

Jo Thori Lind

**Testing av stokastiske
individuelle effekter i
paneldatamodeller**

1 Innledning¹

Paneldata er i mange sammenhenger nyttig siden de gjør det mulig å ta hensyn til individ-spesifikke effekter uten eksplisitt å forklare dem. Derimot er spesielle estimatorene som tar hensyn til panel-strukturen i data ofte noe mere kompliserte enn de ordinære estimatorene. I tillegg til at det kan være av interesse i seg selv er derfor dette en grunn til at det er nyttig å kunne teste om det er tegn i data på at det er individ-spesifikke effekter som gjør seg til kjenne.

Når vi ser på statiske balanserte panel uten målefeil er det to modeller som dominerer litteraturen; modeller med faste effekter og modeller med tilfeldige effekter. I førstnevnte klasse av modeller estimeres (ofte implisitt) en parameter for hvert individ. En test for individuelle effekter kan da konstrueres som en enkel test for den samlede signifikansen av de individ-spesifikke parametrene basert på Wald-testen. I modeller med tilfeldige effekter antar vi at den individspesifikke effekten er gitt ved en stokastisk variabel som vi ikke forsøker å estimere verdien til. I disse modellene er tester for individuelle effekter noe mere komplisert, og dette er tema i dette essayet.

Avsnitt 2 presenterer en paneldatamodel med tilfeldige effekter, mens avsnitt 3 presenterer ulike tester for tilstedeværelse av individuelle effekter. Denne modellen anvendes så på norske forbruksdata i seksjon 4, og resultatene drøftes og utvides i seksjon 5. Avsnitt 6 oppsummerer.

2 En paneldatamodel med tilfeldige effekter

Vi skal her se på en økonometrisk modell som forsøker å forklare budsjettandelen for mat. Vi skal se på N individer indeksert med i som alle observeres over T perioder indeksert med t . Vi skal bare ta for oss balanserte paneler, det vil si at alle individene er observert i alle periodene, men modellene kan utvides til ubalanserte tilfeller også. Vi skal anta modellen

$$y_{it} = x_{it}\beta + u_{it} \tag{1}$$

$$u_{it} = \mu_i + \nu_{it}$$

¹Dette notater er en omarbeidet utgave av et essay til Erik Biørns kurs "Økonometri - Analyse av paneldata" ved UiO høsten 2000. Jeg er takknemlig for svært nyttige kommentarer og innspill fra Jørgen Aasness.

hvor y_{it} er individ i s budsjettandel på mat i periode t og x_{it} er forklaringsvariabler for individ i i periode t . Vi skal anta at x_{it} inneholder et konstantledd. Restleddet u_{it} består av en individ-spesifikk komponent $\mu_i \sim \text{uif}(0, \sigma_\mu^2)$ og et "rent" restledd $\nu_{it} \sim \text{uif}(0, \sigma_\nu^2)$ hvor uif står for "uavhengig identisk fordelt". Vi skal anta at μ_i og ν_{it} er stokastisk uavhengige av x_{it} . Modellen kunne også utvides til å inkludere en periode-spesifikk komponent, men det er utenfor rammene av dette essayet. Definerer vi nå $y_i = (y_{i1}, \dots, y_{iT})'$, $x_i = (x'_{i1}, \dots, x'_{iT})'$ og $u_i = (u_{i1}, \dots, u_{iT})'$ kan vi skrive (1) som

$$y_i = x_i \beta + u_i \quad (2)$$

hvor² $Eu_i = 0$ og $\text{Var}(u_i) = \sigma_\mu^2 (e_T e_T') + \sigma_\nu^2 I_T$. Her er e_T en T -vektor av enere og I_T er den T -dimensjonale identitetsmatrisen. Setter vi nå alle individ-vektorene sammen, kan vi definere $Y = (y'_1, \dots, y'_N)'$, $X = (x'_1, \dots, x'_N)'$ og $u = (u'_1, \dots, u'_N)'$. Nå kan (1) skrives som

$$Y = X\beta + u \quad (3)$$

hvor $Eu = 0$ og $\text{Var}(u) = \sigma_\mu^2 (I_N \otimes e_T e_T') + \sigma_\nu^2 I_{NT} \equiv \Omega$. Vi kan bruke minste kvadraters metode (OLS) på (3), men denne estimatoren vil ikke være BLUE. Det vil derimot generalisert minste kvadraters metode (GLS), definert som

$$\hat{\beta}^{GLS} = (X\Omega^{-1}X)^{-1} X\Omega^{-1}Y, \quad (4)$$

være. Vi trenger altså et uttrykk for Ω^{-1} . La oss definere $A_T = \frac{1}{T}e_T e_T'$ og $B_T = I_T - A_T$. Vi ser at A_T og B_T er symmetriske, idempotente og oppfyller $A_T B_T = B_T A_T = 0$. Videre kan vi skrive $\Omega = I_N \otimes [T\sigma_\mu^2 A_T + \sigma_\nu^2 (A_T + B_T)] = (T\sigma_\mu^2 + \sigma_\nu^2) I_N \otimes A_T + \sigma_\nu^2 I_N \otimes B_T$, som er en spektraldekomponering av Ω . Det følger da at

$$\Omega^{-1} = I_N \otimes \left(\frac{1}{T\sigma_\mu^2 + \sigma_\nu^2} A_T + \frac{1}{\sigma_\nu^2} B_T \right), \quad (5)$$

så GLS-estimatoren kan skrives som

$$\hat{\beta}^{GLS} = (X'WX + \theta X'VX)^{-1} (X'WY + \theta X'VY) \quad (6)$$

hvor $W = I_N \otimes B_T$, $V = I_N \otimes A_T$ og

$$\theta = \frac{\sigma_\nu^2}{\sigma_\nu^2 + T\sigma_\mu^2}. \quad (7)$$

²Merk at 0 betegner en null-vektor eller -matrise av dimensjon som vil framgå av sammenheng.

