Interval censored regression with fixed effects*

Jason Abrevaya† and Chris Muris‡

March 2018

Abstract

This paper considers estimation of a fixed-effects model with an interval-censored dependent variable. In each time period, the researcher observes the interval (with known endpoints) in which the dependent variable lies but not the value of the dependent variable itself. Two versions of the model are considered, a parametric model with logistic errors and a semiparametric model with errors having an unspecified distribution. In both cases, the error disturbances can be heteroskedastic over cross-sectional units as long as they are stationary within a cross-sectional unit; the semiparametric model also allows for serial correlation of the error disturbances. A conditional-logit-type composite likelihood estimator is proposed for the logistic fixed-effects model, and a composite maximum-score-type estimator is proposed for the semiparametric model. In general, the scale of the coefficient parameters is identified by these estimators, meaning that the causal effects of interest are estimated directly in cases where the latent dependent variable is of primary interest (e.g., pure data-coding situations). Monte Carlo simulations and an empirical application to birthweight outcomes illustrate the performance of the parametric estimator.

*This paper has benefited from comments by Ying-Ying Lee and seminar participants at McGill University, the 2017 Texas Econometrics Camp, and the 2017 International Panel Data Conference. Data and code for the empirical application are available from https://github.com/chrismuris/IntervalFixedEffects.

†Department of Economics, The University of Texas at Austin, abrevaya@austin.utexas.edu.

‡Department of Economics, University of Bristol, chris.muris@bristol.ac.uk.

1 Introduction

This paper considers estimation of a fixed-effects model (large n, small T) with an interval-censored dependent variable. In each time period, the researcher observes the interval (with known endpoints) in which the dependent variable lies but not the value of the dependent variable itself. This type of interval-censored variable is commonly encountered in data used by economists, including several well-known longitudinal datasets. For instance, the public-use version of the Medical Expenditure Panel Survey (MEPS) has interval censoring of the annual health-care expenditure variable; in 2014, the intervals for the health-care expenditure variable were as follows: equal to \$0, between \$1 and \$322, between \$323 and \$1,050, between \$1,051 and \$3,877, and above \$3,877. Other longitudinal datasets with interval-censored variables include the British Household Panel Survey (interval-censored investment income), the British Cohort Survey (interval-censored parental income), the General Social Survey (interval-censored earnings variable), and the Health and Retirement Survey (interval-censored earnings variable for some respondents).

Since the actual value of the interval-censored dependent variable is unobserved, the use of standard linear fixed-effects estimators is inappropriate. Instead, we propose a latent-variable fixed-effects model that allows for heteroskedasticity of the error disturbances and, due to the known interval cutoff points, permits identification of the scale of the slope parameters. Both of these features distinguish the model from the ordered-choice fixed-effects model, with unknown interval cutoff points, that has previously been considered in the literature (e.g., Baetschmann, Staub, and Winkelmann, 2015; Das and van Soest, 1999; Muris, 2017). We consider two alternative assumptions on the error disturbances in the latent-variable model. The first is a parametric specification in which the disturbances are assumed to be logistic, with heteroskedasticity over cross-sectional units but stationarity and lack of serial correlation within a cross-sectional unit. The second is a semiparametric specification in which the disturbances have an unspecified distribution. Like the logistic model, the error disturbances in the semiparametric model are permitted to be heteroskedastic over cross-sectional units as long as they are stationary within cross-sectional units; unlike the logistic model, the error disturbances may also exhibit serial correlation within cross-sectional units.

For the parametric (logistic) fixed-effects model with interval-censored outcomes, we propose a conditional-logit-type composite likelihood estimator. The idea of conditional likelihood estimation, as a way of eliminating fixed effects, is now familiar in the literature on non-linear fixed-effects models. This approach was initially proposed for the binary-choice fixed-effects logit model (Andersen, 1970; Chamberlain, 1980) and has been extended to the ordered-choice fixed-effects logit model. Unlike these models, where lack of scale identification necessitates an error-disturbance variance normalization, our model does not require a variance normalization and moreover allows for heteroskastic error disturbances. In this sense, the error-disturbance assumptions that we make are similar to those used in Honoré (1992) for the censored regression fixed-effects model. The estimators proposed in Honoré (1992) allow for heteroskedastic (and stationary) error disturbances but do not allow for serial correlation. In contrast to our parametric estimator, the estimators in Honoré (1992) do not require a parametric assumption to achieve $\sqrt{n}$-consistency since the actual value of the dependent variable is observed for uncensored observations.[1] Finally, we note that conditional likelihood estimators have also been proposed for other non-linear fixed-effects models with underlying exponential likelihoods. Examples of such models include the Poisson model (Hausman, Hall, and Griliches, 1984), the binomial regression model with logistic link function (Machado, 2004), and certain duration models (Chamberlain, 1985).

For the semiparametric fixed-effects model with interval-censored outcomes, we propose a composite maximum-score-type estimator to consistently estimate the slope parameters. Manski (1987) proposed the *maximum score estimator* for a binary-choice fixed-effects model in which the error disturbances have an unspecified and possibly heteroskedastic (but stationary) distribution. The semiparametric maximum-score approach was extended to the ordered-choice fixed-effects model and other non-linear fixed-effects models by Abrevaya (2000). In the models considered by Manski (1987) and Abrevaya (2000), the scale of the slope parameters in the latent-variable model is not identified. In contrast, our proposed semiparametric maximum-score-type estimator identifies the scale of the slope parameters.

The outline of the paper is as follows. Section 2 introduces the interval-censored fixed-effects model with logistic error disturbances. Identification of the model's parameters is shown to hold when there are at least three intervals in each time period. We propose a composite maximum likelihood estimator (CMLE), based upon the conditional-logit approach for the binary-choice fixed-effects model, and we show consistency and asymptotic normality of the CMLE. In a version of the logistic model that allows for heteroskedasticity, we suggest using an exponential specification for the heteroskedasticity function and propose an appropriately modified version of the CMLE. Section 3 considers the semiparametric version of the model, where the distribution of the error disturbances is left unspecified. As in the logit model, we allow for heteroskedasticity and require stationarity of the error disturbances, but we relax the assumption of no serial correlation. For this semiparametric model, we propose a composite maximum-score-type estimator that can be used to consistently estimate the slope parameters from the latent-variable equation. Section 4 provides Monte Carlo evidence on the finite-sample performance of the CMLE estimator. Section 5 considers an empirical application to birthweight outcomes using a very large two-period panel dataset. To compare the estimator's performance to the baseline of the first-difference estimator (under complete observability), we start from a dataset with fully observed birthweight outcomes and then artificially interval-censor the outcome variable. Section 6 concludes.

[1] Some other recent work on panel models with fixed effects that accommodate censored outcomes include Alan et al. (2013), Galvao et al. (2013), and Khan et al. (2016). The methods proposed in this paper use continuous support of the outcome variable on a subset of the real line, which rules out their application to the case of interval-censored outcomes.

2 Interval-censored fixed-effects logit model

We consider an interval-censored fixed-effects model, with the researcher observing the interval in which the dependent variable lies but not the value of the dependent variable itself. The endpoints of the intervals are assumed to be *known*. For simplicity, the theoretical treatment focuses upon the case of two time periods and time-invariant interval endpoints; extensions to additional time periods and time-varying interval endpoints are straightforward. In addition, a homoskedastic version of the logistic model is considered initially, with the extension to heteroskedasticity considered in Section 2.2. With i indexing cross-sectional units and t indexing time periods, the following linear fixed-effects specification is assumed for the underlying latent dependent variables:

$$y_{it}^* = \alpha_i + X_{it}\beta_0 - \sigma_0 u_{it} \ (i = 1, \ldots, n;\ t = 1, 2), \tag{1}$$

with a vector of regressors $X_{it} \in \mathbb{R}^{1\times K}$, an associated vector of regression coefficients $\beta_0 \in \mathbb{R}^{K\times 1}$, and a positive scaling parameter σ_0 on the error terms. The error terms (u_{i1}, u_{i2}) are assumed to be serially independent, conditional on the regressors $X_i \equiv (X_{i1}, X_{i2})$ and the fixed effect (unobserved heterogeneity) $\alpha_i \in \mathbb{R}$, with each following a standard logistic distribution:

$$(u_{i1}, u_{i2})|\,(\alpha_i, X_i) \sim \text{iid logistic}. \tag{2}$$

The cdf function associated with the logistic distribution is denoted $\Lambda(v)$, where $\Lambda(v) = \exp(v)/(1+\exp(v))$.

