



IZA DP No. 5443

Consistent Estimation of the Fixed Effects Ordered Logit Model

Gregori Baetschmann
Kevin E. Staub
Rainer Winkelmann

January 2011

Consistent Estimation of the Fixed Effects Ordered Logit Model

Gregori Baetschmann
University of Zurich

Kevin E. Staub
University of Zurich

Rainer Winkelmann
*University of Zurich,
CESifo and IZA*

Discussion Paper No. 5443
January 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Consistent Estimation of the Fixed Effects Ordered Logit Model*

The paper re-examines existing estimators for the panel data fixed effects ordered logit model, proposes a new one, and studies the sampling properties of these estimators in a series of Monte Carlo simulations. There are two main findings. First, we show that some of the estimators used in the literature are inconsistent, and provide reasons for the inconsistency. Second, the new estimator is never outperformed by the others, seems to be substantially more immune to small sample bias than other consistent estimators, and is easy to implement. The empirical relevance is illustrated in an application to the effect of unemployment on life satisfaction.

JEL Classification: C23, C25, J28, J64

Keywords: ordered response, panel data, correlated heterogeneity, incidental parameters

Corresponding author:

Rainer Winkelmann
University of Zurich
Department of Economics
Zürichbergstr. 14
CH-8032 Zurich
Switzerland
E-mail: rainer.winkelmann@econ.uzh.ch

* We thank Paul Frijters, Arie Kapteyn and participants of 2011 Engelberg Workshop in Labor Economics for very valuable comments on an earlier draft.

1 Introduction

Economists' use of panel data has been increasing steadily over the past years. As a key advantage, such data potentially allow to solve the endogeneity problem arising from correlated time-invariant unobserved individual-specific effects. However, while implementing corresponding estimators is straightforward if the model is linear, no generally valid method exists for non-linear models.

One important application of non-linear models arises in the context of responses that are coded on a discrete and ordinal scale. Such scales are prominent in many household surveys, where they provide information on subjective assessments, judgments, or expectations. Examples are an individuals' satisfaction (with one's job, life in general, etc.) or expectations about the future (of the economy, of one's income, etc.). There are good reasons for using such subjective evaluations in empirical research, for example because they substitute for objective information that is not collected, or because the subjective responses are of their own interest. For instance, subjective health status might be more closely tied to certain behavioral responses than actual health.

The most popular regression-type models for such dependent variables are the ordered probit model and, in particular, the ordered logit model. With cross-section data, these parametric models are very easy to use and to estimate by maximum likelihood. However, extensions to a panel data context are complex and far from obvious. Unlike in the linear model, no simple transformation (such as first-differencing or within-transformation) is available that would purge the ordered response models from the individual-specific fixed effects.

While the situation is hopeless for the ordered probit model, it is more favorable for the ordered logit model, where it has been recognized early on that the estimation problem can be simplified to that of a binary logit model for which a fixed effects estimator exists, by collapsing the J categorical responses into two classes (e.g. Winkelmann and Winkelmann, 1998). The binary logit fixed effects estimator, due to Chamberlain (1980), uses the fact

that conditioning the individual likelihood contribution on the sum of the outcome over time provides an expression which is independent of the fixed effects. The effect of the time-varying regressors can then be estimated by conditional maximum likelihood (CML).

Other popular estimators for the fixed effects ordered logit model proposed in the literature are also built around the same idea of reducing the ordered model to a binary one, but aim at improving over a simple dichotomization by exploiting additional information available in the data. One approach is to estimate fixed effects logits with every possible dichotomizing cutoff point, and then combine the resulting estimates by minimum distance estimation (Das and van Soest, 1999).

A second approach is to dichotomize every individual separately, at some sort of ‘optimal’ or ‘efficient’ cutoff point (Ferrer-i-Carbonell and Frijters, 2004). The most popular variant dichotomizes the dependent variable at the individual mean, which ensures that every individual displaying some time variation in the outcome is included in the estimation. Such fixed effects ordered logit models have been used frequently in the literature. Recent applications to health economics include Jones and Schurer (2009), and Frijters, Haisken-DeNew and Shields (2004 a, b); additions to the satisfaction literature comprise Kassenboehmer and Haisken-DeNew (2009), Booth and van Ours (2008), D’Addio, Eriksson and Frijters (2007), and Frijters, Haisken-DeNew and Shields (2004).

In this article we propose a new consistent estimator for the ordered logit model with fixed effects. We then compare the existing and the new estimators in Monte Carlo simulations. There are two important findings which have implications for future applied research based on panel ordered logit models. First, we observe that individual-specific dichotomized estimators are biased in finite samples. Worse, the bias does not vanish as simulations with increased sample size are considered. We provide reasons for this observation, and show that, in general, these estimators are inconsistent. The problem is that by choosing the cutoff point based on the outcome, they produce a form of endogeneity. Second, we provide evidence on the good finite sample performance of Das and van Soest’s

(1999) estimator and our new estimator. In contrast to Das and van Soest's (1991), the new estimator remains unbiased even in very small samples. Moreover, it can be easily implemented using existing software for CML logit estimation.

The paper proceeds as follows. Section 2 presents the different estimators for the fixed effects ordered logit models. Then, we explain our Monte Carlo simulation setup and discuss its results (Section 3), followed by an application of the estimators to data from the German Socioeconomic Panel which studies the effect of unemployment on life satisfaction (Section 4). Section 5 concludes.

2 Estimators for the FE ordered logit model

2.1 The FE ordered logit model

The fixed effects ordered logit model relates the latent variable y_{it}^* for individual i at time t to a linear index of observable characteristics x_{it} and unobservable characteristics α_i and ε_{it} :

$$y_{it}^* = x_{it}'\beta + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, N \quad t = 1, \dots, T \quad (1)$$

The time-invariant part of the unobservables, α_i , can be statistically dependent of x_{it} . In this case, one can either make an assumption regarding the joint distribution of α_i and x_{it} , or else treat α_i as a fixed effect. This paper considers estimation under the fixed effects approach.

The latent variable is tied to the (observed) ordered variable y_{it} by the observation rule:

$$y_{it} = k \quad \text{if} \quad \tau_k < y_{it}^* \leq \tau_{k+1}, \quad k = 1, \dots, K$$

where thresholds τ are assumed to be strictly increasing ($\tau_k < \tau_{k+1} \quad \forall k$) and $\tau_1 = -\infty$, $\tau_{K+1} = \infty$. It is possible to formulate the model more generally with individual-specific thresholds (Ferrer-i-Carbonell and Frijters, 2004):

$$y_{it} = k \quad \text{if} \quad \tau_{ik} < y_{it}^* \leq \tau_{ik+1}, \quad k = 1, \dots, K \quad (2)$$

The distributional assumption completing the specification of the fixed effects ordered logit model is that conditionally on x_{it} and α_i , ε_{it} are IID standard logistically. I.e., if $F(\cdot)$ denotes the cdf of ε_{it}

$$F(\varepsilon_{it}|x_{it}, \alpha_i) = F(\varepsilon_{it}) = \frac{1}{1 + \exp(-\varepsilon_{it})} \equiv \Lambda(\varepsilon_{it}) \quad (3)$$

Hence, the probability of observing outcome k for individual i at time t using (1), (2), (3) is

$$\Pr(y_{it} = k|x_{it}, \alpha_i) = \Lambda(\tau_{ik+1} - x'_{it}\beta - \alpha_i) - \Lambda(\tau_{ik} - x'_{it}\beta - \alpha_i) \quad (4)$$

which depends not only on β and x_{it} , but also on α_i and τ_{ik}, τ_{ik+1} .

There are two problems with Maximum Likelihood (ML) estimation based on expression (4). The first is a problem of identification: τ_{ik} cannot be distinguished from α_i ; only $\tau_{ik} - \alpha_i \equiv \alpha_{ik}$ is identified and can thus, in principle, be estimated consistently for $T \rightarrow \infty$. The second problem arises, since in most applications, T must be treated as fixed and relatively small. But under fixed- T asymptotics even α_{ik} cannot be estimated consistently, due to the incidental parameter problem (see, for instance, Lancaster, 2000). This does have consequences for estimation of β – the bias in α_{ik} contaminates $\hat{\beta}$. In short panels, the resulting bias in $\hat{\beta}$ can be substantial (Greene, 2004).

We next consider different approaches to estimate β consistently. They all use the same idea of collapsing y_{it} into a binary variable and then applying the sufficient statistic suggested by Chamberlain (1980) to construct a CML estimator.

