

Baetschmann, Gregori; Staub, Kevin E.; Winkelmann, Rainer

**Working Paper**

## Reconsidering the analysis of longitudinal happiness data: With an application to the effect of unemployment

Working Paper, No. 4 [rev.]

**Provided in Cooperation with:**

Department of Economics, University of Zurich

*Suggested Citation:* Baetschmann, Gregori; Staub, Kevin E.; Winkelmann, Rainer (2011) :  
Reconsidering the analysis of longitudinal happiness data: With an application to the effect of  
unemployment, Working Paper, No. 4 [rev.], University of Zurich, Department of Economics, Zurich,  
<https://doi.org/10.5167/uzh-59259>

This Version is available at:

<https://hdl.handle.net/10419/77542>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen  
Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle  
Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich  
machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen  
(insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten,  
gelten abweichend von diesen Nutzungsbedingungen die in der dort  
genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal  
and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to  
exhibit the documents publicly, to make them publicly available on the  
internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content  
Licence (especially Creative Commons Licences), you may exercise  
further usage rights as specified in the indicated licence.*



**University of  
Zurich** <sup>UZH</sup>

University of Zurich  
Department of Economics

Working Paper Series

ISSN 1664-7041 (print)  
ISSN 1664-705X (online)

---

Working Paper No. 4

**Reconsidering the analysis of longitudinal  
happiness data - with an application to the  
effect of unemployment**

Gregori Baetschmann, Kevin E. Staub and Rainer Winkelmann

Revised version, June 2011

---

# Reconsidering the analysis of longitudinal happiness data - with an application to the effect of unemployment\*

Gregori Baetschmann<sup>a</sup>, Kevin E. Staub<sup>a</sup> and Rainer Winkelmann<sup>a,b</sup>

<sup>a</sup>*University of Zurich, Department of Economics*

<sup>b</sup>*CESifo, Munich, and IZA, Bonn*

June 2011

**Abstract:** The paper reconsiders existing estimators for the panel data fixed effects ordered logit model, including one that has not been used in econometric studies before, and studies the small sample 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. Second, the new estimator seems to be 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 happiness. Choosing the right estimator avoids a bias of up to 30 percent in key parameters.

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

**JEL Codes:** C23, C25, J28, J64.

---

\* We thank Paul Frijters, Arie Kapteyn and participants of the 2011 Engelberg Workshop in Labor Economics and the 2011 North American Summer Meetings of the Econometric Society for very valuable comments on an earlier draft.

Address for correspondence: University of Zurich, Department of Economics, Zürichbergstr. 14, CH-8032 Zürich, Switzerland, ☎ +41 44 634 2312 (K.E. Staub), ✉ gregori.baetschmann@econ.uzh.ch, kevin.staub@econ.uzh.ch, rainer.winkelmann@econ.uzh.ch.

# 1 Introduction

The estimation of longitudinal ordered logit models with panel-specific unobserved heterogeneity poses a challenge for applied research in happiness economics and elsewhere. Ignoring such panel effects causes the usual estimator to be inconsistent, regardless of whether the effects are correlated with the regressors or not. Naturally, then, estimators that condition on the panel effects offer a promising solution, as they do not require making further assumptions regarding the distribution of the effects.

A popular method for estimating the fixed effects ordered logit model is due to Ferrer-i-Carbonell and Frijters (2004). The Ferrer-i-Carbonell and Frijters (FF) estimator, or an approximation thereof, is frequently used in empirical studies of health and well-being. Examples include Frijters, Haiken-DeNew and Shields (2004 a, 2004b, 2005), Kassenboehmer and Haiken-DeNew (2009), Booth and van Ours (2008), D’Addio, Eriksson and Frijters (2007), Schmitz (2011) and Jones and Schurer (2011).

We show in this paper that the FF estimator is inconsistent. We also discuss an alternative consistent estimator that has been introduced in the statistics literature by Mukherjee et al. (2008) but not been applied in econometric studies to date. For reasons that become apparent when we introduce the estimator in detail, we refer to it as “blow-up and cluster” (BUC) estimator. The BUC estimator is close in spirit to that of Das and van Soest (1999), but it avoids some small sample problems that can arise for the Das and van Soest (DvS) estimator. The BUC estimator is simple to implement, and Stata code is provided in an appendix.