The latent-variable y_{it}^* is not observed. For each $t = 1, 2$, the observed interval-censored dependent variable $y_{it} \in \{1, \cdots, J\}$ is linked to the latent variable y_{it}^* through *known* cutoff points $-\infty < c_1 < \cdots < c_{J-1} < \infty$ in the following way:

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^* < c_1, \\ 2 & \text{if } c_1 \leq y_{it}^* < c_2, \\ \vdots & \vdots \\ J & \text{if } c_{J-1} \leq y_{it}^*. \end{cases} \tag{3}$$

Consistent with the notation for the regressors, $y_i \equiv (y_{i1}, y_{i2})$ will denote the collection of outcome variables. Taken together, the model in equations (1)-(3) will be called the *interval-censored fixed-effects logit model* in the discussion below.

Conditional on the covariates X_i and the fixed effect α_i, the probability that the censored variable y_{it} assumes a particular value $j \in \{1, \ldots, J\}$ is

$$\Pr(y_{it} = j|\,X_i, \alpha_i) \ = \ \Lambda\left(\frac{1}{\sigma_0}(c_j - \alpha_i - X_{it}\beta_0)\right) - \Lambda\left(\frac{1}{\sigma_0}(c_{j-1} - \alpha_i - X_{it}\beta_0)\right),$$

where $c_0 = -\infty$ and $c_J = +\infty$.

To simplify matters, we introduce a collection of indicator variables to indicate whether the latent variable is above a given cutoff point. Let $\pi \equiv (\pi(1), \pi(2))$ index a pair of cutoff points, one for each time period, with $\pi(t) \in \{1, \ldots, J-1\}$. There are $(J-1)^2$ possible pairs. Then, define the following quantities:

$$d_{it\pi} \equiv 1\left\{y_{it}^* \geq c_{\pi(t)}\right\}, \; t = 1, 2$$

$$d_{i\pi} \equiv (d_{i1\pi}, d_{i2\pi}).$$

Note that the indicator variable $d_{it\pi}$ is determined as follows:

$$\begin{aligned} d_{it\pi} &= 1\left[\alpha_i + X_{it}\beta_0 - \sigma_0 u_{it} \geq c_{\pi(t)}\right] & (4) \\ &= 1\left[u_{it} \leq \frac{1}{\sigma_0}\left(\alpha_i + X_{it}\beta_0 - c_{\pi(t)}\right)\right] & (5) \end{aligned}$$

The number of outcomes above their respective cutoff values is denoted by

$$\bar{d}_{i\pi} \equiv d_{i1\pi} + d_{i2\pi}.$$

The following theorem formalizes that $\bar{d}_{i\pi}$ is a sufficient statistic for α_i in the transformed model.

Theorem 1. *If (y_i, X_i) follows the interval-censored fixed effects logit model, then for any π,*

$$\begin{aligned} p_{i\pi 2}(\beta_0, \sigma_0) &\equiv \Pr\left(d_{i,\pi} = (0,1) \middle| \bar{d}_{i\pi} = 1, X_i, \alpha_i\right) & (6) \\ &= \Lambda\left((X_{i2} - X_{i1})\frac{\beta_0}{\sigma_0} - \left(c_{\pi(2)} - c_{\pi(1)}\right)\frac{1}{\sigma_0}\right). & (7) \end{aligned}$$

Proof. See Appendix A.1. □

2.1 Identification and estimation

The parameters of the interval-censored fixed-effects model with logistic errors show up as regression coefficients in the conditional probability of the transformed binary choice model in (5). As such, the identification and estimation of β_0 and σ_0 can proceed using existing tools for the binary-choice fixed-effects model with logistic errors.

With $\Delta X_i \equiv X_{i2} - X_{i1}$, it is convenient to use the notation

$$\theta_0 \equiv \left(\frac{\beta_0}{\sigma_0}, \frac{1}{\sigma_0}\right)$$

$$W_{i\pi} \equiv \left(\Delta X_i, -\left(c_{\pi(2)} - c_{\pi(1)}\right)\right)$$

so that we can write

$$
\begin{aligned}
p_{i\pi 2}(\theta_0) &\equiv p_{i\pi 2}(\beta_0, \sigma_0) \\
&= \Lambda(Z_{i\pi}\theta_0)
\end{aligned}
$$

based on Theorem 1.

Before the main identification result can be stated, the following full-rank assumption is required:

Assumption 1. *$Var(\Delta X_i)$ is invertible.*

Assumption 1 implies that $E\left(\Delta X_i' \Delta X_i\right)$ is invertible, and guarantees that $E\left(W_{i\pi}' W_{i\pi}\right)$ is invertible whenever $\pi(2) \neq \pi(1)$. This assumption is standard for fixed-effects estimators, requiring some variation in the regressors over time (i.e., lack of perfect colinearity in the differenced regressors).

The main identification result is given by the following theorem:

Theorem 2. *If (y_i, X_i) follows the fixed effects interval-censored logit model with $J \geq 3$ and Assumption 1 holds, then (β_0, σ_0) is identified.*

Proof. See Appendix A.2. □

Turning to estimation, for a given choice of π, the conditional maximum likelihood estimator for the transformed binary choice model (Andersen, 1970; Chamberlain, 1980) is given by

$$
\begin{aligned}
\hat{\theta}_\pi &= \arg \max_{\mathbb{R}^{K+1} \times \mathbb{R}^{n_\pi}} \frac{1}{n} \sum_{i=1}^{n} 1\left\{\bar{d}_{i\pi} = 1\right\} \left(d_{i2\pi} \ln \Lambda(W_{i\pi}\theta) + (1 - d_{i2\pi}) \ln(1 - \Lambda(W_{i\pi}\theta))\right) \qquad (8) \\
&= \arg \max_{\mathbb{R}^{K+1} \times \mathbb{R}^{n_\pi}} \frac{1}{n} \sum_{i=1}^{n} 1\left\{\bar{d}_{i\pi} = 1\right\} l_{i\pi}(W_{i\pi}\theta),
\end{aligned}
$$

where $l_{i\pi}(W_{i\pi}\theta) \equiv d_{i2\pi} \ln \Lambda(W_{i\pi}\theta) + (1 - d_{i2\pi}) \ln(1 - \Lambda(W_{i\pi}\theta))$ and θ denotes a generic element of the parameter space. Well-definedness of the estimator in (8) requires that $\pi(2) \neq \pi(1)$.[2]

We propose to estimate (β_0, σ_0) using all possible transformations jointly, using a composite maximum likelihood estimator (CMLE) as in Baetschmann et al. (2015) and Muris (2017):

$$
\hat{\theta} = \arg \max_{\mathbb{R}^{K+1} \times \mathbb{R}^{n_\pi}} \frac{1}{n} \sum_{\pi} \sum_{i=1}^{n} 1\left\{\bar{d}_{i\pi} = 1\right\} l_{i\pi}(W_{i\pi}\theta). \qquad (9)
$$

Theorem 3. *Let $(\{y_i, X_i\}, i = 1, \cdots, n)$ be a random sample from the interval-censored fixed-effects logit model with $J \geq 3$ and with true parameter values $\theta_0 = \left(\frac{\beta_0}{\sigma_0}, \frac{1}{\sigma_0}\right)$, and let $\hat{\theta}$ be the estimator defined in (9). If Assumption 1 holds, then*

$$
\hat{\theta} \xrightarrow{p} \theta_0 \text{ as } n \to \infty;
$$

[2] If $\pi(1) = \pi(2)$, then $c_{\pi(2)} - c_{\pi(1)}$ and $W_{i\pi}\theta_0 = \Delta X_i \frac{\beta_0}{\sigma_0} + 0$, so that one can only identify β_0/σ_0.

and

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, H^{-1}\right) \quad as\ n \to \infty, \tag{10}$$

where

$$H = -E\left[\sum_{\pi} 1\left\{\bar{d}_{i\pi}=1\right\} \Lambda\left(W_{i\pi}\theta_0\right)\left(1-\Lambda\left(W_{i\pi}\theta_0\right)\right) W_{i\pi}' W_{i\pi}\right].$$

Proof. In Appendix A.3, we show that the limiting objective function is strictly concave, and derive the variance matrix and Hessian. In Appendix A.4, we obtain the stated result by verifying the conditions for consistency and asymptotic normality for maximizers of concave objective functions (Newey and McFadden, 1994, Theorems 2.7 and 3.1). □

The parameters of interest are (β_0, σ_0), and they can be estimated by partitioning

$$\hat{\theta} \equiv \left(\widehat{\left(\frac{\beta_0}{\sigma_0}\right)}, \widehat{\left(\frac{1}{\sigma_0}\right)}\right) \equiv \left(\hat{\theta}_1, \hat{\theta}_2\right)$$

and setting

$$\hat{\beta} = \hat{\theta}_1/\hat{\theta}_2,$$
$$\hat{\sigma} = 1/\hat{\theta}_2.$$

Consistency and asymptotic normality of the parameters of interest follows by the delta method and Theorem 3.

Corollary 1. *Under the conditions of Theorem 3,*

$$\sqrt{n}\begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\sigma}-\sigma_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, MH^{-1}M'\right)$$

as $n \to \infty$, where

$$M = \begin{bmatrix} \sigma_0 \iota_K & -\sigma_0\beta_0 \\ 0 & -\sigma_0^2 \end{bmatrix}$$

and ι_k is a $K \times 1$ vector of ones.

