Peer reviewed version

## University of Bristol - Explore Bristol Research
### General rights

# Interval censored regression with fixed effects[*]

Jason Abrevaya[†] and Chris Muris[‡]

June 2019

## Abstract

This paper considers identification and 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.

**Keywords:** Panel data, fixed effects, interval censoring.

**JEL Classification:** C13, C14, C23, C24.

# 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.

---

[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.

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.

## 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.[2] 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:

---

[2]Muris (2017) treats the case of general $T$ for the closely related ordered-choice fixed-effects model. Another way to deal with multiple time periods is to apply the procedure below to each combination of distinct time periods. The results can then be combined using minimum distance, and inference can be done using the bootstrap.

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

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 $\sigma_0$ 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 $-\infty < c_{1,t} < \cdots < c_{J-1,t} < \infty$ in the following way:[3]

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

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 $\alpha_i$, the probability that the censored variable $y_{it}$ assumes a particular value $j \in \{1, \ldots, J\}$ is

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

---

[3]The number of cutoffs $J$ could also be time-varying $(J_t)$, but the paper focuses on time-invariant $J$ for notational simplicity.

where $c_{0,t} = -\infty$ and $c_{J,t} = +\infty$ for all $t$.

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:

$$d_{it\pi} \equiv 1\{y_{it} \geq \pi(t)\} \tag{4}$$
$$= 1\{y_{it}^* \geq c_{\pi(t),t}\}, \quad t = 1, 2,$$
$$d_{i\pi} \equiv (d_{i1\pi}, d_{i2\pi}). \tag{5}$$

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

$$d_{it\pi} = 1\left[\alpha_i + X_{it}\beta_0 - \sigma_0 u_{it} \geq c_{\pi(t),t}\right] \tag{6}$$
$$= 1\left[u_{it} \leq \frac{1}{\sigma_0}\left(\alpha_i + X_{it}\beta_0 - c_{\pi(t),t}\right)\right]. \tag{7}$$

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 $\alpha_i$ in the transformed model.

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

$$p_{i\pi 2}(\beta_0, \sigma_0) \equiv \Pr\left(d_{i,\pi} = (0,1)\mid \bar{d}_{i\pi} = 1, X_i, \alpha_i\right) \tag{8}$$
$$= \Lambda\left((X_{i2} - X_{i1})\frac{\beta_0}{\sigma_0} - \left(c_{\pi(2),2} - c_{\pi(1),1}\right)\frac{1}{\sigma_0}\right). \tag{9}$$

*Proof.* See Appendix A.1. □

Theorem 1 establishes a conditional probability that, for switchers, does not depend on the fixed effect. It resembles the result for the binary choice case, see e.g. Chamberlain (1980, p. 229). The two main differences with the binary choice case are that: (i) in the interval-censored fixed effects logit model, a switcher is defined relative to cutoff points chosen by the researcher,

6

i.e. $y_{i1} < \pi(1)$ and $y_{i2} \geq \pi(2)$, while in binary choice case $\pi(t) = 1$ for all $t$; (ii) the resulting expression (9) features parameters that are absent from its binary choice analog, namely the variance parameter $\sigma_0$ and the difference in latent cutpoints $c_{\pi(2),2} - c_{\pi(1),1}$. In the binary choice model, normalizations on those parameters are required for identification, and they therefore do not feature in the conditional probability.

## 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 (7). As such, the identification and estimation of $\beta_0$ and $\sigma_0$ 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

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

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

so that we can write

$$p_{i\pi 2}(\theta_0) \equiv p_{i\pi 2}(\beta_0, \sigma_0)$$

$$= \Lambda(W_{i\pi}\theta_0)$$

based on Theorem 1.

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

**Assumption 1.** $E\left( \Delta X_i' \Delta X_i \right)$ *is invertible.*

Assumption 1 guarantees that $E\left( W_{i\pi}' W_{i\pi} \right)$ is invertible whenever $c_{\pi(2),2} \neq c_{\pi(1),1}$. 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).[4]

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 $(\beta_0, \sigma_0)$ is identified.*

*Proof.* See Appendix A.2. □

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

$$
\hat{\theta}_\pi = \arg \max_{\theta \in \mathbb{R}^{K+1}} \frac{1}{n} \sum_{i=1}^{n} 1\left\{\bar{d}_{i\pi} = 1\right\} \left(d_{i2\pi} \ln \Lambda\left(W_{i\pi}\theta\right) + \left(1 - d_{i2\pi}\right) \ln\left(1 - \Lambda\left(W_{i\pi}\theta\right)\right)\right) \quad (10)
$$
$$
= \arg \max_{\theta \in \mathbb{R}^{K+1}} \frac{1}{n} \sum_{i=1}^{n} 1\left\{\bar{d}_{i\pi} = 1\right\} l_{i\pi}\left(W_{i\pi}\theta\right),
$$

where $l_{i\pi}\left(W_{i\pi}\theta\right) \equiv d_{i2\pi} \ln \Lambda\left(W_{i\pi}\theta\right) + \left(1 - d_{i2\pi}\right) \ln\left(1 - \Lambda\left(W_{i\pi}\theta\right)\right)$ and $\theta$ denotes a generic element of the parameter space. Well-definedness of the estimator in (10) requires that $c_{\pi(2),2} \neq c_{\pi(1),1}$.[5]

