ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistics in Society
Series A

J. R. Statist. Soc. A (2015)

Consistent estimation of the fixed effects ordered logit model

Gregori Baetschmann,

University of Zurich, Switzerland

Kevin E. Staub

University of Melbourne, Australia

and Rainer Winkelmann

University of Zurich, Switzerland, and Institute for the Study of Labor, Bonn, Germany

[Received July 2013. Final revision July 2014]

Summary. The paper considers panel data methods for estimating ordered logit models with individual-specific correlated unobserved heterogeneity. We show that a popular approach is inconsistent, whereas some consistent and efficient estimators are available, including minimum distance and generalized method-of-moment estimators. A Monte Carlo study reveals the good properties of an alternative estimator that has not been considered in econometric applications before, is simple to implement and almost as efficient. An illustrative application based on data from the German Socio-Economic Panel confirms the large negative effect of unemployment on life satisfaction that has been found in the previous literature.

Keywords: Fixed effects; Logistic regression; Ordered response data; Panel data; Plant closures; Subjective wellbeing; Unemployment

1. Introduction

When estimating life satisfaction equations, or analysing determinants of job satisfaction or self-assessed health, researchers are often concerned about unobserved heterogeneity. Such heterogeneity can result from omitted variables or from subjective differences in anchoring of responses on the ordered response scale. If unaccounted for, heterogeneity will generally bias the estimated effects. Panel data offer a promising solution to this problem since the longitudinal information can be exploited to construct consistent estimators provided that the unobserved heterogeneity is time invariant.

Unfortunately, there is no consensus in the past literature on how to implement a fixed effects estimator for the ordered logit model. All proposals rely on conditional logit estimation of a dichotomized response (Chamberlain, 1980). In an early application, Winkelmann and Winkelmann (1998) used a single dichotomization at a constant value for all individuals to estimate the effect of unemployment on life satisfaction. This estimator is simple to implement but it does not use all the available information and is thus inefficient. Das and

Address for correspondence: Kevin E. Staub, Department of Economics, University of Melbourne, 111 Barry Street, Melbourne, VIC 3010, Australia.
E-mail: kevin.staub@unimelb.edu.au

 0964–1998/15/178000

van Soest (1999) obtained an efficient estimator for the fixed effects ordered logit model by combining all possible dichotomizations by using a two-step minimum distance (MD) method. Ferrer-i-Carbonell and Frijters (2004) proposed an individual-specific dichotomization to reduce the variance of the estimator. Whereas the MD estimator has not been used in the subsequent literature, applications of the Ferrer-i-Carbonell and Frijters (FF) estimator are quite frequent and include Frijters *et al.* (2004a,b, 2005), Kassenboehmer and Haisken-DeNew (2009), Booth and van Ours (2008), D'Addio *et al.* (2007), Schmitz (2011) and Jones and Schurer (2011).

The methodological contribution of this paper is twofold. First, we show that the FF estimator and related approaches are inconsistent. We construct a simple analytical counterexample demonstrating that the expected score of the log-likelihood function is not equal to 0 at the true parameter values. Moreover, a set of Monte Carlo simulations shows that the bias can be large and can affect coefficients as well as 'trade-off' ratios of coefficients.

Second, we explore three alternatives to MD estimation which are all consistent and efficient, or at least nearly efficient, as well: the generalized method of moments (GMM), empirical likelihood (EL) and an application of composite likelihood estimation due to Mukherjee *et al.* (2008) that we refer to as 'blow-up and cluster' (BUC). The BUC estimator is very simple to implement. Moreover, although not asymptotically efficient, the BUC estimator performs well in small samples. In fact, we do not find any noticeable losses of efficiency relative to the MD, GMM and EL estimators in our Monte Carlo simulations.

The performance of the various estimators is then compared in an application, where the ordered response variable 'life satisfaction' is modelled as a function of a number of socio-economic characteristics, among them current unemployment. The data are from the German Socio-Economic Panel for the years 2001–2011. Our results show substantial differences between the estimated BUC, MD and FF coefficients of the fixed effects ordered logit model. In accordance with our Monte Carlo results, the FF estimate of the unemployment effect is about 10% smaller in absolute value than the BUC estimate. Plant closure adds to the negative effect of unemployment but the difference is not statistically significant.

The paper proceeds as follows. Section 2 reviews the various estimators for the fixed effects ordered logit model. Section 3 reports results from a Monte Carlo study to compare the performance of the various estimators as a function of sample size (the number of individuals and number of time periods) as well as number of ordered categories. In Section 4, the methods are illustrated in an analysis of the effect of unemployment on life satisfaction. Section 5 concludes.

The programs that were used to analyse the data can be obtained from

`http://wileyonlinelibrary.com/journal/rss-datasets`

2. Econometric methods

2.1. 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 α_i and ε_{it}:

$$y^*_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{1}$$

The time invariant part of the unobservables, α_i, may or may not depend on x_{it}. One can either make an assumption regarding the distribution of α_i (or 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 observed ordered variable y_{it} is tied to the latent variable by the observation rule

$$y_{it} = k \qquad \text{if } \tau_{ik} < y_{it}^{*} \leqslant \tau_{ik+1}, \quad k = 1, \ldots, K, \tag{2}$$

where individual-specific thresholds τ_i are increasing ($\tau_{ik} \leqslant \tau_{ik+1} \forall k$), $\tau_{i1} = -\infty$, and $\tau_{iK+1} = \infty$. Moreover, the fixed effects ordered logit model assumes that ε_{it} are independent and identically distributed with logistic cumulative distribution function

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

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

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

which depends not only on β and x_{it}, but also on α_i, τ_{ik} and τ_{ik+1}. It is clear from equation (4) that only $\tau_{ik} - \alpha_i \equiv \alpha_{ik}$ is identified. Moreover, under fixed T asymptotics, estimation of α_{ik} uses information only from a finite number of observations even as the total number of observations NT grows without bound. Thus, the fixed effects α_{ik} cannot be estimated consistently and, in general, their inconsistency spills over to inconsistency of the parameters that are common to all observations. This situation is known as the *incidental parameters problem* (Neyman and Scott, 1948; Lancaster, 2000). In short panels, the resulting bias in $\hat{\beta}$ can be substantial (Abrevaya, 1997; Greene, 2004). Instead, a consistent estimator of β is obtained from collapsing y_{it} into a binary variable and then applying conditional maximum likelihood (CML) estimation (Andersen, 1970; Chamberlain, 1980).

The CML estimator is well known, but we present it nevertheless in detail to fix the notation. Let d_{it}^k denote the binary dependent variable that results from dichotomizing the ordered variable at the cut-off point k: $d_{it}^k = \mathbb{1}(y_{it} \geqslant k)$. k can be any integer between 2 and K. By construction, $\Pr(d_{it}^k = 0) = \Pr(y_{it} < k) = \Lambda(\tau_{ik} - x_{it}'\beta - \alpha_i)$, and $\Pr(d_{it}^k = 1) = 1 - \Lambda(\tau_{ik} - x_{it}'\beta - \alpha_i)$. Now consider the joint probability of observing $d_i^k = (d_{i1}^k, \ldots, d_{iT}^k)' = (j_{i1}, \ldots, j_{iT})'$ with $j_{it} \in \{0, 1\}$. The sum of all the individual outcomes over time is a sufficient statistic for α_i as

$$\mathcal{P}_i^k(\beta) \equiv \Pr\left(d_i^k = j_i \,\middle|\, \sum_{t=1}^{T} d_{it}^k = g_i \right) = \frac{\exp(j_i' x_i \beta)}{\sum\limits_{j \in B_i} \exp(j' x_i \beta)} \tag{5}$$

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

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

i.e. over all possible vectors of length T which have as many elements equal to 1 as the actual outcome of individual i, g_i. The number of j-vectors in B_i is

$$\binom{T}{g_i} = \frac{T!}{g_i!(T - g_i)!}.$$

Chamberlain (1980) showed that maximizing the conditional log-likelihood

$$\mathrm{LL}^k(b) = \sum_{i=1}^{N} \log\{\mathcal{P}_i^k(b)\} \tag{6}$$

gives a consistent estimator for β, denoted by $\hat{\beta}^k$ and henceforth referred to as the *Chamberlain estimator*.