All fixed effects ordered logit estimators rely on conditional logit estimation of dichotomized ordered responses. In the most basic version, a single cut-off is used for all cross-sectional units. However, this leads to a substantial loss of information, and subsequent estimators have attempted to exploit information on as many panel units as possible, increasing the efficiency of the estimators. This can be achieved by using alternative rules for dichotomization. The BUC and DvS estimators are consistent, because they use ex-

ogenous dichotomization rules. The FF estimator is inconsistent, because it dichotomizes endogenously. We prove the inconsistency analytically, and document in a series of Monte Carlo simulations the persistent bias of the FF estimator.

We apply these estimators in a replication study of Winkelmann and Winkelmann (1995, 1998). These papers used data from the German Socio-Economic Panel (GSOEP) for the period 1984-1989 and were the first to establish that the large negative effect of unemployment on life satisfaction observed in earlier analyses of cross-sectional data persists in a fixed effects panel model, countering the argument that the correlation between unemployment and life satisfaction is simply due to time-invariant confounders, such as personality. In the GSOEP, life satisfaction is recorded on a discrete 0-10 scale, and in the original study of 1998, the data were dichotomized at 8. Applying the different estimators discussed in this paper, we find that the use of FF or mean-cutoff estimators leads to an attenuation bias for the key parameters of interest in the order of 20 to 30 percent. The BUC results are close to those obtained from dichotomizing at 8, albeit more precise.

Having established our preferred estimation methods, we move beyond the initial replication and provide genuinely new evidence on the link between unemployment and life satisfaction. First, as the GSOEP is an ongoing survey, the dataset has grown substantially and the latest survey year available for estimation is now 2009. Second, as in Kassenboehmer and Haiken-DeNew (2009), we exploit the fact that since 1991, the unemployed were asked for the reasons for termination of their last job, among others, whether it was due to a plant closure. Plant closures arguably induce an exogenous variation in unemployment, and comparing the unemployment effect on life satisfaction among those affected by a plant closure and all others thus provides an additional check on the potential endogeneity of unemployment. The BUC estimation results for a sample of men in West Germany between 1991 and 2009 largely corroborate the earlier findings in the literature. The adverse effect of unemployment on life satisfaction is large. Interestingly, those affected by plant closures report even higher life satisfaction losses than others.

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

## 2 Econometric methods

### 2.1 The fixed effects 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  $\alpha_i$  and  $\varepsilon_{it}$ :

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

The time-invariant part of the unobservables,  $\alpha_i$ , may or may not depend statistically on  $x_{it}$ . One can either make an assumption regarding the distribution of  $\alpha_i$  (or the joint distribution of  $\alpha_i$  and  $x_{it}$ , respectively), or else treat  $\alpha_i$  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:

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

where thresholds  $\tau_i$  are strictly increasing ( $\tau_{ik} < \tau_{ik+1} \quad \forall k$ ) and  $\tau_{i1} = -\infty$ ,  $\tau_{iK+1} = \infty$ . The specification of the fixed effects ordered logit model is completed by assuming that conditionally on  $x_{it}$  and  $\alpha_i$ ,  $\varepsilon_{it}$  are IID standard logistically. I.e., if  $F(\cdot)$  denotes the cdf of  $\varepsilon_{it}$

$$F(\varepsilon_{it}|x_{it}, \alpha_i) = F(\varepsilon_{it}) = \frac{1}{1 + \exp(-\varepsilon_{it})} \equiv \Lambda(\varepsilon_{it}). \quad (2)$$

Hence, the probability of observing outcome  $k$  for individual  $i$  at time  $t$  is given by

$$\Pr(y_{it} = k | x_{it}, \alpha_i) = \Lambda(\tau_{ik+1} - x'_{it}\beta - \alpha_i) - \Lambda(\tau_{ik} - x'_{it}\beta - \alpha_i) , \quad (3)$$

which depends not only on  $\beta$  and  $x_{it}$ , but also on  $\alpha_i$  and  $\tau_{ik}, \tau_{ik+1}$ .

There are two problems with Maximum Likelihood (ML) estimation based on expression (3). The first is a problem of identification:  $\tau_{ik}$  cannot be distinguished from  $\alpha_i$ ; 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  $\alpha_{ik}$  cannot be estimated consistently, due to the incidental parameter problem (see, for instance, Lancaster, 2000). This does have consequences for estimation of  $\beta$  – the bias in  $\alpha_{ik}$  contaminates  $\hat{\beta}$ . In short panels, the resulting bias in  $\hat{\beta}$  can be substantial (Greene, 2004).

