Interne notater

STATISTISK SENTRALBYRÅ

81/13 25. mars 1981

EN ENKEL ØKONOMETRISK MODELL FOR REGIONALT ARBEIDSTILBUD

Av

Lasse Fridstrøm

INNHOLD

1. BAKGRUNN

I forbindelse med prosjektet DRØM ("Økonomisk-demografisk modellsystem for regional befolkningsfordeling", stat. nr. 6933) er det behov for en delmodell som beregner antall personer i arbeidsstyrken (det "realiserte tilbud av arbeidskraft") i hvert fylke med utgangspunkt i befolkningens størrelse og sammensetning og stramheten på arbeidsmarkedet.

En skisse til en slik delmodell ble framlagt i et notat LEF/AMJ, 15/11-79, jfr. også Brun og Sørensen (1980). Den utforming som der ble valgt reiser imidlertid en del uoverstigelige metodiske problemer, da den forutsetter estimering av en svært komplisert ikke-lineær relasjon.

I dette notatet vil vi legge fram en forenklet utgave bygd på samme tankeskjema, og i detalj beskrive hvordan modellen kan estimeres og testes.

2. EN TEORI FOR TILPASNINGEN PÅ ARBEIDSMARKEDET

En teori om arbeidsmarkedet går ut på at den observerte arbeidsstyrken (sysselsatte + arbeidsledige) varierer i takt med stramheten på arbeidsmarkedet. Når markedet er slapt vil en rekke personer oppgi sin arbeidssøking, idet de ikke har håp om å finne arbeid ("discouraged workers", "hidden unemployment"). Teorien har støtte i empiriske undersøkelser fra bl.a. USA (Mincer 1966 og 1973, Bowen og Finegan 1969, Flaim 1973), Frankrike (Salais 1971), Tyskland (Egle 1979), og også fra Norge (Vannebo 1977, Fridstrøm 1981 a og b). Konsekvensen av dette fenomenet er at dersom en bruker arbeidsstyrken som mål på arbeidstilbudet til enhver tid, så vil en undervurdere både nivået på og endringene i den "egentlige" arbeidsledigheten (åpen+skjult).

Salais (1971) resonnerer som følger: Endringene i den observerte arbeidsstyrken er en funksjon av to ulike mekanismer: en langsiktig utvikling og en kortsiktig tilpasning. Den langsiktige utvikling har sammenheng med endringer i befolkningens sammensetning og størrelse, og med visse langsomme endringer i de enkelte befolkningsgruppers tilbøyelighet til å ta (ønske) lønnet arbeid (generasjonseffekter, holdningsendringer e.l.). Den kortsiktige tilpasning reflekterer stramheten i arbeidsmarkedet, slik at yrkesdeltakingen er høyest når markedet er stramt.

La c være antallet arbeidsledige totalt og la $\gamma = c/b$ være andelen arbeidsledige blant <u>befolkningen</u> i yrkesaktiv alder b. Velg et referansenivå på arbeidsledigheten c^o og tilsvarende $\gamma^o = c^o/b$. Vi vil bruke γ som et mål på stramheten i arbeidsmarkedet. Den arbeidsstyrke som gjelder ved en arbeidsledighet på γ^o vil vi kalle arbeidsstyrken <u>ex ante</u>, z^o. Størrelsen z^o er da renset for konjunkturpåvirkninger og uttrykker kun den langsiktige utvikling i arbeidstilbudet. Den observerte arbeidsstyrken kaller vi z - dette er arbeidsstyrken <u>ex post</u>, dvs. etter at den kortsiktige tilpasningsmekanismen har fått virke. Dersom x betegner sysselsettingen, kan vi skrive

$$(2,1) \qquad z = x + c.$$

La nå p_g være yrkesfrekvensen <u>ex post</u> for persongruppe g, og p_g^o være yrkesfrekvensen <u>ex ante</u>. Vi postulerer nå

$$(2,2) \qquad p_g - p_g^o = k_g(\gamma-\gamma^o) \qquad (k_g<0),$$

altså at avviket mellom yrkesfrekvensene ex post og ex ante er proporsjonalt med forskjellen mellom arbeidsledighetsprosentene ex post og ex ante, men med motsatt fortegn.