The first-order conditions are $\Sigma_i s_i^k(b) = 0$ where

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

and the asymptotic variance of $\hat{\beta}^k$ is given by

$$\text{Avar}(\hat{\beta}^k) = -E\{H_i^k(\beta)\}^{-1} = E\{s_i^k(\beta) s_i^k(\beta)'\}^{-1}, \tag{8}$$

which can be estimated by averaging over individual Hessians

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

An important property of the Chamberlain estimator is that individuals with constant d_{it}^k do not contribute to the conditional likelihood function, since $\Pr(d_{it}^k = 1 | \Sigma_{t=1}^T d_{it}^k = T) = \Pr(d_{it}^k = 0 | \Sigma_{t=1}^T d_{it}^k = 0) = 1$. However, the ordered dependent variable can be dichotomized at different cut-off points, resulting in several consistent Chamberlain estimators. With K ordered outcomes, there are $K-1$ such estimators, and they employ information from different groups of individuals, depending on who crosses the cut-off and thus has variation in the dichotomized variable. For each individual there is at least one $k = \tilde{k}$ such that $d_{it}^{\tilde{k}}$ is not constant, unless $y_{i1} = \ldots = y_{iT}$. This feature is exploited by the individual-specific cut-off estimators that are discussed in the next section.

2.2. Individual-specific cut-off points

Ferrer-i-Carbonell and Frijters (2004) suggested the use of a single but distinct, in some sense 'optimal', cut-off point for each individual. A compact way of writing the FF estimator is by way of a weighted conditional log-likelihood function

$$\text{LL}^{\text{FF}}(b) = \sum_{i=1}^{N} \sum_{k=2}^{K} w_i^k \log\{\mathcal{P}_i^k(b)\}, \tag{10}$$

where $\mathcal{P}_i^k(b)$ is defined as in equation (5), $w_i^k = 0, 1$ and $\Sigma_{k=2}^K w_i^k = 1$. This objective function is maximized with respect to b, conditionally on the individual's weight vector w_i^k, $k = 2, \ldots, K$. The crucial question is where to dichotomize the dependent variable or, equivalently, which w_i^k to set to 1. The FF approach is to calculate for every individual all Hessian matrices under different cut-off points and then to choose the cut-off with the smallest Hessian:

$$w_i^k = 1, \qquad \text{if } k = \arg\min_k H_i^k(\beta).$$

In practice, the Hessians are evaluated at $\hat{\beta}$, a preliminary consistent estimator. By choosing the cut-off point leading to the smallest Hessian, this rule should yield a fixed effects ordered logit estimator with the smallest inverse of minus the sum of the Hessians, and thus minimal variance. Other simpler rules for choosing w_i^k have been used in the literature, trading efficiency for computational convenience. In fact, the standard way in which this estimator is implemented in applications is by choosing the individual mean of the dependent variable as dichotomizing cut-off point. Another possibility is to dichotomize at the median.

The key point is that these procedures determine the dichotomizing cut-off point endogenously, since it depends on y_i. This is problematic and leads to an inconsistent estimator. To provide some intuition for the inconsistency, consider the mean cut-off estimator as an example. In that estimator, it is easily seen that the cut-off is endogenous since $d_{it}^{\mathrm{Mn}}=1$ if and only if $y_{it} \geqslant T^{-1} \Sigma_t y_{it}$. Thus, y_{it} itself is part of the cut-off, and the probability $\Pr(d_{it}^{\mathrm{Mn}}=1)$ can be written as

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

Thus, the probability $\Pr(d_{it}^{\mathrm{Mn}}=1)$ 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 value for every period, and the implicit within-individual correlation between y_{it} and the time varying cut-off is negative. With endogenous cut-offs, the score of these estimators does not converge to 0 at the true parameter. On-line Web appendix A gives a formal proof of inconsistency. In Section 3, we provide quantitative information on the magnitude of the bias in some Monte Carlo simulations.

2.3. *Consistent and efficient estimators*

The estimators that have been discussed so far use only one dichotomization per individual to estimate β. This implies that they do not use all information contained in the variation of the dependent variable, and alternative approaches can provide gains in efficiency. A first possibility is to calculate all Chamberlain estimators separately and then to combine them in a second step using MD estimation (Das and van Soest, 1999). The second approach is based on the composite likelihood method (Varin *et al.*, 2011); it estimates β on the basis of the sum of the likelihood functions of all the different Chamberlain estimators. This method was used for instance by Mukherjee *et al.* (2008). The third approach is to combine the moment restrictions that are implied by the model and to use them in a GMM framework to estimate β.

2.3.1. *Minimum distance estimation*

Since every Chamberlain estimator $\hat{\beta}^k$ is a consistent estimator of β, so is any weighted average of them. The efficient combination can be obtained by MD estimation. Specifically, let M be a matrix of $K-1$ stacked L-dimensional identity matrices, and $\tilde{\beta}$ the $(K-1)L \times 1$ vector containing the $K-1$ Chamberlain estimators. The MD estimator is given by

$$\hat{\beta}^{\mathrm{MD}}=\arg\min_b(\tilde{\beta}-Mb)'\,\mathrm{var}(\tilde{\beta})^{-1}(\tilde{\beta}-Mb), \tag{11}$$

where $\mathrm{var}(\tilde{\beta})$ is the variance–covariance matrix of the stacked Chamberlain estimators (Das and van Soest, 1999). The solution to equation (11) is

$$\hat{\beta}^{\mathrm{MD}}=\{M'\mathrm{var}(\tilde{\beta})^{-1}M\}^{-1}M'\mathrm{var}(\tilde{\beta})^{-1}\tilde{\beta},$$

showing that the MD estimator is a matrix weighted average of the Chamberlain estimators. The asymptotic variance (i.e. the limiting variance of $\sqrt{n}(\hat{\beta}^{\mathrm{MD}}-\beta)$) is

$$\begin{aligned}\mathrm{Avar}(\hat{\beta}^{\mathrm{MD}})&=\{M'\mathrm{Avar}(\tilde{\beta})^{-1}M\}^{-1}\\&=[E(H_i(\beta))'E\{s_i(\beta)s_i(\beta)'\}^{-1}E(H_i(\beta))]^{-1},\end{aligned}$$

where $s_i(\beta)$ denotes individual i's stacked Chamberlain scores evaluated at β, and $H_i(\beta)$ the stacked Hessians.

2.3.2. Restricted conditional maximum likelihood estimation

Alternatively, the information that is associated with the different cut-offs can be combined in a single likelihood function, leading to a one-step estimator of β. The sample (quasi-) log-likelihood function of this restricted CML estimator is

$$\mathrm{LL}^{\mathrm{BUC}}(b)=\sum_{k=2}^{K}\mathrm{LL}^{k}(b), \tag{12}$$

where $\mathrm{LL}^k(b)$ is defined as in equation (6), and $\hat{\beta}^{\mathrm{BUC}}$ is the estimator that maximizes equation (12) and thus imposes the restriction that $\hat{\beta}^2=\ldots=\hat{\beta}^K$. Such an estimator has been suggested by Mukherjee *et al.* (2008). We refer to it as BUC because that describes the way of implementing this estimator by using CML estimation: replace every observation in the sample by $K-1$ copies of itself ('blow up' the sample size), and dichotomize each of the $K-1$ copies of the individual at a different cut-off point. The BUC estimates are obtained by CML estimation using the entire sample. The standard errors need to be clustered at the individual level since observations are dependent by construction.