2.2 Chamberlain's CML estimator for the dichotomized ordered logit model

Let d_{it}^k denote the binary dependent variable that results from dichotomizing the ordered variable at the cutoff point k : $d_{it}^k = \mathbb{1}(y_{it} \geq k)$. By construction, $P(d_{it}^k = 0) = P(y_{it} < k) = \Lambda(\tau_{ik+1} - x'_{it}\beta - \alpha_i)$, and $P(d_{it}^k = 1) = P(y_{it} \geq k) = 1 - \Lambda(\tau_{ik+1} - x'_{it}\beta - \alpha_i)$. Now consider

the joint probability of observing $d_i = (d_{i1}^k, \dots, d_{iT}^k) = (j_{i1}, \dots, j_{iT})$ with $j_{it} \in \{0, 1\}$. The sum of all the individual outcomes over time is a sufficient statistic for α_i as

$$\mathcal{P}_i^k(\beta) \equiv \Pr \left(d_i^k = j_i \left| \sum_{t=1}^T d_{it}^k = a_i \right. \right) = \frac{\exp(j_i' x_i \beta)}{\sum_{j \in B_i} \exp(j' x_i \beta)} \quad (5)$$

does not depend on α_i and the thresholds. In (5), $j_i = (j_{i1}, \dots, j_{iT})$, x_i is the $(T \times L)$ -matrix with t th row equal to x_{it} , L is the number of regressors and $a_i = \sum_{t=1}^T j_{it}$. The sum in the denominator goes over all vectors j which are elements of the set B_i

$$B_i = \left\{ j \in \{0, 1\}^T \left| \sum_{t=1}^T j_t = a_i \right. \right\},$$

i.e., over all possible vectors of length T which have as many elements equal to 1 as the actual outcome of individual i (a_i). The number of j -vectors in B_i , and therefore of terms in the sum in the denominator of (5), is $\binom{T}{a_i} = \frac{T!}{a_i!(T-a_i)!}$.

Chamberlain (1980) shows that maximizing the conditional likelihood

$$\log \mathcal{L}^k(b) = \sum_{i=1}^N \log \mathcal{P}_i^k(b) \quad (6)$$

gives a consistent estimate for β (subject to mild regularity conditions on the distribution of α_i , cf. Andersen, 1970). I.e. the score —the gradient of the log-likelihood with respect to β — converges to zero when evaluated at the true β :

$$\text{plim} \frac{1}{N} \sum_i s_i^k(\beta) = 0, \quad (7)$$

where

$$s_i^k(b) = \frac{\partial \ln \mathcal{P}_i^k(b)}{\partial b} = x_i' \left(d_i^k - \sum_{j \in B_i} j \frac{\exp(j' x_i b)}{\sum_{l \in B_i} \exp(l' x_i b)} \right) \quad (8)$$

The reason why (7) holds is that for every i , conditional on x_i , the expectation of the term in parentheses in (8) is zero as it defines a conditional expectation residual.

Note that conditioning on a_i causes all time-invariant elements in (4) to cancel. I.e., not only α_i and τ_{ik+1} are not estimated, but also elements of the β vector corresponding to observables that do not change over time. Also, individuals with constant d_{it}^k do not

contribute to the conditional likelihood function, as $P(d_i^k = 1 | \sum_{t=1}^T d_{it}^k = T) = P(d_i^k = 0 | \sum_{t=1}^T d_{it}^k = 0) = 1$.

The Hessian is

$$H_i^k(b) = \frac{\partial^2 \ln \mathcal{P}_i^k(b)}{(\partial b)(\partial b)'} = - \sum_{j \in B_i} \frac{\exp(j' x_i b)}{\sum_{l \in B_i} \exp(l' x_i b)} \times \left(x_i' j - \sum_{m \in B_i} \frac{\exp(m' x_i b)}{\sum_{l \in B_i} \exp(l' x_i b)} m' x_i \right) \left(x_i' j - \sum_{m \in B_i} \frac{\exp(m' x_i b)}{\sum_{l \in B_i} \exp(l' x_i b)} m' x_i \right)' \quad (9)$$

2.3 Combining all possible dichotomizations: Das and van Soest's (1999) two-step estimation, and a new approach

The estimator of β based on (6), say $\hat{\beta}^k$, does not use all the variation in the ordered dependent variable y_{it} , as individuals for which either $y_{it} < k$ or $y_{it} \geq k$ for every t do not contribute to the log-likelihood. Since every $\hat{\beta}^k$ for $k = 2, \dots, K$ provides a consistent estimator of β , and every individual with some variation in y_{it} will contribute to at least one log-likelihood $\mathcal{L}^k(b)$, one can perform CML estimation on all possible $K-1$ dichotomizations and then, in a second step, combine the resulting estimates. The efficient combination will weight the $\hat{\beta}^k$ by the inverse of their variance (Das and van Soest, 1999):

$$\hat{\beta}^{DvS} = \arg \min_b (\hat{\beta}^{2'} - b', \dots, \hat{\beta}^{K'} - b') \Omega^{-1} (\hat{\beta}^{2'} - b', \dots, \hat{\beta}^{K'} - b')' \quad (10)$$

The variance Ω has entries ω_{gh} , $g = 2, \dots, K$, $h = 2, \dots, K$, such that

$$\omega_{gh} = \left[\mathbb{E} \left(\frac{\partial \log \mathcal{P}^g}{\partial b} \right) \left(\frac{\partial \log \mathcal{P}^g}{\partial b} \right)' \right]^{-1} \left[\mathbb{E} \left(\frac{\partial \log \mathcal{P}^g}{\partial b} \right) \left(\frac{\partial \log \mathcal{P}^h}{\partial b} \right)' \right] \left[\mathbb{E} \left(\frac{\partial \log \mathcal{P}^h}{\partial b} \right) \left(\frac{\partial \log \mathcal{P}^h}{\partial b} \right)' \right]^{-1}$$

evaluated at $b = \beta$. In practice, the unknown variance Ω is replaced by an estimate $\hat{\Omega}$ which is evaluated at $\hat{\beta}^k$, $k = 2, \dots, K$. The solution to (10) is

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

where H is the matrix of $K-1$ stacked identity matrices of dimension L (the size of each $\hat{\beta}^k$). An estimate of the variance of the estimator can be obtained as

$$\widehat{\text{Var}}(\hat{\beta}^{DvS}) = \left(H' \hat{\Omega}^{-1} H \right)^{-1}$$

Because $\hat{\beta}^{DvS}$ is a linear combination of consistent estimators, it is itself consistent. Ferrer-i-Carbonell and Frijters (2004) discuss some small sample issues which might affect the performance of $\hat{\beta}^{DvS}$. For instance, one concern is that $\hat{\Omega}$ might be estimated very imprecisely when for some g and h there are only few observations with nonzero contributions to $\hat{\omega}_{gh}$. This is the case when there is only a small overlap between the samples contributing to the CML logit estimator dichotomized at g and the one dichotomized at h .

Thus, we propose an alternative to this two-step combination of all possible dichotomizations which avoids such problems by estimating all dichotomizations jointly subject to the restriction $\beta^k = \beta \forall k = 2, \dots, K$. Hence, the sample (quasi-) log-likelihood of this restricted CML estimator is

$$\log \mathcal{L}(b) = \sum_{k=2}^K \log \mathcal{L}^k(b) \quad (11)$$

The score of this estimator is the sum of the scores of the CML logit estimators. Since these estimators are consistent, their scores converge to zero in probability. It follows that the probability limit of the score of the restricted CML estimator is zero as well, establishing its consistency:

$$\text{plim} \sum_{k=2}^K \frac{1}{N(K-1)} \sum_{i=1}^N s_i^k(\beta) = \text{plim} \frac{1}{N} \sum_i s_i^2(\beta) + \dots + \text{plim} \frac{1}{N} \sum_i s_i^K(\beta) = 0, \quad (12)$$

Since some individuals contribute to several terms in the log-likelihood this creates dependence between these terms, invalidating the usual estimate of the estimator variance based on the information matrix equality. Instead, a sandwich variance estimator (White, 1982) should be used. We propose using the cluster-robust variance estimator which allows for arbitrary correlation within the various contributions of any individual:

$$\widehat{\text{Var}}(\hat{\beta}) = \left(\sum_{i=1}^N \hat{h}_i \right)^{-1} \left(\sum_{i=1}^N \hat{s}_i \hat{s}_i' \right)^{-1} \left(\sum_{i=1}^N \hat{h}_i \right)^{-1}$$

where \hat{s}_i are the stacked CML scores of individual i evaluated at $\hat{\beta}$, $\hat{s}_i = (\hat{s}_i^{2'}, \dots, \hat{s}_i^{K'})'$, and \hat{h}_i is the matrix of derivatives of s_i with respect to β evaluated at $\hat{\beta}$.

We will refer to this estimator as the BUC estimator. The acronym stands for “Blow-Up and Cluster” which describes the way of implementing this estimator using the CML estimator: Replace every observation in the sample by $K - 1$ copies of itself (“blow-up” the sample size), and dichotomize every $K - 1$ copy of the individual at a different cutoff point. Estimate CML logit using the entire sample; these are the BUC estimates. Cluster standard errors at the individual level. This implementation requires but a few lines of code in standard econometric software (cf. Appendix A, which contains code for implementation in Stata).