Ved å multiplisere (2,2) med folkemengden b_g i aldersgruppe g og summere over g, får vi

$$(2,3) \qquad \sum_g b_g(p_g-p_g^o) = \bar{a}b(\gamma-\gamma^o)$$

eller

(2,4) $z - z^o = a(c-c^o)$

der $a = \sum_g k_g b_g/b$ og $b = \sum_g b_g$. Ved å kombinere (2,1) med (2,4) finner vi

(2,5) $c - c^o = \alpha(z^o-x-c^o)$

(2,6) $z - z^o = -(1-\alpha)\cdot(z^o-x-c^o)$

der $\alpha = \frac{1}{1-a}$. Parameteren α kalles delingskoeffisienten ("coefficient de partage"): En kortsiktig nedgang i sysselsettingen $-\Delta x$ fører til at den observerte arbeidsledigheten øker med en andel $\alpha\Delta x$, mens resten av nedgangen i sysselsetting $(1-\alpha)\Delta x$ tar veien ut av arbeidsstyrken. Omvendt vil en konjunkturbestemt økning i sysselsettingen ikke redusere arbeidsledigheten med mer enn en andel α - resten rekrutteres utenfra arbeidsstyrken.

Estimering av modellen (2,5) - (2,6) byr på visse problemer, idet kun størrelsene c, z og x er observerbare. Ved å legge på restriksjonen

(2,7) $k_g = \eta p_g^o\ (1-p_g^o)$

for alle g og benytte hver ett-årige aldersgruppe g som en (tverrsnitts)observasjon i hvert folketellingsår, samt ty til diverse aritmetiske knep og triks, finner likevel Salais (op.cit.) et estimat for α. Restriksjonen (2,7) innebærer at persongrupper med yrkesdeltaking nær 50% anses være de mest "konjunkturfølsomme". α-estimatet beregnet på denne måten blir ca. 0.30 for hver av folketellingene 1954, 1962 og 1968.

En alternativ strategi vil være å estimere modellen ved hjelp av tidsserier. Salais gjennomførte også slike beregninger og finner et estimat på 0.28 for perioden 1954-1971, m.a.o. meget godt samsvar mellom de to metoder.

Seinere studier (Eymard-Duvernay og Salais 1975, Salais 1977) gir imidlertid et mindre entydig bilde. Koeffisienten α synes å anta vidt forskjellige verdier i nedgangs- og oppgangstider og gir tilsynelatende ikke uttrykk for noen stabil relasjon i det hele tatt. Den grove modellen (2,5) - (2,6) må nyanseres ved at en tar hensyn til bl.a. følgende forhold: Virkningen på arbeidsledigheten er forskjellig alt etter som sysselsettingsøkingen skjer i sekundær- eller tertiær-næringene; delingskoeffisienten er forksjellig ettersom det dreier seg om vekst eller reduksjon i arbeidskraftetterspørselen; arbeidsstyrkens følsomhet overfor konjunktursvingningene synes å avta over tid; de ulike kjønns- og aldersgrupper (eller generasjoner) reagerer ulikt på de forskjellige arbeidsmarkedsimpulser og må behandles separat.

Det står fast at Salais' modell åpner for en måte å definere - og måle - arbeidstilbudet på som trolig er nokså autonom overfor endringer i etterspørselen etter arbeidskraft: Arbeidsstyrken "ex ante" (z^o) er en slik størrelse. Hvilket nivå vi finner for z^o er avhengig av referansenivået (γ^o) på arbeidsledigheten og er for så vidt vilkårlig, men endringene i z^o kan gis en presis tolkning. Det blir mulig å skille mellom endringene i arbeidsstyrken og endringene i arbeidstilbudet; eller, om man vil, å beregne endringene i den skjulte ledighet ut fra endringene i den observerte ledighet.

3. EN STATISK REGIONMODELL FOR NORGE

Istedenfor å anvende Salais' modell som en teori for variasjonen over tid, vil vi anta at en tilsvarende mekanisme gjør seg gjeldende i forholdet mellom ulike regioner på et gitt tidspunkt.

Vi skriver da, tilsvarende (2,1) - (2,6)

(3,1) $z_r = x_r + c_r$

$$(3,2)\quad p_{gr} - p^o_{gr} = k_g\gamma_r \qquad (k_g < 0)$$

$$(3,3)\quad \sum_g b_{gr}(p_{gr}-p^o_{gr}) = \left(\frac{\Sigma k_g b_{gr}}{b_r}\right) b_r\gamma_r = a_r b_r \gamma_r$$

$$(3,4)\quad z_r - z^o_r = a_r c_r$$

$$(3,5)\quad c_r = \alpha_r(z^o_r - x_r)$$

$$(3,6)\quad z_r - z^o_r = -(1-\alpha_r)\cdot(z^o_r - x_r)$$

der

$$(3,7)\quad \alpha_r = (1-a_r)^{-1} = (1 - \sum_g k_g b_{gr}/b_r)^{-1}.$$

Her har vi valgt $\gamma^o = 0$, dvs. at p^o_{gr} får en tolkning som yrkesfrekvensen ved null arbeidsledighet ($c^o_r=0$). Det er vel ikke unaturlig å betrakte p^o_{gr} som et mål på tilbudet av arbeid.

