A review of estimators for the fixed effects ordered logit model

Andy Dickerson    Arne Risa Hole    Luke Munford
University of Sheffield

UK Stata Users Group Meeting 2011
Background

- There has been an increase in the use of panel data in the social sciences in recent years.
- One advantage of panel data is the ability to control for unobserved time-invariant heterogeneity.
- While random effects estimators exists for a range of limited dependent variable models few fixed effects estimators are available.
- This talk will review the available estimators for the fixed effects ordered logit (FE-OL) model and discuss ways of implementing these in Stata.
- Draws on recent paper by Baetschmann, Staub and Winkelmann (2011).
The starting point is a latent variable model

\[ y_{it}^* = x_{it}' \beta + \alpha_i + \varepsilon_{it}, \quad i = 1, \ldots, N \quad t = 1, \ldots, T \]

- \( \alpha_i \) can be assumed to be uncorrelated with \( x_{it} \) and normally distributed (random effects)
- Or we can allow \( \alpha_i \) to be correlated with \( x_{it} \) (fixed effects)
- We observe \( y_{it} \) which is related to \( y_{it}^* \) as follows

\[ y_{it} = k \quad \text{if} \quad \mu_k < y_{it}^* \leq \mu_{k+1}, \quad k = 1, \ldots, K \]

- The thresholds are assumed to be strictly increasing \((\mu_k < \mu_{k+1} \ \forall k)\) and \( \mu_1 = -\infty \) and \( \mu_{K+1} = \infty \).
\( \varepsilon_{it} \) is assumed to be IID standard logistic

Then the probability of observing outcome \( k \) for individual \( i \) at time \( t \) is

\[
\Pr(y_{it} = k|x_{it}, \alpha_i) = \Lambda(\mu_k + x_{it}'\beta - \alpha_i) - \Lambda(\mu_k - x_{it}'\beta - \alpha_i)
\]

There are two problems with ML estimation of this expression (Baetschmann et al., 2011):

- Identification: only \( \alpha_{ik} = \mu_k - \alpha_i \) can be identified
- Under fixed-\( T \) asymptotics \( \alpha_{ik} \) cannot be estimated consistently due to the incidental parameter problem
- This also affects estimates of \( \beta \) - the bias can be substantial in short panels (Greene, 2004)
The Chamberlain estimator

- Proposed solution: collapse $y_{it}$ to a binary variable and use Chamberlain’s estimator for fixed effects binary logit models
- Define $d_{it}^k = I(y_{it} \geq k)$ and $d_i^k = (d_{i1}^k, \ldots, d_{iT}^k)$
- The sum of all individual outcomes over time is a sufficient statistic for $\alpha_i$

$$P_i^k(\beta) = \Pr(d_i^k = j_i | \sum_{t=1}^{T} d_{it}^k = a_i) = \frac{\exp(j_i' x_i \beta)}{\sum_{j \in B_i} \exp(j' x_i \beta)}$$

- Chamberlain (1980) shows that maximizing the conditional log-likelihood $LL^k(b) = \sum_{i=1}^{N} \ln P_i^k(b)$ gives a consistent estimate of $\beta$
A straightforward way of estimating the FE-OL model is therefore to pick a cutoff point \( k \) and use the Chamberlain estimator.

But note that individuals with constant \( d_{it}^k \) do not contribute to the likelihood function since
\[
Pr(d_i^k = 1 | \sum_{t=1}^T d_{it}^k = T) = Pr(d_i^k = 0 | \sum_{t=1}^T d_{it}^k = 0) = 1
\]

Any particular choice of cutoff is therefore likely to lead to some observations being discarded.

The question is then whether we can do better than choosing a single cutoff.