2.4 Endogenous dichotomization: Ferrer-i-Carbonell and Frijters (2004) and related approaches

The previous approaches used all possible dichotomizations. Ferrer-i-Carbonell and Frijters (2004) proposed an estimator which chooses dichotomizations separately for every individual. The (quasi-) log-likelihood for their estimator can be written as

$$\log \mathcal{L}^{FF}(b) = \sum_{i=1}^N \sum_{k=2}^K w_i^k \log \mathcal{P}_i^k(b), \quad w_i^k \in \{0, 1\}, \quad \sum_{k=2}^K w_i^k = 1 \quad (13)$$

This objective function is maximized with respect to b after choosing the cutoff point at which to dichotomize each y_i , i.e. after deciding which one of the individual’s weight vectors w_i^k is equal to 1.

Ferrer-i-Carbonell and Frijters’ (2004) approach here is to calculate for every individual all Hessian matrices under different cutoff points and choosing the smallest:

$$w_i^k = 1 \quad \text{if} \quad k = \arg \min_{\kappa} \frac{\partial^2 \log \mathcal{P}_i^{\kappa}(b)}{(\partial b)(\partial b)'} \Big|_{b=\beta}$$

In practice, the Hessian is evaluated at $\hat{\beta}$, where $\hat{\beta}$ is a preliminary consistent estimator. Since for every possible dichotomization the choice falls on the cutoff point leading to the

smallest Hessian, this rule should yield the estimator of (13) with minimal variance. Other, simpler rules for choosing w_i^k for (13) have been used, trading efficiency for computational ease. In fact, the standard way in which this estimator is implemented in the applied literature is by choosing the dichotomizing cutoff point as the mean of the dependent variable:

$$w_i^k = 1 \quad \text{if} \quad k = \text{ceil} \left(T^{-1} \sum_t y_{it} \right)$$

where $\text{ceil}(z)$ stands for rounding z up to the nearest integer. This ensures that every individual with time-variation in y_i will be part of the estimating sample. Studies using both rules report little difference in estimates and standard errors, which has led to the view that this way of choosing w_i^k is an approximation to Ferrer-i-Carbonell and Frijters' (2004). An alternative is using the median instead of the mean as a rule to define the individual dichotomization.

Thus, all these procedures choose the dichotomizing cutoff point endogenously, since it depends on y_i . This is obviously problematic and we show in Appendix B that these estimators are, in general, inconsistent. Here we provide some intuition for this result using the mean-cutoff estimator as an example; similar arguments hold for the other estimators.

The problem is not, as one might suspect, that the cutoffs vary between individuals *per se*. For instance, if the variation of the cutoffs between individuals was random, the resulting estimator would be consistent: the score would be a sum of scores of CML logit estimators, much like the BUC estimator (but with $K - 1$ times less observations as each individual would contribute only to exactly one CML logit estimator). I.e., in terms of (8), for every random individual-specific cutoff, the resulting vectors d_i converge to their respective conditional expectation, yielding an expected score of zero at the limit.

The real problem lies in the endogeneity of the cutoff. For the mean estimator, $d_{it}^{\text{Mn}} = 1$ if and only if $y_{it} \geq T^{-1} \sum_t y_{it}$. Thus, y_{it} itself is part of the cutoff and the probability

$\Pr(d_{it}^{\text{Mn}} = 1)$ can be written as

$$\Pr(d_{it}^{\text{Mn}} = 1) = \Pr\left(y_{it} \geq \frac{1}{T} \sum_t y_{it}\right) = \Pr\left(y_{it} \geq \frac{1}{T-1} \sum_{s \neq t} y_{is}\right)$$

The expression after the first equality makes clear that for any t , y_{it} is on both sides of the inequality sign. Solving for y_{it} shows that the probability $\Pr(d_{it} = 1)$ is equal to the probability that the outcome in t is greater than the average outcome in the remaining periods. In general, this is a different dichotomizing cutoff point within the same individual for every period. Thus, although the researcher is setting an individual-specific cutoff, say k , the endogenous way in which this cutoff is chosen implies that it is equivalent to choosing different cutoff points for the same individual. With endogenous cutoffs the conditional distribution of d_i can be shown to differ from the CML terms, and the score of these estimators will, in general, not converge to zero.

3 Monte Carlo simulations

We compare the performance of the estimators discussed in the previous section in finite samples using Monte Carlo simulations. To the best of our knowledge, this is the first investigation of these estimators in a Monte Carlo study. The aim is to assess the small sample biases and efficiency across different data generating processes.

3.1 Experimental design

The setup of the Monte Carlo experiment is as follows. The data generating process (DGP) for the latent variable is

$$y_{it}^* = \beta_x x_{i,t} + \beta_d d_{i,t} + \alpha_i + \varepsilon_{it},$$

and we set $\beta_x = 1$, $\beta_d = 1$. The regressor x is continuous, while d is binary. We follow Greene (2004) in specifying the fixed effects as

$$\alpha_i = \sqrt{T} \bar{x}_i + \sqrt{T} \bar{u}_i, \quad \bar{x}_i = T^{-1} \sum_t x_{it}, \quad \bar{u}_i = T^{-1} \sum_t u_{it}, \quad u_{it} \sim N(0, 1)$$

For the simulations, we use fixed (not individual-specific) thresholds:

$$y_{it} = k \quad \text{if} \quad \tau_k < y_{it}^* \leq \tau_{k+1}, \quad k = 1, \dots, K$$

Finally, ε_{it} is sampled from a logistic distribution as in (3).

The baseline DGP is a balanced panel of $N=500$ individuals observed for $T=4$ periods. The continuous regressor x is distributed as standard normal, the binary regressor’s probability of a 1 is 50%. The latent variable is discretized into $K=5$ categories, choosing the thresholds to yield the marginal distribution depicted in the upper left graph in Fig. 1. We call this distribution of y “skewed”.

The baseline DGP is modified in a number of dimensions, which can be broadly classified into two experiments. First, different kinds of asymptotics are considered by increasing N , T and K . Second, the influence of the data distribution is explored by sampling from different distributions from the regressors, and by shifting the thresholds to yield different marginal distributions for y_{it} . In the following section, we comment on selected results from these experiments. A supplementary appendix containing full simulation output from a comprehensive exploratory study is available from the authors on request.

3.2 Results

Table 1 contains results for the baseline scenario. Columns contain mean and standard deviation of estimated coefficients (labeled M and SD), as well as the mean of standard errors (labeled SE) corresponding to x (first three columns) and d (last three columns). Every row gives these results for a different estimator. All entries have been rounded to two decimal places.

The first row, named DvS, contains results for the two-step estimator of Das and van Soest (1999). With means of 0.99 for $\hat{\beta}_x$ and 1.00 for $\hat{\beta}_d$ DvS is virtually unbiased. The BUC estimator, whose results are displayed in the second row, produces unbiased results, too. There is almost no perceivable difference in efficiency between the two estimators.

Estimation of the coefficient corresponding to the binary variable is less precise than that of the continuous regressor — its standard deviation is around 60% higher.

The next three rows contain results for Ferrer-i-Carbonell and Frijters' (2004) estimator (named FF), as well as for the variants dichotomizing at the individual mean (labeled Mean) and at the individual median (labeled Median). These three estimators display standard deviations of the same size as BUC's and DvS'. However, their means shows a clear downward bias. E.g., for $\hat{\beta}_x$, it ranges from 7% for FF to 4% for Mean. With a standard deviation of 0.07 and 1,000 replications, the margin of error at 99% confidence for these biases is less than 0.6%-points.

The last four rows contain results for CML logit estimators dichotomized at the categories 2 to 5 (named ' $y \geq 2$ ' to ' $y \geq 5$ '). As DvS and BUC, these estimators show little finite sample bias. The standard deviations are at best about 30% larger than BUC's — this corresponds to cases where the dichotomized dependent variable has a distribution which is as balanced as possible. For ' $y \geq 2$ ' and ' $y \geq 3$ ' the percentage of zeros is 40% and 70%. For ' $y \geq 5$ ' this percentage is 95%, and the standard deviation of the estimator is more than double that of BUC.

Comparing columns containing the standard deviations of the estimators (SD) with columns containing average standard errors (SE) shows that standard errors are estimated satisfactorily in all cases.

Taken together, the results of Table 1 contain two important findings. First, there is no evidence that finite sample issues affect the DvS estimator. All estimators exploiting more information in the data than CML logit estimators with fixed cutoffs are, indeed, more efficient than them. However, they all display about the same standard deviation. Second, estimators based on endogenous dichotomizing cutoff points are all biased in this setup.

Next, we want to check whether these results can be generalized to other settings. We start by conducting asymptotic exercises to explore under which conditions the biases of FF, Mean and Median can be expected to vanish. The results are reported in Table 2.

The first panel of Table 2 ('Baseline scenario'), consisting of the first four columns, copies the results from Table 1 for easier comparability. Columns with averages of standard errors (SE) were dropped to avoid clutter; we found that results for SE were similar to Table 1's for all DGPs considered in this paper. In the second panel ('N=1,000', the next four columns) the effect of increasing sample size with fixed T is considered. As expected by the ratio $\sqrt{500}/\sqrt{1,000}$ the standard deviation falls by 30% for all estimators. As before, DvS and BUC are unbiased. However, FF, Mean and Median estimators remain biased. Indeed, their bias is essentially the same with 1,000 individuals as with 500. This suggests that these are not small sample biases, but that they can be attributed entirely to these estimators' inconsistency.