Merk (likning (3,2)) at "konjunkturfølsomhetsparameteren" for persongruppe g er den samme for alle regioner, mens ex ante yrkesdeltaking p^o_{gr} varierer både over persongrupper og region. Dersom alle persongrupper var helt homogene, ville vi kunne anta at p^o_{gr} var lik for alle r. Dette er imidlertid lite realistisk. Uansett hvilke observerbare kjennetegn en nytter til persongruppeinndelingen, må en anta at de mest arbeidsmarkedsorienterte innen gruppen i noen grad vil flytte til områder med stramt arbeidsmarked, slik at p^o_{gr} synker i fraflyttingsregionen og (sannsynligvis) øker i tilflyttingsregionen. Vi venter altså å finne høyere p^o_{gr} i Oslo enn i Finnmark.

Modellen fungerer slik: Vi må kjenne variabelverdiene x_r og b_{gr} og parameterverdiene k_g og p^o_{gr}. Vi kan da regne ut

$$(3,8)\quad z^o_r = \sum_g b_{gr} p^o_{gr}$$

og α_r ifølge (3,7). Ved innsetting i (3,5) og (3,6) kan vi nå finne arbeidsstyrken ex post z_r og arbeidsledigheten ex post c_r for hver region. Videre finner vi

$$(3,9)\quad \gamma_r = \frac{c_r}{b_r}$$

og kan, om ønskelig, regne ut yrkesfrekvensene ex post som løsningen av (3.2) m.h.p. p_{gr}.

Variabelen x_r forutsettes bestemt "fra etterspørselssiden" i modellsystemet, og variablene b_{gr} gjennom den demografiske delmodellen. Parametrene k_g og p^o_{gr} kan estimeres (og testes) som følger.

4. ESTIMERING OG TESTING

Anta at vi har et tilfeldig utvalg fra befolkningen i alle regioner. La

$$(4,1)\quad y_i = \begin{cases} 1 & \text{dersom individ i er i arbeidsstyrken} \\ 0 & \text{ellers} \end{cases}$$

Definer dessuten

$$(4,2)\qquad \delta_{ir} = \begin{cases} 1 & \text{dersom individ i bor i region r} \\ 0 & \text{ellers} \end{cases}$$

$$(4,3)\qquad \varepsilon_{ig} = \begin{cases} 1 & \text{dersom individ i tilhører persongruppe g} \\ 0 & \text{ellers} \end{cases}$$

y_i er da Bernoulli-fordelt med parameter

$$(4,4)\qquad p_i = E(y_i) = \sum_{gr}\sum \varepsilon_{ig}\delta_{ir}p_{gr} = \sum_{gr}\sum \varepsilon_{ig}\delta_{ir}\cdot\left[p^o_{gr} + k_g\gamma_r\right].$$

Variabelen δ_{ir} "plukker ut" de <u>regionkjennetegn</u> som gjelder for individ i, mens ε_{ig} gjør samme jobben med <u>persongruppekjennetegn</u>. $\sum_g \varepsilon_{ig}\sum_r \delta_{ir}p^o_{gr} = p^o_{hs}$ hvis og bare hvis individ i bor i region s og tilhører gruppe h, osv.

Dersom regioninndelingen er den samme for ex ante-yrkesfrekvensene p^o_{gr} og arbeidsledighetsratene γ_r, er parametrene p^o_{gr} og k_g ikke identifiserbare. Vi velger derfor for estimeringsformål <u>fylke</u> som regional enhet for p^o_{gr} og <u>p.p.-region</u> for arbeidsledighetsratene. Dersom befolkningens sammensetning etter persongrupper (g) ikke er den samme i alle p.p.-regioner innen hvert fylke, begår vi her en (mindre) aggregeringsfeil.

La γ^*_i betegne arbeidsledighetsraten i den p.p.-region der individ i er bosatt. (4,4) kan da omskrives til

$$(4,5)\qquad p_i = \sum_g \varepsilon_{ig}\sum_r \delta_{ir}p^o_{gr} + \gamma^*_i\sum_g \varepsilon_{ig}k_g\,.$$

Tettheten til y_i kan skrives

$$(4,6)\qquad f(y_i) = p_i^{y_i}(1-p_i)^{1-y_i}\,.$$

Idet vi antar at alle observasjoner er uavhengige, blir log-sannsynlighetsfunksjonen for hele utvalget (n individer) lik

$$(4,7)\qquad \log L = \sum_{i=1}^{n} f(y_i) = \sum_{i=1}^{n}\left[y_i\cdot\log p_i + (1-y_i)\log(1-p_i)\right]\,.$$

