “within” transformatie die je zou kunnen toepassen in een lineair model. Dit is niet mogelijk met het eerder voorgestelde probit model.

De aanpassing is noodzakelijk omdat R computatieproblemen heeft met de normaalverdeling toegepast op mijn geselecteerde sample, waardoor de schattingen niet convergeren.

Een bijkomend (erg groot) voordeel is dat de schattingen van het logit model consistent zijn (en die van het probit model niet).

Voor zover ik dit nu getest heb, werkt dit naar behoren met het geselecteerde sample.