A different asymptotic experiment holds N fixed and increases the number of time periods. Based on the discussion of the inconsistency of estimators with endogenous dichotomization, we would expect this to have an attenuating impact on their biases: As T increases, the contribution of any y_{it} to the endogenous cutoff (a function of all y_{it} of an individual) decreases. If its contribution was zero, the cutoff would be exogenous. For instance, this is particularly transparent for the mean estimator. In the probability $\Pr(d_{it}^{\text{Mn}} = 1) = \Pr\left(y_{it} \geq \frac{1}{T-1} \sum_{s \neq t} y_{is}\right)$, the threshold consisting of the average y_{is} , $s \neq t$, becomes less variable for different t as T increases.

The results for this experiment are reported in the next panel, labeled 'T=8', where the number of time periods in the simulations were duplicated from T=4 to T=8. The decrease in the standard deviations relative to the Baseline scenario is roughly of the same magnitude as in the experiment with N=1,000. Clearly, the biases of FF, Mean and Median are reduced, consistent with our expectation.

A last kind of informal asymptotic experiment which can be conceived is increasing the number of categories. In the limit, the observed variable would be equal to the continuous latent variable. We increase the number of categories from K=5 to K=10, setting the marginal distribution to the one displayed in the lower right panel of Fig. 1. While this

distribution is skewed, too, it is of course not exactly the same as in the baseline case. The results are displayed in the fourth panel in Table 2, labeled ‘K=10’. There are now 5 additional CML logit estimators ($y \geq 6$ to $y \geq 10$), but for the sake of brevity we omit results for these. While Dvs and BUC are almost invariant to the increase in the number of ordered categories, the three estimators based on endogenous dichotomization worsen in terms of bias. This, too, is to be expected. With increasing K and fixed T, the sensitivity of endogenous cutoffs to a particular y_{it} will increase in general. For the mean estimator, for instance, the variance in the mean y_{is} , $s \neq t$ increases with K. It is interesting to note that the median estimator suffers more severely from increasing K, which is in line with the fact that the variance of the median y_{it} is larger than that of the mean y_{it} in our distributions of y_{it} .

A noteworthy constant in the discussion of results so far has been the equally good performance of DvS and BUC. This is remarkable as previous literature raised the concern that the DvS estimator could show difficulties when confronted with small samples for the different CML logit estimates. In the setup with K=10 and N=500 the last two CML logit estimators ($k=9$ and $k=10$) used on average about 137 and 78 individuals. DvS is only slightly (but statistically significantly) biased downwards. The last panel in Table 2 shows the results from a smaller sample of N=100 while maintaining K=10. This produces a difficult DGP for DvS, as only about 28 and 29 individuals are used in the CML logit estimations of $k=9$ and $k=10$. This resembles the situation in life satisfaction studies, where responses in lower categories are extremely infrequent (Ferrer-i-Carbonell and Frijters, 2004). Here we do find biases of -6% and -7% for DvS (margin of error at 99%: 1%- and 2%-points, respectively). The BUC estimator in contrast remains as unbiased as in previous DGPs. FF, Median and Mean estimators also show little change and are as biased as with N=500.

The influence of the distribution of the data on the performance of the estimators is addressed in the DGPs whose results are shown in Table 3. Again, the first panel repeats

the results for the baseline case from Table 1. The next two panels —with headings ‘bell-shaped y ’ and ‘uniform y ’ — show results for different marginal distributions of y_{it} . I.e., all parameters from the baseline DGP are kept unchanged, except for thresholds τ which have been shifted to yield these distributions (cf. Fig. 1). These changes in y_{it} seem to have close to no impact on the performance of the estimators. Only CML logit estimators are affected in their precision. It is no surprise that, for given distribution of x, d , the more balanced the distribution of the dichotomized variable, the higher the precision of the resulting CML logit estimator. The last panel in Table 3 resets the thresholds to their baseline values and changes the distribution of the explanatory variables. The continuous x is now drawn from a log-normal distribution, standardized to have mean zero and unit variance; the binary d ’s new distribution is highly unbalanced with only 10% of observations having $d = 1$ on average. As before, the picture remains by and large the same: All estimators show higher standard deviations in this DGP, but the ranking is unchanged.

4 Application: Why are the unemployed so unhappy?

The preceding section documented the performance of different estimators for the fixed effects ordered logit model in simulations. In this section, the estimators are used to reestimate the effect of unemployment on life satisfaction using the dataset of Winkelmann and Winkelmann (1998). The data consists of a large sample from the German Socioeconomic Panel, totaling at 20,944 observations; the model includes 9 explanatory variables. With these values, the application provides a typical setting to which the estimators have been put to use and are likely to be applied in the future.

4.1 Data and specification

The sample consists of the first six consecutive waves of the German Socioeconomic Panel going from 1984 to 1989. It includes all observations of persons aged 20-64 years with

participation in at least two waves and non-missing responses for all variables of the model. These are 20,944 person-year observations corresponding to 4,261 individuals. Of these, 1,873 observations corresponding to 303 persons are discarded because they do not display any variation over time in their outcome variable, leaving the dataset with 19,079 observations corresponding to 3,958 individuals.

The outcome variable is satisfaction with life which is measured as the answer to the question “*How satisfied are you at present with your life as a whole?*”. The answer can be indicated in 11 ordered categories ranging from 0, “completely dissatisfied”, to 10, “completely satisfied”.

The key explanatory variables are a set of three dummy variables which indicate current labor market status: *Unemployed*, *Employed* and *Out of labor force*. These dummies exhaust the possible labor market status and are mutually exclusive, so *Employed* is used as the omitted reference category in the model.

Additional information about psychological costs of unemployment might be revealed through the length of the unemployment spell. Thus, the model contains the variables *Duration of unemployment* and *Squared duration of unemployment*.

Demographic control variables include marital status (*Married*), health status (*Good health*), age (*Age* and *Squared age*) and household income (in logarithms, *Log. household income*). We refer to the original source for comprehensive descriptions of data and specification.

4.2 Results

Estimation results are presented in Table 4. Every column depicts results of the same model for a different estimator. The first replicates the original results in Winkelmann and Winkelmann (1998, Table 4, column 2, p.11) who used a CML logit estimator dichotomized at the cutoff 8. This cutoff results in a distribution of the binary dependent variable which is about balanced with around 50% of the responses being equal or greater than 8. In

total 2,573 individuals cross this cutoff resulting in an estimating sample size of 12,980 observations.

To briefly summarize the results, the effect of unemployment is found to be both large and statistically significant; there is no effect of unemployment duration on life satisfaction, so there seems to be no mental adaptation process of unemployed persons to their status. Coefficients of socio-demographic variables display expected signs and magnitudes (cf. Clark and Oswald, 1994).

Moving to the right of the table, the next two columns correspond to results obtained using the DvS and BUC estimators, and the final three columns show results for FF, Mean and Median estimates. The most striking feature of the results as a whole is that the first three columns—which are based on consistent estimators—are remarkably similar, while they differ from the three last columns containing estimates from inconsistent estimators. The marginal effect of unemployment on latent life satisfaction is estimated to be around -1 when using CML logit, DvS or BUC; but it ranges only from -0.84 to -0.66 when using FF, Mean or Median estimators. Similarly, effects for marital status and age are estimated to be larger using either of the consistent estimators. Although estimation is not precise enough to reject equality of coefficients, these results clearly echo patterns from the Monte Carlo simulations. There is only one clear difference between consistent estimators. It relates to the coefficient of *Out of labor force*, which is -0.24 and insignificant for CML logit while being around -0.45 and significant for DvS and BUC. A potential explanation for this is that most of the changes in *Out of labor force* occur at levels of satisfaction lower than the cutoff used by the CML logit estimator, so that this information is lost to the CML logit estimation. DvS and BUC, on the other hand, use all 3,958 persons displaying some time variation in life satisfaction (for BUC the number of persons corresponds to the number of clusters; the number of individuals is the cross-sectional dimension of the “blown-up” or inflated sample).

All estimators display similarly sized standard errors. CML logit estimates are the

largest, but this difference is not pronounced. BUC's standard errors are slightly larger than DvS and the inconsistent estimators', the difference being minimal. There is another negligible difference between estimators. While Mean and Median estimators use the same number of persons as DvS and BUC for estimation (3,958), FF uses 9 individuals (or 0.2%) less. This means that for these 9 individuals, the smallest Hessians (remember that the smallest Hessian determines the dichotomizing cutoff for FF) were to be found for cutoffs which lead to no time variation in y .