We propose to estimate $(\beta_0, \sigma_0)$ using all possible transformations jointly, using a composite maximum likelihood estimator (CMLE) as in Baetschmann et al. (2015) and Muris (2017):

$$
\hat{\theta} = \arg \max_{\theta \in \mathbb{R}^{K+1}} \frac{1}{n} \sum_{\pi} \sum_{i=1}^{n} 1\left\{\bar{d}_{i\pi} = 1\right\} l_{i\pi}\left(W_{i\pi}\theta\right). \quad (11)
$$

**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 (11). If Assumption 1 holds, then*

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

---

[4]Assumption 1 allows for bounded regressors. In the binary choice version of our model ($J = 2$), the logistic assumption is then necessary and sufficient for identification of the regression coefficient; see Chamberlain (2010). In what follows, we will show that the logistic assumption is sufficient for identification in our model when regressors are bounded. Whether the logistic assumption is necessary for identification in the interval-censored fixed effects model with bounded regressors is an interesting question that is left for future research. Note that, as the number of support points increases, the econometrician obtains more information about the value of $y_{it}^*$. In a certain limit, the model becomes a fixed effects linear regression model where logistic errors are no longer necessary. It is unclear whether the logistic assumption can be relaxed for finite $J$.

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

*and*

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, H^{-1}\right) \quad as \ n \to \infty, \tag{12}$$

*where*

$$H = -E\left[\sum_{\pi} 1\left\{\bar{d}_{i\pi} = 1\right\} \Lambda\left(W_{i\pi}\theta_0\right)\left(1 - \Lambda\left(W_{i\pi}\theta_0\right)\right) W_{i\pi}' W_{i\pi}\right]. \tag{13}$$

*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). $\qquad\square$

*Remark* 1. That the variance of this estimator is decreasing in the number of thresholds $J$ can be seen from the expression in (13). An increase in $J$ means that the sum is taken over additional transformations. Each transformation adds a positive definite term

$$1\left\{\bar{d}_{i\pi} = 1\right\} \Lambda\left(W_{i\pi}\theta_0\right)\left(1 - \Lambda\left(W_{i\pi}\theta_0\right)\right) W_{i\pi}' W_{i\pi}$$

to the expression.

*Remark* 2. Jointly, the transformed binary choice models overidentify the parameters in the original interval-censored model. This suggests that we can (1) test the model specification and (2) obtain more efficient estimators by optimal weighting. To do so, consider an alternative to the composite maximum likelihood estimator that starts from the set of scores associated with each of the transformed models. Stack the scores, and view them as overidentifying moment conditions in a GMM setting (see Muris, 2017, Section V, for a similar perspective). One can then use a standard $J$-test for the validity of the moment conditions. For efficiency, one can use an optimal two-step version of this GMM procedure or use a weighted version of (11). However, Muris (2017) found that setting the weighting matrix to the identity matrix works well across a wide range of simulation designs.

The parameters of interest are $(\beta_0, \sigma_0)$, 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,*

$$\sqrt{n} \left( \begin{array}{c} \hat{\beta} - \beta_0 \\ \hat{\sigma} - \sigma_0 \end{array} \right) \xrightarrow{d} \mathcal{N} \left( 0, M H^{-1} M' \right)$$

*as $n \to \infty$, where*

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

*and $\iota_k$ 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). $\square$

## 2.2 Heteroskedasticity

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

$$y_{it}^* = \alpha_i + X_{it}\beta_0 - \sigma_0(Z_i)u_{it} \ (i = 1, \ldots, n; \ t = 1, 2), \tag{14}$$

with the error scaling function $\sigma_0(Z_i) > 0$ of an $L$-dimensional vector $Z_i$ replacing the constant scaling parameter $\sigma_0$. We will call the model with the homoskedastic outcome equation (1) replaced

10

by (14) the *heteroskedastic fixed effects interval-censored logit model.* It is implicit in (14) that we require the heteroskedasticity to be time-invariant, which is restrictive.

Theorem 2 can be extended in order to show identification of $\beta_0$ 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 $E\left(\Delta X_i' \Delta X_i \middle| Z_i = z\right)$ is invertible, then $(\beta_0, \sigma_0(z))$ is identified.*

*Proof.* The proof resembles that of Theorem 2, after conditioning on $Z_i$. It is provided in Appendix A.5. $\qquad\square$

If $E\left(\Delta X_i' \Delta X_i \middle| Z_i = z\right)$ is invertible for all $z$ in the support of $Z$, it follows that the heteroskedasticity function $\sigma_0(\cdot)$ is nonparametrically identified within the model (14).