We next consider different approaches to estimate  $\beta$ . 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 The single cutoff estimator

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  $\alpha_i$  as

$$\mathcal{P}_i^k(\beta) \equiv \Pr \left( d_i^k = j_i \middle| \sum_{t=1}^T d_{it}^k = a_i \right) = \frac{\exp(j_i' x_i \beta)}{\sum_{j \in B_i} \exp(j' x_i \beta)} \quad (4)$$

does not depend on  $\alpha_i$  and the thresholds. In (4),  $j_i = (j_{i1}, \dots, j_{iT})$ ,  $x_i$  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 (4), is  $\binom{T}{a_i} = \frac{T!}{a_i!(T-a_i)!}$ . This calculation shows that the estimator can be computationally demanding, or even intractable, if the number of periods is large. For example, with  $T = 16$  and  $a_i = 8$ , the denominator has 12'870 distinct elements.

Chamberlain (1980) shows that maximizing the conditional likelihood

$$\log \mathcal{L}^k(b) = \sum_{i=1}^N \log \mathcal{P}_i^k(b) \quad (5)$$

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

$$\text{plim} \frac{1}{N} \sum_i s_i^k(\beta) = 0, \quad (6)$$

where

$$s_i^k(b) = \frac{\partial \ln \mathcal{P}_i^k(b)}{\partial b} = x_i' \left( d_i^k - \sum_{j \in B_i} j \frac{\exp(j' x_i b)}{\sum_{l \in B_i} \exp(l' x_i b)} \right). \quad (7)$$

The reason why (6) holds is that for every  $i$ , conditional on  $x_i$ , the expectation of the term in parentheses in (7) is zero as it defines a conditional expectation residual.

Note that conditioning on  $a_i$  causes all time-invariant elements in (3) to cancel. I.e., not only  $\alpha_i$  and  $\tau_{ik}$  are not estimated, but also elements of the  $\beta$  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

$$H_i^k(b) = \frac{\partial^2 \ln \mathcal{P}_i^k(b)}{(\partial b)(\partial b)'} = - \sum_{j \in B_i} \frac{\exp(j' x_i b)}{\sum_{l \in B_i} \exp(l' x_i b)} \times \left( x_i' j - \sum_{m \in B_i} \frac{\exp(m' x_i b)}{\sum_{l \in B_i} \exp(l' x_i b)} m' x_i \right) \left( x_i' j - \sum_{m \in B_i} \frac{\exp(m' x_i b)}{\sum_{l \in B_i} \exp(l' x_i b)} m' x_i \right)' . \quad (8)$$

### 2.3 Multiple cutoffs

The estimator of  $\beta$  based on (5), 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  $\beta$ , 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  $K-1$  dichotimizations 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):

$$\hat{\beta}^{DvS} = \arg \min_b (\hat{\beta}^{2'} - b', \dots, \hat{\beta}^{K'} - b') \Omega^{-1} (\hat{\beta}^{2'} - b', \dots, \hat{\beta}^{K'} - b')' . \quad (9)$$

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

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

evaluated at  $b = \beta$ . In practice, the unknown variance  $\Omega$  is replaced by an estimate  $\hat{\Omega}$  which is evaluated at  $\hat{\beta}^k$ ,  $k = 2, \dots, K$ . The solution to (9) 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

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

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  $\beta^k = \beta \forall k = 2, \dots, K$ . Mukherjee et al. (2008) refer to this estimator as *amalgamated conditional likelihood*. The sample (quasi-) log-likelihood of this restricted CML estimator is

$$\log \mathcal{L}(b) = \sum_{k=2}^K \log \mathcal{L}^k(b) . \quad (10)$$

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:

$$\text{plim} \sum_{k=2}^K \frac{1}{N} \sum_{i=1}^N s_i^k(\beta) = \text{plim} \frac{1}{N} \sum_i s_i^2(\beta) + \dots + \text{plim} \frac{1}{N} \sum_i s_i^K(\beta) = 0, \quad (11)$$

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 cluster-robust variance estimator which allows for arbitrary correlation within the various contributions of any individual should be used:

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

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  $\beta$  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 cutoffs

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

$$\log \mathcal{L}^{FF}(b) = \sum_{i=1}^N \sum_{k=2}^K w_i^k \log \mathcal{P}_i^k(b), \quad w_i^k \in \{0, 1\}, \quad \sum_{k=2}^K w_i^k = 1. \quad (12)$$

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 (12) with minimal variance. Other, simpler rules for choosing  $w_i^k$  for (12) 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. 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 (7), 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, and the implicit within-individual correlation between  $y_{it}$  and the time-varying cutoff is negative. 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 using Monte Carlo simulations in order to assess the small sample biases and efficiency across a number of 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,  $\varepsilon_{it}$  is sampled from a logistic distribution as in (2).

