Estimation in the Fixed Effects Ordered Logit Model

Chris Muris (SFU)

Outline

Introduction

Model and main result

Cut points

Estimation

Simulations and illustration

Conclusion

Setting

1. **Fixed-T panel.** A random sample $\{(y_{it}, X_{it}), i = 1, \cdots, N, t = 1, \cdots, T\}$, with $N \to \infty$
2. **Ordered logit**. y_{it} is an ordered response in $\{1, 2, \cdots, J\}$,

$$
\begin{aligned}
y_{it}^* &= \alpha_i + X_{it}\beta + u_{it}, \\
y_{it} &= \begin{cases} 1 & \text{if } \quad y_{it}^* < \gamma_1, \\ 2 & \text{if } \gamma_1 \le y_{it}^* < \gamma_2, \\ \vdots & \vdots \\ J & \text{if } \gamma_{J-1} \le y_{it}^*, \end{cases}
\end{aligned}
$$

   for cut points γ_j. Errors are logistic.
3. **Fixed effects.** Joint distribution of α_i and X_i is unrestricted.

Contribution

This paper:

- Estimation of differences of the **cut points**
- **More efficient** estimation of the regression coefficient

Why does this matter?

- Cut points: bounds on partial effects
- Model is **heavily used** (BSW, 2015: >150 cites)

Application (1): Allen and Arnutt (WP, 2013)

Effect of **"Teach First"** program on **educational outcomes.**

- y_{it}: **letter grade** student i for subject-year t
- $D_{it} \in \{0, 1\}$: school **enrolled** in "Teach First"?
- Latent variable model:

$$y_{it}^{*} = \alpha_i + \beta_1 D_{it} + X_{it}\beta_2 + u_{it},$$

where

- α_i is **unobserved** student ability
- X_{it} are controls

Application (1): Allen and Arnutt (WP, 2013)

All three model ingredients are present

1. **Fixed-T:** number of subjects per student is much smaller than the number of students
2. **Ordered:** letter grade is an ordered outcome
3. **Fixed effects:** schools with results in the bottom 30% are eligible

Application (2): Frijters et al. (AER, 2004):

Effect of **income** on **life satisfaction**

- y_{it}: **life satisfaction** on scale $\{0, \cdots, 10\}$
  - "completely dissatisfied" to "completely satisfied".
- X_{it}: real household **income**
- Latent variable model: $y_{it}^* = \alpha_i + \beta_1 X_{it} + Z_{it}\beta_2 + u_{it}$
  - α_i: unobserved student ability
  - X_{it} may correlated with α_i
  - Z_{it}: other controls.

More applications

- **Health**
  - Khanam et al. (JHE, 2014): income and child health
  - Carman (AER, 2013): intergenerational transfers and health
  - Frijters et al. (JHE, 2005): income on health
- **Labor**
  - Hamermesh (JHR, 2001): earnings shocks and job satisfaction
  - Das and van Soest (JEBO, 1999): expectations about future income

More applications (2)

- **Happiness**
  - Frijters et al. (AER, 2004): income and life satisfaction
  - Blanchflower and Oswald (JPE, 2004): trends in US life satisfaction
- **Credit / debt ratings**
  - Amato, Furfine (JBF 2003): credit ratings are not procyclical
  - Afonso et al. (IJOFE, 2013): determinants of sovereign debt ratings
- **Education**
  - Allen and Alnutt (2013): effect of "Teach First" program on student achievement

Literature

- Chamberlain (RES, 1980)**:** binary choice and unordered choice
- Das and van Soest (JEBO, 1999): all cutoffs
- Ferrer-i-Carbonell and Frijters (EJ, 2004): individual-specific cutoffs
- Baetschmann et al. (JRSS-A, 2015): small-sample improvements

None of these papers estimate the cut point differences.

Outline

Introduction

Model and main result

Cut points

Estimation

Simulations and illustration

Conclusion

Model

- **Random sample** of size $n \to \infty$, T fixed:

$$\{(y_{i1}, \cdots, y_{iT}, X_{i1}, \cdots, X_{iT}), \; i = 1, \cdots, n\}$$

  - y_{it} is an ordered outcome in $\{1, \cdots, J\}$
  - $X_{it} = (X_{it,1}, \cdots, X_{it,K})$ are covariates

- **Unobserved heterogeneity** in the latent variable:

$$y_{it}^* = \alpha_i + X_{it}\beta + u_{it}$$

- Serially **independent, exogenous logistic** errors

$$u_{i1}, \cdots, u_{iT} | (X_{i1}, \cdots, X_{iT}), \alpha_i \sim \text{iid} LOG(0, 1)$$

- **Link** between latent and observed by **cut points**

$$y_{it} = \begin{cases} 1 & \text{if } \quad y_{it}^* < \gamma_1 \\ 2 & \text{if } \gamma_1 \leq y_{it}^* < \gamma_2 \\ \vdots & \vdots \\ J & \text{if } \gamma_{J-1} \leq y_{it}^*. \end{cases}$$