*Remark* 3. Remember that $X_{it}$ is the $K$-vector of regressors in period $t$. Denote by $X_{it,k}$ the $k$-th regressor in period $t$. The assumption that $E\left(\Delta X_i' \Delta X_i \middle| Z_i = z\right)$ is invertible rules out that $Z_i = (X_{i1}, X_{i2})$, or that $Z_i$ contains $(X_{i1,k}, X_{i2,k})$ for any $k$. One possibility is to use $Z_i = \frac{1}{2}(X_{i1} + X_{i2})$. While this invertibility condition is necessary for nonparametric identification, it can be relaxed with a parametric specification, as below.

For estimation purposes, we recommend a parametric specification for the $\sigma_0(\cdot)$ function. In particular, we will focus on the exponential specification given by

$$\sigma_0(Z_i) = \exp(Z_i \gamma_0). \tag{15}$$

The exponential specification is commonly used in practice (see, e.g., Wooldridge, 2010, Chapter 16). Under exponential heteroskedasticity, the true parameter vector $\theta_0$ of interest is now $(\beta_0, \gamma_0)$ rather than $(\beta_0, \sigma_0)$. For a given choice $\pi$ of the cutoff-value pair, the conditional probability $p_{i\pi2}(\theta_0)$ is

$$p_{i\pi2}(\theta_0) = \Pr\left(d_{i,\pi} = (0,1) \middle| \bar{d}_{i,\pi} = 1, X_i, Z_i, \alpha_i\right) = \Lambda\left(\frac{(X_{i2} - X_{i1})\beta_0 + \left(c_{\pi(2),2} - c_{\pi(1),1}\right)}{\exp(Z_i \gamma_0)}\right).$$

This suggests the composite-likelihood estimator

$$\widehat{\theta}_{het} = \arg\max_{\theta \in \mathbb{R}^{K+L}} \frac{1}{n} \sum_{\pi} \sum_{i=1}^{n} 1\left\{\bar{d}_{i\pi} = 1\right\} \left(d_{i2\pi} \ln p_{i\pi2}(\theta) + (1 - d_{i2\pi}) \ln\left(1 - p_{i\pi2}(\theta)\right)\right). \tag{16}$$

Denoting $\Gamma_i = \begin{bmatrix} \Delta X_i / \exp\left(Z_i \gamma_0\right) \\ -v_i Z_i \end{bmatrix}$, the following theorem establishes consistency and asymptotic normality in the heteroskedastic case.

**Theorem 5.** *Let $\left(\{y_i, X_i, Z_i\}, i = 1, \cdots, n\right)$ be a random sample from the heteroskedastic fixed effects interval-censored logit model with $J \geq 3$ and with true parameter values $\theta_0 = (\beta_0, \gamma_0)$, and let $\widehat{\theta}_{het}$ be the estimator defined in (16). If heteroskedasticity is exponential, as in (15), and $E[\Gamma_i' \Gamma_i]$ is invertible, then*

$$\widehat{\theta}_{het} \xrightarrow{p} \theta_0 \ \text{as } n \to \infty;$$

*and*

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, H_{het}^{-1}\right) \quad \text{as } n \to \infty, \tag{17}$$

*where*

$$H_{het} = -E\left[\sum_{\pi} 1\left\{\bar{d}_{i\pi} = 1\right\} \Lambda\left(v_i\right)\left[1 - \Lambda\left(v_i\right)\right] \Gamma_i \Gamma_i'\right], \tag{18}$$

*where $v_i = \frac{\Delta X_i \beta_0 + \left(c_{\pi(2),2} - c_{\pi(1),1}\right)}{\exp(Z_i \gamma_0)}$.*

*Proof.* The proof is very similar to that of Theorem 3, and is therefore omitted. Details are provided in Appendix A.6. $\square$

# 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

$$y_{it}^* = \alpha_i + X_{it}\beta_0 - u_{it} \ (i = 1, \ldots, n; \ t = 1, 2), \tag{19}$$

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

$$u_{i1}|X_i, \alpha_i \text{ and } u_{i2}|X_i, \alpha_i \text{ are identically distributed.} \tag{20}$$

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. The observed $y_{it}$ is defined as in (3), the same as the parametric model.

For a given choice $\pi$ 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 $c_{\pi(t),t}$:

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

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

$$(X_{i2} - X_{i1})\beta_0 - \left(c_{\pi(2),2} - c_{\pi(1),1}\right) \geq 0 \iff \Pr(d_{i2\pi} = 1|X_i, \alpha_i) \geq \Pr(d_{i1\pi} = 1|X_i, \alpha_i). \tag{21}$$

With $J \geq 3$ and the technical assumptions of Manski (1987), identification of $\beta_0$ can be established based upon the relationship in (21). A more precise formulation now follows. Denote by $F(u_{it}|X_i, \alpha_i)$ 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.** *Assume that the stationarity in* (20) *holds, and that the support of* $F(\cdot|X_i, \alpha_i)$ *is* $\mathbb{R}$ *for all* $(X_i, \alpha_i)$.