It is straightforward to see that this approach leads to a consistent estimator. The score of the BUC log-likelihood function equals the sum of the scores of the Chamberlain estimators. Since these estimators are consistent, their scores converge to 0 in probability at the true parameter. It follows that the probability limit of the score of the restricted CML estimator is 0 as well:

$$\operatorname{plim}\sum_{k=2}^{K}\frac{1}{N}\sum_{i=1}^{N}s_i^k(\beta)=\operatorname{plim}\frac{1}{N}\sum_i s_i^2(\beta)+\ldots+\operatorname{plim}\frac{1}{N}\sum_i s_i^K(\beta)=0 \tag{13}$$

which, together with the concavity of the objective function, implies that $\hat{\beta}^{\mathrm{BUC}}$ converges to β.

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. The formula for the variance can be found in the next section, where it is shown that the BUC estimator can be written as an inefficient GMM estimator. The main difference between MD and BUC estimation is the weighting: by simply summing over the log-likelihood contributions, the BUC estimator implies weights of the Chamberlain estimators that are different from the variance-based weights that are used by the MD estimator.

2.3.3. Generalized method of moments and empirical likelihood

A third approach for achieving gains in efficiency over the simple Chamberlain estimator combines the moment conditions that are implied by the model under the different dichotomizations. With L explanatory variables, each dichotomization leads to L zero-expected score moment conditions. This gives $(K-1)L$ restrictions in total. Since only L parameters are estimated, the system is overidentified. The GMM estimator with weighting matrix W is

$$\hat{\beta}^{\mathrm{GMM}}=\arg\min_b s(b)'W\,s(b), \tag{14}$$

where

$$s(b)'=\frac{1}{N}\sum_{i=1}^{N}(s_i^{2'}(b),\ldots,s_i^{K'}(b)).$$

The first-order conditions of the GMM estimator with weighting matrix W are given by

$$\frac{\partial s(\hat{\beta}^{\text{GMM}})}{\partial \hat{\beta}^{\text{GMM}\prime}} W\, s(\hat{\beta}^{\text{GMM}}) = H(\hat{\beta}^{\text{GMM}})' W\, s(\hat{\beta}^{\text{GMM}}) = 0, \tag{15}$$

where $H(\hat{\beta}^{\text{GMM}})$ denotes the matrix of stacked Hessians of the single Chamberlain estimators evaluated at $\hat{\beta}^{\text{GMM}}$: $H(b)' = (H^2(b), \ldots, H^k(b))$. The efficient GMM estimator uses the inverse of the variance of the moment conditions as weighting matrix: $W^{\text{OPT}} = E\{s(\beta)\, s(\beta)'\}^{-1}$.

The asymptotic variance of the efficient GMM is

$$\begin{aligned}\text{Avar}(\hat{\beta}^{\text{GMM}}) &= \left[E\left(\frac{\partial s_i(\beta)}{\partial \beta'} \right)' E\{s_i(\beta)\, s_i(\beta)'\}^{-1} E\left(\frac{\partial s_i(\beta)}{\partial \beta'} \right) \right]^{-1} \\ &= [E(H_i(\beta))' E\{s_i(\beta) s_i(\beta)'\}^{-1} E(H_i(\beta))]^{-1}.\end{aligned} \tag{16}$$

It equals the asymptotic variance of the MD estimator. The form of the first-order conditions, equation (15), implies a GMM representation of the BUC estimator: setting the weighting matrix to a block diagonal matrix with the inverse of the Chamberlain Hessians on the diagonal yields the first-order conditions of the BUC estimator. Since this matrix, say W^{BUC}, is not equal to W^{OPT}, the weighting matrix of the efficient GMM estimator, the BUC estimator has in general a larger variance than the MD and GMM estimators. Using standard GMM results, the asymptotic variance of the BUC estimator is

$$\begin{aligned}\text{Avar}(\hat{\beta}^{\text{BUC}}) &= (H_i' W^{\text{BUC}} H_i)^{-1} (H_i' W^{\text{BUC}} S_i W^{\text{BUC}} H_i)(H_i' W^{\text{BUC}} H_i)^{-1} \\ &= \left\{ \sum_{k=2}^{K} E(H_i^k(\beta)) \right\}^{-1} \left[\sum_{k=2}^{K} \sum_{l=2}^{K} E\{s_i^k(\beta) s_i^l(\beta)\} \right] \left\{ \sum_{k=2}^{K} E(H_i^k(\beta)) \right\}^{-1},\end{aligned} \tag{17}$$

with $H_i = E(H_i(\beta))$, $S_i = E\{s_i(\beta) s_i(\beta)'\}$ and W^{BUC} denoting the weighting matrix described. The second equality follows since $W^{\text{BUC}} H_i = M$: a matrix of $K-1$ stacked L-dimensioned identity matrices. An estimate of expression (17) can be used to construct optimal weights for a weighted version of the BUC estimator.

As an alternative to the GMM, the EL estimator works directly with moment conditions as well. It has an asymptotic distribution that is identical to that of the efficient GMM estimator. However, EL estimators usually have better small sample properties (see for example Kitamura (2006)). In our set-up, the EL estimator is the result to the optimization problem

$$\max_{p,b} \sum_{i=1}^{N} \log(p_i), \qquad \text{subject to} \sum_{i=1}^{N} p_i = 1 \text{ and } \sum_{i=1}^{N} s_i(b) p_i = 0. \tag{18}$$

The vector $s_i(b)$ is the vector of stacked Chamberlain scores for individual i. p_i denotes the probability of observing individual i's variable realizations. The interest is only in b, whereas p is treated as an auxiliary parameter vector.

3. Monte Carlo study

This section compares the bias, precision and overall robustness of the various estimators of the fixed effects ordered logit model in small samples by using Monte Carlo simulations. First, although all estimators may suffer from bias in small and moderately sized samples due to the non-linearity of the objective functions, this bias, if any, should be minor compared with the bias from inconsistency due to endogenously chosen cut-offs by the FF estimators. Second, although the MD, GMM and EL are more efficient than the Chamberlain and BUC estimators, this is an asymptotic result that requires the use of optimal weights. In practice, the weights are unknown and need to be estimated from the data. This can be problematic if the sample size is small and there is a large number of categories, so that the number of individuals who cross a certain threshold is low. This situation is frequently encountered in applied research. In such cases the performance of the estimators may be poor, or, even worse, empirical counterparts of some of the moments may not be defined owing to a lack of observations.

It is therefore not clear, *ex ante*, whether the efficient estimators dominate the simpler estimators in finite sample settings. Anticipating our results, we find that the estimator which suffers the least from such problems is the BUC estimator. It is approximately unbiased and the loss of efficiency relative to the optimal estimators is very modest in our simulations. These facts, together with the simplicity of the implementation, make the BUC estimator an attractive option.

3.1. Experimental design

The data-generating process (DGP) for the latent variable is

$$y^*_{it} = \beta_1 x_{1it} + \beta_2 x_{2it} + \alpha_i + \varepsilon_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

where we set $\beta_1 = \beta_2 = 1$. The continuous regressor x_1 is normally distributed $N(0, 0.5)$. The error ε_{it} has a standard logistic distribution. The other regressor, x_2, is a binary regressor that is correlated with α_i. To generate correlation, we define two equally sized latent populations of individuals. In population 1, $x_{2it} \sim$ Bernoulli(0.5) and $\alpha_i \sim N(1, 0.5)$. In contrast, members of population 2 do not exhibit any variation in x_{2it} and the distribution of their time invariant heterogeneity has a lower mean: $x_{2it} = 0$ and $\alpha_i \sim N(0, 0.5)$. Thus, overall in the DGP α_i is uncorrelated with x_{1it} but correlated with x_{2it} ($\rho \approx 0.4$).