De deriverte er

$$(4,8)\qquad \frac{\partial \log L}{\partial\theta_j} = \sum_{i=1}^{n}\left\{\frac{\partial p_i}{\partial\theta_j}\cdot\left[\frac{y_i}{p_i} - \frac{1-y_i}{1-p_i}\right]\right\} \qquad (\theta_j = p^o_{11},\ \ldots,\ p^o_{GR},\ k_1,\ldots,\ k_G)$$

der p_i er gitt ved (4,5), og

$$(4,9)\qquad \frac{\partial p_i}{\partial p^o_{gr}} = \varepsilon_{ig}\delta_{ir} \qquad (g=1,\ 2,\ \ldots,\ G;\ r=1,\ 2,\ \ldots,\ R)$$

$$(4,10)\qquad \frac{\partial p_i}{\partial k_g} = \varepsilon_{ig}\gamma^*_i \qquad (g=1,\ 2,\ \ldots,\ G).$$

G og R betegner henholdsvis antall persongrupper og antall regioner. Langt de fleste av disse deriverte blir lik null. Systemet

$$(4,11)\quad \frac{\partial \log L}{\partial \theta_j} = 0 \quad (\theta_j = p^o_{gr},\ k_g;\ g=1,2,..,G;\ r=1,2,..,R)$$

(jfr. (4,8) - (4,10)) danner et ikke-lineært system av $G\cdot(R+1)$ likninger i like mange parametre. Løsningen av likningssystemet er sannsynlighetsmaksimeringsestimatene av p^o_{gr} og k_g.

De annen-deriverte er lik

$$(4,12)\quad \frac{\partial^2 \log L}{\partial p^o_{gr}\partial p^o_{hs}} = \sum_{i=1}^{n} \left\{ \frac{\partial^2 p_i}{\partial p^o_{gr}\partial p^o_{hs}} \left[\frac{y_i}{p_i} - \frac{1-y_i}{1-p_i} \right] - \frac{\partial p_i}{\partial p^o_{gr}} \cdot \frac{\partial p_i}{\partial p^o_{hs}} \cdot q_i \right\} = 0 \qquad \text{for } g\neq h \text{ og/eller } r\neq s,$$

der vi for korthets skyld har skrevet

$$(4,13)\quad q_i = \frac{y_i}{p_i^2} + \frac{1-y_i}{(1-p_i)^2}.$$

q_i er lik p_i^{-2} når $y_i = 1$ og lik $(1-p_i)^{-2}$ når $y_i = 0$. Videre (husk at $\varepsilon_{ig}=\varepsilon_{ig}^2$ og $\delta_{ir}=\delta_{ir}^2$):

$$(4,14)\quad \frac{\partial^2 \log L}{\partial p^o_{gr}\partial p^o_{gr}} = -\sum_{i=1}^{n} \left[\frac{\partial p_i}{\partial p^o_{gr}}\right]^2 \cdot q_i = -\sum_{i=1}^{n} \varepsilon_{ig}\delta_{ir}q_i \equiv \pi_{gr} \ (\text{def.})$$

$$(4,15)\quad \frac{\partial^2 \log L}{\partial p^o_{gr}\partial k_h} = -\sum_{i=1}^{n} \frac{\partial p_i}{\partial p^o_{gr}} \cdot \frac{\partial p_i}{\partial k_h} q_i = 0 \qquad \text{for } g \neq h$$

$$(4,16)\quad \frac{\partial^2 \log L}{\partial p^o_{gr}\partial k_g} = -\sum_{i=1}^{n} \varepsilon_{ig}\delta_{ir}q_i\gamma_i^* \equiv \lambda_{gr} \ (\text{def.})$$

$$(4,17)\quad \frac{\partial^2 \log L}{\partial k_g\partial k_h} = -\sum_{i=1}^{n} \frac{\partial p_i}{\partial k_g} \cdot \frac{\partial p_i}{\partial k_h} q_i = 0 \qquad \text{for } g \neq h$$

$$(4,18)\quad \frac{\partial^2 \log L}{\partial k_g\partial k_g} = -\sum_{i=1}^{n} \left[\frac{\partial p_i}{\partial k_g}\right]^2 q_i = -\sum_{i=1}^{n} \varepsilon_{ig}q_i(\gamma_i^*)^2 \equiv \kappa_g \ (\text{def.}).$$