**Assumption 3.** *(i) The distribution of* $\Delta X_i$ *is such that at least one component of* $\Delta X_i$ *has positive Lebesgue density on* $\mathbb{R}$ *conditional on all the other components of* $\Delta X_i$ *with probability one. The corresponding component of* $\beta_0$ *is non-zero; (ii) The support of* $(\Delta X_i, 1)$ *is not contained in any proper linear subspace of* $\mathbb{R}^{K+1}$.

Condition (i) standard in the analysis of panel binary choice models with unbounded regressors (see e.g. Manski (1987) and Magnac (2004)). The main identification result then follows:

**Theorem 6.** *If $(y_i, X_i)$ follows the fixed effects interval-censored model in (19)-(20) with $J \geq 3$, and Assumptions 2 and 3 hold, then $\beta_0$ is identified.*

*Proof.* See Appendix A.7. □

Note that $\beta_0$ is identified *with scale* here, in contrast to the binary-choice model of Manski (1987). This result mirrors our identification of the variance $\sigma$ in the logit case. Moreover, a composite maximum-score-type estimator can be constructed as follows

$$\widehat{\beta} = \arg\max_{\mathbb{R}^K} \frac{1}{n} \sum_\pi \sum_{i=1}^n \left[ (d_{i2\pi} - d_{i1\pi}) \cdot \operatorname{sgn}\left( (X_{i2} - X_{i1}) \beta - \left( c_{\pi(2),2} - c_{\pi(1),1} \right) \right) \right], \tag{22}$$

$$= \arg\max_{\mathbb{R}^K} S_n(\beta), \tag{23}$$

where $\operatorname{sgn}(v) \equiv 1(v > 0) - 1(v < 0)$ and $\beta$ 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 $\widehat{\beta}$ defined by (22) is a consistent estimator of $\beta_0$.

**Theorem 7.** *Let the conditions of Theorem 6 hold, and assume that a random sample $(y_i, X_i)$, $i = 1, \cdots, n$ from the fixed effects interval-censored model in (19)-(20) with $J \geq 3$ is available. Furthermore, assume that the full-support regressor in Assumption 3(i) is the $K$-th regressor $X_{iK}$, and that there exists a $\eta > 0$ such that $|\beta_{0K}| \geq \eta$. Then the estimator $\widehat{\beta}$ defined in (22) is consistent,*

$$\widehat{\beta} \xrightarrow{p} \beta_0 \ \ as \ n \to \infty.$$

*Proof.* This result follows from a modification of the results in Manski (1985) and Manski (1987). First, by applying Lemma 4 in Manski (1985) for each transformation $\pi$, we obtain

$$S_n(\beta) \xrightarrow{a.s.} S(\beta)$$
$$:= E\left[ \sum_\pi \left[ (d_{i2\pi} - d_{i1\pi}) \cdot \operatorname{sgn}\left( (X_{i2} - X_{i1}) \beta - \left( c_{\pi(2),2} - c_{\pi(1),1} \right) \right) \right] \right]$$
$$=: \sum_\pi S_\pi(\beta).$$

Lemma 3 in Manski (1985) implies that for any $\pi$ such that $c_{\pi(2),2} \neq c_{\pi(1),1}$, $S_\pi(\beta_0) > S_\pi(\beta)$ for

14

any $\beta \neq \beta_0$, and for any $\pi$ with $c_{\pi(2),2} = c_{\pi(1),1}$, that $S_\pi(\beta_0) \geq S_\pi(\beta)$ for all $\beta$. Consistency of our maximum score estimator then follows from Theorem 1 in Manski (1985). See also the results for the binary choice panel model in Manski (1987, p. 361). □

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 (22) 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. Although it is well-documented that the standard bootstrap does not work (Abrevaya and Huang (2005)), inference can be conducted using the bootstrap-based distributional approximation recently proposed by Cattaneo, Jansson, and Nagasawa (2018).