Estimatoren (6) krever kjennskap til σ_μ^2 og σ_ν^2 , noe som vi stort sett ikke har. Vi ser at selv om OLS er en ineffisient estimator på (1) er den allikevel konsistent under vide betingelser, så vi kan her få konsistente estimater på restleddene. Disse kan vi så bruke til å estimere σ_μ^2 og σ_ν^2 ved bruk av følgende estimatorene (se f. eks. Biørn (2000) eller Mátyás (1992)):

$$\hat{\sigma}_\nu^2 = \frac{\hat{u}'W\hat{u}}{N(T-1) - (K-1)} \quad (8)$$

$$\hat{\sigma}_\mu^2 = \frac{\hat{u}'\hat{u}}{NT - K} - \hat{\sigma}_\nu^2 \quad (9)$$

hvor K er antall regressorer inkludert konstantleddet og \hat{u} er vektoren av residualer definert ved $\hat{u} = Y - X\hat{\beta}^{OLS}$. Bruker vi så disse estimatene i likning (6) får vi den såkalte Feasible GLS (FGLS) estimatoren. Vi kan også gjenta operasjonen og fra FGLS-estimatoren lage nye estimater for \hat{u} , som vi igjen bruker til å reestimere variansene til feilkomponentene ved hjelp av (8) og (9), som vi igjen kan bruke i en ny FGLS-estimering. Hvis μ_i og ν_{it} er normalfordelte vil denne iterasjonen konvergere mot maximum likelihood (ML)-estimatoren. I dette tilfellet vil log likelihood-funksjonen være

$$\begin{aligned} \ln L(\beta, \sigma_\mu^2, \sigma_\nu^2; X) &= -\frac{NT}{2} \ln(2\pi) - \frac{1}{2} \ln |\Omega| - \frac{1}{2} (Y - X\beta)' \Omega^{-1} (Y - X\beta) \quad (10) \\ &= -\frac{NT}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma_\nu^2) - \frac{N(T-1)}{2} \ln(\sigma_\nu^2 + T\sigma_\mu^2) \\ &\quad - \frac{1}{2} (Y - X\beta)' \Omega^{-1} (Y - X\beta). \end{aligned}$$

Informasjonsmatrisen er i dette tilfellet blokk-diagonal mellom β og variansene, som gjør at vi kan maksimere funksjonen ved først og maksimere med hensyn på β , så med hensyn til variansene, så β osv.

3 Tester for fravær av individuelle effekter

Vi ønsker her å foreta en statistisk test for å undersøke om det virkelig er individuelle forskjeller i Engel-kurvene, det vil si om μ_i er forskjellig fra individ til individ. Nullhypotesen H_0 om ingen individuelle effekter er da $\sigma_\mu^2 = 0$. Alternativhypotesen H_1 er at det er individuelle forskjeller, dvs. $\sigma_\mu^2 > 0$. Det er enkelte problemer ved å teste denne hypotesen. Likelihoodfunksjonen (10) skal maksimeres over $(\beta', \sigma_\mu^2, \sigma_\nu^2)' \in \mathbb{R}^K \times (\mathbb{R}_+ \cup \{0\})^2$. Vanligvis vises gyldigheten av de klassiske Wald-, Lagrange Multiplikator (LM)- og Likelihood Ratio (LR)-testene ved tester hvor nullhypotesen er et indre punkt i parameter-

rommet. Generelt vil en test mot en nullhypotese som ligger på randa ikke nødvendigvis ha den vanlige χ^2 -fordelingen.

Wald-testen ser vi lett at vil bryte sammen. Som beskrevet over kan vi estimere $\hat{\sigma}_\mu^2$. Denne estimatoren vil ha varians $\text{Var}(\hat{\sigma}_\mu^2) = \sigma_\mu^4$. Tester vi null-hypotesen $\sigma_\mu^2 = 0$ blir altså $\text{Var}(\hat{\sigma}_\mu^2) = 0$, så Z-observatoren blir

$$Z = \frac{\hat{\sigma}_\mu^2}{\sqrt{\text{Var}(\hat{\sigma}_\mu^2)}} \rightarrow \infty \quad (11)$$

for alle $\hat{\sigma}_\mu^2 > 0$. Siden vi kan få positive verdier på $\hat{\sigma}_\mu^2$ selv når H_0 er sann, er det klart at denne test-observatoren ikke vil være normalfordelt.

LR-testen er ikke fullt så innlysende. Denne vil ha en ikke-degenerert fordeling. Selv om denne ikke vil være den teoretiske skal vi se under at feilen ikke nødvendigvis er ekstremt stor.

I motsetning til Wald- og LR-testene kan det vises at Neymans klasse av $C(\alpha)$ -tester er robust ovenfor null-hypoteser på randa av parameterrommet. Siden LM-testen er et spesialtilfelle av $C(\alpha)$ -testene, vil denne derfor kunne brukes uten modifikasjoner. Se Godfrey (1988: 92ff) for flere detaljer.

3.1 Breusch-Pagan

Breusch og Pagan (1980) viser at under H_0 og en forutsetning om at $\nu_{it} \sim N(0, \sigma_\nu^2)$, er testobservatoren

$$\Psi_{BP} = \frac{NT}{2(T-1)} \left[\frac{\tilde{u}'D\tilde{u}}{\tilde{u}'\tilde{u}} - 1 \right]^2 \xrightarrow{d} \chi^2(1) \quad (12)$$

hvor \tilde{u} betegner residualene fra OLS-regresjonen av Y på X , \xrightarrow{d} representerer konvergens i fordeling og $D = TI_N \otimes A_T$. Prinsippet er stort sett at man maksimerer (10) under restriksjonen H_0 , og så tester signifikansen av Lagrange-parameteren på restriksjonen. Denne testen kan virke restriktiv siden den forutsetter at ν_{it} er normalfordelt, som er en grunnleggende forutsetning for at likelihoodfunksjonen skal ha formen (10). Honda (1985) viser derimot at testobservatoren er asymptotisk χ^2 -fordelt under relativt generelle betingelser. Under H_1 vil observatoren ha en ikke-sentral χ^2 -fordeling med eksploderende ikke-sentralitetsparameter, så testen er også konsistent under relativt generelle betingelser.

Det sentrale leddet i observatoren er

$$\left[\frac{\tilde{u}'D\tilde{u}}{\tilde{u}'\tilde{u}} - 1 \right]^2 = \left(T \frac{\tilde{u}'W\tilde{u}}{\tilde{u}'\tilde{u}} \right)^2, \quad (13)$$

som anngir forholdet mellom mellom- og total-variasjonen i restleddene, som kan tolkes som et uttrykk for forholdet mellom σ_μ^2 og $\sigma_\mu^2 + \sigma_\nu^2$.