We will review three estimators that have been proposed in the literature.
The Das and van Soest (DvS) two-step estimator

- Since the estimator of $\beta$ at any cutoff ($\hat{\beta}^k$) is consistent one can estimate the model for all $K - 1$ cutoffs and combine the estimates in a second step.
- The efficient combination weights the estimates by their variance so that

$$\hat{\beta}^{DvS} = \arg \min_{\beta} (\hat{\beta}^2 - b', ..., \hat{\beta}^K - b') \Omega^{-1} (\hat{\beta}^2 - b', ..., \hat{\beta}^K - b')'$$

- The solution to this problem is

$$\hat{\beta}^{DvS} = (H' \Omega^{-1} H)^{-1} H' \Omega^{-1} (\hat{\beta}^2', ..., \hat{\beta}^K')'$$

$H$ is the matrix of $K - 1$ stacked identity matrices of dimension $L$ (number of coefs. in the model).
The DvS estimator can be conveniently implemented in Stata as follows

Step 1: Estimate the model at each (feasible) cutoff and save the results using estimates store. I say "feasible" because some cutoffs may result in very small samples which can lead to convergence problems.

Step 2: Combine the estimates using suest. This provides an estimate of $\Omega$.

Step 3: Calculate $(H'\hat{\Omega}^{-1}H)^{-1}H'\hat{\Omega}^{-1}(\hat{\beta}^{2'},...,\hat{\beta}^{K'})'$ (estimates) and $(H'\hat{\Omega}^{-1}H)^{-1}$ (variance-covariance of estimates) using Stata’s matrix language (or Mata)

The next two slides have some example code. Note that the code assumes that the dependent variable is coded 1, ..., $K$ with no gaps.
local y y // Specify name of dependent variable after the first "y"
local x x1 x2 // Specify names of independent variables after the first "x"
local id id // Specify name of id variable after the first "id"

* Mark estimation sample
marksample touse
markout `touse' `y' `x' `id'

* Run clogit for each cutoff and combine using suest
* Note that with many (most?) datasets this part of the
* code will have to be edited since not all cutoffs can
* be used to estimate the model
qui sum `y' if `touse'
local ymax = r(max)
tempvar esample
gen `esample' = 0
tempname BMAT
forvalues i = 2(1)`ymax' {
    tempvar y`i'
    qui gen `y`i'' = `y' >= `i' if `touse'
    qui clogit `y`i'' `x' if `touse', group(`id')
    qui replace `esample' = 1 if e(sample)
estimates store `y`i''
    local suest `suest' `y`i''
capture matrix `BMAT' = `BMAT', e(b)
    if (_rc != 0) matrix `BMAT' = e(b)
}
qui suest `suest'

Dickerson, Hole, Munford, University of Sheffield
Estimators for the fixed effects ordered logit model
* Calculate Das and Van Soest estimates

tempname VMAT A B COV
local k : word count `x'
matrix `VMAT' = e(V)
matrix `A' = J((`ymax'-1),1,1)#I(`k')
matrix `B' = (invsym(`A''*invsym(`VMAT')*`A')*`A''*invsym(`VMAT')*`BMAT'')'
matrix `COV' = invsym(`A''*invsym(`VMAT')*`A')

* Tidy up matrix names and present results
matrix colnames `B' = `x'
matrix coleq `B' = :
matrix colnames `COV' = `x'
matrix coleq `COV' = :
matrix rownames `COV' = `x'
matrix roweq `COV' = :

qui cou if `esample'
local obs = r(N)
ereturn post `B' `COV', depname(`y') obs(`obs') esample(`esample')
ereturn display

* Calculate the number of individuals

tempvar last
bysort `id': gen `last' = _n==_N if e(sample)
cou if `last'==1
As an alternative to the DvS estimator Baetschmann et al. (2011) propose estimating all dichotomisations jointly subject to the restriction that $\beta^2 = \beta^3 = \cdots = \beta^K$.

This can be done by creating a dataset where each individual is repeated $K - 1$ times, each time using a different cutoff to collapse the dependent variable.

Baetschmann et al. (2011) suggests that the standard errors should be adjusted for clustering as some individuals contribute to several terms in the log-likelihood function.

This estimator does not suffer from the potential problems associated with some cutoffs resulting in small sample sizes.
The next slide has an example of how the BUC estimator can be implemented as an ado-file.