Matrisen av annen-deriverte kan skrives

$$(4,19)\quad M = \begin{bmatrix} \Pi & \Lambda' \\ \Lambda & K \end{bmatrix}$$

der

$$(4,20)\quad \Pi = \begin{bmatrix} \Pi_1 & 0 & \dots 0 \\ 0 & \Pi_2 & \dots 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & \dots \Pi_R \end{bmatrix}$$

og delmatrisene Π_r igjen er gitt ved

(4,21) $\Pi_r = \text{diag}\ (\pi_{1r}\ \pi_{2r} \ldots \pi_{Gr}) \quad (r=1,2,..,R).$

Π er således en RG x RG diagonal matrise bestående av alle π_{gr}. K er tilsvarende den diagonale matrisen av $\kappa_1, \kappa_2, .., \kappa_G$, mens Λ består av R diagonale blokker

(4,22) $\Lambda = \left[\Lambda_1\ \Lambda_2 \ldots \Lambda_R\right],$

der

(4,23) $\Lambda_r = \text{diag}\ (\lambda_{1r}\ \lambda_{2r} \ldots \lambda_{Gr}) \quad (r=1,2,...,R).$

Informasjonsmatrisen er gitt ved

(4,24) $\Psi = -E(M) = -M\Big|_{y_i=p_i} = -M\Big|_{q_i=\left[p_i(1-p_i)\right]^{-1}}.$

Et konsistent estimat for Ψ er gitt ved

(4,25) $\hat{\Psi} = -M\Big|_{q_i=\left[\hat{p}_i(1-\hat{p}_i)\right]^{-1}},$

der $\hat{p}_i$ er den predikerte yrkesdeltakingssannsynlighet for individ i, dannet ved innsetting av sannsynlighetsmaksimeringsestimatene $\hat{p}^o_{rg}$ og $\hat{k}_g$ i (4,5).

La Σ betegne kovariansmatrisen for estimatorene $\hat{p}^o_{gr}$ og $\hat{k}_g$, ordnet på samme måte som M (første indeks g varierer raskest). Siden sannsynlighetsmaksimeringsestimatorene er asymptotisk effisiente (og normalfordelte), er

(4,26) $\hat{\Sigma} = \hat{\Psi}^{-1}$

en konsistent estimator for Σ, (jfr. Fridstrøm 1980, avsn. A.3.6.-A.3.7). $\hat{\Sigma}$ kan om ønskelig regnes ut slik:

(4,27) $\hat{\Sigma} = -M^{-1}\Big|_{q_i=\left[\hat{p}_i(1-\hat{p}_i)\right]^{-1}},$

der

(4,28) $M^{-1} = \begin{bmatrix} S & -S\Lambda' K^{-1} \\ -K^{-1}\Lambda S & T \end{bmatrix}$

(4,29) $T = \left[K - \Lambda\Pi^{-1}\Lambda'\right]^{-1} = \left[K - \sum_{r=1}^{R} \Lambda_r^2 \Pi_r^{-1}\right]^{-1}$

og

(4,30) $S = \Pi^{-1} + \Pi^{-1}\Lambda' T \Lambda \Pi^{-1}$

(se Theil (1971, s. 18)). T er diagonal.

La Ξ være en νxG(R+1) matrise med rang ν, og la $\underset{\sim}{\xi}$ være en G(R+1)-dimensjonal vektor. La dessuten $\underset{\sim}{\theta}$ betegne vektoren av alle G(R+1) parametre, ordnet på samme måte som M. Uttrykket

$$(4,31)\quad H_0 : \Xi\underset{\sim}{\theta} = \underset{\sim}{\xi}$$

betegner da ν lineært uavhengige hypoteser om parametrene

$$(4,32)\quad \underset{\sim}{\theta}' = (\theta_1\ \theta_2 \ldots \theta_{G(R+1)}) = (p_{11}^o\ p_{21}^o \ldots p_{GR}^o\ k_1 \ldots k_G).$$

En asymptotisk test av H_0 mot

$$(4,33)\quad H_1 : \Xi\underset{\sim}{\theta} \neq \underset{\sim}{\xi}$$

på nivå α består i å forkaste H_0 dersom

$$(4,34)\quad (\Xi\hat{\underset{\sim}{\theta}}-\underset{\sim}{\xi})'\left[\Xi\hat{\Sigma}\Xi'\right]^{-1}(\Xi\hat{\underset{\sim}{\theta}}-\underset{\sim}{\xi}) > \chi_\nu^2(1-\alpha).$$

Her er $\hat{\underset{\sim}{\theta}}$ vektoren av sannsynlighetsmaksimeringsestimater og $\chi_\nu^2(1-\alpha)$ betegner $(1-\alpha)$-fraktilen i kji-kvadrat-fordelingen med ν frihetsgrader (jfr. Fridstrøm 1980, avsnitt 4.6, 5.4 og A.3).