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 show 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 transmit two important findings. First, the DvS and BUC estimators are indeed more efficient than the single cutoff estimators but do not differ much in terms of standard errors among themselves. 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 vary  $N$ ,  $T$  and  $K$  in order 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 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 bias: 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. 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 experiment increases the number of categories. In the limit, the observed variable becomes equal to the continuous latent variable. We increase the number of categories from  $K=5$  to  $K=10$ , using the marginal distribution 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  $\tau$  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 investigate the empirical relationship between individual unemployment and life satisfaction. We first replicate results of an earlier study by Winkelmann and Winkelmann (1998). We then extend the scope of the analysis substantially by considering a new sample with 17 years of data and, as in Kassenboehmer and Haisken-DeNew (2009), by estimating a separate effect for those whose unemployment spell was initiated by a plant closure. This is the closest we can get, with this kind of data, to the notion of a truly exogenous variation in unemployment.

### 4.1 Data and specification

The sample used in Winkelmann and Winkelmann (1998) was drawn from the first six (1984-1989) waves of the German Socio-economic Panel. It includes observations on men aged 20-64 years who are observed for at least two waves and have non-missing responses for all variables of the model. There are 20,944 person-year observations corresponding to 4,261 individuals. Of these, 1,873 observations for 303 persons are discarded because they do not display any variation over time in their outcome variable, leaving a final dataset with 19,079 person-year observations on 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.

The original paper was also concerned with habituation effects, the hypothesis being that unemployment might “hurt less” the longer the duration of the unemployment spell. To allow for a possibly non-linear habituation pattern, the specification includes the variables *Duration of unemployment* and *Squared duration of unemployment*. Marital status (*Married*), health status (*Good health*), age (*Age* and *Squared age*) and household income (in logarithms, *Log. household income*) are added as control variables. 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 for a different estimator. The first column reproduces 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 that there is no evidence that people adapt and get used to unemployment. 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 show the 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 Table 4 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).

### 4.3 Testing for parameter heterogeneity

Jones and Schurer (2011) argued that the baseline fixed effect ordered logit model is often misspecified, because it posits a constant  $\beta^k = \beta$  for all  $k$  whereas one would expect parameter heterogeneity ( $\beta^k \neq \beta$  for some  $k$ ) in many applications. Parameter heterogeneity is an important issue also for the estimators discussed in this paper, as it would imply that the different fixed cutoff estimators are no longer consistent for the same underlying (and constant)  $\beta$ . Intuitively, this hypothesis can be easily tested within our BUC framework where the regressors are interacted with a full set of dummies representing the different cutoff points. The relevant restrictions can then be tested by means of a Wald test. This can be done separately for each regressor or jointly for the full parameter vector.

A very similar approach is suggested by Jones and Schurer (2011). However, because

they propose estimating the stacked model without accounting for the interdependence between the observations for the same individual at the different cutoffs, they underestimate the covariances between the estimators, which means that the effective size of their test is distorted.

In our application, we find that the null-hypothesis of constant effects is not rejected for most regressors. If a separate test is performed for each variable (e.g.  $\beta_{unemployed}^k = \beta_{unemployed}$  for all  $k \in \{2, 3, \dots, K\}$ ), evidence for heterogeneity is found only for the variable “age”, where the null hypothesis can be rejected at the 10% level. We conclude that the constant effects assumption seems like a reasonable approximation in this particular application.

#### 4.4 Plant closure unemployment and life satisfaction

Our extended analysis relates to two recent papers, one by Kassenboehmer and Haisken-DeNew (2009) on the causal effect of unemployment on life satisfaction and another one by Schmitz (2011) on the causal effect of unemployment on health. The key assumption of both papers was that unemployment due to plant closure provides a truly exogenous source of variation in unemployment, and that the causal effect of unemployment thus can be determined by comparing outcomes in employment with those in unemployment due to plant closure. We follow this reasoning, although one could argue that plant closure *and* still being unemployed at the time of interview is not exogenous, and instrumental variable approaches should be used.