3.2 Honda

Hovedproblemet med Breusch-Pagan-testen er at den ikke bruker den kjente informasjonen at $\sigma_\mu^2 \geq 0$, dvs at testen er en test av H_0 mot $H_2 : \sigma_\mu^2 \neq 0$. Dette innebærer at det finnes tester med høyere teststyrke; se Rogers (1986) for en relativt generell drøfting av dette problemet. Honda (1985) har foreslått testobservatoren

$$\Psi_H = \sqrt{\frac{NT}{2(T-1)}} \left[\frac{\tilde{u}'D\tilde{u}}{\tilde{u}'\tilde{u}} - 1 \right] \xrightarrow{d} N(0, 1). \quad (14)$$

Den asymptotiske fordelingen her er ganske robust ovenfor forutsetninger om fordelingen til restleddet, og testen er konsistent under ganske vide betingelser. Testen går da ut på å forkaste H_0 hvis $\Psi_H > Z_\alpha$ hvor Z_α er den øvre $1 - \alpha$ kvartilen til standard normalfordelingen. Denne testen er den uniformt sterkeste (uniformly most powerful - UMP) testen. Vi ser at uttrykket i hakeparentes i (12) i endelige utvalg kan få relativt store negative verdier under H_0 . Dette er mere sannsynlig under H_0 enn under H_1 , men vil gi en høy Ψ_{BP} -observator. Honda-testen retter opp i dette ved bare å gi en høy verdi på testobservatoren ved positive verdier i hakeparentesen.

3.3 Baltagi-Chang-Li/Gouriéroux-Holly-Montfort

Et alternativ til Honda-testen er følgende observator foreslått av Baltagi et al. (1992):

$$\Psi_{BCL} = \begin{cases} \Psi_H^2 & \text{hvis } \Psi_H > 0 \\ 0 & \text{hvis } \Psi_H \leq 0 \end{cases}. \quad (15)$$

Ved å bruke resultater fra Gouriéroux et al. (1982) viser de at denne observatoren har fordelingen

$$\Psi_{BCL} \sim \frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1), \quad (16)$$

det vil si en blandet χ^2 -fordeling hvor $\chi^2(0)$ er en variabel som med sannsynlighet 1 er lik 0 mens $\chi^2(1)$ er en sentral χ^2 -fordeling med 1 frihetsgrad. Det vil si at med sannsynlighet $\frac{1}{2}$ er $\Psi_{BCL} = 0$ og ellers er den χ^2 -fordelt med 1 frihetsgrad. Med et signifikansnivå på α forkaster vi H_0 til fordel for H_1 hvis vi observerer $\Psi_{BCL} > c_\alpha$ hvor c_α er definert ved $\Pr(\chi^2(1) > c_\alpha) = \frac{\alpha}{2}$. Det følger at denne testen også er UMP (Gouriéroux et al. 1982: 68), selv om testen ikke er ekvivalent med Honda-testen.

3.4 Moulton-Randolph

Et annet problem med Honda-testen er at den har en fordeling som kan avvike sterkt fra den asymptotiske fordelingen i moderate utvalg (Moulton og Randolph 1989). Moulton og Randolph (1989) bruker resultater om kvadratiske former til å beregne $E\Psi_H$ og $\text{Var}(\Psi_H)$ i endelige utvalg. De foreslår da å bruke den normaliserte testobservatoren

$$\Psi_{MR} = \frac{\Psi_H - E\Psi_H}{\sqrt{\text{Var}(\Psi_H)}} \xrightarrow{d} N(0, 1). \quad (17)$$

Definerer vi $d = \tilde{u}'D\tilde{u}/\tilde{u}'\tilde{u}$, $p = NT - (K + 1)$ og $M_X = I_{NT} - X(X'X)^{-1}X'$ har vi (jf. Baltagi 1995: 63)

$$\Psi_{MR} = \frac{d - Ed}{\sqrt{\text{Var}(d)}} = \frac{d - \frac{\text{tr}(DM_X)}{p}}{\sqrt{2 \frac{p \text{tr}[(DM_X)^2] - [\text{tr}(DM_X)]^2}{p^2(p+2)}}}. \quad (18)$$

4 Estimering og testresultater

La oss nå se på et eksempel hvor vi konkret skal ta for oss estimering av Engel-kurver for mat. Vi skal bruke data fra Statistisk sentralbyrås forbruksundersøkelser for årene 1993 og 1994. Disse dataene består av opplysninger om husholdninger som fører regnskap over alle sine utgifter i en to-ukers periode. Disse skaleres så opp til å tilsvare årlige forbruk. Dette er i utgangspunktet en tverrsnittsundersøkelse, men enkelte husholdninger fører også regnskap i samme to-ukers periode to år på rad. Vi har data om $N = 193$ husholdninger, som da er over $T = 2$ perioder. Regressanden er andelen utgift til mat har av total forbruksutgift (se Statistisk sentralbyrå [1996] for definisjoner av variablene og utdypende informasjon om dataene). Vi skal først se på en enkel modell hvor vi kun har med et konstantledd og total forbrukutgift i 1000 1998-kroner som forklaringvariabler. Tabell 1 viser estimeringsresultater. I tillegg til FGLS- og ML-estimatorene som er drøftet over har jeg også tatt med fixed effects- og between-estimatorene, som er utledet og drøftet i Biørn (2000). Alle estimeringer er gjort i matrise-programmeringsspråket Ox (Doornik 1999); se appendikset for et utdrag av koden.

Vi ser fra estimatene at økt total forbruksutgift ser ut til å trekke i retning av en nedgang i budsjettandelen som brukes på mat, noe som er i overensstemmelse med Engels lov. Effekten er også signifikant negativ ved bruk av alle estimatorene bortsett fra fixed effects. Fixed effects er en estimator som i vårt tilfelle, hvor T er lav, vil ha få frihetsgrader.

Table 1: Estimeringsresultater (standardfeil i parentes)

Variabel	OLS	Fixed effect	Between	FGLS	ML
Konstantledd	0.226 (0.00916)		0.208 (0.0101)	0.238 (0.00989)	0.240 (0.0100)
Total forbruksutgift (1000 kr)	-0.000274 ($3.74 \cdot 10^{-5}$)	-0.000607 (0.000418)	-0.000190 ($4.21 \cdot 10^{-5}$)	-0.000333 ($3.94 \cdot 10^{-5}$)	-0.000340 ($3.95 \cdot 10^{-5}$)
s^2	0.00775	0.196	0.00786		
s_ν^2				0.00508	0.00484
s_μ^2				0.00268	0.00297

Dette er trolig mye av forklaringen på at denne estimatoren gir estimater med høye standardavvik.