Proof. The result follows immediately from Theorem 3 and a delta method for vector-valued statistics, e.g. van der Vaart (1998, Theorem 3.1). □

2.2 Heteroskedasticity

To incorporate heteroskedasticity, the latent-variable model in (1) can be modified as follows:

$$y_{it}^* = \alpha_i + X_{it}\beta_0 - \sigma_0(Z_i)u_{it} \ (i = 1, \ldots, n;\ t = 1, 2), \tag{11}$$

with the error scaling function $\sigma_0(Z_i) > 0$ replacing the constant scaling parameter σ_0. We will call the model with the homoskedastic outcome equation (1) replaced by (11) the *heteroskedastic fixed effects interval-censored logit model.*

Theorem 2 can be extended in order to show identification of β_0 and the $\sigma_0(\cdot)$ function.

Theorem 4. *Assume that (y_i, X_i, Z_i) follows the heteroskedastic fixed effects interval-censored logit model with $J \geq 3$. If $Var\left(\Delta X_i | Z_i = z\right)$ is invertible, then $\left(\beta_0, \sigma_0\left(z\right)\right)$ is identified.*

Proof. The proof resembles that of Theorem 2, after conditioning on Z_i. It is provided in Appendix A.5. □

If $Var\left(\Delta X_i | Z_i = z\right)$ is invertible for all z in the support of Z, it follows that the heteroskedasticity function $\sigma()$ is nonparametrically identified within the model (11).

Remark 1. The assumption that $Var\left(\Delta X_i | Z_i = z\right)$ is invertible rules out $Z_i = X_i$, since conditioning on X_i leaves the random variable ΔX_i degenerate.

For estimation purposes, we recommend a parametric specification for the $\sigma()$ function. In particular, we will focus on the exponential specification given by

$$\sigma(Z_i) = \exp(Z_i \gamma_0).$$

The exponential specification is commonly used in practice (see, e.g., Wooldridge, 2010 (Chapter 16)). Under exponential heteroskedasticity, the true parameter vector θ_0 of interest is now (β_0, γ_0) rather than (β_0, σ_0). For a given choice π of the cutoff-value pair, the conditional probability $p_{i\pi 2}(\theta_0)$ is

$$p_{i\pi 2}(\theta_0) = \Pr\left(d_{i,\pi} = (0,1) | \bar{d}_{i,\pi} = 1, X_i, Z_i, \alpha_i\right) = \Lambda\left(\frac{\left(X_{i2} - X_{i1}\right)\beta_0 + \left(c_{\pi(2)} - c_{\pi(1)}\right)}{\exp(Z_i \gamma_0)}\right).$$

The composite-likelihood function can then be modified as follows:

$$\hat{\theta} = \arg \max_{\mathbb{R}^{\dim(\theta_0)} \times \mathbb{R}^{n_\pi}} \frac{1}{n} \sum_{\pi} \sum_{i=1}^{n} 1\left\{\bar{d}_{i\pi} = 1\right\} \left(d_{i2\pi} \ln p_{i\pi 2}(\theta) + \left(1 - d_{i2\pi}\right) \ln\left(1 - p_{i\pi 2}(\theta)\right)\right), \tag{12}$$

where θ denotes a generic element of the parameter space.

3 Semiparametric interval-censored fixed-effects model

In this section, we consider a semiparametric version of the interval-censored fixed-effects model, where the distribution of the error disturbance is left unspecified. Using the same notation as above, the latent-variable model is given by

$$y_{it}^* = \alpha_i + X_{it}\beta_0 - u_{it} \ (i = 1, \ldots, n;\ t = 1, 2), \tag{13}$$

where we assume only stationarity of u_{it} for each cross-sectional unit i:

$$u_{i1}|X_i, \alpha_i \ \text{ and } u_{i2}|X_i, \alpha_i \text{ are identically distributed.} \tag{14}$$

This assumption allows for heteroskedasticity across cross-sectional units and for serial correlation within cross-sectional units. This assumption is also used in Manski (1987) for the binary-choice fixed-effects model and Honoré (1992) for the censored fixed-effects model.

For a given choice π of the cutoff-value pair, the indicator variable $d_{it\pi}$ is still used to indicate whether or not y_{it}^* is greater than the cutoff value $c_{\pi(t)}$:

$$d_{it\pi} = 1\left[y_{it}^* \geq c_{\pi(t)}\right] = 1\left[u_{it} \leq \alpha_i + X_{it}\beta_0 - c_{\pi(t)}\right].$$

Since u_{i1} and u_{i2} have the same distribution, conditional on X_i and α_i, it follows immediately that

$$\left(X_{i2} - X_{i1}\right)\beta_0 - \left(c_{\pi(2)} - c_{\pi(1)}\right) \geq 0 \iff \Pr\left(d_{i2\pi} = 1|X_i, \alpha_i\right) \geq \Pr\left(d_{i1\pi} = 1|X_i, \alpha_i\right). \tag{15}$$

With $J \geq 3$ and the technical assumptions of Manski (1987), identification of β_0 can be established based upon the relationship in (15). A more precise formulation now follows. Denote by $F\left(u_{it}|\, X_i, \alpha_i\right)$ the conditional distribution of the error term in both periods. The following technical assumptions, analogous to those made in Manski (1987), are made:

Assumption 2. *The support of $F\left(\cdot|\, X_i, \alpha_i\right)$ is $\mathbb{R}$ for all (X_i, α_i).*

Assumption 3. *(i) The distribution of ΔX_i is such that at least one component of ΔX_i has positive Lebesgue density on $\mathbb{R}$ conditional on all the other components of ΔX_i with probability one. The corresponding component of β_0 is non-zero; (ii) The support of $(\Delta X_i, 1)$ is not contained in any proper linear subspace of $\mathbb{R}^{K+1}$.*

The main identification result then follows:

Theorem 5. *If (y_i, X_i) follows the fixed effects interval-censored model in (13)-(14) with $J \geq 3$, the distribution of (y_i, X_i) is identified, and Assumptions 2 and 3 hold, then β_0 is identified.*

Note that β_0 is identified *with scale* here, in contrast to the binary-choice model of Manski (1987). This result mirrors our identification of the variance σ in the logit case. Moreover, a composite maximum-score-type estimator can be constructed as follows

$$\hat{\beta} = \arg \max_{\mathbb{R}^K \times \mathbb{R}^{n_\pi}} \frac{1}{n} \sum_{\pi} \sum_{i=1}^{n} \left[\left(d_{i2\pi} - d_{i1\pi}\right) \cdot \operatorname{sgn}\left(\left(X_{i2} - X_{i1}\right)\beta - \left(c_{\pi(2)} - c_{\pi(1)}\right)\right)\right], \tag{16}$$

where $\text{sgn}(v) \equiv 1(v > 0) - 1(v < 0)$ and β is a generic element of $\mathbb{R}^{K\times 1}$. Again, with $J \geq 3$ and the technical assumptions of Manski (1987), it can be shown that the estimator $\hat{\beta}$ defined by (16) is a consistent estimator of β_0.

As with the binary-choice model, the tradeoff for the semiparametric approach involves less restrictive assumptions on the error disturbances in exchange for a slower rate of convergence. The convergence rate of the estimator defined by (16) is $\sqrt[3]{n}$ (Kim and Pollard, 1990), slower than the parametric $\sqrt{n}$ rate achieved by the composite conditional-likelihood estimator for the logit model.

4 Monte Carlo simulations

In this section, we present Monte Carlo results for the composite maximum likelihood estimator (CMLE) introduced in Section 2. We focus on the parametric (logit) version of our model, with homoskedastic errors (constant σ_0), two time periods ($t = 1, 2$), interval-censoring into three intervals ($J = 3$), and one covariate ($K = 1$). For the simulations, we consider the following data-generating process. First, regressors are generated from a standard normal distribution,

$$X_{it} \sim \mathcal{N}(0, 1) \text{ iid.}$$

Second, individual fixed effects are constructed as linear in the average regressor, plus a logistic error, centered at 65:

$$\alpha_i \sim LOG\left(65 + \frac{1}{2}(X_{i1} + X_{i2}), 1\right).$$

Third, the latent variable is constructed according to our parametric model with standard logistic errors,

$$y^*_{it} \sim LOG\left(\alpha_i + \beta_0 X_{it}, \sigma_0^2\right),$$

where β_0 and σ_0 are varied across designs. Finally, the interval-censored dependent variable is generated through

$$y_{it} = \begin{cases} 1 & \text{if } y^*_{it} < 60 \\ 2 & \text{if } 60 \leq y^*_{it} < 70 \\ 3 & \text{if } y^*_{it} \geq 70. \end{cases}$$

Note that y^*_{it} is symmetrically distributed around 65. Therefore, the design parameters (β_0, σ_0) influence not only the relationship between X_{it} and y_{it}, but also the amount of variation that is observed in the interval-censored dependent variable. Furthermore, we vary the number of observations n to be 250, 500, or 750.