# 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 $\sigma_0$), 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 $\beta_0$ and $\sigma_0$ are varied across designs. Finally, the interval-censored dependent variable is generated through

$$
y_{it} = \begin{cases} 1 & \text{if } y_{it}^* < 60 \\ 2 & \text{if } 60 \le y_{it}^* < 70 \\ 3 & \text{if } y_{it}^* \ge 70. \end{cases}
$$

Note that $y_{it}^*$ is symmetrically distributed around 65. Therefore, the design parameters $(\beta_0, \sigma_0)$ 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 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 $\sigma_0$ 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 $\beta_0$. Large values of $\beta_0$ 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 $\{1, 3\}$.

| $n$ | $\beta_0$ | $\sigma_0$ | 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 $y_{it}^*$.
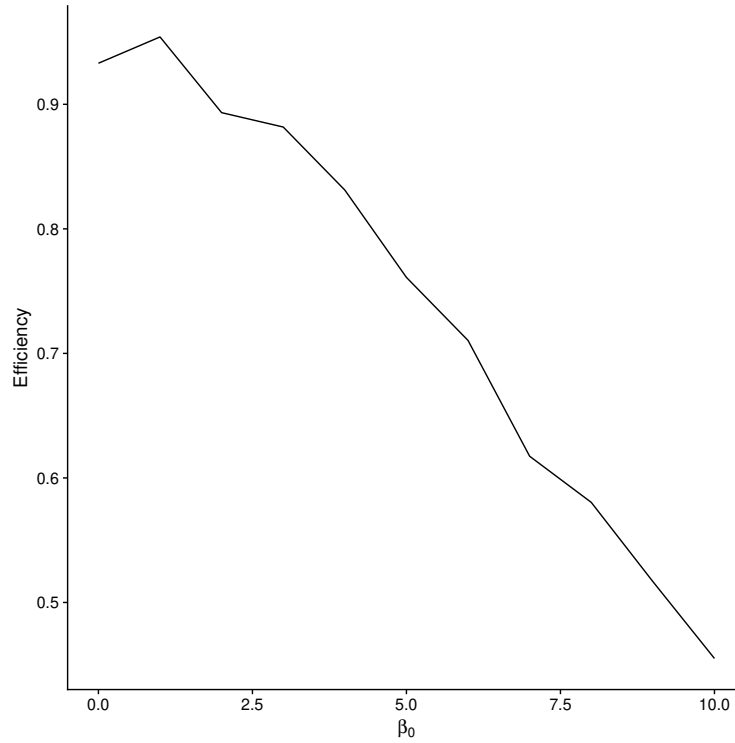


Figure 1: Efficiency of $\hat{\beta}$ as a function of $\beta_0$.

| $\gamma_1$ | CMLE | | CMLE-HET | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $\beta$ | | $\beta$ | | $\gamma_0$ | | $\gamma_1$ | |
| | Bias | SD | Bias | SD | Bias | SD | Bias | SD |
| -2 | 0.17 | 0.35 | 0.02 | 0.37 | 0.03 | 0.16 | 0.12 | 0.35 |
| 0 | 0.00 | 0.16 | 0.00 | 0.16 | 0.08 | 0.06 | 0.00 | 0.06 |
| 1 | 0.04 | 0.27 | 0.03 | 0.28 | 0.07 | 0.12 | 0.03 | 0.15 |
| 2 | 0.17 | 0.33 | 0.02 | 0.37 | 0.04 | 0.16 | 0.10 | 0.34 |

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

## 4.1 Heteroskedasticity

In the presence of heteroskedasticity, the CMLE is no longer guaranteed to estimate $\beta_0$ 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

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

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 biased 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 a small price in terms of standard deviation. CMLE-HET estimates the skedasticity parameters well, as can be seen from the low bias and standard deviations in Table 2.

# 5 Empirical application

In this section, we consider an empirical application of the parametric (CMLE) estimator introduced in Section 2.[6] 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

---

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

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.[7] 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 $c_1 = 2500$, $c_2 = 3000$, $c_3 = 3500$, $c_4 = 4000$, $c_5 = 4500$). In both cases, $c_{j,t} = c_j$ for all $(j, t)$. 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 in the second trimester, $pretri3$ for first visit in the third trimester; a first-trimester visit is the omitted category).

Table 3 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 also consider a heteroskedastic model with

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

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 $\log(e_i^2)/2$ on the variables in the heteroskedasticity function. For the interval-censored estimator,