5 Conclusions

This article studied extant estimators for the fixed effects ordered logit model and proposed a new one. All these estimators are based on CML binary logit estimation. Estimators most represented in the literature are characterized by selecting the dichotomizing cutoff point endogenously, i.e. as a function of the outcome of the dependent variable. In general, this will lead to inconsistency of the estimator, a result which was extensively documented in Monte Carlo simulations. The consistent estimators, Das and van Soest's (1999) minimum distance estimator and our BUC estimator, clearly outperformed simple CML logit estimation in terms of efficiency. Their performance in several DGPs of the Monte Carlo simulations and in a large-scale application using survey data from Germany indicates that they are recommendable for applied work. If the ordered dependent variable displays extremely low responses for some categories, our simulation evidence suggest that BUC estimation is preferable.

References

- Andersen, Erling B. (1970), “Asymptotic Properties of Conditional Maximum-likelihood Estimators”, *Journal of the Royal Statistical Society, Series B (Methodological)*, **32**, 283-301.
- Booth, Alison L., and Jan C. van Ours (2008), “Job Satisfaction and Family Happiness: The Part-time Work Puzzle”, *Economic Journal*, **118** F77-F99.
- Chamberlain, Gary (1980), “Analysis of covariance with qualitative data”, *Review of Economic Studies*, **47**, 225-238.
- Clark, Andrew E. and Andrew J. Oswald (1994), “Unhappiness and unemployment”, *Economic Journal*, **104**, 648-659.
- Das, Marcel, and Arthur van Soest (1999), “A panel data model for subjective information on household income growth”, *Journal of Economic Behavior & Organization*, **40**, 409-426.
- D’Addio, Anna Cristina, Tor Eriksson and Paul Frijters (2007), “An Analysis of the Determinants of Job Satisfaction when Individuals’ Baseline Satisfaction Levels May Differ”, *Applied Economics*, **39**, 2413-2423.
- Ferrer-i-Carbonell, Ada, and Paul Frijters (2004), “How important is methodology for the estimates of the determinants of happiness?”, *Economic Journal*, **114**, 641-659.
- Frijters, Paul, John P. Haisken-DeNew and Michael A. Shields (2004 a), “Investigating the Patterns and Determinants of Life Satisfaction in Germany Following Reunification”, *Journal of Human Resources*, **39**, 649-674.
- Frijters, Paul, John P. Haisken-DeNew and Michael A. Shields (2004 b), “Money Does Matter! Evidence from Increasing Real Income and Life Satisfaction in East Germany Following Reunification”, *American Economic Review*, **94**, 730-740.
- Frijters, Paul, John P. Haisken-DeNew and Michael A. Shields (2005), “The causal effect of income on health: Evidence from German reunification”, *Journal of Health Economics*, **24**, 997-1017.

Greene, William H. (2004), “The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects”, *Econometrics Journal*, **7**, 98-119.

Jones, Andrew M., and Stephanie Schurer, “How does heterogeneity shape the socioeconomic gradient in health satisfaction?”, forthcoming in *Journal of Applied Econometrics*, published online December 14 2009, DOI: 10.1002/jae.1133.

Kassenboehmer, Sonja C., and John P. Haisken-DeNew (2009), “You’re Fired! The Causal Negative Effect of Unemployment on Life Satisfaction”, *Economic Journal*, **119**, 448-462.

White, Halbert (1982), “Maximum likelihood estimation of misspecified models”, *Econometrica*, **50**, 1-25.

Winkelmann, Liliana, and Rainer Winkelmann (1998), “Why are the unemployed so unhappy? Evidence from panel data”, *Economica*, **65**, 1-15.

A Implementing the BUC estimator in Stata

To perform BUC estimation in Stata, run the following code, replacing `ivar` `yvar` `xvar` in the last program line as follows:

`ivar` is the individual identifier,
`yvar` is the ordered dependent variable, and
`xvars` is the list of explanatory variables.

```
capture program drop feologit_buc
program feologit_buc, eclass
    version 10
    gettoken gid 0: 0
    gettoken y x: 0
    tempvar iid id cid gidcid dk

    qui sum `y'
    local lk= r(min)
    local hk= r(max)
    bys `gid': gen `iid'=_n
    gen long `id'=`gid'*100+`iid'
    expand `='hk'-'lk''
    bys `id': gen `cid'=_n
    qui gen long `gidcid'=`gid'*100+`cid'
    qui gen `dk'=`y'>=`cid'+1

    clogit `dk' `x', group(`gidcid') cluster(`gid')
end

feologit_buc ivar yvar xvars
```


B Inconsistency of estimators with endogenous cut-offs for $T=3$, $K=3$

Here we analytically examine consistency of the estimators in a particular setup: $T=3$, $K=3$, $x_i = x$ for all i . We show inconsistency of the mean estimator in this case. Thus, the mean estimator is inconsistent, in general. This setup is particularly convenient for two reasons. First, it is simple and tractable. Second, in this setup the mean estimator is equal to the median estimator, thus extending the inconsistency result to the median estimator. Finally, for particular values of x and β , the mean estimator is also equivalent to Ferrer-i-Carbonell and Frijters' (2004) estimator (FF), showing that the FF estimator, too, is inconsistent, in general.

The x 's change within an individual over time, but only the individual fixed effect α_i is allowed to change between individuals. We treat x_i and α_i as fixed. If a particular estimator is consistent for arbitrary fixed x 's and α 's, it is also consistent for varying x 's and α 's.

B.1 Probability limit of the score

First we derive the probability limit of the score of the estimators to be examined. These are the CML logit estimators dichotomized at 2 and at 3, the mean, median and FF estimators. Since all estimators have the same score structure and differ only by the dichotomization rule, we index estimators by $c \in \{k = 2, k = 3, \text{Mn}, \text{Md}, \text{FF}\}$, respectively. Then, the probability limit of estimator's c score is

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N s_i^c(b) &= x' \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[d_i^c - \sum_{\mathcal{I}(j)=\mathcal{I}(d_i^c)} j \frac{\exp(j'xb)}{\sum_{\mathcal{I}(l)=\mathcal{I}(d_i)} \exp(l'xb)} \right] \\ &= x' \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[\sum_{a=1}^2 \mathbb{1}(\mathcal{I}(d_i^c) = a) \left(d_i^c - \sum_{\mathcal{I}(j)=a} j \frac{\exp(j'xb)}{\sum_{\mathcal{I}(l)=a} \exp(l'xb)} \right) \right] \\ &= x' \sum_{a=1}^2 \Pr(\mathcal{I}(d^c) = a) \left[\mathbb{E}(d^c | \mathcal{I}(d^c) = a) - \sum_{\mathcal{I}(j)=a} j \frac{\exp(j'xb)}{\sum_{\mathcal{I}(l)=a} \exp(l'xb)} \right], \end{aligned}$$

where d_i^c is the binary dependent variable obtained by using dichotomizing rule c . $\mathbb{1}(z)$ denotes the indicator function (equal to 1 if z is true, 0 otherwise), and $\mathcal{I}(j) = \sum_t \mathbb{1}(j_t = 1)$. I.e., $\mathcal{I}(j)$ is the function that returns the number of elements in j that are equal to one. We use $\sum_{\mathcal{I}(j)=a} f(j)$ to denote the sum of $f(j)$ over all vectors j satisfying $\mathcal{I}(j) = a$.

Setting the score to zero yields an implicit function for estimator c ($\hat{\beta}^c$). If for all relevant values of a (here: 1,2; $a=0$ and $a=3$ do not contribute to the score) it holds that

$$\mathbb{E}(d^c | \mathcal{I}(d^c) = a) - \sum_{\mathcal{I}(j)=a} j \frac{\exp(j'x\beta)}{\sum_{\mathcal{I}(l)=a} \exp(l'x\beta)} = 0 \quad (\text{B.1})$$

it follows that estimator c is consistent. If this is the case, the score is zero if and only if $b = \beta$ because the score is monotonic in b . I.e., if we show that the conditional expectation of the dependent variable dichotomized using rule c is

$$\mathbb{E}(d^c | \mathcal{I}(d^c) = a) = \sum_{\mathcal{I}(j)=a} j \frac{\exp(j'x\beta)}{\sum_{\mathcal{I}(l)=c} \exp(l'x\beta)} \quad \forall a \in 1, 2 \quad (\text{B.2})$$

then estimator c is consistent.

