DREAM

Uddannelse og evne

Peter Stephensen
Den økonomiske modelgruppe DREAM

DREAM Arbejdspapir 2017:1
September 2017

Abstract

I papiret opstilles en simpel model for valg af uddannelse givet en persons evner. Modellen tager udgangspunkt i en population opdelt på et antal uddannelsesgrupper og evne-grupper. Modellen kan anvendes til at modellere skift i populationens uddannelsessammensætning. Den præsenterede metode er relativt let at parametrisere, idet metoden kun forudsætter kendskab til den initiale fordeling af uddannelse og evne i populationen. Metoden vil uden større besværligheder kunne implementeres i en CGE-model.

Den økonomiske modelgruppe DREAM, Amaliegade 44, 1256 København K
www.dreammodel.dk

Uddannelse og evne

Peter Stephensen, DREAM

September 6, 2017

Antag vi betragter en population opdelt på uddannelser $e = 1, ..., E$ og evne-grupper $a = 1, ..., A$. Vi vil opstille en simpel model for valg af uddannelse givet evne og anvende denne til at modellere effekten af et skift i uddannelsessammensætningen. Parametriseringsmæssigt kræver denne metode "kun" kendskab til den initiale fordeling i populationen af uddannelse og evne. Metoden vil relativt let kunne implementeres i en CGE-model.

Antag et en person med evne a skal vælge mellem E uddannelser. Hendes nytte antages at være

$$U_{ae}^0 = u_{ae}^0 + \varepsilon_{ae}$$

Parameteren u_{ae}^0 er det initiale nytte-niveau der er fælles for alle personer med evne a og ε_{ea} er et extreme-value-fordelt restled der beskriver uobserveret heterogenitet i gruppen af personer med evne a[1]. Under disse antagelser vil sandsynligheden for at en person med evne a vælger uddannelsen e være givet ved logit-funktionen:

$$p_{ae}^0 = \frac{e^{u_{ae}^0}}{\sum_s e^{u_{as}^0}}$$

Hvis vi antager at evne-fordelingen er fast og givet ved $\overline{N}_a$ kan vi beregne antallet af personer med uddannelse e og evne a :

$$n_{ae}^0 = \frac{e^{u_{ae}^0}}{\sum_s e^{u_{as}^0}} \overline{N}_a$$

Uddannelses-fordelingen N_e^0 kan da beregnes som:

$$N_e^0 = \sum_a n_{ae}^0 = \sum_a \frac{e^{u_{ae}^0}}{\sum_s e^{u_{as}^0}} \overline{N}_a$$

Antag nu at vi starter med de initiale nytte-niveauer u_{ae}^0 og som følge heraf den initiale uddannelsesfordeling N_e^0. Der sker nu en ændring i uddannelsessammensætningen til N_e, $e = 1, ..., E$. Vi vil gerne beregne de underliggende ændringer i fordelingen n_{ae}.

[1] Extreme-value-fordelingen minder meget om en normalfordeling og anvendes ofte i stedet for normalfordelingen fordi den giver anledning til analytiske løsninger i diskret-valg-problemer.

For at beregne dette bliver vi nødt til at have en teori om hvordan ændringen er sket. Vi antager at den enkelte agents nytte ændrer sig således:

$$U_{ae} = \alpha_e + u_{ae}^0 + \varepsilon_{ae} \tag{1}$$

hvor

$$\sum_e \alpha_e = 0. \tag{2}$$

Vi antager at nytte-ændringerne er additive og kun afhænger af uddannelse. Hvis flere tager en given uddannelse antager vi at det skyldes en additiv evne-uafhængig nytte-stigning der modsvares af tilsvarende nytte-fald for andre uddannelser. Vi tager ikke stilling til hvad nytteændringen skyldes (lønændringer, mode, adgangsbegrænsning osv.). Sum-restriktionen (2) er en harmløs identificerende antagelse der skyldes at hvis man adderer en værdi til alles nytte, da påvirker det ikke adfærden.

Givet en ny uddannelsesfordeling $(N_1,...,N_E)$ skal vi altså finde parametre $(\alpha_1,...,\alpha_E)$ således at

$$n_{ae} = \frac{e^{\alpha_e + u_{ae}^0}}{\sum_s e^{\alpha_s + u_{as}^0}} \overline{N}_a \tag{3}$$

$$N_e = \sum_a n_{ae},\ e = 1,...,E-1 \tag{4}$$

og

$$\sum_e \alpha_e = 0 \tag{5}$$

Bemærk at (3) kan omskrives til

$$n_{ae} = \frac{e^{\alpha_e} n_{ae}^0}{\sum_s e^{\alpha_s} n_{as}^0} \overline{N}_a \tag{6}$$

Dette er det centrale resultat. Den nye fordeling på uddannelse og evne kan beregnes ud fra den initiale fordeling og parametrene $(\alpha_1,...,\alpha_E)$. Det ville være relativt simpelt at indføre denne modellering i en CGE-model.