---

[7]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.
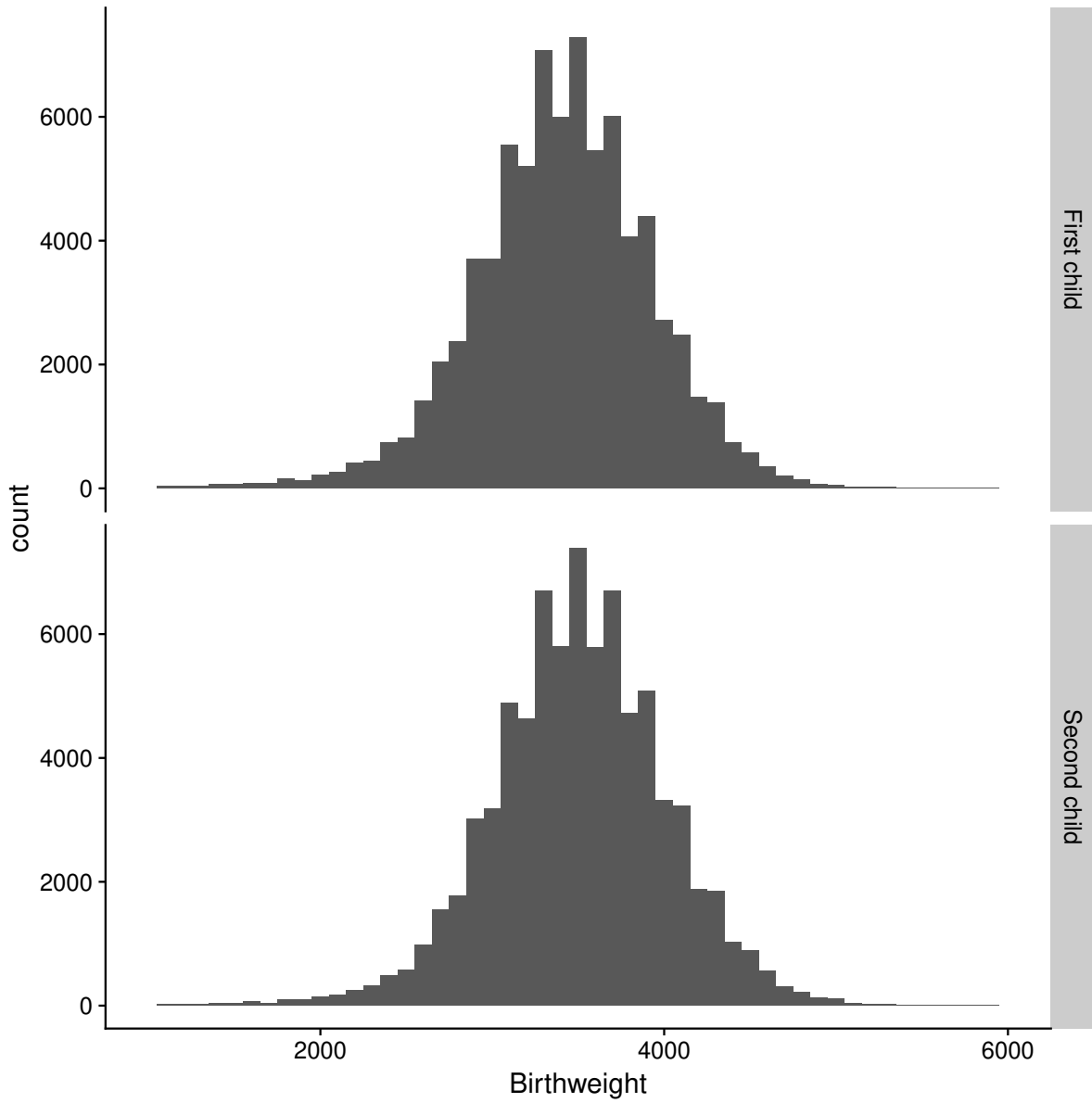
Figure 2: Histogram of birthweights (in grams)

|  | 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) |
| $\log(\sigma)$ | 6.01 | | 5.49 | | 5.31 | |

Table 3: Estimation results. Standard errors in parentheses. Standard errors for the CMLE are based on 1000 bootstrap replications.

we use the procedure outlined in Section 2.2. We present the results in Table 4. 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, the CMLE 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.

The results in this section assume that the error terms are logistic. An inspection of the empirical distribution of the differenced residuals for the heteroskedastic case, not reported here, does not yield evidence against that assumption.

Overall, in this application, CMLE provides estimates similar to those from the first-difference estimator, even 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 dis-

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

turbances, with the semiparametric model also allowing for serially correlation in the error disturbances. 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 $\beta$ and $\sigma_0(\cdot)$ within the logit model for a fully nonparametric specification for $\sigma_0(\cdot)$ . Furthermore, it may be possible to formulate a specification test based on estimators that use differing levels of coarseness. For example, one estimator could be based on the observed outcome with, say, $J = 8$ support points; and another could be based on a collapsed version of the outcome variable with $J' = 4$ support points. Under the assumptions of the model above, the probability limit of the two estimators should coincide.[8]

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

---

[8]We are grateful to an anonymous referee for pointing this out.