Incidental parameters

For each category j,

$$P\left(y_{it}=j|\, X_{it},\alpha_i\right)=\Lambda\left(\gamma_j-\alpha_i-X_{it}\beta\right)-\Lambda\left(\gamma_{j-1}-\alpha_i-X_{it}\beta\right),$$

where $\Lambda=\exp\left(x\right)/\left(1+\exp\left(x\right)\right)$. Likelihood is

$$\prod_{i=1}^{n}\prod_{t=1}^{T}\prod_{j=1}^{J}\left[\Lambda\left(\gamma_j-\alpha_i-X_{it}\beta\right)-\Lambda\left(\gamma_{j-1}-\alpha_i-X_{it}\beta\right)\right]^{1\{y_{it}=j\}}.$$

- Fixed T: maximum likelihood estimator (MLE) is inconsistent

Incidental parameters (logit)

$\hat{\beta}_{ML}$: maximum likelihood estimator for $T = J = 2$

- **Inconsistent** (Abrevaya, 1997)
  - $\hat{\beta}_{ML} \xrightarrow{p} 2\beta$ as $n \to \infty$
- **Solution** (Chamberlain, 1980)
  - $y_{i1} + y_{i2}$ is a sufficient statistic for α_i
  - conditional MLE (CMLE) with

$$P\left(y_i = (1,0) \mid y_{i1} + y_{i2} = 1, X_i, \alpha_i\right) = \frac{1}{1 + \exp\left(\left(X_{i2} - X_{i1}\right)\beta\right)}$$

  is consistent

- **Drawback:** CMLE uses only **switchers**

Incidental parameters (Ordered logit)

- Solution for incidental parameters problem is **model-specific**
- No sufficient statistic (yet?) for ordered logit
- No **exponential** form:

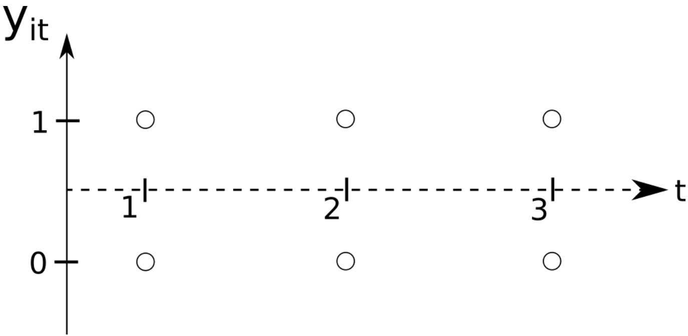
$$P\left(y_{it} = j \middle| X_{it}, \alpha_i\right) = \Lambda\left(\gamma_j - \alpha_i - X_{it}\beta\right) - \Lambda\left(\gamma_{j-1} - \alpha_i - X_{it}\beta\right)$$

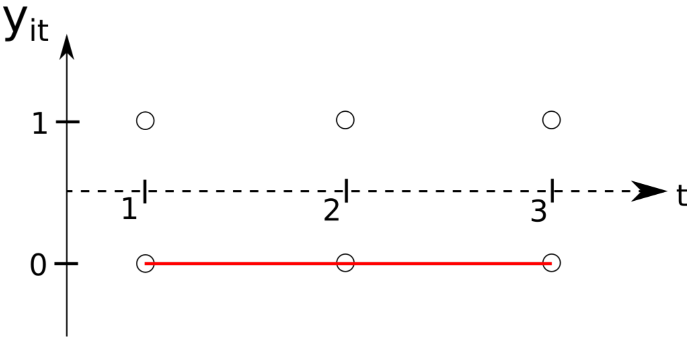
Incidental parameters (Takeaway)

- Unobserved heterogeneity can cause **inconsistency**
- **Solution** exists for the case of **binary** logit
  - Solution uses only switchers
  - Does not extend to ordered logit model

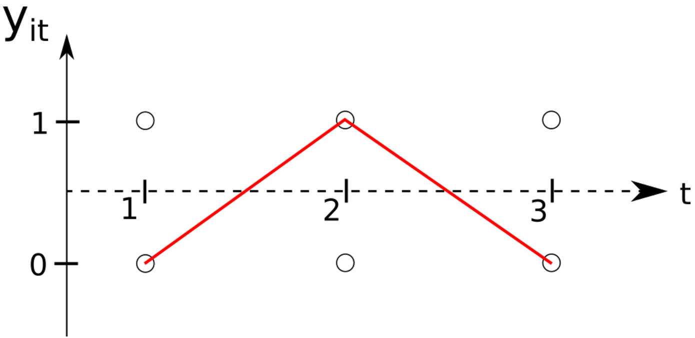
Ordered choice

- Consider **ordered** choice with $y_{it} \in \{1, \cdots, J\}$
- **Dichotomization**:
  - Pick some $j \in \{1, \cdots, J-1\}$ and define the binary variable
    $$d_{it,j} = \begin{cases} 1 & \text{if } y_{it} \leq j, \\ 0 & \text{otherwise.} \end{cases}$$
  - Apply Chamberlain's CMLE to $y_{it,j}$
- **Consistent but inefficient**:
  - Information is lost by discarding more precise measurement y_{it}
  - Winkelmann and Winkelmann (1998):
    - $\{0, \cdots, 10\}$ collapsed to $\{0, 1\}$ by cutting at 7
    - Out of 10000 observations, only 2523 are switchers