The observed ordered response variable y is obtained from the threshold mechanism (2). Although thresholds could be individual specific, we vary them only between the two populations. The number of categories K is equal to 5. In the first population, thresholds 2, 3 and 4 are equal to 0. Thus, only outcomes $y = 1$ and $y = 5$ are observed. (Whereas standard (cross-sectional) ordered logit models require $\tau_{ik} < \tau_{ik+1}$ as a regularity condition for identification of the thresholds, the CML approach that is discussed here does not estimate thresholds and so only requires the weaker regularity condition that $\tau_{ik} \leqslant \tau_{ik+1} \; \forall \, k$ and $\tau_{ik} < \tau_{ik+1}$ for at least one k.) In the second population, the thresholds are chosen such that y follows a discrete uniform distribution.

The DGP reflects a situation where the variation in the independent variables (which identifies the coefficients) differs across segments of the population that also differ in the distribution of their unobserved heterogeneity and the way that they translate the latent outcome into ordered responses. In the context of our empirical application, for instance, only a small fraction of the individuals exhibit variation in their employment status, and one could imagine that such individuals who differ from the remaining population in terms of observables also do so in terms of fixed effects and reporting.

In our simulation DGP, β_2 will be estimated from population 1 since x_{2it} exhibits only within-individual variation in this group, whereas β_1 is estimated from both populations. Hence, for inconsistent estimators, biases in the latter should be an average of the type of biases that are obtained in population 1 and 2, whereas the bias in estimates of β_2 should depend mainly on the distribution of y_{it} in population 1. Since there is more within-individual variance in the outcome in population 1 the sensitivity of endogenous cut-offs with respect to a particular y_{it} should be higher here and induce larger biases.

The baseline DGP is a balanced panel of $N = 500$ individuals observed for $T = 3$ periods. In a second step, the DGP is modified by increasing N, T and K.

3.2. Results

Table 1 contains results for the Monte Carlo simulations, based on 1000 replications of each DGP. Columns with heading $\hat{\beta}_1$ show means of estimated coefficients corresponding to x_1, and columns labelled $\hat{\beta}_2$ show means for those corresponding to x_2. The numbers in parentheses are the standard deviations of the estimates. The first group of four columns provides the results of the baseline DGP: $N = 500$, $T = 3$ and $K = 5$; the next group the results for a scenario where the number of individuals is increased to 1000; the following group the results for twice the number of time periods; and the last four columns show the results with 10 instead of five categories. Each row of Table 1 refers to a different estimator: the top four rows display results for Chamberlain estimators with cut-off points 2–5, followed by the three estimators with endogenous cut-offs. The last four rows contain the results of the four estimators combining the information of different cut-offs.

The simple Chamberlain estimators perform well in the simulations. They appear not to suffer from small sample bias and the estimation procedures always converge. However, they have a higher variance compared with the other proposed estimators. The loss of efficiency, as we would expect, becomes larger as the number of categories increases.

The estimators with endogenous cut-offs, in contrast, are clearly distorted. In the baseline scenario, the order of the biases is between 5% and 14%. The margin of error at the 99% level ($\pm 2.57\,\mathrm{SD}/\sqrt{1000}$) is always smaller than 0.02 and the distortion is therefore substantial. The bias is larger for the original FF estimator than for the mean or median cut-off estimators. To confirm the hypothesis that the deviation of the estimator's mean from the true parameter is caused by the procedure's inconsistency and not just a result of the small sample, we doubled the number of independent observations. The deviation is the same and illustrates therefore that these estimators are inconsistent.

Expanding the number of time periods, in contrast, reduces the distortion. The reason for the bias is not the use of individual-specific cut-offs *per se*, but the dependence of these cut-offs on y_i. Increasing the number of time periods decreases the dependence between cut-offs and realized error terms, and leads therefore to less biased estimators. In contrast, the size of the bias is exacerbated by adding categories. This is to be expected since all estimators degenerate to the same consistent Chamberlain estimator if the number of categories shrinks to 2. For example, for 10 categories, a standard number in research on job satisfaction and happiness, the mean of the FF estimator for β_2 is 0.80, which is well below the true value of 1.

Regarding the estimators which combine the available information of the Chamberlain estimators, it is noteworthy how well the BUC estimator performs. Although it is asymptotically less efficient, we find that the actual loss of efficiency of the BUC estimator is small to negligible in our Monte Carlo simulations. Regarding distortions, neither the BUC nor the EL estimator seems to suffer from an observable small sample bias. The GMM and the MD estimators in contrast show signs of distortions. These are most accentuated if there are few observations and many categories. The bias for β_1^{MD}, for example, is around 5% in the scenario with 500 individuals, three time periods and 10 categories. The bias of the GMM estimator in this setting is about 6%. Another problem is that the GMM and EL estimators do not always converge, at least with our implementation. It is known that this sort of convergence difficulties tends to be more pervasive with a higher number of explanatory variables. Thus, the BUC estimator can be a useful alternative for applied work with high dimensional x. (For instance, in 100 replications of the baseline DGP but with $x_1 \sim N(0, 0.5)$ replaced by five regressors distributed as $N(0, 0.1)$, the rate of convergence of our implementation decreased from 99.3% and 99.9% for GMM and EL reported in the footnotes of Table 1 to 84% and 97% respectively. BUC's convergence rate remained at 100%.)

Table 1. Monte Carlo simulation results (1000 replications)†

	Results for N = 500, T = 3 and K = 5		*Results for N = 1000, T = 3 and K = 5*		*Results for N = 500, T = 6 and K = 5*		*Results for N = 500, T = 3 and K = 10*	
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
Chamberlain estimators								
$y \geqslant 2$	1.00 (0.16)	1.01 (0.26)	1.01 (0.11)	1.01 (0.18)	1.00 (0.10)	1.00 (0.16)	1.01 (0.18)	1.02 (0.36)
$y \geqslant 3$	1.00 (0.15)	1.02 (0.21)	1.01 (0.11)	1.01 (0.14)	1.00 (0.09)	1.00 (0.13)	1.00 (0.16)	1.01 (0.26)
$y \geqslant 4$	1.00 (0.15)	1.01 (0.21)	1.00 (0.11)	1.01 (0.14)	1.00 (0.09)	1.00 (0.13)	1.00 (0.16)	1.01 (0.23)
$y \geqslant 5$	1.00 (0.16)	1.03 (0.26)	1.00 (0.11)	1.01 (0.18)	1.00 (0.10)	1.01 (0.16)	1.00 (0.15)	1.02 (0.21)
Estimators with endogenous cut-offs								
FF	0.92 (0.13)	0.86 (0.18)	0.92 (0.10)	0.86 (0.12)	0.96 (0.08)	0.92 (0.11)	0.88 (0.13)	0.80 (0.17)
Median	0.92 (0.13)	0.88 (0.18)	0.93 (0.10)	0.87 (0.12)	0.96 (0.09)	0.92 (0.11)	0.89 (0.13)	0.83 (0.17)
Mean	0.95 (0.13)	0.92 (0.18)	0.95 (0.10)	0.91 (0.12)	0.97 (0.09)	0.95 (0.12)	0.94 (0.13)	0.90 (0.18)
Estimators which combine all information								
MD	0.98 (0.14)	1.00 (0.18)	1.00 (0.10)	1.00 (0.12)	0.99 (0.09)	1.00 (0.11)	0.95 (0.13)	0.98 (0.17)
BUC	1.00 (0.14)	1.01 (0.18)	1.00 (0.10)	1.01 (0.12)	1.00 (0.09)	1.01 (0.11)	1.00 (0.14)	1.01 (0.17)
GMM	1.02 (0.14)	1.04 (0.19)	1.01 (0.10)	1.02 (0.12)	1.00 (0.09)	1.00 (0.11)	1.03 (0.15)	1.07 (0.19)
EL	1.00 (0.14)	1.01 (0.18)	1.01 (0.10)	1.01 (0.12)	1.00 (0.09)	1.00 (0.11)	1.00 (0.14)	1.01 (0.18)

†$\beta_1 = \beta_2 = 1$. Columns contain the mean of the estimated coefficients over all replications where the estimator has converged, with the standard deviation of the estimated coefficients in parentheses. Number of cases where an estimator did not converge, for $N = 500, T = 3$ and $K = 5$, GMM 7 and EL 1; for $N = 1000, T = 3$ and $K = 5$, GMM 3; for $N = 500, T = 6$ and $K = 5$, GMM 9; for $N = 500, T = 3$ and $K = 10$, GMM 8.