Fra opplysningene i Tabell 1 kan vi også gjennomføre en Durbin-Wu-Hausman-test for å teste om μ_i virkelig er stokastisk uavhengig av regressorene, som vi forutsatte³. Hvis dette ikke er sant, vil bare fixed effects-estimatoren være konsistent, mens GLS vil være konsistent og mere effisient hvis vi har uavhengighet. Baltagi (1995: Kap. 4.3) viser at Durbin-Wu-Hausman-testen i dette tilfellet kan skrives som

$$\Psi_{DWH} = \frac{\hat{\beta}_{\text{Fixed effects}} - \hat{\beta}_{\text{Between}}}{\sqrt{\text{Var}(\hat{\beta}_{\text{Fixed effects}}) + \text{Var}(\hat{\beta}_{\text{Between}})}}, \quad (19)$$

og under null-hypotesen om stokastisk uavhengighet vil $\Psi_{DWH} \sim N(0, 1)$. Vi får $\Psi_{DWH} = 0.993$, så vi kan ikke forkaste på noe ordinært signifikansnivå at det er stokastisk uavhengighet. Vi kan derfor trolig bruke Between-, FGLS- og ML-estimatorene.

For å testen hypotesen om ingen individuelle effekter bruker vi residualene fra OLS-regresjonen til å beregne testobservatorene som ble beskrevet over. Disse er presentert i Tabell 2. Resultater fra en LR-test er også tatt med. p-verdiene er her beregnet som om observatoren var χ^2 -fordelt. Selv om de asymptotiske p-verdiene varierer noe fra test til test er det helt klart at vi kan forkaste H_0 på alle vanlige signifikansnivåer. Følgelig er det svært usannsynlig å få realisert det datasettet vi har hvis det i virkeligheten ikke er individuelle effekter gitt at modellen er riktig spesifisert.

³Det kan selvsagt også være av interesse å teste om ν_{it} er stokastisk uavhengig av regressorene. Hvis vi har instrumenter for regressorene som vi vet er uavhengige av ν_{it} kan vi bruke en Durbin-Wu-Hausman-test til dette også. Dette er derimot utenfor rammene av dette essayet, og jeg skal anta at det her er stokastisk uavhengighet.

Table 2: Testresultater

Test	Testobservator	Asymptotisk p-verdi
Breusch og Pagan	23.0	$1.59 \cdot 10^{-6}$
Honda	4.80	$7.97 \cdot 10^{-7}$
Baltagi, Chang og Li	23.0	$7.97 \cdot 10^{-7}$
Moulton og Randolph	4.86	$5.84 \cdot 10^{-7}$
LR	27.0	$2.02 \cdot 10^{-7}$

Table 3: Estimeringsresultater med utvidet sett av regressorer (standardfeil i parentes)

Variabel	OLS		Between		FGLS		ML	
Konstantledd	0.156	(0.0518)	0.166	(0.0553)	0.155	(0.0544)	0.154	(0.0546)
Total forbruksutgift (1000 kr)	-0.000418	($4.31 \cdot 10^{-5}$)	-0.000330	($5.35 \cdot 10^{-5}$)	-0.000435	($4.38 \cdot 10^{-5}$)	-0.000437	($4.38 \cdot 10^{-5}$)
Antall barn	0.0316	(0.00476)	0.0309	(0.00501)	0.0316	(0.00502)	0.0316	(0.00505)
Antall voksne	0.0257	(0.00554)	0.0247	(0.00616)	0.0258	(0.00578)	0.0258	(0.00580)
Alder	-0.000344	(0.00218)	-0.00118	(0.00233)	-0.000171	(0.00229)	-0.000153	(0.00230)
Alder kvadrert	$2.01 \cdot 10^{-5}$	($2.28 \cdot 10^{-5}$)	$3.00 \cdot 10^{-5}$	($2.44 \cdot 10^{-5}$)	$1.79 \cdot 10^{-5}$	($2.40 \cdot 10^{-5}$)	$1.77 \cdot 10^{-5}$	($2.41 \cdot 10^{-5}$)
Hovedinntektsyter mann	0.000856	(0.0110)	-0.000862	(0.0125)	0.000858	(0.0114)	0.000853	(0.0114)
Østlandet utenom Oslo/Akershus	-0.000955	(0.0129)	0.000196	(0.0132)	-0.00106	(0.0137)	-0.00107	(0.0138)
Agder og Rogaland	-0.0250	(0.0149)	-0.0248	(0.0153)	-0.0246	(0.0159)	-0.0246	(0.0160)
Vest-Norge	-0.00765	(0.0148)	-0.00380	(0.0152)	-0.00821	(0.0158)	-0.00826	(0.0159)
Trøndelag	-0.00341	(0.0156)	-0.000497	(0.0160)	-0.00390	(0.0166)	-0.00394	(0.0167)
Nord-Norge	0.0319	(0.0162)	0.0333	(0.0166)	0.0317	(0.0172)	0.0316	(0.0173)
Faglært arbeider	-0.0201	(0.0172)	-0.0371	(0.0206)	-0.0164	(0.0176)	-0.0160	(0.0176)
Funksjonær, lavt nivå	-0.0340	(0.0193)	-0.0397	(0.0228)	-0.0323	(0.0198)	-0.0321	(0.0198)
Funksjonær, middelnivå	-0.0153	(0.0151)	-0.0293	(0.0170)	-0.0123	(0.0156)	-0.0120	(0.0157)
Funksjonær, høyt nivå	-0.00662	(0.0162)	-0.0228	(0.0182)	-0.00357	(0.0168)	-0.00326	(0.0169)
Landbruk og fiske	-0.0449	(0.0229)	-0.0509	(0.0248)	-0.0434	(0.0239)	-0.0432	(0.0240)
Annen selvstendig næringsdrivende	0.0465	(0.0214)	0.0453	(0.0239)	0.0462	(0.0222)	0.0461	(0.0223)
Student og elev	-0.0437	(0.0284)	-0.0618	(0.0369)	-0.0395	(0.0287)	-0.0391	(0.0287)
Pensjonist	-0.00391	(0.0175)	-0.011	(0.0193)	-0.00225	(0.0182)	-0.00207	(0.0182)
Hjemmeværende	-0.0526	(0.0386)	-0.0892	(0.0568)	-0.0467	(0.0383)	-0.0462	(0.0382)
Annet eog uklassifisert	-0.0134	(0.0378)	0.0173	(0.0542)	-0.0159	(0.0375)	-0.0161	(0.0375)
s^2	0.00628		0.00641					
s_p^2					0.00540		0.00533	
s_μ^2					0.000873		0.000958	