Results are based on 1000 simulations for each design and are presented in Table 1. Both bias and RMSE generally decrease when n increases. There does not seem to be an effect on the relative efficiency of the CMLE versus the infeasible fixed-effects estimator that regresses $y^*_{i2} - y^*_{i1}$ on $X_{i2} - X_{i1}$. This can be seen from the column "Eff," which reports the ratio of the RMSE of the

n	β_0	σ_0	100xBias		100xRMSE		Eff
			$\hat{\beta}$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{\sigma}$	
250	1	5	0.33	-2.58	0.63	0.41	0.94
		10	-2.99	1.69	1.23	0.98	0.94
	2	5	0.11	-2.82	0.61	0.42	0.92
		10	0.34	-1.34	1.23	1.04	0.96
500	1	5	-0.41	-1.06	0.43	0.29	0.94
		10	-0.35	-2.20	0.85	0.70	0.96
	2	5	1.62	-2.13	0.44	0.29	0.92
		10	3.06	-0.53	0.89	0.73	0.93
750	1	5	0.28	-0.44	0.37	0.24	0.91
		10	-0.65	-2.49	0.68	0.59	0.95
	2	5	0.45	-1.06	0.36	0.24	0.92
		10	1.59	1.85	0.71	0.58	0.93

Table 1: Monte Carlo simulation results. Eff ("Efficiency") is the ratio of the root mean squared error of $\hat{\beta}$ relative to that of an infeasible fixed effects estimator that uses y^*_{it}.

infeasible first-difference estimator versus the RMSE of the CMLE. Note that the CMLE achieves a relative efficiency greater than 90% in all designs considered here.

The results from increasing the error term variance, from $\sigma_0 = 5$ to $\sigma_0 = 10$, are also clear from Table 1. While the larger value for σ_0 increases both bias and variance of $\hat{\beta}$, it decreases the bias for $\hat{\sigma}$ and increases its RMSE. The effect of changing the slope parameter ($\beta_0 = 1$ versus $\beta_0 = 2$) are hard to infer from Table 1. To get a clearer picture of the effect of changing the slope parameter, Figure 1 presents results for $n = 1000$, $\sigma_0 = 5$, and a range of slope-parameter values ($\beta_0 \in \{0, 1, 2, \cdots, 10\}$). We plot the efficiency of the CMLE relative to the infeasible first-difference estimator as a function of β_0. Large values of β_0 lead to very few observations in the middle category ($y_{it} = 2$), leading to an obvious reduction in the relative efficiency of the CMLE. If the slope parameter were to grow even larger, the variation in y_{it} would eventually be reduced to that of a binary-choice model with outcomes $\{1, 3\}$.

4.1 Heteroskedasticity

In the presence of heteroskedasticity, the CMLE is no longer guaranteed to estimate β_0 consistently. In Section 2.2, we proposed an estimator that is consistent under the assumption of exponential heteroskedasticity, call it CMLE-HET. We use the same DGP as above, except that

$$\sigma(X_i) = \exp\left(\log(2) + \gamma_1 (X_{i1} + X_{i2})\right).$$

We consider four designs, $\gamma_1 \in \{-2, 0, 1, 2\}$, with $n = 1000$. The design $\gamma_1 = 0$ corresponds to the homoskedastic case. Results are presented in Table 2. We conclude that: (1) CMLE is bias in the presence of heteroskedasticity, and the bias seems to increase in the amount of heteroskedasticity; (2) CMLE-HET shows virtually no bias, even with strong heteroskedasticity; (3) CMLE-HET pays

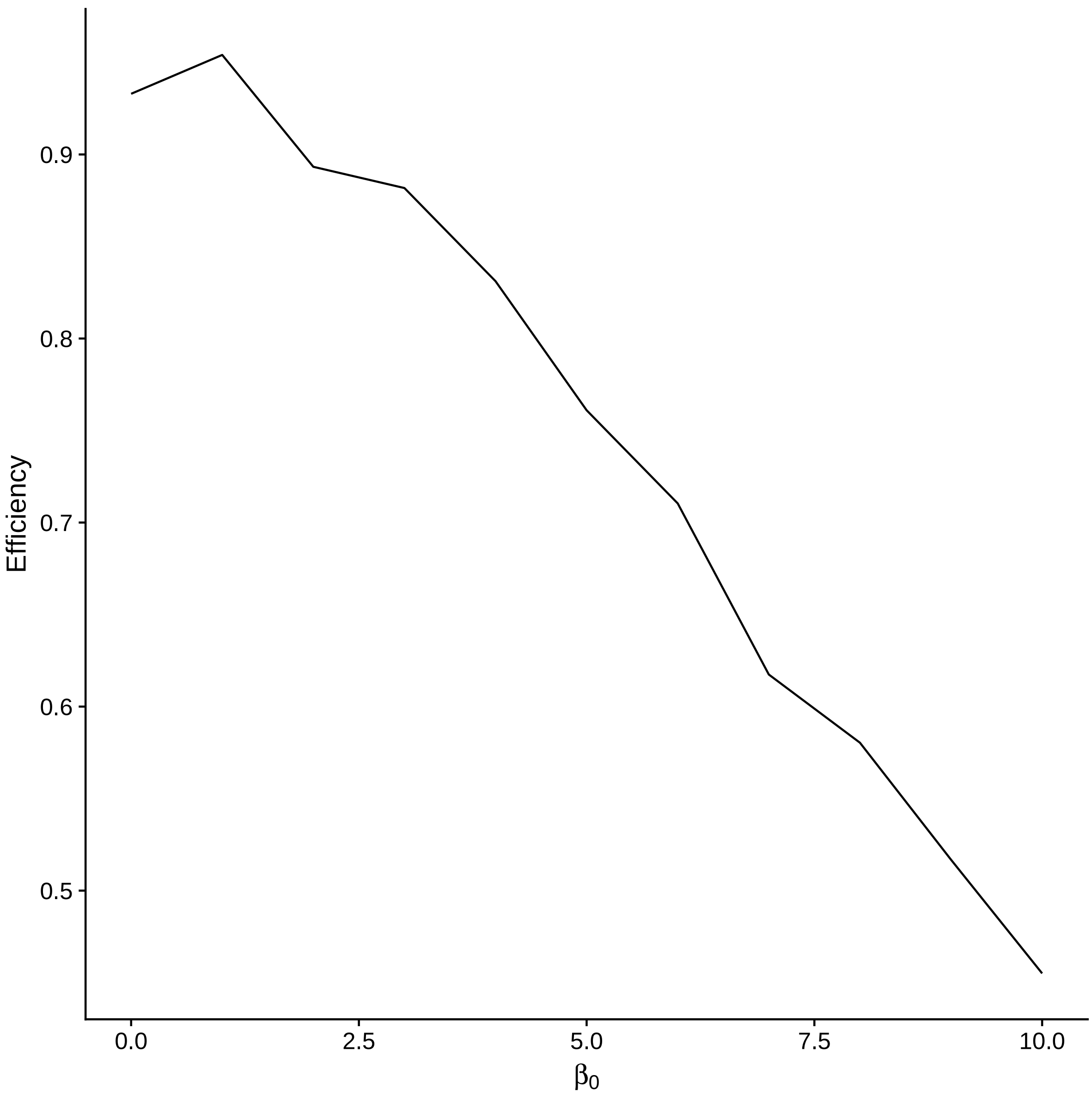

Figure 1: Efficiency of $\hat{\beta}$ as a function of β_0.

γ_1	CMLE		CMLE-HET	
	Bias	SD	Bias	SD
-2	0.17	0.35	0.02	0.37
0	0.00	0.16	0.00	0.16
1	0.04	0.27	0.03	0.28
2	0.17	0.33	0.02	0.37

Table 2: Monte Carlo simulation results in the presence of heteroskedasticity: estimation of β. Different rows in the table correspond to results for different values of $\gamma_1 \in \{-2, 0, 1, 2\}$.

γ_1	γ_0		γ_1	
	Bias	SD	Bias	SD
-2	0.03	0.16	0.12	0.35
0	0.08	0.06	0.00	0.06
1	0.07	0.12	0.03	0.15
2	0.04	0.16	0.10	0.34

Table 3: Monte Carlo simulation results in the presence of heteroskedasticity: γ_0,γ_1. Different rows in the table correspond to results for different values of $\gamma_1 \in \{-2, 0, 1, 2\}$.

a small price in terms of its standard deviation. CMLE-HET estimates the skedasticity parameters well, as can be seen from the low bias and standard deviations in Table 3.

