Linear Regression for Dependently Censored Panel Duration Models with Nonadditive Fixed Effects

By: Mitali Das* and Zhiliang Ying**

Columbia University, New York NY 10027

This Version: August 2005

This paper proposes estimators for a class of panel duration models with induced dependent censoring and fixed effects which may appear nonadditively in the model. No parametric assumptions are made about the distribution of the random error. The dependent censoring implies that the usual differencing approach common in fixed-effects models will lead to disproportionate weighting of the errors, and will not yield an unbiased estimating equation. Under mild assumptions on the conditional distribution of the errors, an estimator is proposed by proportionately trimming the marginal error distributions. It is proved that the estimators are consistent and asymptotically normal. Standard error estimators are constructed by using a resampling approach that repeatedly perturbs the minimand, thus avoiding the need for density estimation or numerical differentiation. Extensive simulations show that the estimators can have very good finite sample properties, and the resampling approach is reliable and efficient. The proposed approach is also applied to an unemployment duration problem using NLSY data.

---

* Associate Professor, Department of Economics, 420 West 118 st., Columbia University, New York, NY 10027 (mitali.das@columbia.edu).

** Professor, Department of Statistics, MC 4690, Columbia University, 1255 Amsterdam Avenue, 10th Floor, New York, NY 10027 (zying@stat.columbia.edu).

We would like to thank Sokbae (Simon) Lee for graciously sharing and helping with use of some NLSY data and Yu Zheng for exceptional research assistance. We also thank the Institute for Mathematical Sciences at the National University of Singapore for a stimulating atmosphere where part of the work was carried out. This research was supported in part by grants from the Ford Foundation (Das), the National Science Foundation and the National Institutes of Health (Ying).

1 Introduction

The analysis of duration data has a long tradition in both methodological and empirical research. In keeping with the features of most duration data and the requirements of empirical analysis, the recent important developments in duration research have been inference methods for multi-spell (panel) durations, which permit either individual heterogeneity or censored durations or both. Two broad classes of estimators have dominated this analysis. Using the Cox (1972) model and partial likelihood as a foundation, estimators for the panel proportional hazards model with "fixed effects" have been suggested in, among others, Chamberlain (1985), Ridder and Tunali (1999), Lancaster (2000), Horowitz and Lee (2003). The accelerated failure time (AFT) model is an important complementary specification, and has been similarly extended to the panel setting (see, e.g., Lin, Wei and Ying (1998), Abrevaya (1999) and Lee (2004)).

Despite such methodological advances, important gaps remain between the features of many panel duration data and the available analytical methods. In particular, in most analysis of duration data a key assumption is of independent censoring, i.e., independence of the censoring and the duration variables. However, since the sampling in panel durations very likely induces dependence across successive durations, the assumption of independent censoring rules out any successive accumulation of data (Lancaster (2000)). Such data arise frequently in empirical research, including in the work histories of the National Longitudinal Surveys (NLS), sequentially collected over the 1979-1996 period, as well as many health data where spells are durations between the successive occurrences of a disease. Despite the pervasiveness of such data, analytical methods for dependent censoring induced by latent "fixed effects" are largely unknown.

The defining feature of dependently censored data is that the sums of successive duration are censored, rather than individual durations themselves; see e.g., Visser (1996), Wang and Wells (1998), Lin, Sun and Ying (1999). For example, in J durations $y_1, ..., y_J$ that occur consecutively over a period of random length c, the jth spell ($j > 1$) is uncensored only if $y_j \leq c - \Sigma_{k=1}^{j-1} y_k$. Thus, the longer the preceding spells, the shorter the observation period for subsequent spells. In a fixed-effects model, where durations are linked by the common individual specific effects, this censoring invariably leads to a dependence between the censoring and duration variables, for all but the first spell. The most important consequence of dependent censoring is that well known methods for multi-spell duration data, including the partial likelihood estimator and other estimators for uncensored or independently censored data, do not yield consistent parameter estimates. Thus, estimators for such data, particularly reliable semiparametric methods, would have a valuable role in a range of empirical work.