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 $d_{i1\pi} \equiv 1\{y_{i1}^* \geq c_{\pi(1),1}\}$ and $d_{i2\pi} \equiv 1\{y_{i2}^* \geq c_{\pi(2),2}\}$. The (conditional) probability of these indicator variables being equal to one are given by

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

and, similarly,

$$P\left(d_{i2\pi} = 1 | X_i, \alpha_i\right) = \Lambda\left(\frac{1}{\sigma_0}\left(\alpha_i + X_{i2}\beta_0 - c_{\pi(2),2}\right)\right). \tag{24}$$

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

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

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

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 $\alpha_i$. 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:

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

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

$\square$

## A.2 Proof of Theorem 2 (Identification)

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

$$p_{i\pi 2}(\beta_0, \sigma_0) = \Lambda \left( \frac{1}{\sigma_0} \left( \Delta X_i \beta_0 - \left( c_{\pi(2),2} - c_{\pi(1),1} \right) \right) \right)$$
$$= \Lambda \left( W_{i\pi} \theta_0 \right).$$

Pick $\pi(1)$ and $\pi(2)$ such that $c_{\pi(1),1} \neq c_{\pi(2),2}$. For the purpose of this proof, change the notation to:

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

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

For any $W_{i\pi}$, we have $\Lambda^{-1} \left( p_\pi \left( W_{i\pi} \right) \right) = W_{i\pi} \theta_0$ and $W'_{i\pi} \Lambda^{-1} \left( p_\pi \left( W_{i\pi} \right) \right) = W'_{i\pi} W_{i\pi} \theta_0$. Taking expectations yields

$$E \left[ W'_{i\pi} \Lambda^{-1} \left( p_\pi \left( W_{i\pi} \right) \right) \right] = E \left( W'_{i\pi} W_{i\pi} \right) \theta_0. \tag{25}$$

Assumption 1 implies that $E \left[ W'_{i\pi} W_{i\pi} \right]$ is invertible for our case with $c_{\pi(1),1} \neq c_{\pi(2),2}$. We conclude that

$$\theta_0 = \left[ E \left( W'_{i\pi} W_{i\pi} \right) \right]^{-1} E \left[ W'_{i\pi} \Lambda^{-1} \left( p_\pi \left( W_{i\pi} \right) \right) \right].$$

Therefore, $\beta_0/\sigma_0$ and $1/\sigma_0$ are identified since all RHS quantities are identified from the data. Since $\beta_0 = \frac{\beta_0/\sigma_0}{1/\sigma_0}$ and $\sigma_0 = \frac{1}{1/\sigma_0}$, the result follows. $\square$

## 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

$$Q_n(\theta) = \frac{1}{n} \sum_i \sum_\pi 1\left\{\bar{d}_{i\pi} = 1\right\} l_{i\pi}(W_{i\pi}\theta). \tag{26}$$

Therefore, the population (limiting) objective function is

$$Q_0(\theta) = \text{plim } Q_n(\theta)$$
$$= E\left[\sum_\pi 1\left\{\bar{d}_{i\pi} = 1\right\} l_{i\pi}(W_{i\pi}\theta)\right]. \tag{27}$$

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 $W_{i\pi}\theta$:

$$\frac{\partial l_{i\pi}(v)}{\partial v} = d_{i\pi2}\frac{\Lambda(v)(1-\Lambda(v))}{\Lambda(v)} - (1-d_{i\pi2})\frac{\Lambda(v)(1-\Lambda(v))}{1-\Lambda(v)}$$
$$= d_{i\pi2}(1-\Lambda(v)) - (1-d_{i\pi2})\Lambda(v)$$
$$= d_{i\pi2} - \Lambda(v). \tag{28}$$

It follows that the second derivative is

$$\frac{\partial^2 l_{i\pi}(v)}{(\partial v)^2} = -\Lambda(v)(1-\Lambda(v)) < 0, \tag{29}$$

i.e. $l_{i\pi}$ is strictly concave in $W_{i\pi}\theta$. Therefore, it is concave in $\theta$ for any $\pi$, and strictly concave in $\theta$ 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 $\pi$, the sample objective function in (11) 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 (29) and the chain rule, we obtain

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

$$\equiv H.$$

**Jacobian.** Using the expression in (28) we obtain for the variance of the score from $l_{i\pi}$:

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

so that

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

Finally, note that

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

$$= \Lambda\left(W_{i\pi}\theta_0\right)\left(1 - \Lambda\left(W_{i\pi}\theta_0\right)\right)$$

where the first step follows because $E\left[d_{i\pi 2}\middle| W_{i\pi}, \bar{d}_{i\pi} = 1\right] = \Lambda\left(W_{i\pi}\theta_0\right)$, see (24); and the second step follows because the variance of a Bernoulli random variable with probability $p$ is $p\left(1 - p\right)$. It therefore follows that $H = -\Sigma$.