The discussion so far has focused on bias and precision of the different estimators for β. In ordered logit models, β-parameters have a straightforward interpretation as marginal effects of regressors on the log-odds, $\ln\{P(y>k)/P(y\leqslant k)\}$. However, the main interest is often not in the β-parameters *per se* but rather in derived statistics, such as ratios of coefficients or average marginal effects. Ratios are interesting, because they determine the compensating change of one variable required to offset a change in another, such that response probabilities remain unchanged. Ratios therefore quantify trade-offs between variables and, if one variable is income or price, monetary compensation.

Table 2 presents means and standard deviations for the estimated ratio $\hat{\beta}_1/\hat{\beta}_2$ over the 1000 replications corresponding to the four DGPs in Table 1. The results are in line with those of Table 1. The means of the ratio for the endogenous cut-off estimators are far off the true value (which equals 1). The Chamberlain estimators display acceptable results when either N or T is large, but they show some bias in the baseline DGP and in the last DGP with an increased number of categories ($K=10$). (One estimator which performs poorly in that DGP is the Chamberlain estimator with $y\geqslant 2$. However, some of the bias is due to outliers: the 1% trimmed mean is 1.11.) The estimators combining all available information, in contrast, estimate the ratio closely in all scenarios. In the results of Table 2 the efficiency advantage of MD, BUC, GMM and EL over Chamberlain and endogenous cut-off estimators is also much more visible than in Table 1. For instance, in the DGP with $K=10$ the standard deviation of the ratio is up to 40% larger for endogenous cut-off estimators than for the consistent and efficient estimators.

The average marginal effect of the lth regressor on the probability of outcome k has the form

$$\mathrm{AME}_l^k=(NT)^{-1}\sum_{i,t}[-\{\Lambda_{ik+1}(1-\Lambda_{ik+1})-\Lambda_{ik}(1-\Lambda_{ik})\}\,\beta_l],$$

Table 2. Monte Carlo simulation results (1000 replications): mean ratio $\hat{\beta}_1/\hat{\beta}_2$†

	$\hat{\beta}_1/\hat{\beta}_2$ *for the following DGPs:*			
	Baseline	*N = 1000*	*T = 6*	*K = 10*
Chamberlain estimators				
$y\geqslant 2$	1.08 (0.57)	1.03 (0.22)	1.02 (0.19)	1.27 (6.07)
$y\geqslant 3$	1.03 (0.28)	1.02 (0.18)	1.02 (0.16)	1.08 (0.57)
$y\geqslant 4$	1.04 (0.27)	1.02 (0.17)	1.02 (0.16)	1.04 (0.31)
$y\geqslant 5$	1.05 (0.35)	1.02 (0.22)	1.02 (0.19)	1.03 (0.28)
Estimators with endogenous cut-offs				
FF	1.11 (0.29)	1.09 (0.19)	1.06 (0.16)	1.15 (0.33)
Median	1.09 (0.28)	1.08 (0.18)	1.06 (0.16)	1.13 (0.32)
Mean	1.07 (0.27)	1.06 (0.18)	1.04 (0.16)	1.08 (0.27)
Estimators which combine all information				
MD	1.01 (0.23)	1.01 (0.15)	1.01 (0.14)	1.00 (0.23)
BUC	1.02 (0.23)	1.01 (0.15)	1.01 (0.14)	1.02 (0.23)
GMM	1.01 (0.23)	1.01 (0.15)	1.01 (0.13)	1.00 (0.23)
EL	1.02 (0.23)	1.01 (0.15)	1.01 (0.13)	1.02 (0.23)

†$\beta_1/\beta_2=1$. Columns contain the mean of the ratio of the estimated coefficients over all replications; the standard deviation of the estimated ratio is in parentheses. Baseline DGP: $N=500$, $T=3$ and $K=5$. Other DGPs differ from baseline only as indicated in the column header. See also the footnotes to Table 1.

Table 3. Monte Carlo simulation results (1000 replications): mean ratio $\hat{\beta}_1/\hat{\beta}_2$ under error misspecification†

	$\hat{\beta}_1/\hat{\beta}_2$ *for the following error specifications:*				
	$\varepsilon \sim N$	$\varepsilon \sim t(6)$	$\varepsilon \sim LN$	$\varepsilon \sim Po$	$\varepsilon \sim B$
Chamberlain estimators					
$y \geqslant 2$	0.76 (8.21)	1.09 (0.54)	0.76 (0.13)	0.89 (0.28)	0.31 (0.11)
$y \geqslant 3$	1.05 (0.31)	1.05 (0.29)	0.88 (0.12)	1.02 (0.28)	0.66 (0.91)
$y \geqslant 4$	1.05 (0.31)	1.04 (0.28)	1.03 (0.16)	1.07 (0.33)	0.65 (0.58)
$y \geqslant 5$	1.03 (0.38)	1.08 (0.39)	1.41 (0.41)	1.16 (0.64)	0.31 (0.11)
Estimators which combine all information					
MD	1.01 (0.25)	1.03 (0.23)	0.89 (0.14)	0.98 (0.24)	0.41 (0.14)
BUC	1.02 (0.25)	1.03 (0.24)	0.98 (0.12)	1.00 (0.24)	0.45 (0.15)

†$\beta_1/\beta_2 = 1$. Columns contain the mean of the ratio of the estimated coefficients over all replications; the standard deviation of the estimated ratio is in parentheses. Baseline DGP with $N = 500$, $T = 3$ and $K = 5$ but with error distributions as indicated in the column headers: N, normal; $t(6)$, Student's t; LN, log-normal; Po, Poisson; B, Bernoulli. The mean and variance of the error distributions are normalized to 0 and $\pi^2/3$.

where $\Lambda_{ik} = \Lambda(\tau_{ik} - x_{it}'\beta - \alpha_i)$. Because these effects depend on α_i, τ_{ik} and τ_{ik+1} they cannot be estimated consistently. However, the effect can be computed at a specific value of the linear index, such as that resulting in the sample probabilities. In the baseline DGP, this effect calculated for the binary regressor on the highest outcome is 24.4 percentage points. Since the effect is proportional to β_l, its relative bias is equal to the relative bias in β_l. Using the FF estimate, the effect is underestimated by 14%.

3.3. Sensitivity to misspecification of the error distribution

The CML estimator is inconsistent if the error distribution is not logistic. Some theoretical results (Czado and Santner, 1992) and simulation evidence (Cramer, 2007) suggest a certain robustness of the cross-sectional maximum likelihood logit estimator to error misspecification. In this subsection we explore the performance of some of the estimators under misspecification.