5 Drøfting

Det bør ikke komme som noen stor overraskelse at vi ikke kan forkaste H_0 i modellen over. Her har vi antatt at total forbruksutgift er den eneste variabelen som forklarer forbruk av mat, noe som selvsagt er en sterk overforenkling. En del forklaringvariable vil være mer eller mindre konstante over tid men husholdningsspesifikke. Disse effektene vil da gi seg utslag i μ_i . Tabell 3 gir estimeringsresultater fra en modell med et større antall forklaringsvariable. Siden mange av disse er konstante over tid kan ikke fixed effects estimatoren brukes i dette tilfellet. Merk at bostedsregion og sosio-økonomisk status er dummy-variable hvor ”Oslo og Akershus” og ”Ufaglært arbeider” er utelatte kategorier.

Gjennomfører vi testene beskrevet over på residualene fra denne modellen får vi resul-

Table 4: Testresultater i utvidet modell

Test	Testobservator	Asymptotisk p-verdi
Breusch og Pagan	6.70	0.00963
Honda	2.59	0.00481
Baltagi, Chang og Li	6.70	0.00482
Moulton og Randolph	3.12	0.000920
LR	6.14	0.0132

tatene i Tabell 4. Selv om vi fortsatt kan forkaste H_0 på de fleste ordinære signifikansnivåer er forkastningen mindre klar. Det vil si at det fortsatt er usannsynlig å observere dataene vi har gitt H_0 og en riktig spesifisert modell, men det er mindre usannsynlig enn gitt modellen over.

En annen kritikk av framgangsmåten som kan reises er at resultatene kun er gyldig asymptotisk, mens vi har et endelig utvalg. For å forsøke å få et inntrykk av hvordan testene oppfører seg i et endelig utvalg har jeg gjennomført noen enkle Monte Carlo-simuleringer i den enkle modellen med total forbruksutgift som eneste regressor. Simuleringene er gjort betinget på de observerte regressorene, og parameterverdiene beregnet ved hjelp av OLS er sett på som de "sanne" verdiene. En framgangsmåte er å trekke nye restledd som er uavhengig normalfordelte med forventning null og varians s^2 , hvor s^2 er beregnet fra OLS-residualene. Det er derimot ikke sikkert at restleddene reelt sett er normalfordelte. For å ta hensyn til dette er det også gjennomført en ikke-parametrisk simulering som beveger seg mot bootstrap-tankegangen. Vi har her beregnet residualene fra OLS. For å trekke nye tilfeldige restledd trekkes disse da uavhengig fra disse residualene. Totalt ble det simulert 10000 tilfeldige datasett for hver spesifikasjon av restleddet. Resultatene er presentert i Tabell 5 . Figur 1 og 2 presenterer histogrammer over de simulerte fordelingene. De heltrukne kurvene i figurene er de teoretiske fordelingene. Selv om fordelingen til de simulerte testobservatorene ikke svarer helt til den teoretiske, spesielt i tilfellet med ikke-parametriske restledd, er ikke avviket stort. Også LR-testen ser ut til å følge den vanlige χ^2 -fordelingen i relativt stor grad.

Dette innebærer at testene som ble gjennomført over neppe er helt eksakte, men tross alt kan sees på som en relativt god tilnærming til virkeligheten. Det må dog bemerkes at gyldigheten til alle testene er betinget på at modellen er riktig spesifisert. Hvis vi for eksempel har heteroskedastisitet, vil ikke de teoretiske fordelingene til testobservatorene

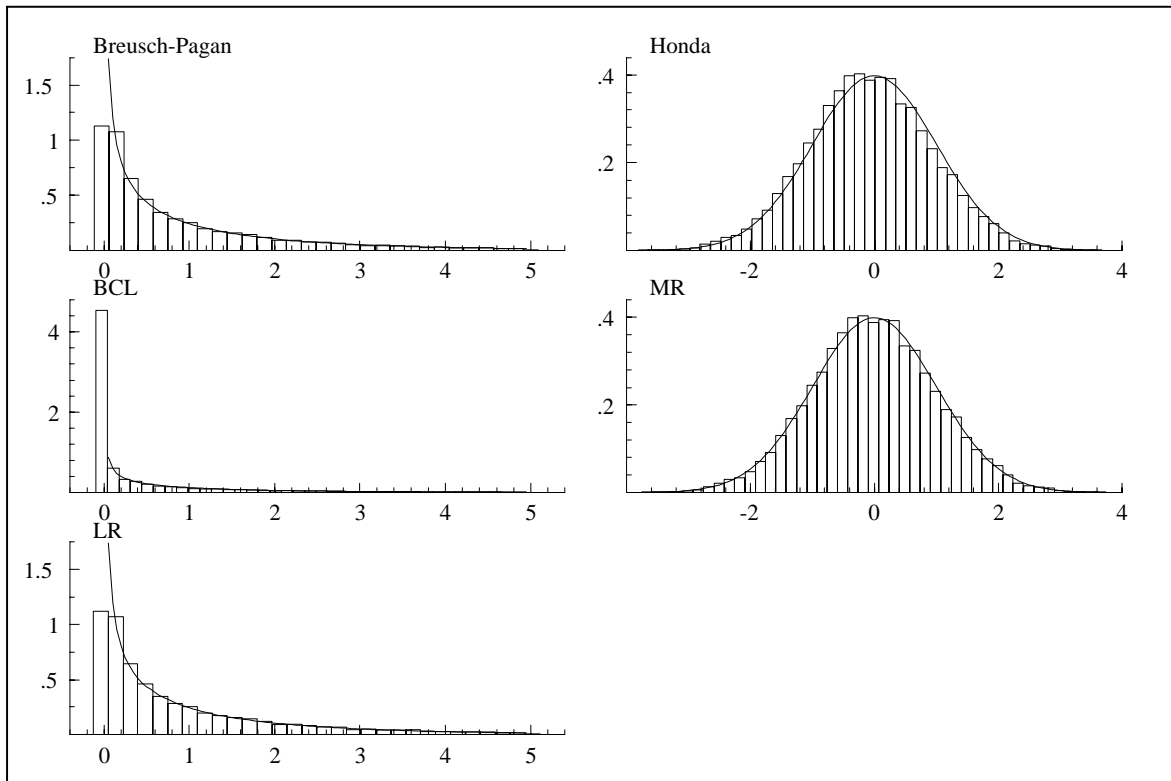


Figure 1: Fordeling av testobservatorene med normalfordelte restledd

generelt være gyldige lenger.