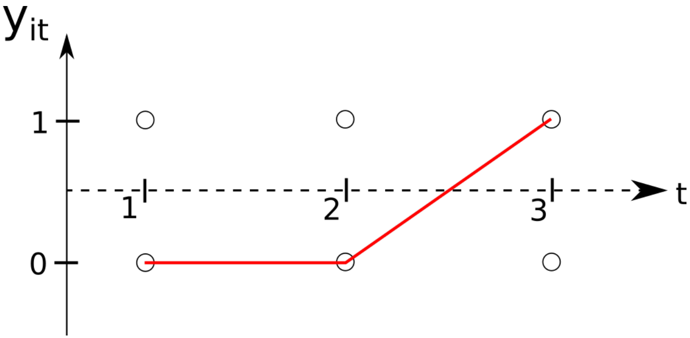

y_{it}
1
0
1
2
3
t



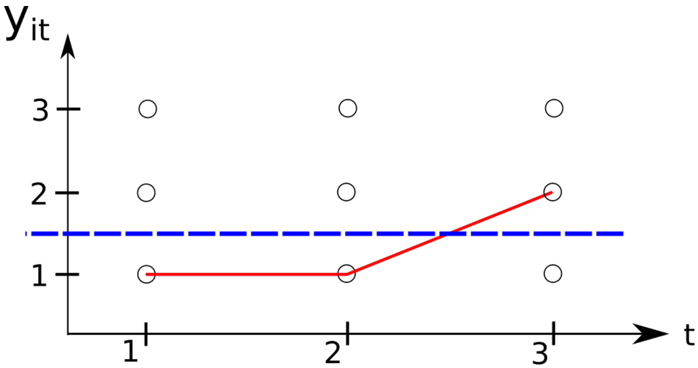
Non-switcher: not informative

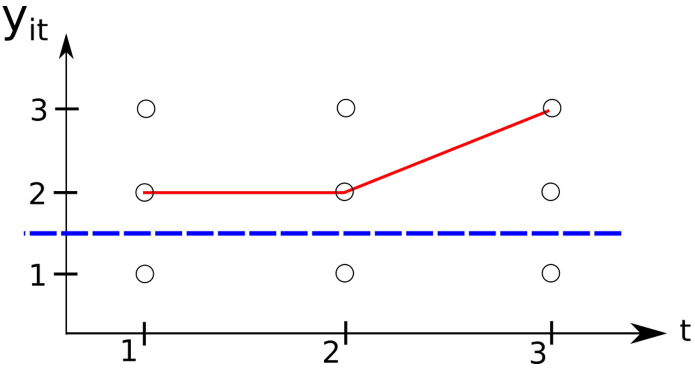


Switcher: informative

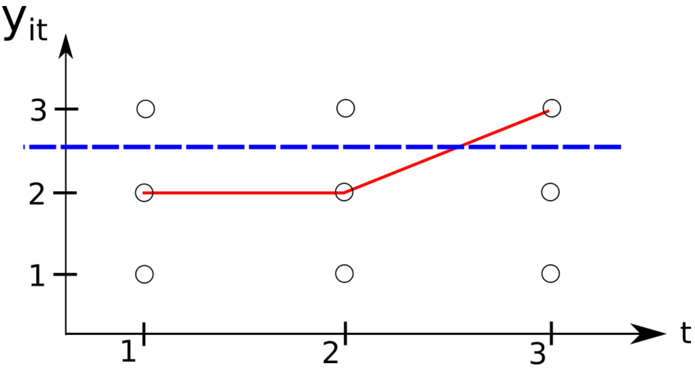


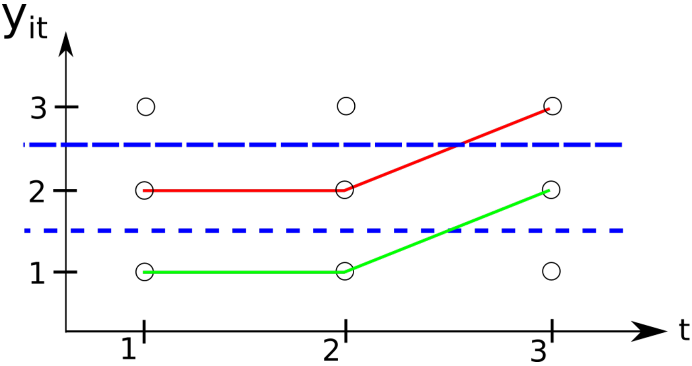
y_{it}
1
0
1
2
3
t


y_{it}
3
2
1
t
1
2
3

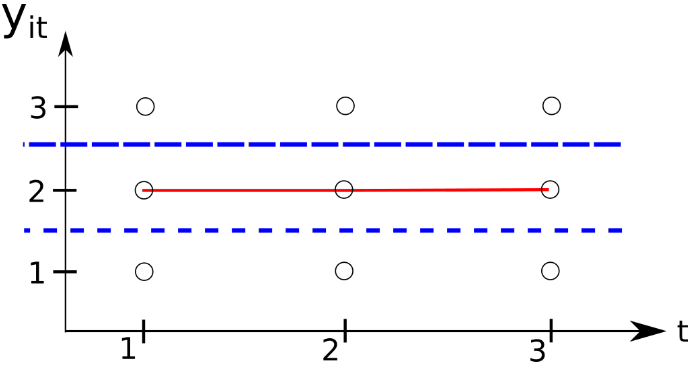


y_{it}
3
2
1
1
2
3
t

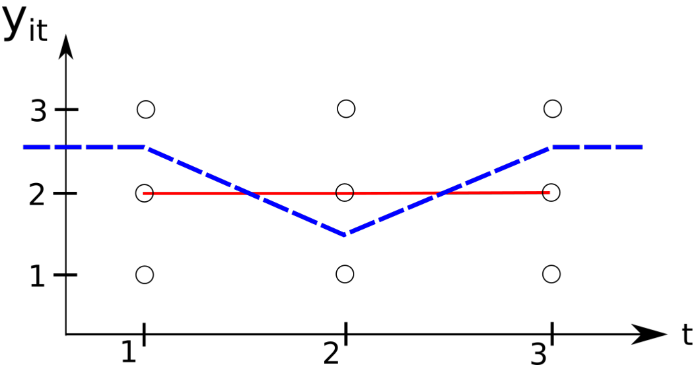

y_{it}
3
2
1
1
2
3
t



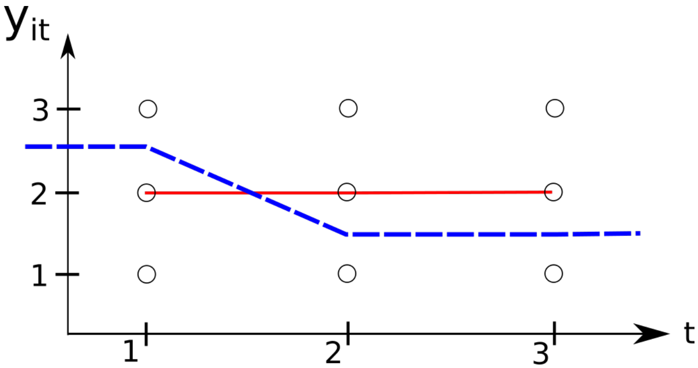
Das and van Soest: multiple cutoffs



Time-invariant transformations do not catch flat patterns



Time-varying transformations catch flat patterns



There are $(J-1)^T \geq (J-1)$ time-varying transformations

Main result (notation)

- **Cutoff categories** $\pi_t \leq J - 1$
- $\pi = (\pi_1, \cdots, \pi_T)$ is a **transformation**
- $d_{it,\pi} = 1\{y_{it} \leq \pi_t\}$ is the $\pi-$**transformed** dependent variable
    - time series for unit i: $d_{i,\pi} \in \{0, 1\}^T$
- $\bar{d}_{i,\pi} = \sum_t d_{it,\pi}$: number of times below cutoff
- $F_{\bar{d}}$ is the set of all binary $T-$vectors f with sum $\bar{d}$

Main result

Theorem

If the random vector (y_i, X_i) follows the fixed effects ordered logit model, then for any transformation π, the conditional probability distribution of the π-transformed dependent variable $d_{i,\pi}$ is given by

$$
\begin{aligned}
p_{i,\pi}\left(d \mid \beta, \gamma\right) &\equiv P\left(d_{i,\pi} = d \mid \bar{d}_{i,\pi} = \bar{d}, X_i, \alpha_i\right) && (1)\\
&= \frac{1}{\sum_{f \in F_{\bar{d}}} \exp\left\{\sum_t \left(f_t - d_t\right)\left(\gamma_{\pi(t)} - X_{it}\beta\right)\right\}} && (2)
\end{aligned}
$$

for any $d \in \{0,1\}^T$.