We focus on the simple Chamberlain CML estimators for a dichotomized y, as well as on two of the estimators which combine all available information: one efficient estimator—the MD estimator—and the inefficient BUC. Table 3 contains the results from the baseline DGP but with ε_{it} drawn from different, non-logistic distributions. In binary latent variable models, coefficients are identified only up to scale, and a normalization is required. We keep the logit normalization in all of the following DGPs, setting the variance of ε_{it} to $\pi^2/3$ (and its mean to 0). We follow Cramer (2007) in the selection of the distributions. We report the mean over 1000 replications of the estimated ratio β_1/β_2 which is equal to 1 in all DGPs, and its standard deviation in parentheses.

The first distribution that we consider is the Gaussian distribution (the column labelled '$\varepsilon_{it} \sim N$'). The choice between specifying a logistic or a normal error distribution is a matter of convenience. For instance, the sum of the outcome is a sufficient statistic only in the logit but not in the probit fixed effects model. The results from the simulations show, however, that the simple CML estimates of the coefficients' ratio seem to be robust to this error misspecification. The first Chamberlain estimator (dichotomized at 2) is an exception, exhibiting a large distortion.

However, this is mainly due to about 0.1% of the replications, where the coefficient estimates were extraordinarily large. One could argue that these observations should be excluded, which would lead to results that are comparable with those for the other Chamberlain estimators. Nevertheless, we report the results including these replications to illustrate the bias reducing effect of the estimators combining the available information: indeed, both MD and BUC show no significant distortion and estimate the ratio closely.

The normal is quite similar to the logistic distribution. Thus, as the next step, we consider a distribution with considerably higher kurtosis than that of the logistic: Student's t-distribution with 6 degrees of freedom (the column '$\varepsilon_{it} \sim t(6)$'). The bias in the Chamberlain estimators is now somewhat higher, but again MD's and BUC's performance is quite acceptable with biases of about 3%.

Both the normal distribution and the t-distribution are symmetric. The column '$\varepsilon \sim$ LN' reports results from a DGP with a log-normal—and thus skewed—distribution. The bias in the Chamberlain estimators is now asymmetric, also, going from −24% (for $y \leqslant 2$) to 41% (for $y \leqslant 5$). Whereas the MD estimator exhibits a bias of −11%, the BUC estimator remains virtually unbiased. The following column again uses a skew distribution. Taking the misspecification even further, the distribution is discrete: the errors in the DGP from the column '$\varepsilon \sim$ Po' follow a Poisson distribution. The pattern of the biases in the Chamberlain estimators is the same as in the log-normal case. Here again both MD and BUC can estimate the ratio without bias.

The previous evidence shows that the proposed estimators for the fixed effects ordered logit model seem to be quite robust to light and moderate misspecification of the error distribution. The final column acts as a *caveat* and shows that this result cannot be extended to cases of severe misspecification. In the column '$\varepsilon \sim B$' errors are drawn from a Bernoulli distribution and are therefore radically different from logistic errors. The ratios are now far off the true value not only for the Chamberlain estimators but also for the MD and BUC estimators.

4. Effect of unemployment on life satisfaction

To illustrate the functioning of the various estimators in a real data example, we revisit the empirical modelling of the determinants of life satisfaction by using household panel data for Germany, extracted from the German Socio-Economic Panel (Wagner *et al.*, 2007). This application provided the initial context for the development of some of the fixed effects ordered logit estimators that are discussed in this paper (Winkelmann and Winkelmann, 1998; Ferrer-i-Carbonell and Frijters, 2004). Moreover, it is an area where inconsistent estimators have been used in the past (e.g. Kassenboehmer and Haisken-DeNew (2012) and Knabe and Rätzel (2011)).

The main purpose of this section is illustrative, comparing the various estimators in a typical application of fixed effects ordered logit models, rather than contributing to the frontier of research on the wellbeing cost of unemployment. For instance, by considering the contemporaneous effect only, we do not allow for anticipation or adaptation effects. Nor do we consider the importance of social work norms, aggregate or household unemployment, or peer effects that have been discussed in the more recent literature.

Our specification is similar to that of the early panel applications. Regressors include indicators of labour force status, marital status, health and logarithmic household income, all referring to the time of the interview. Macroeconomic shocks are controlled for by the inclusion of time effects that may or may not vary by state ('*Bundesland*'). It is impossible, in a fixed effects model, to identify the linear effect of age and the linear component of a time trend separately. Hence, we follow Ferrer-i-Carbonell and Frijters (2004) and exclude the variable 'age' from our model.

We also consider a slightly extended specification, where we include an indicator for plant

Table 4. Sample averages by year†

	Results for the following years:										
	2001	*2002*	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*
Life satisfaction	7.20	7.16	7.05	6.92	7.06	7.04	7.09	7.09	7.01	7.12	7.08
Age	42.4	43.3	43.1	43.1	43.2	43.6	43.6	43.8	44.0	44.4	44.7
Logarithm of net household income	10.36	10.52	10.52	10.52	10.53	10.52	10.53	10.56	10.56	10.58	10.59
Years in panel	7.8	7.6	8.7	9.4	10.2	9.8	10.8	11.6	10.9	11.8	12.0
Unemployed	0.050	0.046	0.058	0.064	0.065	0.062	0.049	0.043	0.049	0.050	0.041
Unemployment due to plant closure	0.044	0.050	0.085	0.082	0.041	0.047	0.030	0.041	0.050	0.051	0.027
Not in labour force	0.137	0.138	0.137	0.131	0.122	0.129	0.121	0.119	0.125	0.128	0.130
Married	0.672	0.678	0.661	0.647	0.640	0.644	0.636	0.630	0.616	0.609	0.610
Good health	0.842	0.841	0.843	0.846	0.85	0.845	0.85	0.851	0.844	0.834	0.826
Number of observations	6309	6897	6311	6063	5646	5934	5464	5046	5278	4747	4582

†Data source: German Socio-Economic Panel, waves 2001–2011.

closure unemployment and control for panel learning effects. Repeated participation in the survey appears to be associated with lower reported life satisfaction, *ceteris paribus* (Kassenboehmer and Haisken-DeNew, 2012). One explanation is that increasing familiarity with the interviewer changes the reporting behaviour by reducing social desirability bias. In a fixed effects model, the panel learning effect is identified from individuals with gaps in their participation history. Moreover, people who lost their job during the 12 months before the interview were asked for the reasons for termination of their employment, one of them being 'plant closure'. Kassenboehmer and Haisken-DeNew (2009) interpreted plant closure as an exogenous shock, adding support to the view that corresponding estimates identify the causal effect of plant closure unemployment on life satisfaction (see also Schmitz (2011)). Our discussion of results will centre on the estimated coefficient of the unemployment indicator variable, as well as on the income coefficient, the ratio of the two providing an estimate of the 'shadow cost' of unemployment.

4.1. Data

The sample that was used for estimation has been extracted from the 2001–2011 waves of the German Socio-Economic Panel. We start in 2001, since consecutive information on all variables of interest is available only from that year onwards. The analysis focuses on men in the West German subsample, aged between 21 and 64 years. The sampling results in an unbalanced panel with a total of 62277 observations on 11563 individuals.

The outcome variable is satisfaction with life measured by the question 'How satisfied are you at present with your life as a whole?'. The answer has 11 ordered categories ranging from 0, 'completely dissatisfied', to 10, 'completely satisfied'. Sample averages for the main variables employed in the analysis are given in Table 4.