Av spesiell interesse for oss er hypoteser av formen

$$(4,35)\quad H_{rs} : \underset{\sim}{p}_{.r}^o = \underset{\sim}{p}_{.s}^o \quad (r \neq s)$$

der $\underset{\sim}{p}_{.r}^o = (p_{1r}^o\ p_{2r}^o \ldots p_{GR}^o)'$, altså at ex-anteyrkesfrekvensene er de samme <u>for alle persongrupper</u> i to ulike regioner. Det finnes i alt $\binom{R}{2} = R(R-1)/2$ par av regioner som kan testes mot hverandre, mens antallet <u>uavhengige</u> hypoteser av dette slaget ikke er større enn R-1.

For testing av slike hypoteser som (4,35) er det nok å arbeide med øvre venstre del av matrisen $\hat{\Sigma}$, med andre ord med

$$(4,36)\quad -\hat{S} = -\hat{\Pi}^{-1} - \hat{\Pi}^{-1}\hat{\Lambda}'\hat{T}\hat{\Lambda}\hat{\Pi}^{-1} = -\hat{\Pi}^{-1} - \hat{\Pi}^{-1}\hat{\Lambda}'(\hat{K}-\hat{\Lambda}\hat{\Pi}^{-1}\hat{\Lambda}')^{-1}\hat{\Lambda}\hat{\Pi}^{-1},$$

der tegnet ^ som vanlig betyr at vi har satt inn for sannsynlighetsmaksimeringsestimatene $\hat{q}_i = \left[\hat{p}_i(1-\hat{p}_i)\right]^{-1}$, jfr. (4,27) - (4,30). Matrisen Ξ består her av en rad av R+1 blokker av dimensjon GxG, der blokk nr. r er identitetsmatrisen, blokk nr. s er minus identitetsmatrisen og de andre blokkene består av bare nuller. Vektoren $\underset{\sim}{\xi}$ blir lik null.

Testkriteriet (4,34) kan da skrives

$$(4,37)\quad \chi_G^2(1-\alpha) < -(\hat{\underset{\sim}{p}}_{.r}^o-\hat{\underset{\sim}{p}}_{.s}^o)'\left[\hat{\Pi}_r^{-1} + \hat{\Pi}_s^{-1} + (\hat{\Pi}_r^{-1}\hat{\Lambda}_r-\hat{\Pi}_s^{-1}\hat{\Lambda}_s)^2\hat{T}\right]^{-1}(\hat{\underset{\sim}{p}}_{.r}^o-\hat{\underset{\sim}{p}}_{.s}^o)$$

$$= -\sum_{g=1}^{G} \frac{(\hat{p}_{gr}^o-\hat{p}_{gs}^o)^2}{\hat{\pi}_{gr}^{-1} + \hat{\pi}_{gs}^{-1} + \left[(\hat{\pi}_{gr}^{-1}\hat{\lambda}_{gr}-\hat{\pi}_{gs}^{-1}\hat{\lambda}_{gs})^2/(\hat{\kappa}_g-\sum_{t=1}^{R}\hat{\lambda}_{gt}^2\hat{\pi}_{gt}^{-1})\right]}$$

idet alle matrisene $\hat{\Pi}_r$, $\hat{\Pi}_s$, $\hat{\Lambda}_r$, $\hat{\Lambda}_s$ og $\hat{T}$ er diagonale.

Variansestimater for de enkelte parameterestimatorer er gitt ved

$$(4,38)\quad \hat{\text{var}}(\hat{p}_{gr}^o) = -\hat{\pi}_{gr}^{-1} - \left[\hat{\pi}_{gr}^{-2}\cdot\hat{\lambda}_{gr}^2/(\hat{\kappa}_g-\sum_{t=1}^{R}\hat{\lambda}_{gt}^2\hat{\pi}_{gt}^{-1})\right]$$

$$(4,39)\quad \hat{\text{var}}\ (\hat{k}_g) = 1/(\hat{\kappa}_g-\sum_{t=1}^{R}\hat{\lambda}_{gt}^2\hat{\pi}_{gt}^{-1})$$

5. DATA

Det er opprettet en file, heretter kalt file C, med de data som trenges for å gjennomføre estimeringen.

Grunnlaget for file C er AKU 1977, der vi imidlertid har valgt ut halvparten av materialet, på en slik måte at ingen intervjuobjekter er med mer enn én gang. Filen inneholder i underkant av 20 000 records.

De ulike variablene som inngår i sannsynlighetsfunksjonen (4,5) har en posisjon på filen som framgår av tabell 1.