Main result (remarks)

1. Conditional probability does not depend on α_i
2. Sufficient statistic exists for $(J-1)^T$ transformations of y_i
3. Existing approaches use **at most** $(J-1)$ of those transformations

Main result ($T = 2$)

Evaluate the conditional probability for $d = (1, 0)$

- For any time-invariant transformation:

$$\frac{1}{1 + \exp\left\{-\left(X_{i2} - X_{i1}\right)\beta\right\}}$$

- For time-varying transformation $\pi = (j, k)$, $j \neq k$

$$\frac{1}{1 + \exp\left\{\left(\gamma_k - \gamma_j\right) - \left(X_{i2} - X_{i1}\right)\beta\right\}}$$

Identification of $\gamma_k - \gamma_j$. Intuition: subpopulation with $X_{i2} = X_{i1}$

Outline

Introduction

Model and main result

Cut points

Estimation

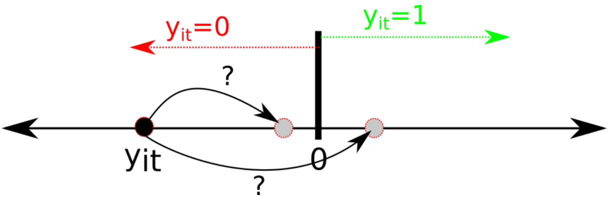
Simulations and illustration

Conclusion

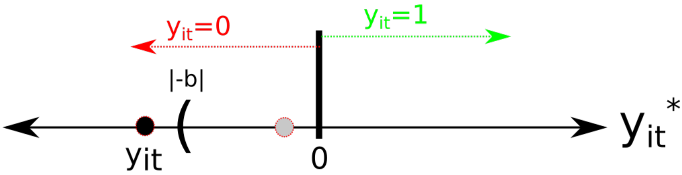
Cut points: binary

- Panel data **binary choice** ($J = 2$)**:**
  - no interpretation of the magnitude of β
  - evaluation of partial effects requires value/distribution α_i
- Existing estimators for ordered choice inherit this problem by **eliminating** thresholds
- Marginal effect of a ceteris paribus change in regressor m with coefficient β_m:

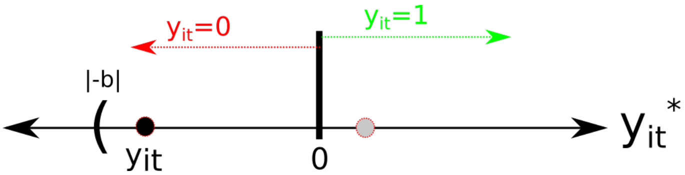
$$\frac{\partial P\left(y_{it} \leq j \middle| X_{it}, \alpha_i\right)}{\partial X_{it,m}} = \beta_m \Lambda\left(\alpha_i + X_{it}\beta - \gamma_j\right)\left[1 - \Lambda\left(\alpha_i + X_{it}\beta - \gamma_j\right)\right]$$



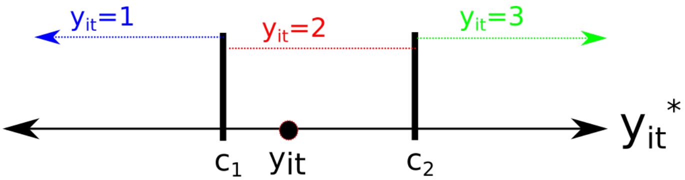
Change in y_{it} for unit change in $X_{it,m}$?



If $y_{it}^{*} = \alpha_i + X_{it}\beta + u_{it} < -\beta_m$, then y_{it} is unchanged.



No marginal effects without info on α_i or $\alpha_i | X_{it}$.



$y_{it}=1$
$y_{it}=2$
$y_{it}=3$
y_{it}^*
c_1
yit
c_2

Bounds (notation)

- Consider a ceteris paribus change in X_{it} of Δx
- The counterfactual latent dependent variable is

$$\tilde{y}_{it}^{*} = y_{it}^{*} + (\Delta x)\,\beta;$$

- $\tilde{y}_{it}$: the counterfactual ordered outcome.

Bounds

Conditional probability for the observed counterfactual outcome:

$$P\left(\tilde{y}_{it} > j \middle| y_{it} = j, X_{it}\right) = \begin{cases} 1 \text{ if } \left(\Delta x\right)\beta > \gamma_j - \gamma_{j-1}, \\ 0 \text{ if } \left(\Delta x\right)\beta < 0, \\ \frac{F_v\left(\gamma_j - X_{it}\beta\right) - F_v\left(\gamma_j - (X_{it} + \Delta x)\beta\right)}{F_v\left(\gamma_j - X_{it}\beta\right) - F_v\left(\gamma_{j-1} - X_{it}\beta\right)} & \text{else} \end{cases}$$

Paper presents a more general result along the same lines. Note: intermediate category.