To interessante forhold kan vises angående relationen (6). Antag at vi har et afstandsmål $d\left((n_{ae}),(n_{ae}^0)\right)$ som måler afstanden mellem den oprindelige fordeling n_{ae}^0 og den nye fordeling n_{ae} og at vi ønsker at finde den fordeling n_{ae} som netop minimerer afstanden mellem den nye og den gamle fordeling givet de lineære restriktioner

$$\sum_e n_{ae} = \overline{N}_a,\ e = 1,...,E$$

og

$$\sum_a n_{ae} = N_e,\ a = 1,...,A$$

I så fald har man modeleret tilpasningen til den nye uddannelsesfordeling som den mindst mulige ændring i den samlede fordeling. Det kan det vises at der findes et ofter anvendt afstandsmål på fordelinger (kaldet Relativ entropi eller Kullbeck-Leibler Information Criterion (KLIC); se fx Stutzer, 1996; Kitamura & Stutzer, 1997; Buchen & Kelly, 1996; Robertson et al., 2005; Stephensen, 2016) som netop giver (6) som løsning.

For det andet kan det vises at der findes en algoritme der overraskende let finder løsningen til (6). Dette er mindre interessant i dette tilfælde idet vi har tænkt os at benytte GAMS til at løse CGE-modellen (se evt. Stephensen, 2016).

Et eksempel

Lad os lave et eksempel på kunstigt data. Vi antager vi har 3 uddannelser og 20 evne-grupper (mange evne-grupper for at få pæne figurer. Typisk ville man nok bruge 10 eller færre). Vi antager at nytte-funktionerne har formen

$$u_{ae}^{0} = \delta_e a + \eta_e$$

og vælger parametrene således at uddannelsessandsynlighederne bliver som vist i Figur 1 (til venstre). En af uddannelserne (den mørke) tiltrækker personer med gode evner (højtuddannede), mens det modsatte er tilfældet for en af de andre uddannelser (den hvide: lavtuddannede). I mellem ligger en gruppe af mellemuddannede.

Den initiale evnefordeling er vist i figuren til højre. Den fede kurve viser populationens samlede evnefordeling (antaget normalfordelt). Sandsynlighederne i figuren til venstre giver anledning til uddannelsesspecifikke evne-fordelinger som vist i figuren til højre. Den røde kurve gælder for de lavtuddannde, den blå kurve gælder for de mellemuddannede og den sorte kurve gælder for de højt uddannede. Disse tre kurver summerer til den samlede fordeling (den fede).

Vi forøger nu antallet af mellemuddannede på bekostning af de lavtuddannede. Den samlede population og overordnede evnefordelig er uændret. Resultatet af et sådant stød er vist i Figur 2. De stiplede kurver viser fordelingen efter støddet.

Referencer

Buchen, P. W., & Kelly, M. (1996). The maximum entropy distribution of an asset inferred from option prices. Journal of Financial and Quantitative Analysis, 31(01), 143–159.

Kitamura, Y., & Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation. Econometrica: Journal of the Econometric Society, 861–874

Robertson, J. C., Tallman, E. W., & Whiteman, C. H. (2005). Forecasting using relative entropy. Journal of Money, Credit, and Banking, 37(3), 383–401.

Stutzer, M. (1996). A simple nonparametric approach to derivative security valuation. The Journal of Finance, 51(5), 1633–1652

Stephensen, P. (2016). Logit Scaling: A General Method for Alignment in Microsimulation models. International Journal of Microsimulation, International Microsimulation Association, vol. 9 (3), side 89-102.

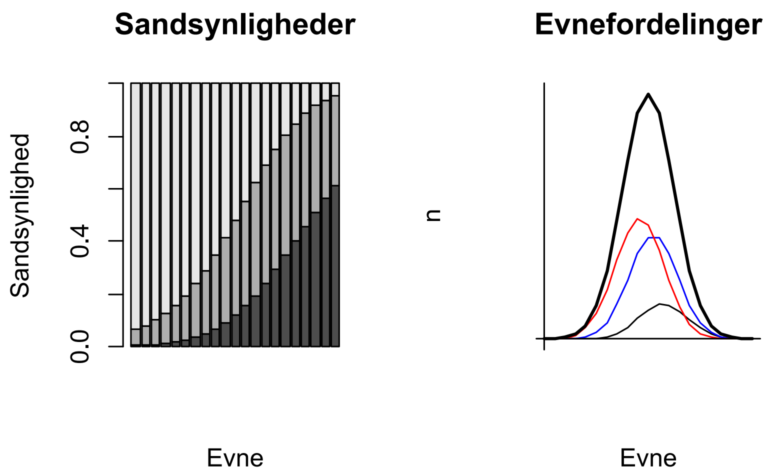

Figure 1: Initiale fordelinger

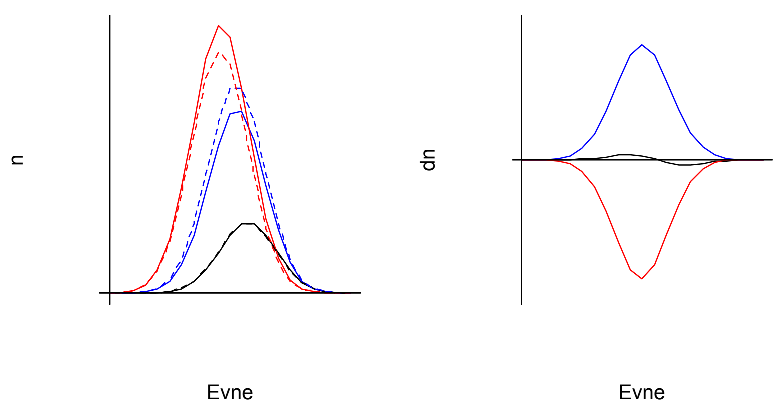

Figure 2: Effekt af ændret uddanelsesfordeling