Average life satisfaction is slightly above 7 in most of the 11 years. Both average age and average duration in the panel increase over time, from 42.4 to 44.7 and from 7.8 to 12.0 years respectively. The 'unemployment rate', defined as the fraction of unemployed in the total working age population, peaks at 6.4% in 2004 (corresponding to 7.4% of the labour force). Plant closure unemployment is a relatively rare event during the observation period: only between 3% and 8% of all unemployed people state that they lost their previous job because of plant closure. As a consequence, we expect that it will be difficult with these data to estimate the effect of plant closure unemployment precisely. Income is calculated as the logarithm of household income after taxes and transfers. 'Good health' combines two questions: the value is 1 if the respondent indicates having no chronic health condition, in addition to having spent no days in a hospital during the preceeding year, and 0 otherwise. The proportion of people in good health remains relatively stable over time, with a total average of about 84% of the responses. Table 4 does not contain any variables that are time invariant, such as completed education or personality traits. These factors would be absorbed into the fixed effect in our models and are thus irrelevant for the analysis.

4.2. Fixed effects ordered logit results

Table 5 presents estimation results for a benchmark ordered logit model with individual and year fixed effects by using six different estimators: a fixed cut-off estimator (cut-off at 6), the BUC estimator, the MD estimator and three estimators with individual-specific endogenous cut-off: the mean, median and minimum Hessian (FF). There are 62277 observations in total, distributed over 11563 distinct individuals. Among them, 3357 individuals have no time variation in life satisfaction, a majority (2521 of them) because they are present in the sample for a single wave only. Regardless of estimation method, such observations cannot contribute to estimation in a fixed effects ordered logit model. Of the remaining 8206 people the Chamberlain estimator with dichotomization at value 6 shown in column (1) of Table 5 uses only 3418 individuals with 25981 person-year observations. This dichotomization entails therefore a large loss of information.

By contrast, the BUC and MD estimators that are shown in columns (2) and (3) of Table 5 use more than twice as many observations, 57274 entries for 8 206 individuals. The resulting gains in

Table 5. Fixed effects ordered logit estimates of life satisfaction†

	(1), $y\geqslant 6$	*(2), BUC*	*(3), MD*	*(4), Mean*	*(5), Median*	*(6), FF*
Unemployed	−1.096	−1.035	−1.005	−0.898	−0.854	−0.941
	(0.074)	(0.059)	(0.056)	(0.059)	(0.058)	(0.065)
Out of labour force	−0.471	−0.330	−0.255	−0.278	−0.240	−0.254
	(0.075)	(0.054)	(0.047)	(0.048)	(0.048)	(0.056)
Married	0.333	0.302	0.302	0.359	0.342	0.405
	(0.090)	(0.064)	(0.060)	(0.059)	(0.059)	(0.065)
Good health	0.318	0.318	0.298	0.264	0.241	0.286
	(0.057)	(0.040)	(0.036)	(0.037)	(0.037)	(0.042)
Age squared	0.0007	0.0008	0.0008	0.0007	0.0007	0.0006
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
log(net household income)	0.270	0.278	0.257	0.214	0.216	0.203
	(0.054)	(0.038)	(0.038)	(0.035)	(0.035)	(0.037)
Individual fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo-log-likelihood	−9486.6	−61219.0		−25725.4	−26047.8	−19401.4
Observations	25981	57274	57274	57274	57274	57274
Individuals	3418	8206	8206	8206	8206	8206

†Dependent variable, life satisfaction; data source, German Socio-Economic Panel, waves 2001–2011; cluster robust standard errors are in parentheses; 'Observations' denotes the number of person-years in the estimation sample; 'Individuals' denotes the number of unique people in the estimation sample.

efficiency are substantial. For example, the standard error of the unemployment effect drops by 20% from 0.074 to 0.059 relatively to that of the Chamberlain estimator. MD estimation reduces the standard errors further, to 0.056 in the case of the unemployment effect. As to substantive results, the effect of unemployment is found to be large and statistically significant, confirming the results of the previous literature. Since the regressions control for log(household income), the measured unemployment effect is purely non-pecuniary, or psychological. To obtain the total reduction in life satisfaction due to unemployment, we would need to add the additional cost of reduced household income if benefits replace income less than fully.

The most striking feature of Table 5 as a whole is that columns (1)–(3)—which are based on consistent estimators—give broadly similar results, whereas they differ from the three last columns based on inconsistent estimators. The unemployment coefficient lies between -1 and -1.1 when using Chamberlain, MD or BUC estimators but it ranges from -0.85 to -0.94 when using FF, mean or median estimators. Attenuation bias is observed for the effects of non-participation, health and household income as well, though not for marital status. Indeed, a tendency for attenuation bias was already present in the life satisfaction models of Ferrer-i-Carbonell and Frijters (2004), when they compared their estimator with the MD approach, though they did not interpret it that way.

4.3. *Marginal effects and shadow values*

In the ordered logit model, coefficients β_j are equal to the relative change in the continuation odds $\Pr(y>k)/\Pr(y\leqslant k)$ that are associated with a small change in x_j. The effects of x on other quantities, such as $\Pr(y>k)$ or $\Pr(y=k)$, depend on the fixed effects α_i and are thus not determined. Of course, since probabilities are a monotone function of odds, one can compute marginal probability effects for given odds (e.g. at the sample mean).

In practice, life satisfaction equations are frequently employed to estimate shadow values of non-traded goods or 'bads' (including airport noise, pollution, health and unemployment) as an input into cost–benefit analyses. For example, minus the ratio of the unemployment coefficient and the coefficient of logarithmic income indicates the relative change in income that is required to keep overall life satisfaction of an unemployed person equal to that of an otherwise similar employed person.

On the basis of the estimates in Table 5, we find that the shadow value of unemployment varies between a minimum of 3.72 (BUC) and a maximum of 4.63 (FF). Estimation of a linear fixed effects panel data model would yield a shadow value of 4.25. Even larger differences are found for other variables. For example, the implied shadow value of being married is equal to a 1.1-fold increase in income on the basis of the BUC estimator, compared with a 2.0-fold increase in income based on the FF estimator. The application thus corroborates the simulation evidence on trade-off distortions, with the corresponding adverse consequences of using inconsistent estimators in policy analysis. Note that all the shadow values are very high, perhaps unrealistically so, which is a common result in the literature resulting from the relative insensitivity of reported life satisfaction to income changes.

4.4. *Plant closure unemployment*

Does it matter how the unemployment spell was initiated? To estimate the effect of plant closure unemployment, we include an interaction term Unemployed $\times$ Plant Closure, in addition to the main effect Unemployed. This interaction term is 1 if the respondent lost his or her job during the previous 12-month period due to plant closure *and* is unemployed at the time of the interview, and is 0 otherwise. People may become unemployed disproportionately in periods and

Table 6. Plant closure, unemployment and life satisfaction†

	(1)	*(2)*	*(3)*	*(4)*
Unemployed	−1.039	−1.046	−1.023	−1.032
	(0.059)	(0.059)	(0.061)	(0.060)
Unemployed × Plant Closure			−0.222	−0.186
			(0.192)	(0.192)
Out of labour force	−0.333	−0.336	−0.330	−0.334
	(0.054)	(0.054)	(0.054)	(0.054)
Married	0.302	0.300	0.302	0.299
	(0.064)	(0.064)	(0.064)	(0.064)
Good health	0.321	0.324	0.321	0.324
	(0.040)	(0.040)	(0.040)	(0.040)
Age squared	0.0006	0.0006	0.0006	0.0006
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
log(household income)	0.275	0.278	0.277	0.279
	(0.038)	(0.038)	(0.038)	(0.038)
State effects, year effects	Yes	No	Yes	No
State × year effects	No	Yes	No	Yes
Pseudo-log-likelihood	−61130.8	−60950.4	−61128.3	−60948.7

†Dependent variable, life satisfaction; all models are estimated with the BUC estimator; 57274 observations; all models include individual-specific fixed effects and a second-order polynomial in panel duration. Cluster robust standard errors are given in parentheses. Data source: the German Socio-Economic Panel, 2001–2011.

places in which the economy does not do well. If people are less satisfied in such economically depressed times for reasons other than own unemployment, the analysis confounds those effects. We therefore include in an extended specification a full set of dummy variables for all state–year combinations.