5 Empirical application

In this section, we consider an empirical application of the parametric (CMLE) estimator introduced in Section 2.[3] Our goal is to compare the estimator's performance to the baseline of complete observability and the linear fixed-effects estimator. As such, we consider an application in which the continuous outcome variable is fully observed in the original data, and then we artificially interval-censor the outcome variable. Specifically, we consider the birth-outcome panel dataset from Abrevaya (2006), where the sample that we use has $n = 78330$ mothers and two children ($t = 1, 2$) for each mother in the data.[4] The outcome of interest is child birthweight (measured in grams). The histogram of birthweights is shown in Figure 2.

We consider two artificially interval-censored versions of the birthweight outcome variable, one with four intervals ($J = 4$ with $c_1 = 2500$, $c_2 = 3500$, $c_3 = 4500$) and one with six intervals ($J = 6$ with $c_1 = 2500$, $c_2 = 3000$, $c_3 = 3500$, $c_4 = 4000$, $c_5 = 4500$). The regressors that we consider are as follows: second-birth indicator ($1(t = 2)$), year of birth ($year$), mother's age (age), mother's age squared (age^2), indicator of smoking during pregnancy ($smoke$), indicator of a male birth ($male$), and categorical indicators for prenatal care ($novisit$ for no prenatal care visit, $pretri2$ for first visit

[3]Data and code for the empirical application are available from https://github.com/chrismuris/IntervalFixedEffects.

[4]The sample is constructed as a subsample from the "Panel #3" dataset in Abrevaya (2006). Black mothers are dropped. Non-black mothers are retained in the sample if their first two births are observed. Any additional births are dropped from the sample so that we have a balanced (two-period) panel.

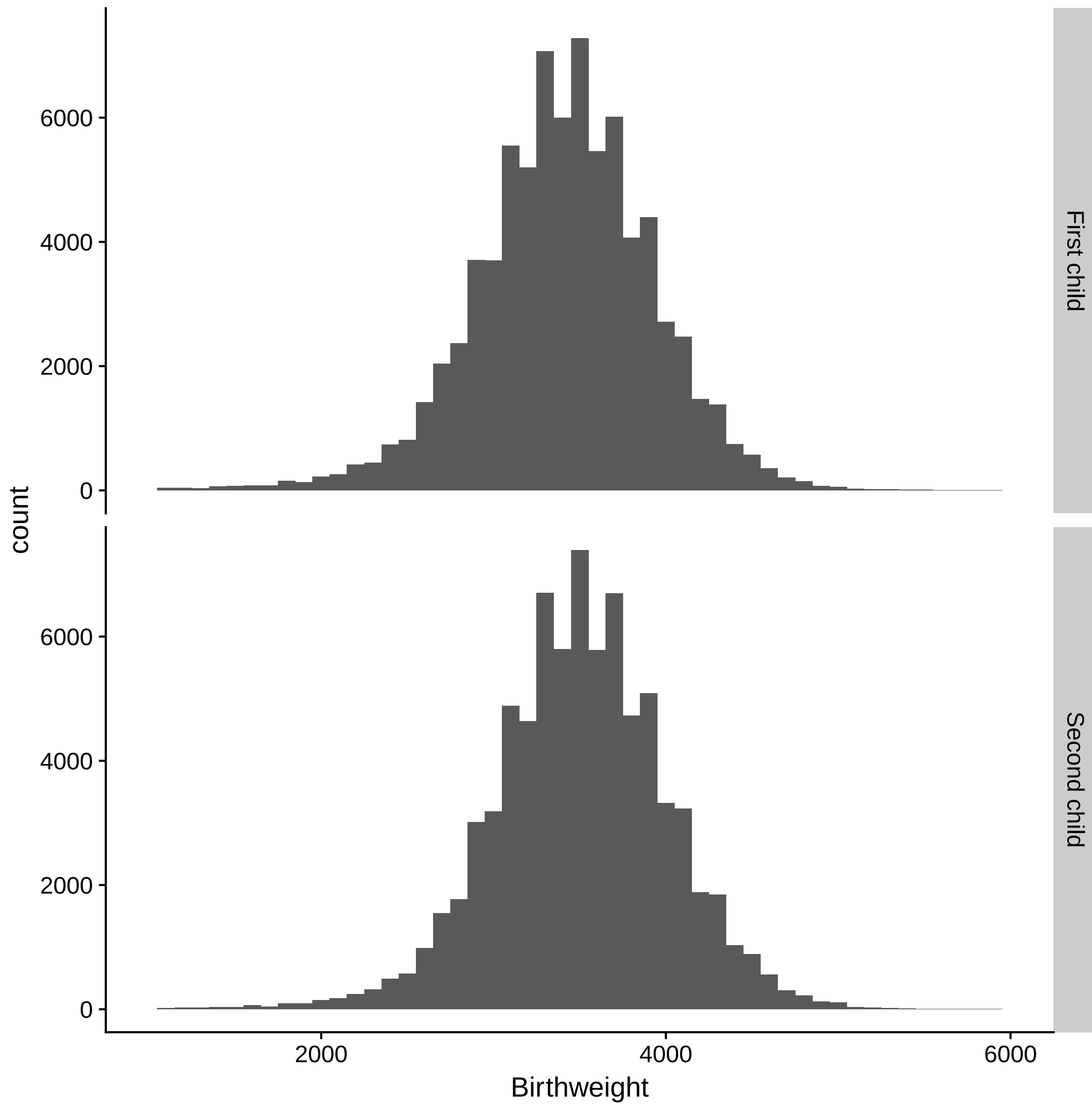

Figure 2: Histogram of birthweights (in grams)

	Linear FE		6 intervals		4 intervals	
1{t=2}	98.74	(4.77)	98.52	(5.72)	100.70	(6.21)
year	-7.77	(3.57)	-6.70	(4.20)	-6.45	(4.77)
age	-23.51	(5.54)	-26.76	(6.52)	-27.80	(7.75)
age^2	0.54	(0.08)	0.57	(0.09)	0.59	(0.11)
smoke	-129.93	(6.83)	-137.73	(8.64)	-130.84	(10.08)
male	134.97	(2.92)	130.41	(3.30)	136.43	(3.97)
novisit	-205.40	(28.47)	-180.53	(35.54)	-188.80	(41.32)
pretri2	-6.64	(5.80)	1.16	(7.14)	-4.81	(7.96)
pretri3	-21.00	(13.99)	-30.86	(18.04)	-35.40	(18.91)
log(σ)	6.01		5.49		5.31	

Table 4: Estimation results. Standard errors in parentheses. Standard errors for the CMLE are based on 1000 bootstrap replications.

in the second trimester, *pretri*3 for first visit in the third trimester; a first-trimester visit is the omitted category).

Table 4 shows the estimation results for the linear fixed-effects estimator (which is just the first-difference estimator here) and the CMLE estimators for the two artificially interval-censored samples (using $J = 4$ and $J = 6$, as specified above). For the CMLE estimators, homoskedasticity is assumed. The CMLE standard errors were estimated using the bootstrap, with 1000 replications. The estimated slope parameters are fairly comparable for the three estimators, at least within ranges that would be considered reasonable given their respective standard errors. In terms of efficiency, the standard errors for the CMLE estimators are up to 50% higher than the corresponding standard errors for the first-difference estimator. When six intervals are used, the standard errors are much closer to those of the first difference estimator.

We repeat the results for a heteroskedastic model with

$$\sigma\left(Z_i\right) = \exp\left\{\gamma_0 + \gamma_1 \overline{age_i} + \gamma_2 \overline{smoke_i} + \gamma_3 \overline{male_i}\right\},$$

where a bar denotes the average across the two children of that variable. For the first difference estimator, we obtain estimates for the parameters in this function based on linear regression of $\log\left(e_i^2\right)/2$ on the variables in the heteroskedasticity function. For the interval-censored estimator, we use the procedure outlined in Section 2.2. We present the results in Table 5. The CMLE results for the slope estimates are similar to those found for the homoskedastic case. The standard errors for the heteroskedastic CMLE are slightly higher than those for the homoskedastic CMLE, as would be expected given the joint estimation of the heteroskedasticity-function parameters. For the heteroskedasticity function, our estimates are also close to those from the first-difference estimator, with the exception of the coefficient on $\overline{smoke_i}$, which is off by more than a standard error.

	Linear FE		6 intervals		4 intervals	
Regression function						
1{t=2}	98.74	(4.77)	104.25	(6.56)	102.16	(6.45)
year	-7.77	(3.57)	-7.73	(4.92)	-6.65	(4.97)
age	-23.51	(5.54)	-28.61	(7.52)	-26.04	(7.70)
age^2	0.54	(0.08)	0.63	(0.10)	0.55	(0.11)
smoke	-129.93	(6.83)	-131.17	(8.89)	-131.77	(9.93)
male	134.97	(2.92)	120.75	(3.82)	136.21	(3.88)
novisit	-205.40	(28.47)	-198.84	(39.66)	-188.37	(41.62)
pretri2	-6.64	(5.80)	-4.60	(7.81)	-4.79	(7.54)
pretri3	-21.00	(13.99)	-34.22	(17.41)	-35.19	(19.68)
Heteroskedasticity function						
constant	5.7118	(0.0242)	5.5284	(0.0303)	5.4905	(0.0318)
age	-0.0060	(0.0008)	-0.0058	(0.0010)	-0.0050	(0.0011)
smoke	0.0793	(0.0117)	0.1085	(0.0150)	0.1030	(0.0154)
male	0.0807	(0.0156)	0.0357	(0.0174)	0.0495	(0.0189)

Table 5: Estimation results for heteroskedastic model. Standard errors in parentheses. Standard errors for the CMLE are based on 1000 bootstrap replications.