6 Oppsummering

Vi har sett at det finnes flere måter å teste for tilstedeværelse av individ-spesifikke effekter i paneldatamodeller med tilfeldige effekter. Det er et generelt et problem når nullhypotesen er at varians-parameteren til de individuelle effektene er null siden dette innebærer at den ligger på randa av parameterrommet. Ved å bruke tester basert på LM-prinsippet kan det vises at dette ikke fører til problemet. Derimot vil en direkte LM-test som Breusch-Pagan testen ikke være effisient siden den ikke tar med at vi tester en en-sidig hypotese. Tre andre tester som tar dette inn over seg ble presentert. På norske forbruksdata gjennomførte vi så en test, og fant at det er mye som tyder på at det er individ-spesifikke effekter som gjør seg gjeldende. Dette er også tilfelle når vi innfører en serie med husholdningskarakteristika som forklaringsvariable. Det ble også argumentert for at testen er tilnærmet gyldig i et ganske moderat utvalg. Det tyder derfor på at vi må ta hensyn til panel-strukturen i disse dataene hvis vi ønsker effisiente estimater.

Table 5: Resultater fra Monte Carlo-simulering (Normal er med normalfordelte restledd mens NP er med bruk av ikke-parametrisk simulering)

Breusch og Pagan				Honda			
Kvantil	Teoretisk	Normal	NP	Kvantil	Teoretisk	Normal	NP
0.25	0.10	0.101	0.0893	0.25	-0.68	-0.737	-0.718
0.5	0.46	0.444	0.412	0.5	0	-0.0790	-0.0789
0.75	1.32	1.31	1.20	0.75	0.68	0.596	0.564
0.9	2.71	2.68	2.53	0.9	1.29	1.21	1.16
0.95	3.84	3.79	3.60	0.95	1.65	1.59	1.56
0.975	5.02	5.03	4.68	0.975	1.96	1.89	1.90
0.99	6.63	6.60	6.46	0.99	2.33	2.24	2.34

Baltagi, Chang og Li				Moulton og Randolph			
Kvantil	Teoretisk	Normal	NP	Kvantil	Teoretisk	Normal	NP
0.25	0	0	0	0.25	-0.68	-0.680	-0.660
0.5	0	0	0	0.5	0	-0.0213	-0.0212
0.75	0.46	0.355	0.318	0.75	0.68	0.654	0.622
0.9	1.64	1.47	1.35	0.9	1.29	1.27	1.22
0.95	2.71	2.52	2.45	0.95	1.65	1.65	1.62
0.975	3.84	3.58	3.59	0.975	1.96	1.95	1.95
0.99	5.41	5.01	5.49	0.99	2.33	2.30	2.40

Likelihood Ratio			
Kvantil	Teoretisk	Normal	NP
0.25	0.10	0.0958	0.0842
0.5	0.46	0.442	0.409
0.75	1.32	1.32	1.20
0.9	2.71	2.71	2.55
0.95	3.84	3.83	3.64
0.975	5.02	5.12	4.76
0.99	6.63	6.78	6.59

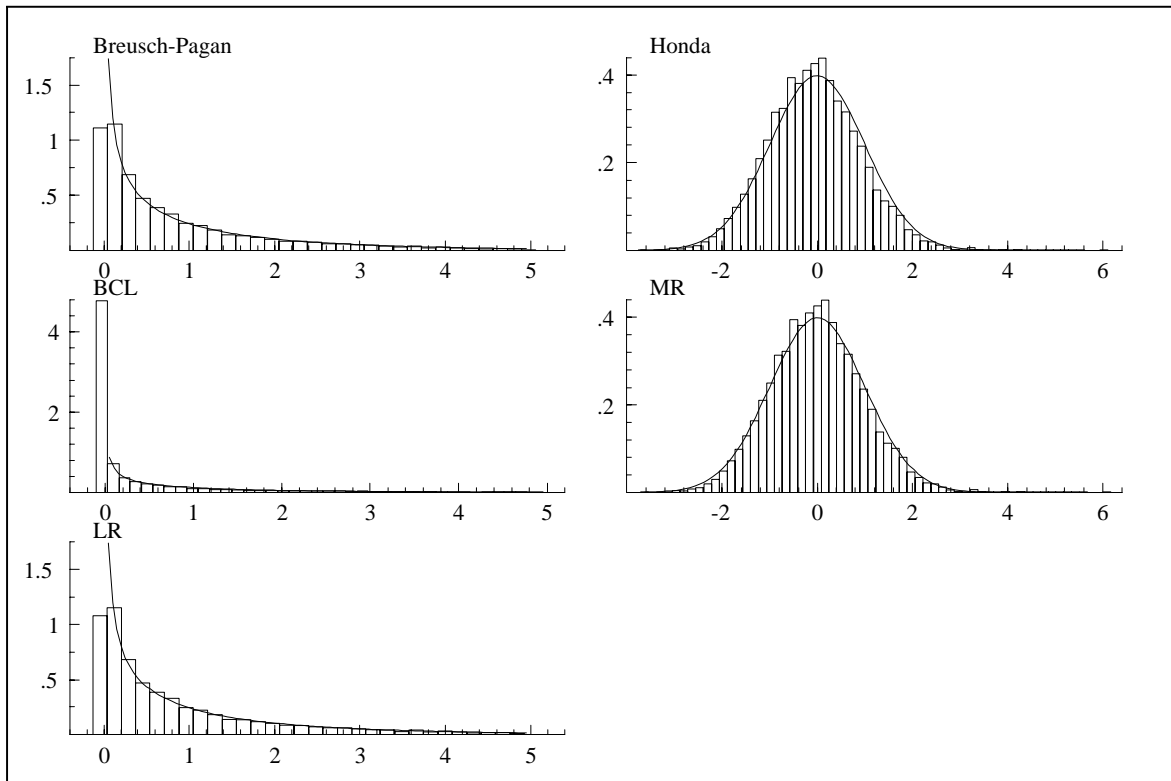


Figure 2: Fordeling av testobservatorene med ikke-parametrisk fordeling på restleddene