To derive $\mathbb{E}(d^c | \mathcal{I}(d^c))$ for the estimators in question, it is helpful to be aware of some simple ordered logit formulas

$$\frac{\Pr(y_{it} \geq k)}{\Pr(y_{it} < k)} = \frac{1 - \frac{\exp(\tau_k - x'_t\beta - \alpha_1)}{1 + \exp(\tau_k - x'_t\beta - \alpha_1)}}{\frac{\exp(\tau_k - x'_t\beta - \alpha_1)}{1 + \exp(\tau_k - x'_t\beta - \alpha_1)}} = \frac{\exp(x'_t\beta + \alpha_1)}{\exp(\tau_k)} = \frac{\exp(x'_t\beta)}{\exp(\tau_k) / \exp(\alpha_1)} \quad (\text{B.3})$$

$$\frac{\Pr(y_{it} = 3)}{\Pr(y_{it} = 2)} = \frac{1 - \frac{\exp(\tau_3 - x'_t\beta - \alpha_i)}{1 + \exp(\tau_3 - x'_t\beta - \alpha_i)}}{\frac{\exp(\tau_3 - x'_t\beta - \alpha_i)}{1 + \exp(\tau_3 - x'_t\beta - \alpha_i)} - \frac{\exp(\tau_2 - x'_t\beta - \alpha_i)}{1 + \exp(\tau_2 - x'_t\beta - \alpha_i)}} = \frac{\exp(x'_t\beta) + \exp(\tau_2) / \exp(\alpha_i)}{\exp(\tau_3) / \exp(\alpha_i) - \exp(\tau_2) / \exp(\alpha_i)} \quad (\text{B.4})$$

For notational ease we use, for example, $\Pr(1, > 1, \geq 2)$ to denote $\Pr(y_1 = 1, y_2 > 1, y_3 \geq 2)$.

Note that the y_t 's within an individual are independent if we either condition on α_i and x_i , or treat α_i and x_i as fixed: $\Pr(y_1 = 1, y_2 > 1, y_3 \geq 2) = \Pr(y_1 = 1) \cdot \Pr(y_2 > 1) \cdot \Pr(y_3 \geq 2)$.

B.2 Consistency of estimators with exogenous cutoff

We begin by showing that estimators dichotomizing at a fixed cutoff point ($k=2,3$) are consistent in this setup. The procedure is as follow: We derive $\mathbb{E}(d^k | \mathcal{I}(d^k) = a)$, for $a = 1, 2$. If both expressions are equal to the right hand side of (B.2) for each a , we have shown that the estimator is consistent.

B.2.1 $a = 1$

$$\begin{aligned}
& \mathbb{E}(d^k | \mathcal{I}(d^k) = 1) \\
&= \frac{\binom{1}{0} \Pr(\geq k, < k, < k) + \binom{0}{1} \Pr(< k, \geq k, < k) + \binom{0}{1} \Pr(< k, < k, \geq k)}{\Pr(\geq k, < k, < k) + \Pr(< k, \geq k, < k) + \Pr(< k, < k, \geq k)} \\
&= \frac{\binom{1}{0} \frac{\Pr(y_1 \geq k)}{\Pr(y_1 < k)} + \binom{0}{1} \frac{\Pr(y_2 \geq k)}{\Pr(y_2 < k)} + \binom{0}{1} \frac{\Pr(y_3 \geq k)}{\Pr(y_3 < k)}}{\sum_{t=1}^3 \frac{\Pr(y_t \geq k)}{\Pr(y_t < k)}} \\
&= \binom{1}{0} \frac{\exp(x'_1 \beta)}{\sum_{t=1}^3 \exp(x'_t \beta)} + \binom{0}{1} \frac{\exp(x'_2 \beta)}{\sum_{t=1}^3 \exp(x'_t \beta)} + \binom{0}{1} \frac{\exp(x'_3 \beta)}{\sum_{t=1}^3 \exp(x'_t \beta)} \tag{B.5}
\end{aligned}$$

where $k \in \{2, 3\}$ denotes the fixed cutoff. The last expression is equal to the right hand side of (B.2) for $a = 1$.

B.2.2 $a = 2$

$$\begin{aligned}
& \mathbb{E}(d^k | \mathcal{I}(d^k) = 2) \\
&= \frac{\binom{0}{1} \Pr(< k, \geq k, \geq k) + \binom{1}{0} \Pr(\geq k, < k, \geq k) + \binom{1}{0} \Pr(\geq k, \geq k, < k)}{\Pr(< k, \geq k, \geq k) + \Pr(\geq k, < k, \geq k) + \Pr(\geq k, \geq k, < k)} \\
&= \frac{\binom{0}{1} \frac{\Pr(y_1 < k)}{\Pr(y_1 \geq k)} + \binom{1}{0} \frac{\Pr(y_2 < k)}{\Pr(y_2 \geq k)} + \binom{1}{0} \frac{\Pr(y_3 < k)}{\Pr(y_3 \geq k)}}{\sum_{t=1}^3 \frac{\Pr(y_t < k)}{\Pr(y_t \geq k)}} \\
&= \frac{\binom{0}{1} \exp(x'_1 \beta)^{-1} + \binom{1}{0} \exp(x'_2 \beta)^{-1} + \binom{1}{0} \exp(x'_3 \beta)^{-1}}{\sum_{t=1}^3 \exp(x'_t \beta)^{-1}} \\
&= \binom{0}{1} \frac{\exp(x'_2 \beta + x'_3 \beta)}{\sum_t \exp(\sum_{m \neq t} x'_m \beta)} + \binom{1}{0} \frac{\exp(x'_1 \beta + x'_3 \beta)}{\sum_t \exp(\sum_{m \neq t} x'_m \beta)} + \binom{1}{0} \frac{\exp(x'_1 \beta + x'_2 \beta)}{\sum_t \exp(\sum_{m \neq t} x'_m \beta)} \tag{B.6}
\end{aligned}$$

The last expression is equal to the right hand side of (B.2) for $a = 2$. Because a can be only either 1 or 2, we have shown that the conditional logit estimator with a fixed cutoff is consistent in this setup.

B.3 Inconsistency of estimators with endogenous cutoff

Now we show that estimators with endogenous cutoff are inconsistent, in general. It is sufficient to show this for the mean estimator, because with $K=3$ and $T=3$, mean and median estimators produce the same dichotomized binary variable. Furthermore, for some

values of x and β , the mean estimator will produce the same dichotomized binary variable than the FF estimator. We give examples of such cases at the end of this section.

To study the mean estimator, we further partition the score into mutually exclusive sets.

$$\begin{aligned} E(d^{\text{Mn}} | \mathcal{I}(d^{\text{Mn}}) = a) &= \Pr(\mathcal{I}(y) = v | \mathcal{I}(d^{\text{Mn}}) = a) \cdot E(d^{\text{Mn}} | \mathcal{I}(d^{\text{Mn}}) = a, \mathcal{I}(y) = v) \\ &+ \Pr(\mathcal{I}(y) \neq v | \mathcal{I}(d^{\text{Mn}}) = a) \cdot E(d^{\text{Mn}} | \mathcal{I}(d^{\text{Mn}}) = a, \mathcal{I}(y) \neq v) \end{aligned} \quad (\text{B.7})$$

The first set consists of cases with v 1's in the y -vector. The second set consists of the remaining cases.

The procedure is the following: First we consider $E(d^{\text{Mn}} | \mathcal{I}(d^{\text{Mn}}) = 1)$. We will partition the expectation in those cases with $\mathcal{I}(y) = 2$ —for instance, $y=(1,2,1)'$ or $y=(3,1,1)'$ — and those with $\mathcal{I}(y) \neq 2$. We show that the expectation of the first set has the desired form (B.2), while the second set does not. Therefore, the score contribution evaluated at the true β is not zero for $a = 1$ if we dichotomize at the individual mean. Then we repeat the analysis for $a = 2$ and $\mathcal{I}(y) = 1$, finding the same pattern. Finally, we will show that, in general, the two score contributions which are different from (B.2) do not add to zero; this implies that the mean estimator is not consistent.

B.3.1 $a = 1$

Consider the case when the vector d^{Mn} has one 1 and two 0's ($a = 1$) and the associated y -vector has two 1's ($\mathcal{I}(y) = 2$).

$$\begin{aligned} E(d^{\text{Mn}} | \mathcal{I}(d^{\text{Mn}}) = 1, \mathcal{I}(y) = 2) &= \frac{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \Pr(\geq 2, 1, 1) + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \Pr(1, \geq 2, 1) + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \Pr(1, 1, \geq 2)}{\Pr(\geq 2, 1, 1) + \Pr(1, \geq 2, 1) + \Pr(1, 1, \geq 2)} \\ &= \frac{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \Pr(\geq 2, < 2, < 2) + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \Pr(< 2, \geq 2, < 2) + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \Pr(< 2, < 2, \geq 2)}{\Pr(\geq 2, < 2, < 2) + \Pr(< 2, \geq 2, < 2) + \Pr(< 2, < 2, \geq 2)} \\ &= \frac{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \frac{\Pr(y_1 \geq 2)}{\Pr(y_1 < 2)} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \frac{\Pr(y_2 \geq 2)}{\Pr(y_2 < 2)} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \frac{\Pr(y_3 \geq 2)}{\Pr(y_3 < 2)}}{\sum_{t=1}^3 \frac{\Pr(y_t \geq 2)}{\Pr(y_t < 2)}} \end{aligned} \quad (\text{B.8})$$

The last expression is equal to right hand side of (B.2). Now we look at the remaining part of $E(d^{\text{Mn}} | \mathcal{I}(d^{\text{Mn}}) = 1)$. The only y -vectors satisfying $\mathcal{I}(y) \neq 2$ and $\mathcal{I}(d^{\text{Mn}}) = 1$ are cases

with two 2's and one 3.