Bounds (2)

Using the first component:

- **Minimum required change in** X_{itm} to move everybody with $y_{it} = j$ up:

$$\delta_m^j \equiv \frac{\gamma_j - \gamma_{j-1}}{\beta_m}$$

- Let Δx_m be the ceteris paribus change in $X_{it,m}$, then

$$\Delta x_m > \delta_m^j \Rightarrow P\left(\tilde{y}_{it} > j \middle| y_{it} = j, X_{it}\right) = 1$$

Outline

Introduction

Model and main result

Cut points

Estimation

Simulations and illustration

Conclusion

Estimation (one transformation)

- $\gamma_{\pi,\Delta}$ are the n_π cut point **differences** that show up
- $\theta_{\pi,0} = (\beta_0, \gamma_{\pi,\Delta,0})$: true **parameter** value

The **CMLE** for transformation π is

$$\hat{\theta}_\pi = \left(\hat{\beta}, \hat{\gamma}_{\pi,\Delta}\right) = \arg \max_{\mathbb{R}^K \times \mathbb{R}^{n_\pi}} \frac{1}{n} \sum_{i=1}^{n} 1\left\{d_i = d\right\} \ln p_{i\pi}\left(d \mid \theta_\pi\right)$$

Estimation (one transformation)

Assumption

The variance matrix of the regressors, $Var\left(\begin{bmatrix} X_{i1}' \\ \vdots \\ X_{iT}' \end{bmatrix}\right)$*, exists and is positive definite.*

Theorem

Let $(\{y_i, X_i\}, i = 1, \cdots, n)$ *be a random sample from the fixed effects ordered logit model, and let* π *be an arbitrary transformation. If the above assumption holds, then* $\hat{\theta}_\pi$ *is consistent and*

$$\sqrt{n}\left(\hat{\theta}_{\pi,n} - \theta_{\pi,0}\right) \xrightarrow{d} \mathcal{N}\left(0, H_\pi^{-1}\Sigma_\pi H_\pi^{-1}\right) \quad \text{as } n \to \infty, \tag{3}$$

where H_π *and* Σ_π *are the variance and expected derivative of (**??**).*

Estimation (one transformation)

The **score**

$$
\begin{aligned}
s_{i,\pi}\left(d|\,\theta_{\pi}\right) &= \begin{bmatrix} \frac{\partial \ln p_{i,\pi}(d|\theta_{\pi})}{\partial \beta} \\ \frac{\partial \ln p_{i,\pi}(d|\theta_{\pi})}{\partial \gamma_{\pi,\Delta}} \end{bmatrix} \\
&= \begin{bmatrix} s_{i,\pi,\beta}\left(\beta, \gamma_{\pi,\Delta}\right) \\ s_{i,\pi,\gamma}\left(\beta, \gamma_{\pi,\Delta}\right) \end{bmatrix}
\end{aligned}
$$

can be used to show **global concavity**

- Identification is guaranteed by condition on $var\left(vec\left(X\right)\right)$
- Assumption are as for **linear** panel model

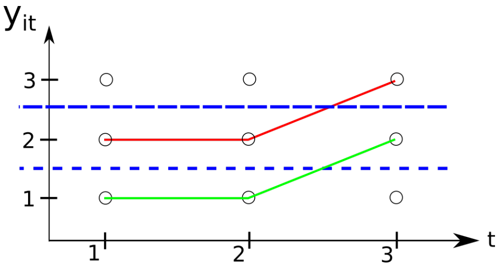
Estimation (more transformations)

CMLE is **equivalent** to solving the moment conditions

$$E\begin{bmatrix} s_{i,\pi,\beta}\left(\beta_0, \gamma_{\pi,\Delta,0}\right) \\ s_{i,\pi,\gamma}\left(\beta_0, \gamma_{\pi,\Delta,0}\right) \end{bmatrix} = 0.$$

GMM provides framework for combining information from multiple transformations.

Estimation (more transformations)



- For time-invariant transformations, $\gamma_{\pi,\Delta}$ is empty
- These are transformations used by **existing procedures**

Estimation (more transformations)

Time-invariant π.

- Combine moment conditions for β_0:

$$E\left[s_{i,1,\beta}\left(\beta_0\right)\right] = E\begin{bmatrix} s_{i,(1,\cdots,1)}\left(\beta_0\right) \\ \vdots \\ s_{i,(J-1,\cdots,J-1)}\left(\beta_0\right) \end{bmatrix} = 0 \qquad (4)$$

- GMM estimator based on (4) is

$$\tilde{\beta}_{W_{1,n}} = \arg\min \bar{s}_{1,n}\left(\beta\right)' W_{1,n} \bar{s}_{1,n}\left(\beta\right)$$

  where $\bar{s}_{1,n}\left(\beta\right) = \frac{1}{n}\sum_{i=1}^{n} s_{i,1,\beta}\left(\beta\right)$
- **Existing** procedure corresponds to choice for $W_{1,n}$
- $\tilde{\beta}^{*}$ is **optimal** estimator in this class