## A.4   Proof of Theorem 3

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

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

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 (28) 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 $\Sigma$ 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. $\left\|\sum_\pi 1\left\{\bar{d}_{i\pi} = 1\right\}\Lambda\left(W_{i\pi}\theta\right)\left(1 - \Lambda\left(W_{i\pi}\theta\right)\right)W'_{i\pi}W_{i\pi}\right\| \leq \|W'_{i\pi}W_{i\pi}\|$, and $E\|W'_{i\pi}W_{i\pi}\| < \infty$. Therefore, a uniform law of large numbers applies on a suitably chosen neighbourhood of $\theta_0$, see e.g. Lemma 2.4 in Newey and McFadden (1994). $\qquad\square$

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

*Proof.* This proof is very close to that in Section A.2. The only difference is that $\sigma_0$ depends on $Z$, and that the expectation in (25) is replaced by a conditional expectation. In other words, we define

$$p_\pi\left(\Delta X_i, Z_i\right) \equiv \Lambda\left(\frac{1}{\sigma_0\left(Z_i\right)}\left(\Delta X_i\beta_0 - \left(c_{\pi(2),2} - c_{\pi(1),1}\right)\right)\right) \equiv \Lambda\left(W_{i\pi}\theta_0\left(Z_i\right)\right),$$

where the parameter of interest is $\theta_0\left(Z_i\right) = \left(\beta_0/\sigma_0\left(Z_i\right), 1/\sigma_0\left(Z_i\right)\right)$. Then, following the steps in Appendix A.2, we obtain

$$\theta_0\left(z\right) = \left[E\left(W'_{i\pi}W_{i\pi}\middle|\, Z_i = z\right)\right]^{-1} E\left[W_{i\pi}\Lambda^{-1}\left(p_\pi\left(\Delta X_i, Z_i\right)\right)\middle|\, Z_i = z\right].$$

where the inverse is guaranteed to exist because of the condition in the statement of the result. The conclusion then follows.

## A.6 Proof of Theorem 5 (Asymptotics for heteroskedastic estimator)

For an individual with $\bar{d}_{i\pi} = 1$, denote the contribution of individual $i$ to the Jacobian for transformation $\pi$ by

$$J_i = \frac{\partial\left(d_{i2\pi}\ln p_{i\pi2}(\theta) + \left(1 - d_{i2\pi}\right)\ln\left(1 - p_{i\pi2}(\theta)\right)\right)}{\partial\theta},$$

and let

$$v_i = \frac{\Delta X_i\beta_0 + \left(c_{\pi(2),2} - c_{\pi(1),1}\right)}{\exp\left(Z_i\gamma_0\right)}$$

Differentiation with respect to $v_i$ yields $(d_{i2\pi} - \Lambda(v_i))$, so that application of the chain rule yields

$$J_{i\pi} = (d_{i2\pi} - \Lambda(v_i)) \frac{1}{\exp(Z_i\gamma_0)} \begin{bmatrix} \Delta X_i \\ -\left(\Delta X_i \beta_0 + \left(c_{\pi(2),2} - c_{\pi(1),1}\right)\right) Z_i \end{bmatrix}.$$

By the information matrix equality, the contribution to the Hessian is

$$\begin{aligned} H_{i\pi} &= -E\left[(d_{i2\pi} - \Lambda(v_i))^2 \Gamma_i \Gamma_i'\right] \\ &= -E\left[\Lambda(v_i)\left[1 - \Lambda(v_i)\right] \Gamma_i \Gamma_i'\right], \end{aligned}$$

where

$$\Gamma_i := \begin{bmatrix} \frac{\Delta X_i}{\exp(Z_i\gamma_0)} \\ -v_i Z_i \end{bmatrix}$$

The Hessian of the log-likelihood is therefore negative definite, so that the objective function is again concave. A proof of consistency and asymptotic normality is therefore very similar to the proof of 3, and is therefore omitted. $\qquad\square$

## A.7 Proof of Theorem 6 (Identification, semiparametric case)

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

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

Rewrite this as

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

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

Then, for a given choice of cut points such that $c_{\pi(2),2} - c_{\pi(1),1} \neq 0$, 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:

$$y_t = d_{it\pi},$$

$$x_t = [X_{it}, -c_{\pi(t),t}],$$

$$c = \alpha_i,$$

$$u_t = u_{it}.$$

Then, equation (20) and Assumption 2 imply Assumption 1 of Manski (1987); Assumption 3 and the choice of cutoffs so that $c_{\pi(2),2} - c_{\pi(1),1} \neq 0$ 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 $\theta_0$ equals 1), the result is obtained. $\square$

# 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] Abrevaya, J. and J. Huang (2005), "On the Bootstrap of the Maximum Score Estimator." *Econometrica,* 73 (4), 1175–1204.

[4] 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.

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

[6] 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.

[7] Cattaneo, M., M. Jansson, and K. Nagasawa (2018), "Bootstrap-Based Inference for Cube Root Consistent Estimators," *Manuscript.*

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

[9] 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.

[10] Chamberlain, G. (2010), "Binary Response Models for Panel Data: Identification and Information," *Econometrica,* 78(1): 159–68.

[11] 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.

[12] 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.

[13] 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.

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

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

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

[17] 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.

[18] Magnac, T. (2004), "Panel Binary Variables and Sufficiency: Generalizing Conditional Logit," *Econometrica,* 72(6), 1859–76.

[19] Manski, C.F. (1985), "Semiparametric Analysis of Discrete Response - Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 313–333.

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

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

[22] 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.

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