$$\begin{aligned}
& \mathbb{E}(d^{\text{Mn}} | \mathcal{I}(d^{\text{Mn}}) = 1, \mathcal{I}(y) \neq 2) \\
&= \frac{\binom{1}{0} \Pr(3, 2, 2) + \binom{0}{1} \Pr(2, 3, 2) + \binom{0}{1} \Pr(2, 2, 3)}{\Pr(3, 2, 2) + \Pr(2, 3, 2) + \Pr(2, 2, 3)} \\
&= \frac{\binom{1}{0} \frac{\Pr(y_1=3)}{\Pr(y_1=2)} + \binom{0}{1} \frac{\Pr(y_2=3)}{\Pr(y_2=2)} + \binom{0}{1} \frac{\Pr(y_3=3)}{\Pr(y_3=2)}}{\sum_{t=1}^3 \frac{\Pr(y_t=3)}{\Pr(y_t=2)}} \\
&= \frac{\binom{1}{0} (\exp(x'_1\beta) + \varkappa_2) + \binom{0}{1} (\exp(x'_2\beta) + \varkappa_1) + \binom{0}{1} (\exp(x'_3\beta) + \varkappa_1)}{\sum_{t=1}^3 (\exp(x'_t\beta) + \varkappa_1)}, \tag{B.9}
\end{aligned}$$

where $\varkappa_2 \equiv \exp(\tau_2)\mathbb{E}(\exp(-\alpha_i))$. This expression is only equal to the right hand side of (B.2) if $\exp(\tau_2) = 0$. This is only possible if τ_2 goes to minus infinity which means that the probability if $y_{it} = 1$ is zero (i.e., this is the limiting case with two categories: $K=2$). Thus, the score contribution for $a = 1$ evaluated at $b = \beta$ is not equal to zero if we dichotomize at the individual mean.

B.3.2 $a = 2$

Now we consider cases where the number of 1's in the d^{Mn} -vector is 2 ($a = 2$). We divide these cases into those satisfying $\mathcal{I}(y) = 1$ and the rest. If we dichotomize at the individual mean, the only y -vectors for which $\mathcal{I}(y) \neq 1$ and $\mathcal{I}(d) = 2$ are those with one 2 and two 3's.

$$\begin{aligned}
& \mathbb{E}(d^{\text{Mn}} | \mathcal{I}(d^{\text{Mn}}) = 2, \mathcal{I}(y) = 1) \\
&= \frac{\binom{0}{1} \Pr(1, \geq 2, \geq 2) + \binom{1}{0} \Pr(\geq 2, 1, \geq 2) + \binom{1}{0} \Pr(\geq 2, \geq 2, 1)}{\Pr(1, \geq 2, \geq 2) + \Pr(\geq 2, 1, \geq 2) + \Pr(\geq 2, \geq 2, 1)} \\
&= \frac{\binom{0}{1} \frac{\Pr(y_1 < 2)}{\Pr(y_1 \geq 2)} + \binom{1}{0} \frac{\Pr(y_2 < 2)}{\Pr(y_2 \geq 2)} + \binom{1}{0} \frac{\Pr(y_3 < 2)}{\Pr(y_3 \geq 2)}}{\frac{\Pr(y_1 < 2)}{\Pr(y_1 \geq 2)} + \frac{\Pr(y_2 < 2)}{\Pr(y_2 \geq 2)} + \frac{\Pr(y_3 < 2)}{\Pr(y_3 \geq 2)}} \tag{B.10}
\end{aligned}$$

This is equivalent to the right hand side of (B.2).

$$\begin{aligned}
& \mathbb{E}(d^{\text{Mn}} | \mathcal{I}(d^{\text{Mn}}) = 2, \mathcal{I}(y) \neq 1) \\
&= \frac{\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \Pr(2, 3, 3) + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \Pr(3, 2, 3) + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \Pr(3, 3, 2)}{\Pr(2, 3, 3) + \Pr(3, 2, 3) + \Pr(3, 3, 2)} \\
&= \frac{\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \frac{\Pr(y_1=2)}{\Pr(y_1=3)} + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \frac{\Pr(y_2=2)}{\Pr(y_2=3)} + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \frac{\Pr(y_3=2)}{\Pr(y_3=3)}}{\sum_{t=1}^3 \frac{\Pr(y_t=2)}{\Pr(y_t=3)}} \\
&= \frac{\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} (\exp(x'_1\beta) + \varkappa_2)^{-1} + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} (\exp(x'_2\beta) + \varkappa_2)^{-1} + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} (\exp(x'_3\beta) + \varkappa_2)^{-1}}{\sum_{t=1}^3 (\exp(x'_t\beta) + \varkappa_2)^{-1}} \tag{B.11}
\end{aligned}$$

This expression is only equivalent to the right hand side of (B.2) if \varkappa_2 vanishes. Thus the score contribution for $a = 2$ evaluated at $\hat{\beta} = \beta$ is not equal to zero if we dichotomize at the individual mean.

If, for instance, $\varkappa_2=1$, $\beta = 1$, and x_t are scalar with $x_t = \ln(t)$, it is easy to verify that both score contributions (B.1) for $a = 1$ and $a = 2$ are negative. Thus, in general, the two non-zero score contributions do not cancel out, because the probability weights are necessarily positive. This implies that the mean estimator is inconsistent. Moreover, it is easy to verify that mean and FF estimator coincide in this DGP. This implies that the FF estimator is inconsistent, in general, too.

Figure 1: Marginal distribution of y in Monte Carlo experiments

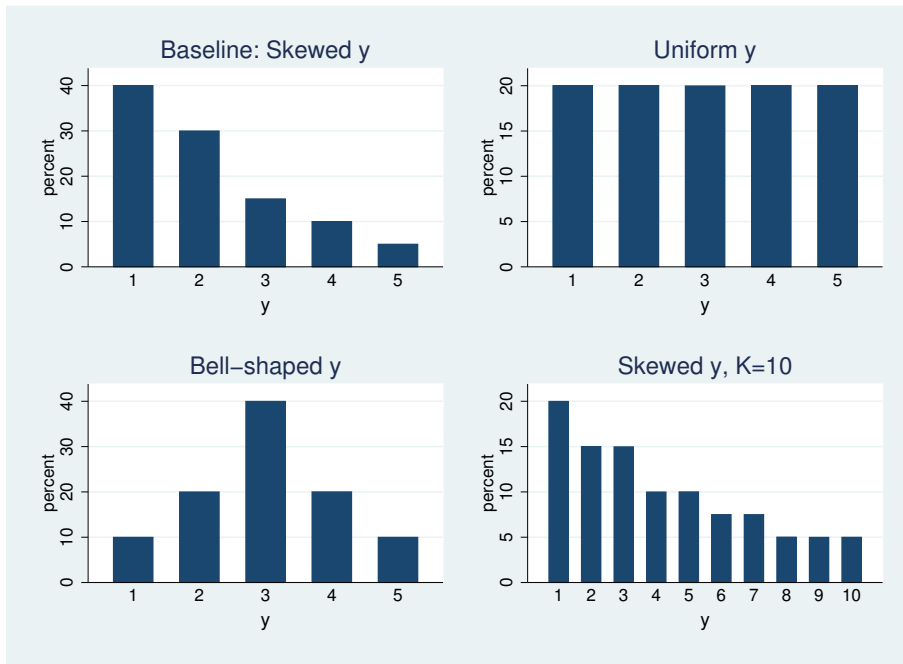


Table 1: Monte Carlo simulation results (1,000 replications): Baseline scenario

Estimators	$\hat{\beta}_x$			$\hat{\beta}_d$		
	M	SD	SE	M	SD	SE
DvS	1.00	0.07	0.07	0.99	0.11	0.11
BUC	1.00	0.07	0.07	1.00	0.12	0.12
FF	0.93	0.07	0.07	0.92	0.12	0.12
Median	0.94	0.07	0.07	0.94	0.12	0.12
Mean	0.96	0.07	0.07	0.95	0.12	0.12
$y \geq 2$	1.00	0.09	0.09	1.00	0.15	0.15
$y \geq 3$	1.01	0.09	0.09	1.00	0.15	0.16
$y \geq 4$	1.01	0.12	0.11	1.00	0.20	0.20
$y \geq 5$	1.03	0.18	0.18	1.02	0.32	0.32

Notes: $\beta_x = \beta_d = 1$. Columns labeled M contain the mean of the estimated coefficients over all replications, columns SD the standard deviation of the estimated coefficients, and columns SE contain the mean of the estimated standard errors. Baseline scenario is $N=500$, $T=4$, $K=5$, $x \sim Normal(0, 1)$, $d \sim Bernoulli(0.5)$, skewed distribution for y .