In this paper, we study the inference problem in a class of panel duration models with possibly non-additive fixed effects and dependent censoring. Underlying our analysis is the *panel transformation* model:

$$g(y_{ij}) = x'_{ij}\beta_0 + h(\alpha_i, \varepsilon_{ij}), \qquad (i = 1, ..., n;\ j = 1, .., J) \tag{1.1}$$

where i and j indicate subject and duration respectively, y_{ij} is a latent duration variable, g is a completely specified monotone increasing function, h is an arbitrary and unspecified function and x_{ij} is a covariate vector. The duration variables are subject to dependent censoring in a manner described in detail below. Here, β_0 is the parameter of interest while $\{\alpha_i\}$ are treated as a sequence of nuisance parameters. The fixed effect is constant across spells and may have arbitrary statistical relation with x. There are two generalities in the sources of heterogeneity in (1.1): one is the nonseparability of the fixed-effect from the idiosyncratic errors, and the other is the arbitrary link function h. This specification includes as a special case the additively separable case traditionally maintained in panel estimation, $h(\alpha_i, \varepsilon_{ij}) = \alpha_i + \varepsilon_{ij}$.[1]

It is well known that (1.1) subsumes both the Cox as well as the AFT models. In the former, g is the logarithm of the cumulative baseline hazard function, while ε has Type 1 extreme value distribution. The AFT is complementary, where the distribution of ε is completely unspecified while g is known ($g = \log$).

This paper considers the case when g is any known and strictly monotone transform, not necessarily the logarithmic transform, the distribution of ε is unspecified and the durations are subject to dependent censoring. In this way, (1.1) is a generalization of the AFT model. A leading special case is panel Tobit regression ($g(u) = u$) that is analyzed in Honore (1992) for the independent (and fixed) censoring case, with linear and additive fixed-effects. Analysis of (1.1) is complementary to one where the baseline hazard is nonparametric (g is unknown) but the distribution of ε is completely specified; see e.g., Horowitz and Lee (2003), Lee (2004). Neither model nests the other, and each is robust to different assumptions about the data generating process. There is, however, a connection between the two in the context of hypothesis testing where a comparison of the two methods can be made in terms of the power, as discussed below.

The simultaneous presence of fixed effects and dependent censoring introduces certain difficulties in developing a valid estimator of β_0 in (1.1). The presence of the nuisance parameters $\{\alpha_i\}$, which has unspecified distribution conditional on the covariates, implies that a marginal analysis of the model would not generally be possible. Further, even if the random errors had specified probability distribution, the partial likelihood would not produce a consistent estimator

[1] Few exceptions exist. Two notable ones are Abrevaya (1999) and Altonji and Matzkin (2001), both of which permit nonlinear and multiplicative fixed-effects. Abrevaya's method does not extend to censoring and requires an i.i.d assumption on the errors; Altonji and Matzkin's exchangeability assumption includes an i.i.d assumption "within" group, which is inessential to this paper. These important differences are elaborated upon below.

In other panel limited dependent variable models the fixed-effect nevertheless enter linearly and additively in the latent model which facilitates the usual first-difference approach to estimation, e.g. Honore (1992).

of β_0 due to the dependence in the data. Frequently, fixed effects models are estimated under a within-individual transform of the model which eliminates the fixed effects. In (1.1), however, the censoring indicates that the first-difference approach extends to only the subset of observed durations, while the dependent censoring implies that conditioning the optimization on only the observed durations induces a bias due to asymmetrical weighting of the errors. The key insight of this paper is that by trimming the marginal distribution of the errors, an unbiased estimating equation can be derived using proportionately weighted errors, which simultaneously addresses the presence of the nuisance parameters as well as the dependent censoring.

2 Dependent Censoring Transformation Model

Initially, the model is described and the results derived for the two spell ($J = 2$) case, with the generalization for the multi-spell given in Section 3.4. Suppose that $\{(y_{i1}, y_{i2}, x_{i1}, x_{i2}, \alpha_i, \varepsilon_{i1}, \varepsilon_{i2}) : i = 1, .., n\}$ are independent and identically distributed. The model of interest is

$$\begin{aligned} g(y_2) &= x_2'\beta_0 + h(\alpha, \varepsilon_2) \\ g(y_1) &= x_1'\beta_0 + h(\alpha, \varepsilon_1). \end{aligned} \tag{2.2}$$

where dim $(x_j) = p$ $(j = 1, 2)$. As our method will involve within-individual differences, where coefficients of time-invariant covariates are not identified, it is assumed that the rows of x_j include only time-varying covariates.