Estimation (even more transformations)

Main result: $(J-1)^T - (J-1)$ **additional,** time-varying transformations

- Scores involve n_γ cut point differences $\gamma_\Delta = (\gamma_{\pi,\Delta})_\pi$
- Collect the scores for γ_Δ in the $n_\gamma \times 1$ vector

$$s_{i,2,\gamma}(\beta, \gamma_\Delta) = (s_{i,\pi,\gamma}(\beta, \gamma_{\pi,\Delta}), \pi : n_\pi \geq 1)$$

- Scores for β from **time-varying** π:

$$s_{i,2,\beta}(\beta, \gamma_\Delta) = (s_{i,\pi,\beta}(\beta, \gamma_{\pi,\Delta}), \pi : n_\pi \geq 1)$$

Estimation (even more πs)

- Proposal: estimation using

$$E\left[s_i\left(\beta_0, \gamma_{\Delta,0}\right)\right] = E\begin{bmatrix} s_{i,1,\beta}\left(\beta_0\right) \\ s_{i,2,\beta}\left(\beta_0, \gamma_{\Delta,0}\right) \\ s_{i,2,\gamma}\left(\beta_0, \gamma_{\Delta,0}\right) \end{bmatrix} = 0$$

- Optimal estimator in this class: $\left(\hat{\beta}^*, \hat{\gamma}_\Delta^*\right)$
- **Question**: Is $\hat{\beta}^*$ more efficient than $\tilde{\beta}^*$?

Efficiency (result)

Theorem

Let $(\{y_i, X_i\}, i = 1, \cdots, n)$ *be a random sample from [...] Then, as* $n \to \infty$,

$$\sqrt{n}\left(\tilde{\beta}^* - \beta_0\right) \overset{d}{\to} \mathcal{N}(0, V_1),$$

$$\sqrt{n}\left(\begin{pmatrix} \hat{\beta}^* \\ \hat{\gamma}^*_\Delta \end{pmatrix} - \begin{pmatrix} \beta_0 \\ \gamma_{\Delta,0} \end{pmatrix}\right) \overset{d}{\to} \mathcal{N}(0, V),$$

where [...]. Furthermore, let V_β *be the top-left* $K \times K$ *block of* V.
Then $V_1 - V_\beta$ **is positive semidefinite**.

Efficiency (proof sketch)

- $\left(\hat{\beta}^{*}, \hat{\gamma}_{\Delta}^{*}\right)$ is based on

$$E\begin{bmatrix} s_{i,1,\beta}\left(\beta_0\right) \\ s_{i,2,\beta}\left(\beta_0, \gamma_{\Delta,0}\right) \\ s_{i,2,\gamma}\left(\beta_0, \gamma_{\Delta,0}\right) \end{bmatrix}$$

- Information from $s_{i,2,\gamma}\left(\beta_0, \gamma_{\Delta,0}\right)$ **exactly identifies** $\gamma_{\Delta,0}$
- $\beta-$estimation based on $s_{i,1,\beta}\left(\beta_0\right)$ is **unaffected** by adding $s_{i,2,\gamma}\left(\beta_0, \gamma_{\Delta,0}\right)$
- $s_{i,2,\beta}\left(\beta_0, \gamma_{\Delta,0}\right)$ yields efficiency gains for β_0

Efficiency (OMD)

The efficient minimum distance estimator based on all $\hat{\beta}_\pi$ is asymptotically equivalent to the optimal GMM estimator $\hat{\beta}^*$

Outline

Introduction

Model and main result

Cut points

Estimation

Simulations and illustration

Conclusion

Implementation

- Stata's **clogit** command implements $\hat{\theta}_\pi$
  - augment regressors with indicator for π
- Apply **clogit** for each π
- Optimally combine the results using **suest**

Simulations

$J = 3, K = 1, T = 2, N = 5000$

	β		$\gamma_2 - \gamma_1$	
Estimator	%Bias	RelSD	%Bias	RelSD
Oracle	0.0	1.00	0.03	1.00
CSLogit	14.2	0.93	6.29	0.95
$\pi = (1,1)$	0.0	1.89	-	-
$\pi = (2,2)$	0.2	2.28	-	-
$\pi = (1,2)$	0.3	4.09	0.18	3.30
$\pi = (2,1)$	0.2	1.90	0.28	3.70
DvS	0.6	1.52	-	-
OMD	**0.8**	**1.35**	**0.13**	**1.51**

Simulations (sensitivity)