Referanser

- Baltagi, B. (1995): *Econometric analysis of panel data*. Chichester: Wiley.
- Baltagi, B. H., Y.-J. Chang og Q. Li (1992): Monte Carlo results on several new and existing tests for the error component model. *Journal of Econometrics* 54, 95-120.
- Biørn, E. (2000): *Økonometrisk analyse av paneldata - en innføring*. Oslo: Unipub.
- Breusch, T. S. og A. R. Pagan (1980): The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* 47, 239-53.
- Doornik, J. A. (1999): *Object-oriented matrix programming using Ox version 2.1*. London: Timberlake.
- Godfrey, L. G. (1988): *Misspecification tests in econometrics. The Lagrange multiplier principle and other approaches*. Cambridge: Cambridge University Press.

Gouriéroux, C., A. Holly og A. Montfort (1982): Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* 50, 63-80.

Honda, Y. (1985): Testing the error components model with non-normal disturbances. *Review of Economic Studies* 52, 681-90.

Mátyás, L. (1992): Error component models. Kap. 4 i L. Mátyás og P. Sevestre: *The econometrics of panel data. Handbook of theory and applications*. Dordrecht: Kluwer.

Moulton, B. R. og W. C. Randolph (1989): Alternative tests of the error component model. *Econometrica* 57, 685-93.

Rogers, A. J. (1986): Modified Lagrange multiplier tests for problems with one-sided alternatives. *Journal of Econometrics* 31, 341-61.

Statistisk sentralbyrå (1996): *Forbruksundersøkelse 1992-1994*. NOS C317

Vedlegg: Ox-program

Under er hoveddelen av et Ox-program som ble brukt til å gjennomføre estimeringene og testene ovenfor gjengitt. Det må bemerkes at programmet er skrevet for et datasett av moderat størrelse. For bruk på større datasett må det gjennomføres visse omskrivninger for å unngå numerisk ustabilitet på grunn av matriser av høy dimensjon.

```
#include <oxstd.h>

panel_est(const y, x, const N, const T, const est_method, const outBeta, const outCov, const outll)
// est_method: 0 Between, 1 Within, 2 FGLS, 3 ML, -1 OLS
{ decl Wxx, Wxy, Bxx, Bxy, beta, cov, u, Sny, Smy, K=columms(x), theta, beta_, stderr, Zvals, ll,
    eps=1e-12,
    A=ones(T, T) ./ T,
    B=unit(T) - A;

    if (est_method==1 && x[][0]==1)
        x=x[][1:];

    Wxx=x'(unit(N)**B)*x;
    Bxx=x'(unit(N)**A)*x;
    Wxy=x'(unit(N)**B)*y;
    Bxy=x'(unit(N)**A)*y;

// OLS
    if (est_method==-1)
    { beta=invertsym(Bxx+Wxx)*(Bxy+Wxy);
      u=y-x*beta;
      Sny=u'u/(N*T-K);
      cov=Sny*invertsym(Bxx+Wxx);
      println("Estimering med OLS");
```

```

    println("%r",{"S^2"},Sny);
}

// Between estimator
if (est_method==0)
{ beta=invertsym(Bxx)*(Bxy);
  u=y-x*beta;
  Sny=u'u/(N*T-K);
  cov=Sny*invertsym(Bxx);
  println("Estimering med Between-estimator");
  println("%r",{"S^2"},Sny);
}

// Within estimator (fixed effects)
if (est_method==1)
{ beta=invertsym(Wxx)*Wxy;
  u=y-x*beta;
  Sny=u'u/(N*T-K-N+1);
  cov=Sny*invertsym(Wxx);
  println("Estimering med Fixed effects");
  println("%r",{"S^2"},Sny);
}

// FGLS
if (est_method==2)
{ beta=invertsym(Bxx+Wxx)*(Bxy+Wxy);
  u=y-x*beta;
  Sny=u'(unit(N)**B)*u./(N*(T-1)-K+1);
  Smy=u'u/(N*T-K)-Sny;
  theta=Sny/(Sny+T*Smy);
  beta=invertsym(Wxx+theta*Bxx)*(Wxy+theta*Bxy);
  cov=Sny*invertsym(Wxx+theta*Bxx);
  ll=-0.5*N*((T-1)*log(Sny)+log(Sny+T*Smy))-0.5*u'(unit(N)**(B+theta*A))'u/Sny;
  println("Estimering med FGLS");
  println("%r",{"Sny", "Smy", "Log likelihood"}, (Sny|Smy|ll));
}

// ML
if (est_method==3)
{ beta=invertsym(Bxx+Wxx)*(Bxy+Wxy);
  do
  { beta_=beta;
    u=y-x*beta;
    Sny=u'(unit(N)**B)*u./(N*(T-1)-K+1);
    Smy=u'u/(N*T-K)-Sny;
    theta=Sny/(Sny+T*Smy);
    beta=invertsym(Wxx+theta*Bxx)*(Wxy+theta*Bxy);
  } while (norm(beta-beta_)/norm(beta+1)>eps);
  cov=Sny*invertsym(Wxx+theta*Bxx);
  ll=-0.5*N*((T-1)*log(Sny)+log(Sny+T*Smy))-0.5*u'(unit(N)**(B+theta*A))'u/Sny;
  println("Estimering med ML");
  println("%r",{"Sny", "Smy", "Log likelihood"}, (Sny|Smy|ll));
}
stderr=sqrt(diagonal(cov))';
println("%c",{"Coeficient", "Std. error", "Z value"}, (beta~stderr~fabs(beta./stderr)));
print("\n\n");
if (outBeta)
  outBeta[0]=beta;
if (outCov)
  outCov[0]=cov;
if (outll)
  outll[0]=ll;
}

ind_eff_test(const y, const x, const N, const T, const prt_scr, const outRes)
// Tester for individuelle effekter
{ decl beta, u, d, DMx, K, p, MRnevner, llML, llOLS, s,