Overall, we conclude that the CMLE provides estimates similar to those from the first-difference estimator, using a relatively coarse version of the birthweight variable as compared to the fully-observed data.

6 Extensions and conclusion

This paper has considered estimation of the interval-censored fixed-effects regression model. The parametric (logit) and semiparametric models considered both allow for heteroskedastic error disturbances, with the semiparametric model also allowing for serially correlation in the error disturbances. A trivial generalization of our model would allow the interval endpoints to vary over time and/or individuals. Time-varying interval endpoints are a feature of some datasets, like the MEPS. And, while we have considered the two-period case for simplicity of exposition, the proposed estimators can easily be generalized to more than two periods, as in Muris (2017) for the logit model and Abrevaya (2000) for the semiparametric model.

Future research could proceed in various directions. The use of appropriate weights in the composite likelihood estimator could increase its efficiency relative to the proposed unweighted estimator. It would also be interesting to consider estimation of β and $\sigma()$ within the logit model for a fully nonparametric specification for $\sigma()$.

Finally, while we focus on estimation of the slope parameters (and heteroskedasticity function) in this paper, it is worth mentioning that the ability to estimate partial effects will depend upon the form of censoring in the data. A well-known drawback associated with estimation of non-linear fixed-effects models is the difficulty in estimating partial effects, even after consistently estimating slope parameters. For instance, in the binary-choice fixed-effects model, the partial effects are themselves functions of the fixed effect; since the fixed effects can not be estimated consistently, the partial effects can not be estimated consistently. Such a problem does not arise in the linear fixed-effects model since the partial effect is not a function of the fixed effect. Likewise, in the censored fixed-effects model of Honoré (1992) or the fixed-effects model considered in this paper, the partial effects of interest do not depend upon the fixed effect when the censoring arises purely due to data-coding issues. Censoring, however, may arise due to a corner-solution outcome (e.g., health-care expenditures of \$0 in MEPS), in which case the partial effects of interest would depend upon the fixed effect. Of course, even for fixed-effects models of corner-solution outcomes, the *relative* partial effects of different variables upon the outcome can be consistently estimated as in the binary-choice model and other models. Wooldridge (2010) provides an excellent discussion of the difference between data-coding and corner-solution situations.

A Appendix: Proofs and derivations

A.1 Proof of Theorem 1 (Sufficiency)

Proof. Consider a pair of cutoff-value indices given by $\pi = (\pi(1), \pi(2))$. Recall the transformed variables $d_{i1\pi} \equiv 1\{y_{i1}^* \geq c_{\pi(1)}\}$ and $d_{i2\pi} \equiv 1\{y_{i2}^* \geq c_{\pi(2)}\}$. The (conditional) probability of these indicator variables being equal to one are given by

$$\begin{aligned}
P\left(d_{i1\pi} = 1 \middle| X_i, \alpha_i\right) &= P\left(y_{i1}^* \geq c_{\pi(1)} \middle| X_i, \alpha_i\right) \\
&= P\left(\alpha_i + X_{i1}\beta_0 - \sigma_0 u_{i1} \geq c_{\pi(1)} \middle| X_i, \alpha_i\right) \\
&= P\left(u_{i1} \leq \frac{1}{\sigma_0}\left(\alpha_i + X_{i1}\beta_0 - c_{\pi(1)}\right) \middle| X_i, \alpha_i\right) \\
&= \Lambda\left(\frac{1}{\sigma_0}\left(\alpha_i + X_{i1}\beta_0 - c_{\pi(1)}\right)\right)
\end{aligned}$$

and, similarly,

$$P\left(d_{i2\pi} = 1 \middle| X_i, \alpha_i\right) = \Lambda\left(\frac{1}{\sigma_0}\left(\alpha_i + X_{i2}\beta_0 - c_{\pi(2)}\right)\right). \tag{17}$$

Due to the (conditional) serial independence of u_{i1} and u_{i2}, we have

$$
\begin{aligned}
p_{1\pi}(X_i,\alpha_i) &\equiv P\left(d_{i\pi}=(1,0)|\, X_i,\alpha_i\right)\\
&= \Lambda\left(\frac{1}{\sigma_0}\left(\alpha_i + X_{i1}\beta_0 - c_{\pi(1)}\right)\right)\left[1-\Lambda\left(\frac{1}{\sigma_0}\left(\alpha_i + X_{i2}\beta_0 - c_{\pi(2)}\right)\right)\right],\\
p_{2\pi}(X_i,\alpha_i) &\equiv P\left(d_{i\pi}=(0,1)|\, X_i,\alpha_i\right)\\
&= \left[1-\Lambda\left(\frac{1}{\sigma_0}\left(\alpha_i + X_{i1}\beta_0 - c_{\pi(1)}\right)\right)\right]\Lambda\left(\frac{1}{\sigma_0}\left(\alpha_i + X_{i2}\beta_0 - c_{\pi(2)}\right)\right).
\end{aligned}
$$

We use the sufficient statistic for the binary choice fixed-effects logit model (Andersen, 1970; Chamberlain, 1980) to obtain a conditional probability that is free of α_i. With $\bar{d}_{i\pi} \equiv d_{i1\pi} + d_{i2\pi}$ and $\Delta X_i \equiv X_{i2} - X_{i1}$, we obtain the final result:

$$
\begin{aligned}
p_{i\pi 1}\left(\beta_0,\sigma_0\right) &\equiv P\left(d_{i\pi}=(1,0)|\, \bar{d}_{i\pi}=1, X_i,\alpha_i\right)\\
&= 1/(1+p_{0\pi}(X_i,\alpha_i)/p_{1\pi}(X_i,\alpha_i))\\
&= \frac{1}{1+\exp\left\{\frac{1}{\sigma_0}\left(\Delta X_i\beta_0 - \left(c_{\pi(2)}-c_{\pi(1)}\right)\right)\right\}}\\
&= 1-\Lambda\left(\frac{1}{\sigma_0}\left(\Delta X_i\beta_0 - \left(c_{\pi(2)}-c_{\pi(1)}\right)\right)\right),\\
p_{i\pi 2}\left(\beta_0,\sigma_0\right) &\equiv P\left(d_{i\pi}=(0,1)|\, \bar{d}_{i\pi}=1, X_i,\alpha_i\right)\\
&= \Lambda\left(\frac{1}{\sigma_0}\left(\Delta X_i\beta_0 - \left(c_{\pi(2)}-c_{\pi(1)}\right)\right)\right).
\end{aligned}
$$

□

A.2 Proof of Theorem 2 (Identification)

Proof. Because (y_i, X_i) follows the fixed effects interval censored regression model, Theorem 1 holds:

$$
p_{i\pi 2}(\beta_0,\sigma_0) = \Lambda\left(\frac{1}{\sigma_0}\left(\Delta X_i\beta_0 - \left(c_{\pi(2)}-c_{\pi(1)}\right)\right)\right).
$$

For the purpose of this proof, change the notation to:

$$
p_{\pi}\left(\Delta X_i\right) \equiv p_{i\pi 2}\left(\beta_0,\sigma_0\right),
$$

which emphasizes the dependence of the conditional probability on the regressors, and the fact that $p_{\pi}\left(\Delta X_i\right)$ is identified from the distribution of (y_i, X_i).

Pick a cutoff-value pair π such that $\pi\left(2\right) = \pi\left(1\right)$, which implies $c_{\pi(2)} = c_{\pi(1)}$ and $p\left(\Delta X_i\right) = \Lambda\left(\Delta X_i \frac{\beta_0}{\sigma_0}\right)$. Thus, for any ΔX_i, we have $\Lambda^{-1}\left(p\left(\Delta X_i\right)\right) = \Delta X_i \frac{\beta_0}{\sigma_0}$ and $\Delta X_i' \Lambda^{-1}\left(p\left(\Delta X_i\right)\right) = \Delta X_i' \Delta X_i \frac{\beta_0}{\sigma_0}$. Taking expectations yields $E\left[\Delta X_i' \Lambda^{-1}\left(p\left(\Delta X_i\right)\right)\right] = E\left(\Delta X_i' \Delta X_i\right)\frac{\beta_0}{\sigma_0}$. Since Assumption 1 holds, we have

$$\frac{\beta_0}{\sigma_0} = \left[E(\Delta X_i' \Delta X_i)\right]^{-1} E\left[\Delta X_i \Lambda^{-1} p(\Delta X_i)\right],$$

similar to the argument in Lewbel (2017, p. 37). Therefore, β_0/σ_0 is identified since all RHS quantities are identified from the data.