Estimator	$(J=3,\ K=3)$ %Bias	$(J=3,\ K=3)$ RelSD	$(J=3,\ K=5)$ %Bias	$(J=3,\ K=5)$ RelSD	$(J=5,\ K=5)$ %Bias	$(J=5,\ K=5)$ RelSD
Coefficient β						
$\pi=(1,1)$	0.17	1.78	0.90	1.70	0.90	1.80
$\pi=(1,2)$	0.30	1.49	0.80	1.33	0.80	1.40
DvS	0.03	1.43	0.61	1.36	0.41	1.37
OMD	**0.24**	**1.22**	**0.01**	**1.16**	**1.69**	**1.28**
Cut point $\gamma_2-\gamma_1$						
$\pi=(1,2)$	0.47	4.20	0.57	5.02	0.57	5.03
$\pi=(2,1)$	0.69	11.01	2.02	14.17	2.02	14.95
OMD	**0.21**	**1.42**	**0.50**	**1.40**	**1.76**	**1.40**

Simulations: many-π bias

- Estimation of weight matrix affects small samples
- Proposal: **composite likelihood estimator (CLE)**

$$\hat{\theta}_{CLE} = \mathbf{arg}\max \sum_{\pi} \sum_{i=1}^{n} 1\{d_i = d\} \ln p_{i\pi}\left(d \middle| \theta_{\pi} \right)$$

  - sacrifices efficiency
  - robust finite-sample performance (large J, T)
  - even easier to implement in Stata (**expand** + **clogit**)

Simulations: many π results

Other parameters unchanged

	$T = 4$		$T = 6$		$T = 8$	
Estimator	%Bias	RelSD	%Bias	RelSD	%Bias	RelSD
Oracle	1.29	1.00	0.38	1.00	1.17	1.00
$\pi = (1, 1)$	2.45	1.57	0.59	1.48	1.70	1.46
BUC	1.64	1.18	0.16	1.15	1.14	1.10
DvS	1.30	1.20	0.00	1.13	0.99	1.09
CLE	**1.87**	**1.15**	**0.20**	**1.10**	**1.05**	**1.07**
OMD	**1.90**	**1.20**	**10.70**	**1.18**	**18.85**	**1.21**

Family income and children's health

- Relationship between reported (subjective) children's health status and total household income
- Seminal paper: Case et al. (2002)
  - 1. children's health is positively related to household income
  - 2. relationship is stronger for older children.
- Currie and Stabile (2003) replicate using Canadian panel
- Murasko (2008) replicates using Medical Expenditure Panel Survey (MEPS)
- Currie et al. (2007) use British data:
  - confirm finding #1
  - no evidence for #2
- Khanam et al. (2014) use Australian data
  - first to control for unobserved heterogeneity
  - no evidence for #2

Illustration (data)

- Data: Panel 16 of the Medical Expenditure Panel Survey (MEPS). US data.
  - MEPS is a rotating panel (Agency for Healthcare Research Quality, 1996)
  - Demographic and socioeconomic variables (survey)
  - Data on health and healthcare usage (admin)
- 4131 children in 2011 and 2012.
- Dependent variable: self-reported health status (*RTHLTH*)
  - "1" = "Poor" - "5" = "Excellent".
- Explanatory variables (de-meaned)
  - total household income
  - interaction age and income.
  - year dummies, family size

Illustration: results

	RE	CRE	BUC	**CLE**
$\log(\text{Income})_{it}$	-0.38	-0.10	-0.09	-0.06
	(0.03)	(0.06)	(0.07)	(0.08)
$\text{Age} \times \log(\text{Income})_{it}$	-0.014	0.017	0.021	0.035
	(0.007)	(0.013)	(0.015)	(0.016)
Family size	0.09	-0.19	-0.20	-0.23
	(0.04)	(0.14)	(0.15)	(0.16)
$\gamma_2 - \gamma_1$	2.01	2.02	-	1.87
	(0.05)	(0.05)		(0.05)
$\gamma_3 - \gamma_2$	2.96	2.97	-	2.92
	(0.09)	(0.09)		(0.12)
$\gamma_4 - \gamma_3$	2.73	2.74	-	2.61
	(0.23)	(0.23)		(0.29)

Illustration: discussion (1)

- Controlling for unobserved heterogeneity is important
- Not enough evidence for income effect at average age
- Sufficient evidence for age-dependent income-health effect
  - CLE is the only estimator to detect it.

Illustration: discussion (2)

- BUC: no cut points
- CLE: income increase $>$ 900% to move a 15-year old from Fair to Good or higher.
- Correlated random effects (CRE):
  - close to correctly specified
  - standard errors are only slightly smaller
  - 100% income increase changes probability of “Good” or above from 0.1415 to 0.1447

Outline

Introduction

Model and main result

Cut points

Estimation

Simulations and illustration

Conclusion

Conclusion

Estimation for **fixed effects ordered logit model**

- Using $(J-1)^T$ fixed effects binary choice logit models
- **Cut point differences** for bounds on partial effects
- Regression coefficient**: increased efficiency**

Extensions

1. Better bounds
2. Other panel data models
   2.1 transformation model
   2.2 interval-censored model
3. Semiparametric ordered choice
4. Dynamic ordered choice
5. Time-varying cut points