Som regional enhet har vi (forsøksvis) valgt fylke, med Oslo og Akershus sammenslått, dvs. R=18. Persongruppeinndelingen følger Arbeidsdirektoratets "10 homogene arbeidskraftgrupper" (dvs. G=10), med noen tillempinger. (Omfatter bare personer 16 - 74 år; en del pensjonister, trygdede og studenter grupperes trolig feil.)

Variablene $\delta_{i1}, .., \delta_{iR}$ dannes ved omkoding av fylkesnummeret. Variablene $\varepsilon_{i1}, .., \varepsilon_{iG}$ står ferdig omkodet på filen.

Variabelen γ_i^* krever en litt mer komplisert omkoding, idet vi på filen kun har arbeidsledige i promille av en beregnet arbeidsstyrke i p.p.-regionen, mens modellen forutsetter ledigheten regnet som andel av folkemengden 16-74 år i p.p.-regionen. Denne siste størrelsen må hentes fra befolkningsstatistikken.

Registrert ledighet i p.p.-regionen bør dessuten justeres med en korreksjonsfaktor $\frac{m_i}{1000}$, lik for alle p.p.-regioner i samme fylke, for å sikre konsistens mellom yrkesdeltaking etter AKU's definisjon og arbeidsledighetsmålet γ_i^*. Variabel m_i står på file C, og er beregnet som et gjennomsnitt for årene 1976 - 1979.

Tabell 1. Plassering av variable på file C.

Variabel		Posisjon	Format	Omkoding
y_i	Dummy for yrkesdeltaking	16	I1	
f_i	Fylkesnummer	221 - 222	I2	
δ_{i1}	Dummy for Østfold			= 1 hvis f_i = 01
δ_{i2}	" " Oslo/Akershus			= 1 hvis f_i = 02 el. 03
δ_{i3}	" " Hedmark			=1 hvis f_i = 04
:	:			: :
δ_{i11}	" " Hordaland			= 1 hvis f_i = 12
δ_{i12}	" " Sogn og Fjordane			= 1 hvis f_i = 14
:	:			: :
δ_{i18}	" " Finnmark			= 1 hvis f_i = 20
ε_{i1}	" " persongruppe 1	851	I1	
ε_{i2}	" " " 2	852	I1	
:	:	:	:	
ε_{i10}	" " " 10	860	I1	
a_i	Antall registrert arbeidsledige i p.p.-regionen i undersøkelsesmåneden, pr. 1000 arbeidsstyrke	140 - 143	I4	
z_i	Beregnet arbeidsstyrke i p.p.-reg.	209 - 214	I6	
m_i	Korreksjonsfaktor for arbeidsledighet i fylket. Arbeidssøkere ifølge AKU pr. 1 000 registrerte ledige.	870 - 873	I4	
b_i	Folkemengde 16-74 år i p.p.-reg.	Hentes fra befolkningsstatistikken pr. 31/12-77. Koples via p.p.-reg.-nr., pos. 243-246, til individrecorden		
γ_i^*	Arbeidsledighetsrate i p.p.-reg.			$= \frac{z_i \cdot a_i \cdot m_i}{1000 \cdot 1000 \cdot b_i}$

6. PRAKTISK GJENNOMFØRING

Beregningene foreslås gjennomført ved hjelp av programbiblioteket NAG. En kan enten benytte rutiner for maksimering av en funksjon (4,7), der både 1.- og 2.-deriverte kan utnyttes, eller en algoritme for løsning av det ikke-lineære systemet (4,11).

Dersom R=18 og G=10 inneholder systemet (R+1)G = 190 parametre å estimere simultant. Dette kan vise seg å være for mye selv for NAG/Honeywell Bull. I så fall bør en forsøke med en grovere regioninndeling for parametrene p_{gr}^{o}. Dette oppnår en ganske enkelt ved å summere grupper av fylkesvariablene δ_{ir} til nye regiondummies δ_{ir}^{*}, summen $\delta_{i16} + \delta_{i17} + \delta_{i18}$ blir f.eks. en dummy for "Nord-Norge". Ved f.eks. å redusere antall regioner til 5 får vi antall parametre ned i 60.

Gangen i analysen blir da som følger:

(1) Maksimer (4,7) m.h.p. p_{gr}^{o} og k_g ($r=1,2, .., R$; $g=1,2, .., G$), dvs. løs likningssystemet (4,11).

(2) Sett inn estimatene $\hat{p}_{gr}^{o}$ og $\hat{k}_g$ i funksjon (4,5) og regn ut $\hat{p}_i$ for alle individer i i utvalget.

(3) Regn ut $\hat{q}_i = \left[\hat{p}_i(1-\hat{p}_i)\right]^{-1}$ for alle i.