All models are estimated by using the BUC estimator. Columns (1) and (2) of Table 6 include only the Unemployed main effect, whereas columns (3) and (4) show the results with the plant closure–unemployment interaction. Columns (1) and (2) largely corroborate the findings of the previous subsection, showing that the unemployment finding is robust to the inclusion of state and time effects, as well as state-specific trends, and thus unrelated to macrofactors that are captured by these variables.

The main effect in columns (3) and (4) now measures the effect for those entering unemployment for reasons that are unrelated to plant closure (such as resulting from individual dismissal or employee-initiated quitting). The point estimate is virtually unchanged. The plant closure interaction is negative; it adds an additional effect of about a fifth, for a total effect of about −1.22. Thus, the detrimental effect of unemployment on life satisfaction seems to be especially large for arguably exogenous unemployment shocks. However, the standard error of the interaction terms is large, and the hypothesis of no extra effect cannot be rejected at conventional levels of significance. As pointed out before, there are just not that many unemployment spells due to plant closure in the data, which is a limitation of this approach.

5. Conclusions

The ordered logit model has several desirable features that make it the first choice in regression analyses of discrete, ordinally measured variables, as they arise in the elicitation of life and job satisfaction or self-assessed health. It has a parsimonious yet flexible parameterization that exploits the ordering information while allowing inferences to be made on the entire distribution of outcomes. However, applications of the ordered logit model to panel data with fixed effects have been hampered so far by the lack of a unified discussion of various possible estimators and their respective advantages and shortcomings.

We show in this paper that two of the existing approaches that have been used in the prior literature cannot be recommended. The first is a simple dichotomization of the ordered response into a binary variable. This approach is inefficient and misses individuals in the sample who do not cross the dichotomizing cut-off over time. The second approach is endogenously choosing individual-specific cut-off values. This approach is inconsistent. The bias can be substantial, as shown both in Monte Carlo simulations and in the illustrative application to the effect of unemployment on life satisfaction.

However, we derive the consistent and asymptotically efficient GMM and EL estimators that use all available information and have the same asymptotic covariance matrix as an MD estimator. We also study a modified estimator that, although not efficient, may be more robust in finite samples. This estimator, which we call the BUC estimator, might be especially attractive to practitioners, as it is simple to implement and its maximization process is stable.

Another computationally simple estimator that we did not discuss in this paper is the ordinary least squares linear fixed effects estimator. If the underlying model is generated by mechanisms akin to those discussed in this paper, ordinary least squares can yield severely biased estimates. The reason is that the ordinary least squares estimand is based on the covariance between the observed ordered variable and the regressors, whereas the true coefficients are defined by the covariance between the unobserved latent variable and the regressors. Depending on the distribution of the thresholds, the two can be quite different. Thus, unless a cardinal interpretation of the ordered data is warranted, ordinary least squares cannot be recommended either.

Acknowledgements

We thank the Joint Editor, Associate Editor, two referees, Paul Frijters and Arie Kapteyn as well as participants of the Engelberg Workshop in Labor Economics, the North American Summer Meetings of the Econometric Society in St Louis and the International Panel Data Conference in Montreal for very valuable comments on an earlier version.

References

Abrevaya, J. (1997) The equivalence of two estimators of the fixed-effects logit model. *Econ. Lett.*, **55**, 41–43.

Andersen, E. B. (1970) Asymptotic properties of conditional maximum-likelihood estimators. *J. R. Statist. Soc.* B, **32**, 283–301.

Booth, A. L. and van Ours, J. C. (2008) Job satisfaction and family happiness: the part-time work puzzle. *Econ. J.*, suppl., **118**, F77–F99.

Chamberlain, G. (1980) Analysis of covariance with qualitative data. *Rev. Econ. Stud.*, **47**, 225–238.

Cramer, J. S. (2007) Robustness of logit analysis: unobserved heterogeneity and mis-specified disturbances. *Oxf. Bull. Econ. Statist.*, **69**, 545–555.

Czado, C. and Santner, T. J. (1992) The effect of link misspecification on binary regression inference. *J. Statist. Planng Inf.*, **33**, 213–231.

D'Addio, A. C., Eriksson, T. and Frijters, P. (2007) An analysis of the determinants of job satisfaction when individuals' baseline satisfaction levels may differ. *Appl. Econ.*, **39**, 2413–2423.

Das, M. and van Soest, A. (1999) A panel data model for subjective information on household income growth. *J. Econ. Behav. Organzn*, **40**, 409–426.

Ferrer-i-Carbonell, A. and Frijters, P. (2004) How important is methodology for the estimates of the determinants of happiness? *Econ. J.*, **114**, 641–659.

Frijters, P., Haisken-DeNew, J. P. and Shields, M. A. (2004a) Investigating the patterns and determinants of life satisfaction in Germany following reunification. *J. Hum. Resour.*, **39**, 649–674.

Frijters, P., Haisken-DeNew, J. P. and Shields, M. A. (2004b) Money does matter!: evidence from increasing real income and life satisfaction in East Germany following reunification. *Am. Econ. Rev.*, **94**, 730–740.
Frijters, P., Haisken-DeNew, J. P. and Shields, M. A. (2005) The causal effect of income on health: evidence from German reunification. *J. Hlth Econ.*, **24**, 997–1017.
Greene, W. H. (2004) The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econmetr. J.*, **7**, 98–119.
Jones, A. M. and Schurer, S. (2011) How does heterogeneity shape the socioeconomic gradient in health satisfaction. *J. Appl. Econmetr.*, **26**, 549–579.
Kassenboehmer, S. C. and Haisken-DeNew, J. P. (2009) You're fired!: the causal negative effect of unemployment on life satisfaction. *Econ. J.*, **119**, 448–462.
Kassenboehmer, S. C. and Haisken-DeNew, J. P. (2012) Heresy or enlightenment?: the well-being age U-shape effect is flat. *Econ. Lett.*, **117**, 235–238.
Kitamura, Y. (2006) Empirical likelihood methods in econometrics: theory and practice. *Discussion Paper 1569*. Cowles Foundation, New Haven.
Knabe, A. and Rätzel, S. (2011) Scarring or scaring?: the psychological impact of past unemployment and future unemployment risk. *Economica*, **78**, 283–293.
Lancaster, T. (2000) The incidental parameter problem since 1948. *J. Econmetr.*, **95**, 391–413.
Mukherjee, B., Ahn, J., Liu, I., Rathouz, P. J. and Sanchez, B. N. (2008) Fitting stratified proportional odds models by amalgamating conditional likelihoods. *Statist. Med.*, **27**, 4950–4971.
Neyman, J. and Scott, E. (1948) Consistent estimation based on partially consistent observations. *Econometrica*, **16**, 1–32.
Schmitz, H. (2011) Why are the unemployed in worse health?: the causal effect of unemployment on health. *Lab. Econ.*, **18**, 71–78.
Varin, C., Reid, N. and Firth, D. (2011) An overview of composite likelihood methods. *Statist. Sin.*, **21**, 5–42.
Wagner, G. G., Frick, J. R. and Schupp, J. (2007) The German Socio-Economic Panel Study (SOEP) scope, evolution and enhancements. *Schmoll. Jahrb.*, **127**, 139–169.
Winkelmann, L. and Winkelmann, R. (1998) Why are the unemployed so unhappy?: evidence from panel data. *Economica*, **65**, 1–15.

Supporting information
Additional 'supporting information' may be found in the on-line version of this article.