The key shortcoming of these two papers is, however, that they employ the mean-cutoff estimator to eliminate the fixed effects. The resulting downward bias can perhaps explain, why Schmitz (2011) finds no effect of plant closure unemployment on health, whereas we in our application find a large and statistically significant effect. The results are shown in Table 5. We use data on men living in West Germany for 1991 to 2009. Two years, 1999 and 2000, had to be dropped from the analysis, since the plant closure variable was not

recorded in these two years. This leaves 17 years for the analysis, with a total of 82'395 person-year observations.

The BUC estimation results largely corroborate the earlier findings in the literature. The adverse effect of unemployment on life satisfaction is large. We use a specification with an unemployment main effect and a plant closure unemployment interaction. The main effect thus gives the effect for those entering unemployment for reasons unrelated to a plant closure (such as individual dismissal or quit). The point estimate for  $\beta$  is -1.13, slightly larger in absolute value than the BUC estimate of the unemployment effect in the earlier dataset. Plant closure unemployment leads to an additional effect of -0.33, so that the estimated cumulative effect of plant closure unemployment is -1.46. This number can be interpreted as the predicted change in the cumulative log odds of reporting a life satisfaction greater or equal to  $k$ , rather than less than  $k$ , if a person is plant closure unemployed rather than employed. This is by all standards a very substantial effect.

## 5 Conclusions

The main conclusion from the research presented in this paper is prescriptive: there are a number of different estimators that practitioners wanting to estimate the fixed effects ordered logit model can choose from. However, only the BUC and DvS estimators can be recommended. The other approaches are either inconsistent or inefficient. This finding raises a caveat over a number of recent studies listed in the introduction that used an endogenously dichotomized cutoff point, and hence an inconsistent estimator. We showed in the application on the effect of unemployment on life satisfaction that the bias in the key parameters of interest can be substantial.

The class of consistent estimators considered in this paper is not complete. For example, the BUC estimator could be modified by weighing the  $k$  cutoffs differently. Moreover, all estimators in this paper choose dichotomizations which are constant over individuals. As

an alternative, one could let the dichotomizing cutoff vary within individuals. As long as this is done exogenously, the resulting estimators are consistent as well. One therefore can in principle generate a huge number of different estimators for  $\beta$  that are all consistent. A challenging question for future research is whether one can define optimal weights for all possible estimators such that the estimation procedure is asymptotically efficient.

# 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 (2011), “How does heterogeneity shape the socioeconomic gradient in health satisfaction?”, *Journal of Applied Econometrics*, **26**, 549-714.
- 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.
- Lancaster, Tony (2000), “The incidental parameter problem since 1948”, *Journal of Econometrics*, **95**, 391-413.
- Mukherjee, Bhramar, Jaeil Ahn, Ivy Liu, Paul J. Rathouz and Brisa N. Sanchez (2008), “Fitting stratified proportional odds models by amalgamating conditional likelihoods”, *Statistics in Medicine*. **27**, 4950-4971.
- Schmitz, Hendrik (2011), “Why are the unemployed in worse health? The causal effect of unemployment on health”, *Labour Economics*, **18**, 71-78
- Winkelmann, Liliana, and Rainer Winkelmann (1995), “Happiness and Unemployment: a Panel Data Analysis for Germany”, *Konjunkturpolitik*, 41(4), 293-307.
- 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 the consistency of fixed effects ordered logit 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  $\beta$ , 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  $\alpha_i$  is allowed to change between individuals. We treat  $x_i$  and  $\alpha_i$  as fixed. If a particular estimator is consistent for arbitrary fixed  $x$ 's and  $\alpha$ 's, it is also consistent for varying  $x$ 's and  $\alpha$ '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

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

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

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

then estimator  $c$  is consistent.

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

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

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

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  $\alpha_i$  and  $x_i$ , or treat  $\alpha_i$  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  $E(d^k | \mathcal{I}(d^k) = a)$ , 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 $a = 1$

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

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 $a = 2$

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

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  $\beta$ , the mean estimator will produce the same dichotomized binary variable as 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  $\beta$  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 similar results. Finally, we 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  $d^{\text{Mn}}$  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.

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

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  $\tau_2$  goes to minus infinity which means that the probability of  $y_{it} = 1$  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  $d^{\text{Mn}}$ -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.