(4) Regn ut de annenderiverte $\hat{\pi}_{gr}$ (4,14), $\hat{\lambda}_{gr}$ (4,16) og $\hat{\kappa}_g$ (4,18) for alle g og r (bruk $q_i = \hat{q}_i$).

(5) For hvert par (r,s) av regioner, regn ut testobservatoren (4,37) ($\binom{R}{2}=R(R-1)/2$ observatorer).

(6) Regn ut standardavvik for parameterestimatene ved å bruke formlene (4,38) og (4,39), og ta kvadratroten.

(7) Vurder, på grunnlag av pkt. (5) og (6), om enkelte regioner bør slås sammen for å oppnå mer presise estimater. Gjenta eventuelt pkt. (1) - (7) med ny regioninndeling.

Merk at uansett hvilken regioninndeling som nyttes ved estimeringen, vil resultatene kunne anvendes på fylkesnivå, eller for den inndeling (grovere enn fylke) som DRØM måtte benytte.

Sannsynlighetsmaksimeringsmetoden sikrer at alle $\hat{p}_i$ ligger innenfor $[0,1]$-intervallet. Vi kan imidlertid risikere å få $\hat{p}_{gr}^{o}<0$ eller $\hat{p}_{gr}^{o}>1$.

En alternativ og (på visse vilkår) enklere estimeringsmetode ville være vanlig minste kvadraters regresjon. Denne metoden har imidlertid to hovedsvakheter: (1) vi risikerer $\hat{p}_i>1$ eller $\hat{p}_i<0$, og (2) enhver form for testing er vanskelig, da restleddene verken er homoskedastiske eller normalfordelte. Vanlige standard regresjonsprogrammer gir derfor galt mål på presisjonen. Utregning av de "riktige" variansestimatene er en heller komplisert affære, jfr. Fridstøm (1980, likning (4.69)-(4.71)), og er dessuten bare mulig dersom vi er heldige og får $0<\hat{p}_i<1$ for alle i. Vanlig minste kvadraters metode kan imidlertid være en siste utvei.

7. HVA VENTER VI Å FINNE?

Alle "konjunkturfølsomhetsparametre" $\hat{k}_g$ bør være negative. Tallverdien vil ventelig være minst for grupper som ligger nær 0 eller 100 prosent yrkesdeltaking.

Vi venter gjennomgående høyere ex-ante-frekvenser $\hat{p}_{gr}^{o}$ i pressområder (tilflyttingsområder) enn i utkantdistrikt (fraflyttingsområder). Ellers venter vi ikke svært store variasjoner i $\hat{p}_{gr}^{o}$ for ulike r, men derimot store forskjeller for ulike g.

LITTERATUR

Bowen, William G. og T. Aldrich Finegan (1969): "The economics of labor force participation". Princeton University Press, Princeton.

Brun, Stein Erland og Knut Ø. Sørensen (1980): "Økonomisk-demografisk modellsystem for regional befolkningsfordeling. Årsrapport for 1979". IN 80/4, Statistisk Sentralbyrå.

Egle, Franz (1979): "Ansätze für eine systematische Beobachtung und Analyse der Arbeitslosigkeit". Beitr AB 36, Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit, Nürnberg.

Eymard-Duvernay, F. og R. Salais (1975): "Une analyse des liens entre l'emploi et le chômage". Economie et Statistique, no. 69.

Flaim, Paul O. (1973): "Discouraged workers and changes in unemployment". Monthly Labor Review, March 1973.

Fridstrøm, Lasse (1980): "Lineære og loglineære modeller for kvalitative avhengige variable". RAPP 80/26, Statistisk Sentralbyrå.

Fridstrøm, Lasse (1981 a): "Framskriving av arbeidsstyrken 1979 - 2000". RAPP 81/xx (under arbeid), Statistisk Sentralbyrå.

Fridstrøm, Lasse (1981 b): "Logit and tobit models of individual labour supply". Statistisk Sentralbyrå (under arbeid).

Mincer, Jacob (1966): "Labor Force Partipation and Unemployment: A Review of Recent Evidence" i R. A. Gordon & M. S. Gordon (eds.): "Prosperity and Unemployment". Wiley, New York.

Mincer, Jacob (1973): "Determining who are the 'hidden unemployed'". Monthly Labor Review, March 1973.

Salais, R. (1971): "Sensibilité de l'activité par sexe et âge aux variations du chômage". Annales de l'INSEE, no. 8.

Salais, R. (1977): "Analyse des mécanismes de détermination du chômage". Economie et Statistique, no. 93.

Theil, Henri (1971): "Principles of econometrics". Wiley, New York.

Vannebo, Olav (1977): "Regionale forskjeller i yrkesdeltakingen 1970". ART 92, Statistisk Sentralbyrå.