Table 2: Monte Carlo simulation results (1,000 replications): The effects of increasing N, T and K

		<i>Baseline scenario:</i> $N=500, T=4, K=5$						$N=1,000$						$T=8$						$K=10$						$K=10, N=100$					
		$\hat{\beta}_x$			$\hat{\beta}_d$			$\hat{\beta}_x$			$\hat{\beta}_d$			$\hat{\beta}_x$			$\hat{\beta}_d$			$\hat{\beta}_x$			$\hat{\beta}_d$			$\hat{\beta}_x$			$\hat{\beta}_d$		
Estimators	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD			
DvS	1.00	0.07	0.99	0.11	1.00	0.05	1.00	0.08	1.00	0.04	1.00	0.08	1.00	0.08	0.99	0.06	0.98	0.10	0.93	0.14	0.94	0.24	0.93	0.14	0.94	0.24					
BUC	1.00	0.07	1.00	0.12	1.00	0.05	1.00	0.08	1.00	0.04	1.00	0.08	1.00	0.08	1.00	0.06	1.00	0.11	1.00	0.14	1.00	0.24	1.00	0.14	1.00	0.24					
FF	0.93	0.07	0.92	0.12	0.93	0.05	0.93	0.09	0.96	0.04	0.96	0.08	0.86	0.11	0.86	0.06	0.86	0.11	0.86	0.13	0.86	0.26	0.86	0.13	0.86	0.26					
Median	0.94	0.07	0.94	0.12	0.94	0.05	0.94	0.09	0.97	0.04	0.97	0.08	0.88	0.11	0.88	0.06	0.87	0.11	0.88	0.14	0.88	0.25	0.88	0.14	0.88	0.25					
Mean	0.96	0.07	0.95	0.12	0.96	0.05	0.96	0.09	0.98	0.05	0.97	0.08	0.94	0.07	0.94	0.07	0.93	0.11	0.95	0.15	0.94	0.26	0.95	0.15	0.94	0.26					
$y \geq 2$	1.00	0.09	1.00	0.15	1.00	0.06	1.00	0.11	1.00	0.05	1.00	0.10	1.01	0.10	1.01	0.10	1.00	0.18	1.04	0.23	1.02	0.41	1.04	0.23	1.02	0.41					
$y \geq 3$	1.01	0.09	1.00	0.15	1.00	0.06	1.00	0.11	1.00	0.06	1.00	0.10	1.01	0.10	1.01	0.09	1.00	0.15	1.02	0.20	1.01	0.34	1.02	0.20	1.01	0.34					
$y \geq 4$	1.01	0.12	1.00	0.20	1.00	0.08	1.00	0.14	1.00	0.07	1.00	0.13	1.00	0.08	1.00	0.08	1.00	0.14	1.02	0.19	1.01	0.33	1.02	0.19	1.01	0.33					
$y \geq 5$	1.03	0.18	1.02	0.32	1.01	0.13	1.01	0.22	1.01	0.11	1.01	0.20	1.01	0.09	1.01	0.09	1.00	0.15	1.01	0.20	1.00	0.34	1.01	0.20	1.00	0.34					

Notes: $\beta_x = \beta_d = 1$. Columns labeled M contain the mean of the estimated coefficients over all replications, columns SD the standard deviation of the estimated coefficients. Baseline scenario is $N=500, T=4, K=5, x \sim Normal(0, 1), d \sim Bernoulli(0.5)$, skewed distribution for y . Departures from baseline scenario are noted in the top row.

Table 3: Monte Carlo simulation results (1,000 replications): Changing the distributions of y , x and d .

Estimators	<i>Baseline: skewed y, $x \sim N(0, 1), d \sim B(0.5)$</i>						<i>bell-shaped y</i>						<i>uniform y</i>						<i>$x \sim LN(0, 1), d \sim B(0.1)$</i>					
	$\hat{\beta}_x$		$\hat{\beta}_d$		$\hat{\beta}_x$		$\hat{\beta}_d$		$\hat{\beta}_x$		$\hat{\beta}_d$		$\hat{\beta}_x$		$\hat{\beta}_d$		$\hat{\beta}_x$		$\hat{\beta}_d$		$\hat{\beta}_x$		$\hat{\beta}_d$	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
DvS	1.00	0.07	0.99	0.11	1.00	0.06	0.99	0.11	0.99	0.06	0.99	0.11	0.99	0.08	0.99	0.11	0.99	0.08	0.99	0.08	0.99	0.08	0.99	0.17
BUC	1.00	0.07	1.00	0.12	1.00	0.06	1.00	0.11	1.00	0.06	1.00	0.11	1.00	0.08	1.00	0.11	1.00	0.08	1.00	0.08	1.00	0.08	1.00	0.17
FF	0.93	0.07	0.92	0.12	0.93	0.07	0.93	0.12	0.91	0.07	0.91	0.12	0.91	0.08	0.91	0.12	0.91	0.08	0.91	0.08	0.91	0.08	0.91	0.18
Median	0.94	0.07	0.94	0.12	0.94	0.07	0.93	0.12	0.92	0.07	0.92	0.12	0.91	0.09	0.94	0.11	0.94	0.09	0.94	0.09	0.94	0.09	0.94	0.18
Mean	0.96	0.07	0.95	0.12	0.95	0.07	0.95	0.12	0.94	0.07	0.94	0.12	0.94	0.09	0.96	0.11	0.96	0.09	0.95	0.09	0.95	0.09	0.95	0.18
$y \geq 2$	1.00	0.09	1.00	0.15	1.01	0.14	1.02	0.23	1.01	0.10	1.01	0.23	1.01	0.13	1.01	0.18	1.01	0.13	1.00	0.13	1.00	0.13	1.00	0.23
$y \geq 3$	1.01	0.09	1.00	0.15	1.01	0.09	1.00	0.16	1.00	0.09	1.00	0.16	1.00	0.11	1.01	0.15	1.01	0.11	1.01	0.11	1.01	0.11	1.01	0.22
$y \geq 4$	1.01	0.12	1.00	0.20	1.01	0.09	1.00	0.15	1.01	0.09	1.00	0.15	1.01	0.12	1.01	0.15	1.01	0.12	1.01	0.12	1.01	0.12	1.01	0.30
$y \geq 5$	1.03	0.18	1.02	0.32	1.01	0.13	1.01	0.23	1.01	0.10	1.01	0.23	1.01	0.18	1.03	0.18	1.03	0.19	1.01	0.19	1.01	0.19	1.01	0.57

Notes: $\beta_x = \beta_d = 1$. Columns labeled M contain the mean of the estimated coefficients over all replications, columns SD the standard deviation of the estimated coefficients. Baseline scenario is $N=500, T=4, K=5, x \sim Normal(0, 1), d \sim Bernoulli(0.5)$, skewed distribution for y (first four columns). Departures from baseline scenario are noted in the top row. $LN()$ stands for log-normal distribution.

Table 4: Fixed Effects Ordered Logit Estimates of Life Satisfaction

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.: <i>Life Satisfaction</i>	$y \geq 8$	DvS	BUC	FF	Mean	Median
<i>Unemployed</i>	-0.96** (0.20)	-0.98** (0.14)	-1.03** (0.16)	-0.77** (0.15)	-0.84** (0.15)	-0.66** (0.15)
<i>Out of labor force</i>	-0.24 (0.12)	-0.42** (0.09)	-0.45** (0.11)	-0.25** (0.09)	-0.25** (0.10)	-0.25** (0.09)
<i>Duration of unemployment</i>	-0.01 (0.02)	-0.01 (0.01)	-0.02 (0.01)	-0.02 (0.01)	-0.01 (0.01)	-0.01 (0.01)
<i>Squared duration of unemp.</i> $\times 10,000^{-1}$	0.60 (2.79)	2.44 (1.56)	2.75 (2.30)	3.18 (1.87)	2.17 (1.88)	2.12 (1.86)
<i>Married</i>	0.67** (0.12)	0.52** (0.09)	0.56** (0.11)	0.37** (0.09)	0.39** (0.09)	0.37** (0.09)
<i>Good health</i>	0.34** (0.06)	0.33** (0.05)	0.36** (0.05)	0.24** (0.05)	0.29** (0.05)	0.24** (0.05)
<i>Age</i>	-0.12** (0.04)	-0.12** (0.03)	-0.12** (0.03)	-0.12** (0.03)	-0.11** (0.03)	-0.12** (0.03)
<i>Squared age</i> $\times 100^{-1}$	-0.84 (4.27)	-2.46 (3.24)	-1.15 (3.82)	-1.30 (3.36)	-2.91 (3.38)	-1.58 (3.35)
<i>Log. household income</i>	0.13* (0.06)	0.12** (0.04)	0.13* (0.05)	0.10* (0.04)	0.10* (0.05)	0.10* (0.04)
$\log L$	-4,996	—	-21,802	-8,003	-7,911	-8,054
Observations	12,980	—	59,535	19,053	19,071	19,071
Individuals	2,573	3,958	11,864	3,949	3,958	3,958
Clusters	—	—	3,958	—	—	—

Notes: Data Source GSOEP, waves 1984-1989. * statistical significance at 5% level, ** statistical significance at 1% level. Observations denotes the number of person-year observations in estimation sample; Individuals denotes number of unique persons in estimation sample; Clusters denotes the number of groups in cluster-robust standard errors.