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

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

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

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

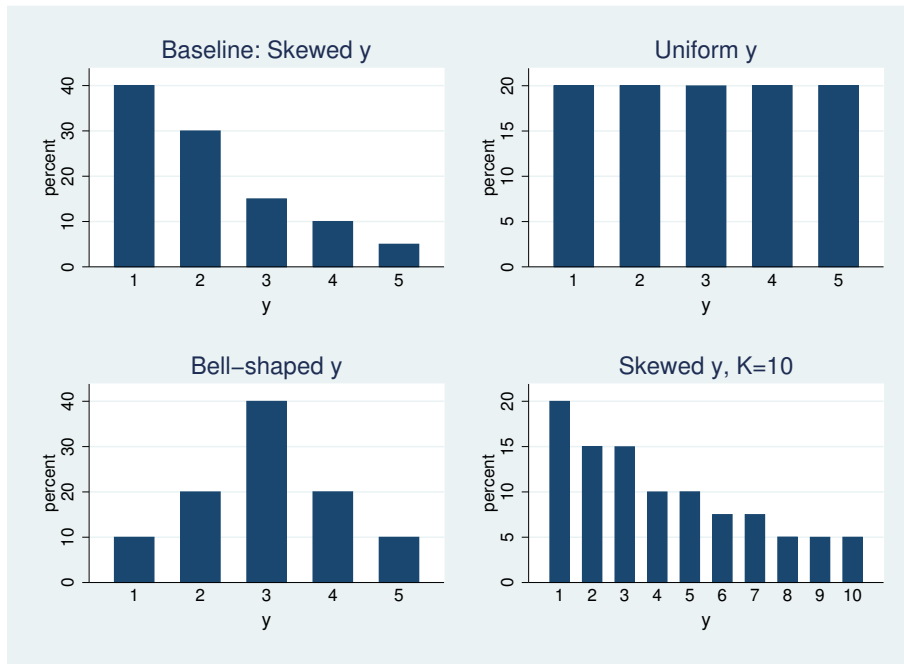


Table 1: Monte Carlo simulation results (1,000 replications): Baseline scenario

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

Baseline scenario: N=500, T=4, K=5			N=1,000						T=8						K=10						K=10, N=100					
			$\hat{\beta}_x$		$\hat{\beta}_d$		$\hat{\beta}_x$		$\hat{\beta}_d$		$\hat{\beta}_x$		$\hat{\beta}_d$		$\hat{\beta}_x$		$\hat{\beta}_d$		$\hat{\beta}_x$		$\hat{\beta}_d$		$\hat{\beta}_x$		$\hat{\beta}_d$	
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
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	0.99	0.06	0.98	0.10	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	0.06	1.00	0.11	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	0.06	0.86	0.11	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	0.06	0.87	0.11	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	0.98	0.08	0.94	0.07	0.93	0.11	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	1.00	0.18	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	0.09	1.00	0.15	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	0.14	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	1.00	0.15	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  $N=500, T=4, K=5, x \sim \text{Normal}(0, 1), d \sim \text{Bernoulli}(0.5)$ , 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  $d$ .

Baseline: skewed $y$ , $x \sim N(0, 1), d \sim B(0.5)$												$x \sim LN(0, 1), d \sim B(0.1)$											
$\hat{\beta}_x$						$\hat{\beta}_d$						$\hat{\beta}_x$						$\hat{\beta}_d$					

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; cluster robust standard errors in parenthesis; \* indicates 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 used for cluster-robust standard errors.

Table 5: FE Ordered Logit Estimates for GSOEP 1991-2009

Dep. var.: <i>Life Satisfaction</i>	$y \geq 8$	BUC
<i>Unemployed</i>	-0.97** (0.06)	-1.13** (0.05)
<i>Unemployed</i> $\times$ <i>Plant Closed</i>	-0.53* (0.22)	-0.33* (0.16)
<i>Out of labor force</i>	-0.30** (0.05)	-0.45** (0.04)
<i>Log. household income</i>	0.29** (0.03)	0.27** (0.03)
<i>Age</i>	-0.11** (0.01)	-0.11** (0.01)
<i>Squared age</i> $\times 100^{-1}$	0.05** (0.01)	0.05** (0.02)
<i>Married</i>	0.35** (0.04)	0.33** (0.05)
<i>Years of schooling</i> $\times 10^{-1}$	-0.04 (0.14)	-0.15 (0.18)
$\log L$	23'664	91'036
Observations	58'514	249'631
Individuals	6'823	9'835

*Notes:* GSOEP, waves 1991-1998 & 2001-2009. see Table 4