Note that the way the ID variable is created in Baetschmann et al.’s code can cause precision problems with some datasets.
*! bucologit 1.0.1 2Sept2011
*! author arh

program bucologit
    version 11.2
    syntax varlist [if] [in], Id(varname)

    preserve

    marksample touse
    markout `touse' `id'

    gettoken yraw x : varlist
    tempvar y
    qui egen int `y' = group(`yraw')

    qui keep `y' `x' `id' `touse'
    qui keep if `touse'

    qui sum `y'
    local ymax = r(max)
    forvalues i = 2(1)`ymax' {
        qui gen byte `yraw'`i' = `y' >= `i'
    }
    drop `y'

    tempvar n cut newid
    qui gen long `n' = _n
    qui reshape long `yraw', i(`n') j(`cut')
    qui egen long `newid' = group(`id' `cut')
    sort `newid'
    clogit `yraw' `x', group(`newid') cluster(`id')

    restore

end

exit
BUC example with simulated data

set more off
set seed 12345

* Generate simulated data
drop _all
set obs 1000
gen id = _n
gen u = 0.5*invnormal(uniform())
expand 10
sort id
matrix means = 0,0
matrix sds = 1,1
drawnorm x1 x2, mean(means) sd(sds)
replace x1 = 0.5*x1 + 0.5*u
gen e = logit(uniform())
gen y_star = x1 + 0.5*x2 + u + e
gen y = 1 if y_star < -4
replace y = 2 if y_star >= -4 & y_star < -2.5
replace y = 3 if y_star >= -2.5 & y_star < -1.5
replace y = 4 if y_star >= -1.5 & y_star < -0.5
replace y = 5 if y_star >= -0.5 & y_star < 0.5
replace y = 6 if y_star >= 0.5 & y_star < 2
replace y = 7 if y_star >= 2

*Run BUC model using the -bucologit- command
bucologit y x1 x2, i(id)
*Note: the i() option is equivalent to group() in the -clogit- syntax

*Compare results with standard ordered logit
ologit y x1 x2
Ferrer-i-Carbonell and Frijters (2004) have proposed an estimator where an optimal cutoff is defined for each individual. This is in contrast to the previous estimators which use all possible dichotomisations. The optimal cutoff is the one that minimises the (individual) Hessian matrix at a preliminary estimate of $\beta$. Many applied papers have instead used a simplified rule for choosing the cutoff, such as the individual-level mean or median of $y_{it}$. Baetschmann et al. (2011) show that the FF-type estimators are in general inconsistent. Stata code for implementing the FF estimator is available on request.
Empirical application

- We use the various estimators to estimate the relationship between commuting time and satisfaction with life overall and satisfaction with leisure time.
- Sample of working age individuals from the BHPS (2002-2008).
- The dependent variable is ordered and ranges from 1-7 (1=Not satisfied at all, 7=Completely satisfied).
- We use all three estimators and compare the results to a standard ordered logit model.
### Satisfaction with life overall

<table>
<thead>
<tr>
<th>Ordered Logit</th>
<th>DvS</th>
<th>BUC</th>
<th>FF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commuting Time</td>
<td>-0.102**</td>
<td>0.048</td>
<td>0.091</td>
</tr>
<tr>
<td>(0.043)</td>
<td>(0.064)</td>
<td>(0.065)</td>
<td>(0.059)</td>
</tr>
<tr>
<td>N</td>
<td>34035</td>
<td>33105</td>
<td>33302</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* p < 0.10, ** p < 0.05, *** p < 0.01

Controls: HH income, education, FT/PT work, marital status, savings, commuting mode and age. In the ordered logit model we also control for gender.
Controls: HH income, education, FT/PT work, marital status, savings, commuting mode and age. In the ordered logit model we also control for gender.
Concluding remarks

- In a simulation experiment Baetschmann et al. (2011) find that the DvS and BUC estimators generally perform well.
- The FF estimator is found to be biased.
- BUC is preferred when the number of responses in some response categories is very low.
- In our empirical application the difference between the estimators is fairly minor.