Pick a different cutoff-value pair π' such that $c_{\pi'(2)} - c_{\pi'(1)} \neq 0$, which is always possible for $J \geq 3$. Using a similar derivation to that above, we obtain

$$\frac{1}{\sigma_0} = \frac{1}{c_{\pi(2)} - c_{\pi(1)}} \left(\Delta X_i \frac{\beta_0}{\sigma_0} - \Lambda^{-1}\left(p(\Delta X_i)\right) \right)$$

for any ΔX_i. This relationship identifies σ_0 because all quantities on the RHS are either known (interval endpoints), identified from the data (functions of ΔX_i), or previously identified (β_0/σ_0).

Note that instead of the two-step proof above, identification of $(\beta_0/\sigma_0, 1/\sigma_0)$ can be obtained directly with $W_{i\pi}$ instead of ΔX_i for $\pi(1) \neq \pi(2)$. However, the proof above provides insight for the generalization of the identification results to the heteroskedastic case. □

A.3 Concavity, score, and Hessian calculations

Recall that

$$l_{i\pi}\left(Z_{i\pi}\theta\right) = d_{i\pi 2} \ln \Lambda\left(W_{i\pi}\theta\right) + \left(1 - d_{i\pi 2}\right) \ln\left(1 - \Lambda\left(W_{i\pi}\theta\right)\right),$$

and that the sample objective function is

$$Q_n(\theta) = \frac{1}{n} \sum_i \sum_\pi 1\left\{\bar{d}_{i\pi} = 1\right\} l_{i\pi}\left(W_{i\pi}\theta\right). \tag{18}$$

Therefore, the population (limiting) objective function is

$$\begin{aligned} Q_0(\theta) &= \text{plim } Q_n(\theta) \\ &= E\left[\sum_\pi 1\left\{\bar{d}_{i\pi} = 1\right\} l_{i\pi}\left(W_{i\pi}\theta\right)\right]. \end{aligned} \tag{19}$$

In this section, we establish concavity of the sample objective function, and derive the Jacobian and Hessian for the population objective function.

Concavity. The likelihood contribution for a given transformation and a given i is concave in the index $W_{i\pi}\theta$:

$$\begin{aligned} \frac{\partial l_{i\pi}(v)}{\partial v} &= d_{i\pi 2} \frac{\Lambda(v)\left(1 - \Lambda(v)\right)}{\Lambda(v)} - \left(1 - d_{i\pi 2}\right) \frac{\Lambda(v)\left(1 - \Lambda(v)\right)}{1 - \Lambda(v)} \\ &= d_{i\pi 2}\left(1 - \Lambda(v)\right) - \left(1 - d_{i\pi 2}\right)\Lambda(v) \\ &= d_{i\pi 2} - \Lambda(v). \end{aligned} \tag{20}$$

It follows that the second derivative is

$$\frac{\partial^2 l_{i\pi}(v)}{(\partial v)^2} = -\Lambda(v)(1-\Lambda(v)) < 0, \tag{21}$$

i.e. $l_{i\pi}$ is strictly concave in $W_{i\pi}\theta$. Therefore, it is concave in θ for any π, and strictly concave in θ when $\pi(2) \neq \pi(1)$ and $\bar{d}_{i\pi} = 1$. Because the sum of concave and strictly concave functions is strictly concave, and because the logit model implies that $E\left[1\left\{\bar{d}_{i\pi} = 1\right\}\right] > 0$ for each π, the sample objective function in (9) is strictly concave for large enough n. Therefore, Q_o is also concave (see e.g. Theorem 2.7 in Newey and McFadden, 1994).

Hessian. Using the expression in (21) and the chain rule, we obtain

$$\begin{aligned} \left.\frac{\partial^2 E\left[\sum_\pi 1\left\{\bar{d}_{i\pi}=1\right\} l_{i\pi}(W_{i\pi}\theta)\right]}{\partial\theta\partial\theta'}\right|_{\theta=\theta_0} &= -E\left[\sum_\pi 1\left\{\bar{d}_{i\pi}=1\right\}\Lambda(W_{i\pi}\theta_0)(1-\Lambda(W_{i\pi}\theta_0))W_{i\pi}'W_{i\pi}\right] \\ &\equiv H. \end{aligned}$$

Jacobian. Using the expression in (20) we obtain for the variance of the score from $l_{i\pi}$:

$$E\left((d_{i\pi 2} - \Lambda(W_{i\pi}\theta_0))^2 W_{i\pi}'W_{i\pi}\right)$$

so that

$$\Sigma = E\left[\sum_\pi 1\left\{\bar{d}_{i\pi}=1\right\}(d_{i\pi 2}-\Lambda(W_{i\pi}\theta_0))^2 W_{i\pi}'W_{i\pi}\right].$$

Finally, note that

$$\begin{aligned} E\left[(d_{i\pi 2}-\Lambda(W_{i\pi}\theta_0))^2\middle| W_{i\pi}, \bar{d}_{i\pi}=1\right] &= Var\left(d_{i\pi 2}\middle| W_{i\pi}, \bar{d}_{i\pi}=1\right) \\ &= \Lambda(W_{i\pi}\theta_0)(1-\Lambda(W_{i\pi}\theta_0)) \end{aligned}$$

where the first step follows because $E\left[d_{i\pi 2}\middle| W_{i\pi}, \bar{d}_{i\pi}=1\right] = \Lambda(W_{i\pi}\theta_0)$, see (17); and the second step follows because the variance of a Bernoulli random variable with probability p is $p(1-p)$. It therefore follows that $H = -\Sigma$.

A.4 Proof of Theorem 3

Proof. The CMLE is an extremum estimator with sample criterion function

$$Q_n(\theta) = \frac{1}{n}\sum_i\sum_\pi 1\left\{\bar{d}_{i\pi}=1\right\}(d_{i\pi 2}\ln\Lambda(W_{i\pi}\theta) + (1-d_{i\pi 2})\ln(1-\Lambda(W_{i\pi}\theta)))$$

It follows from the derivation in Appendix A.3 that Q_n is concave.

Consistency. Here, we verify conditions (i) (identification) and (iii) (pointwise convergence) for Theorem 2.7 in Newey and McFadden (1994), which is a consistency theorem for extremum estimators with concave sample criterion functions. A mean-value expansion of $l_{i\pi}$ around $\theta = 0$ yields

$$l_{i\pi}\left(W_{i\pi}\theta\right) - l_{i\pi}\left(0\right) = \left(d_{i\pi 2} - \Lambda\left(W_{i\pi}\tilde{\theta}\right)\right) W_{i\pi}\theta,$$

using the score in (20) in Section A.3. Boundedness of the RHS follows from Jensen's inequality and Assumption 1, which guarantees that $E\left(|W_{i\pi}|\right) < \infty$. A law of large numbers then implies pointwise convergence of Q_n to Q_0. Therefore, condition (iii) in Newey and McFadden's Theorem 2.7 is satisfied.

Identification was established in Theorem 2. To see that identification is not lost by using the objective function Q_0, note that we established in Section A.3 that the objective function is *strictly* concave. Therefore, condition (i) in Theorem 2.7 is satisfied.

Asymptotic normality. This proof proceeds by verifying the conditions in Theorem 3.1 of Newey and McFadden (1994). Consistency was established above. With concavity of the objective function, there is no need to compactify the parameter space, so that condition (i) is trivially satisfied. Twice continuous differentiability, condition (ii), is easily seen to be satisfied; see the results in Section A.3. Condition (iii) requires that a central limit theorem applies to the score. In Section A.3, we showed that the variance of the score is

$$\Sigma = E\left[\sum_{\pi} 1\left\{\bar{d}_{i\pi} = 1\right\}\left(d_{i\pi 2} - \Lambda\left(W_{i\pi}\theta_0\right)\right)^2 W_{i\pi}' W_{i\pi}\right].$$