```

```

    BreuschPagan, Honda, BCL, MoultonRandolph, LR, Wald;
    olsc(y,x,&beta);
    u=y-x*beta;
    s=u'u/(N*T);
    llOLS=-0.5*N*T*log(s)-0.5*u'u/s;
    d=(u'(unit(N)**ones(T,T))*u)/(u'u);
    panel_est(y,x,N,T,3,0,0,&llML);

    BreuschPagan=(0.5*N*T/(T-1))*(d-1)^2;
    Honda=sqrt(0.5*N*T/(T-1))*(d-1);
    BCL=Honda>0 ? Honda^2 : 0;
    K=columns(x);
    p=N*T-K;
    DMx=(unit(N)**ones(T,T))*(unit(N*T)-x*invertsym(x'x)*x');
    MRnevrner=2*(p*trace(DMx^2)-trace(DMx)^2)/((p+2)*p^2);
    MoultonRandolph=(d-trace(DMx)/p)/sqrt(MRnevrner);
    LR=2*(llML-llOLS);
    if (prt_scr)
    { println("%c",{"Test-observator","p-verdi"},"%r",
      {"Breusch-Pagan","Honda","Baltagi-Chang-Li","Moulton-Randolph","LR"},
      (BreuschPagan|Honda|BCL|MoultonRandolph|LR)
      ~ (1-probchi(BreuschPagan,1)|1-probn(Honda)
      |0.5-0.5*probchi(BCL,1)|1-probn(MoultonRandolph)|1-probchi(LR,1)));
    }
    if (outRes)
      outRes[0]=BreuschPagan~Honda~BCL~MoultonRandolph~LR;
}

main()
{ decl T=2, N=386/T,
  KPI93=0.906,
  KPI94=0.919,
  KPIs=ones(N,1)**(KPI93|KPI94);
  decl m=loadmat("d:\\kurs\\paneldatakurs\\data.mat"),
  y=m[] [0] ./m[] [1],
  x=1^(m[] [1].*KPIs)/1000);

  panel_est(y,x,N,T,-1,0,0,0);
  panel_est(y,x,N,T,0,0,0,0);
  panel_est(y,x,N,T,1,0,0,0);
  panel_est(y,x,N,T,2,0,0,0);
  // panel_est(y,x,N,T,3,0,0,0);

  ind_eff_test(y,x,N,T,TRUE,0);
}

```


De sist utgitte publikasjonene i serien Notater

- 2000/50 J.F. Bjørnstad: En innføring i utvalgsundersøkelser. 91s.
- 2000/51 A.G. Pedersen: Oversikt over dødsårsaksregisterets koderegler 1996-1997. 39s.
- 2000/52 O. Klungshør: Sammenligning av mikroformler for prisindekser og modelltilpassing. 36s.
- 2000/53 G. Dahl og B.R. Joneid: FD - Trygd: Dokumentasjonsrapport: Inntekt og formue, 1992-1997. 36s.
- 2000/54 K. Flugsrud og G. Haakonsen: Utslipp av klimagasser i norske kommuner: En gjennomgang av datakvaliteten i utslippsregnskapet. 44s.
- 2000/56 T. Høgeland: "Ny" vekstteori: Et nytt forskningsprogram eller naturlig progresjon? Utviklingen innenfor økonomisk vekstteori vurdert i forhold til Imre Lakatos' vitenskapsfilosofi. 19s.
- 2000/57 K.-G. Lindquist: SAS-programmer for korrigerende av data fra industristatistikken og beregning av variable for analyseformål. 53s.
- 2000/58 A. Akselsen: FD - Trygd: Dokumentasjon av uttak til Sandmanutvalget (SHD). 28s.
- 2000/59 J. Johansen og ØSivertstøl FD - Trygd: Dokumentasjonsrapport: Fødsels- og sykepenger, 1992-1998. 109s.
- 2000/60 K-G. Lindquist: Database for energintensiv næringer: Tall fra industristatistikken: Oppdatert versjon av notat 97/30. 17s.
- 2000/61 O. Haugen: Utrekning av vekter til inntekts- og formuesundersøkingane 1998. 24s.
- 2000/62 ØKleven: Ferieundersøkelsen 1999 Panel: Dokumentasjonsrapport. 50s.
- 2000/63 P.G. Larssen: Overvakingssystemet for bedrifter i BoF. 29s.
- 2000/64 R. N. Johnsen: Undersøking om foreldrebetaling i barnehagar, august 2000. 36s.
- 2000/65 A. Thomassen: Byggekostnadsindeks for røggerarbeid i kontor- og forretningsbygg. 14s.
- 2000/67 A.G. Hustoft og G. Olsen: Metadata for statistikk om personer og husholdninger : Forprosjektrapport. 34s.
- 2000/68 A. Bruvoll, K. Flugsrud og H. Medin: Dekomponering av endringer i utslipp til luft i Norge - dokumentasjon av data. 19s.
- 2000/69 M. Vik Dysterud og E. Engelién: Tettstedsavgrønsing: Teknisk dokumentasjon 2000. 53s.
- 2000/70 A. Akselsen, G. Dahl, J. Lajord og Ø Sivertstøl FD - Trygd: Variabelliste. 48s.
- 2000/71 B.O. Lagerstrøm: Kompetanse i grunnskolen , del 2: Dokumentasjonsrapport. 19s.
- 2000/72 B.O. Lagerstrøm: Kompetanse i grunnskolen: Hovdresultater 1999/2000 170s.
- 2000/73 J.H. Wang: Kvartalsvis investeringsstatistikk. 57s.
- 2000/74 P.O. Lande og T. Hoel: Dødsårsaksregisteret: Systemdokumentasjon. 90s.
- 2000/75 A.G. Pedersen, P.O. Lande og T. Hoel: Dødsårsaksregisteret: Brukerdokumentasjon. 99s.
- 2000/76 A.G. Hustoft, B. Vannebo: En undersøkelse av frafallet i utvalgsundersøkelser i perioden 1997-2000. 56s.
- 2000/77 P.O. Lande og J. Kittelsen: Forbruksundersøkinga 2000. Innlasting/Innsjekking: Brukerdokumentasjon. 17s.
- 2000/78 J. Fosen, A.K. Johnsen og G. Rønne: Frafall blant innvandrere. En undersøkelse av frafall i Utdanningsundersøkelsen 1999 og i valgundersøkelser blant innvandrere. 53s.
- 2000/79 J. Kittelsen og P.O. Lande: OPPSLAG - Forbruksundersøkelsen. Brukerdokumentasjon. 39s.