Note that Σ is bounded because $1\left\{\bar{d}_{i\pi} = 1\right\} \in [0,1]$, $\left(d_{i\pi 2} - \Lambda\left(W_{i\pi}\theta_0\right)\right)^2 \in [0,1]$, and the second moment of $W_{i\pi}$ is bounded (Assumption 1). Because the score is mean-zero and sampling is random, a Lindeberg-Levy CLT yields condition (iii). Conditions (iv) and (v) are related to the Hessian H derived in Section A.3. There, we established strict concavity so that condition (v) is satisfied (invertibility of H). It is also clear that the second derivative is continuous. What remains is to establish uniform convergence of the estimated Hessian in a neighbourhood around the true value of the parameters. Note that

$$H(\theta) = -E\left[\sum_{\pi} 1\left\{\bar{d}_{i\pi} = 1\right\}\Lambda\left(W_{i\pi}\theta\right)\left(1 - \Lambda\left(W_{i\pi}\theta\right)\right) W_{i\pi}' W_{i\pi}\right],$$

so that $1\left\{\bar{d}_{i\pi} = 1\right\} \in [0,1]$, $\Lambda\left(W_{i\pi}\theta\right)\left(1 - \Lambda\left(W_{i\pi}\theta\right)\right) \in [0, 0.25]$, and Assumption 1 imply that the second moments of $W_{i\pi}$ are bounded. $\|\sum_{\pi} 1\left\{\bar{d}_{i\pi} = 1\right\}\Lambda\left(W_{i\pi}\theta\right)\left(1 - \Lambda\left(W_{i\pi}\theta\right)\right) W_{i\pi}' W_{i\pi}\| \leq \|W_{i\pi}' W_{i\pi}\|$, and $E\|W_{i\pi}' W_{i\pi}\| < \infty$. Therefore, a uniform law of large numbers applies on a suitably chosen neighbourhood of θ_0, see e.g. Lemma 2.4 in Newey and McFadden (1994). □

A.5 Proof of Theorem 4 (Identification under heteroskedasticity)

Proof. This proof differs from that in Section A.2 by conditioning on Z_i. Since (y_i, X_i, Z_i) follows the heteroskedastic fixed effects interval censored regression model, a modification of Theorem 1 says that:

$$p_{i\pi 2}(\beta_0, \sigma_0) = \Lambda\left(\frac{1}{\sigma_0\left(Z_i\right)}\left(\Delta X_i\beta_0 - \left(c_{\pi(2)} - c_{\pi(1)}\right)\right)\right).$$

For the purpose of this proof, change the notation to:

$$p_\pi\left(\Delta X_i, Z_i\right) \equiv p_{i\pi 2}\left(\beta_0, \sigma_0\right),$$

which emphasizes the dependence of the conditional probability on the regressors, and the fact that $p_\pi\left(\Delta X_i, Z_i\right)$ is identified from the distribution of (y_i, X_i, Z_i).

Pick a cutoff-value pair π such that $\pi\left(2\right) = \pi\left(1\right)$, which implies $c_{\pi(2)} = c_{\pi(1)}$ and a value z that satisfies the condition in the theorem that $E(\Delta X'\Delta X|\, Z = z)$ is invertible. Then $p_\pi\left(\Delta X_i, z\right) = \Lambda\left(\Delta X_i \frac{\beta_0}{\sigma_0(z)}\right)$. Following the same arguments as in Appendix A.2, we have Assumption 1 leading to

$$\frac{\beta_0}{\sigma_0\left(z\right)} = \left[E(\Delta X_i'\Delta X_i|\, Z_i = z)\right]^{-1} E\left[\Delta X_i \Lambda^{-1} p(\Delta X_i)\middle|\, Z_i = z\right],$$

Therefore, $\beta_0/\sigma_0\left(z\right)$ is identified since all RHS quantities are identified from the data and the inverses are well-defined.

Pick a different cutoff-value pair π' such that $c_{\pi'(2)} - c_{\pi'(1)} \neq 0$, which is always possible for $J \geq 3$. Similar to the proof in Appendix A.2, we have

$$\frac{1}{\sigma_0\left(z\right)} = \frac{1}{c_{\pi(2)} - c_{\pi(1)}} E\left(\Delta X_i \frac{\beta_0}{\sigma_0\left(z\right)} - \Lambda^{-1}\left(p(\Delta X_i)\right)\middle|\, Z_i = z\right).$$

This identifies $\sigma_0\left(z\right)$ because all quantities on the RHS are either known (interval endpoints), identified from the data (functions of ΔX_i), or previously identified ($\beta_0/\sigma_0\left(z\right)$). □

A.6 Proof of Theorem 5 (Identification, semiparametric case)

Proof. We first remark that, under the assumptions in the model.

$$\left(X_{i2} - X_{i1}\right)\beta_0 - \left(c_{\pi(2)} - c_{\pi(1)}\right) \geq 0 \iff \Pr\left(d_{i2\pi} = 1|X_i, \alpha_i\right) \geq \Pr\left(d_{i1\pi} = 1|X_i, \alpha_i\right).$$

Rewrite this as

$$W_i\theta_0 \geq 0 \iff \Pr\left(d_{i2\pi} = 1|X_i, \alpha_i\right) \geq \Pr\left(d_{i1\pi} = 1|X_i, \alpha_i\right),$$

with $\theta_0 \equiv (\beta_0, 1)$ and $W_i \equiv \left(X_{i2} - X_{i1}, -\left(c_{\pi(2)} - c_{\pi(1)}\right)\right)$.

Then, for a given choice of cut points such that $c_{\pi(2)} - c_{\pi(1)} \neq 0$, this model maps directly into the model of Manski (1987). To see this correspondence, here are the relevant model components, with those from Manski (1987) on the LHS and those from our model on the RHS:

$$\begin{aligned} y_t &= d_{it\pi}, \\ x_t &= [X_{it}, -c_{\pi(t)}], \\ c &= \alpha_i, \\ u_t &= u_{it}. \end{aligned}$$

Then, equation (14) and Assumption 2 imply Assumption 1 of Manski (1987); Assumption 3 and the choice of cutoffs so that $c_{\pi(2)} - c_{\pi(1)} \neq 0$ imply Assumption 2 of Manski (1987). Lemma 2 of Manski (1987) therefore applies to our case and guarantees identification of $\theta_0 = (\beta_0, 1)$ up to scale. With the scale normalization that is implied by our model (i.e. the last element of θ_0 equals 1), the result is obtained. □

References

[1] Abrevaya, J. (2000), "Rank Estimation of a Generalized Fixed-Effects Regression Model," *Journal of Econometrics*, 95(1), 1-23.

[2] Abrevaya, J. (2006), "Estimating the Effect of Smoking on Birth Outcomes Using a Matched Panel Data Approach," *Journal of Applied Econometrics*, 21(4), 489-519.

[3] Alan, S., B. Honoré, L. Hu, and S. Leth-Petersen (2013), "Estimation of Panel Data Regression Models with Two-Sided Censoring or Truncation," *Journal of Econometric Methods*, 3(1), 1-20.

[4] Andersen, E.B. (1970), "Asymptotic Properties of the Conditional Maximum-Likelihood Estimators," *Journal of the Royal Statistical Society, Series B*, 32, 283-301.

[5] Baetschmann, G., K.E. Staub, and R. Winkelmann (2015), "Consistent Estimation of the Fixed-Effects Ordered Logit Model," *Journal of the Royal Statistical Society, Series A*, 178(3), 685-703.

[6] Chamberlain, G. (1980), "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47, 225-238.

[7] Chamberlain, G. (1985), "Heterogeneity, Omitted Variable Bias, and Duration Dependence," in *Longitudinal Analysis of Labor Market Data*, ed. by J.J. Heckman and B. Singer, pp. 3-38, Cambridge University Press.

[8] Das, M. and A. van Soest (1999), "A Panel Data Model for Subjective Information on Household Income Growth," *Journal of Economic Behavior and Organization*, 40, 409-426.

[9] Galvao, A.F., C. Lamarche, and L.R. Lima (2013), "Estimation of Censored Quantile Regression for Panel Data With Fixed Effects," *Journal of the American Statistical Association*, 108(503), 1075-1089.

[10] Hausman, J.A., B.H. Hall, and Z. Griliches (1984), "Econometric Models for Count Data With an Application to the Patents-R&D Relationship," *Econometrica*, 52(4), 909-938.

[11] Honoré, B. (1992), "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60(3), 533-565.

[12] Khan, S., M. Ponomareva, and E. Tamer (2016), "Identification of Panel Data Models with Endogenous Censoring," *Journal of Econometrics,* 194 (1): 57-75.

[13] Kim, J. and D. Pollard (1990), "Cube Root Asymptotics," *Annals of Statistics*, 18(1), 191-219.

[14] Lewbel, A. (2017), "Identification Zoo: Meanings of Identification in Econometrics," working paper.

[15] Machado, M.P. (2004), "A Consistent Estimator for the Binomial Distribution in the Presence of 'Incidental Parameters': An Application to Patent Data," *Journal of Econometrics*, 119(1), 73-98.

[16] Manski, C.F. (1987), "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55(2), 357-362.

[17] Muris, C. (2017), "Estimation in the Fixed Effects Ordered Logit Model," *Review of Economics and Statistics*, 99(3), 465-477.

[18] Newey, W.K. and J.L. Powell (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, Vol. 4*, ed. by R.F. Engle and D.L. McFadden, pp. 2111-2245, North Holland Elsevier.

[19] Wooldridge, J.M. (2010), *Econometric Analysis of Cross Section and Panel Data*, Second Edition, MIT Press.