Suppose y_1 and y_2 occur over an observed interval $(0, c)$ which is random with unspecified probability distribution. The durations y_1, y_2 are both subject to right censorship, in the sense that y_1 is censored by c in the usual way, while y_2 is censored by $(c - y_1)^+$, where for any integer a, $a^+ = \max(0, a)$. Denote

$$z_1 = \min(c, y_1), \qquad z_2 = \min((c - y_1), y_2), \tag{2.3}$$

and $\delta_1 = I\{y_1 \leq c\}, \delta_2 = I\{y_2 \leq c - y_1\}$ where $I\{\cdot\}$ is the indicator function. The observed data consist of $(x_1, x_2, z_1, z_2, \delta_1, \delta_2)$.

This model is relevant for the sampling in many duration data, including work history data (see Horowitz and Lee (2003) for an application) and treatment data in health studies (see, e.g., Aalen and Husbye (1991), Visser (1996)). In work history data, for example, the censoring variable c represents the time elapsed between the start of a respondent's first employment spell and the survey point. In the event that both spells are completed at the survey point, $y_1 + y_2 \leq c$ and no censoring occurs. If the first spell is completed but the second censored, $y_1 + y_2 > c \geq y_1$. Observe that in this case y_2 is censored by $c_2 \equiv I\{y_1 \leq c\}(c - y_1)$, where y_2 is invariably correlated with its censoring variable due to the presence of the fixed effects. The remaining possibility is that the respondent is still in the first spell of employment where y_1 is censored by c and y_2 is not recorded. Similarly, in some health data c represents the duration between the inception of treatment and the survey point, while y_1 and y_2 denote the successive recurrences of a disease (see, e.g., Visser (1996) for an application).

The key assumption underlying the results in this paper is the following.

Assumption A *Conditional on x_1, x_2 and c, the random variables $h(\alpha, \varepsilon_1)$ and $h(\alpha, \varepsilon_2)$ are exchangeable in the sense that*

$$(h(\alpha, \varepsilon_1), h(\alpha, \varepsilon_2)) \stackrel{d}{=} (h(\alpha, \varepsilon_2), h(\alpha, \varepsilon_1)).$$

This conditional exchangeability assumption permits dependence between ε_1 and ε_2, as well as conditional heteroskedasticity in the errors, which is precluded under the i.i.d assumptions used commonly in panel duration models (e.g., Ridder and Tunali (1999), Abrevaya (1999)) including the dependent censoring models in Horowitz and Lee (2003) and Lee (2004). Further, a censoring distribution that has dependence on covariates, including on the fixed effect, is permitted under Assumption A. No assumptions are made about the nature of this dependence. It is a useful generality that could be important in many settings. For instance, stylized facts from search theory imply that low-ability workers have the highest incidence of unemployment durations and the lowest exit rates from unemployment spells (see Singer and Spilerman (1976)). In survey data, it could be precisely such workers with the largest censoring times. Unobserved health quality could similarly affect both the incidence of censoring, as well as the duration of a treatment spell in a health study. The estimators proposed in this paper are robust to this possibility.

The exchangeability condition of Assumption A should be contrasted with one in the non-separable panel models of Altonji and Matzkin (2001),[2] which stipulates that the distribution of (α, ε_j) $(j = 1, 2)$ conditional on (x_1, x_2), is exchangeable in the *covariates*. That is, using f to denote a conditional density function, $f(\alpha, \varepsilon_j | x_1 = \bar{x}_1, x_2 = \bar{x}_2) = f(\alpha, \varepsilon_j | x_1 = \bar{x}_2, x_2 = \bar{x}_1)$. This assumption is essential to their estimator, because the marginal effect of x_1 is obtained by comparing the distributional change given $(x_1 = \bar{x}_1, x_2 = \bar{x}_2)$, with the change given $(x_1 = \bar{x}_2, x_2 = \bar{x}_1)$, as x_1 changes. Our assumption, however, is about exchangeability of the *errors* given the covariates, i.e., $f(\varepsilon_1, \varepsilon_2 | x_1, x_2) = f(\varepsilon_2, \varepsilon_1 | x_1, x_2)$ and involves no restriction on the statistical relation of the fixed-effect conditional on covariates. To better illustrate the difference between the two assumptions, consider the case where $(\varepsilon_1, \varepsilon_2)$ are independent of all other variables. Then, the assumption of Altonji and Matzkin (2001) amounts to $f(\alpha | x_1 = \bar{x}_1, x_2 = \bar{x}_2) = f(\alpha | x_1 = \bar{x}_2, x_2 = \bar{x}_1)$, while in our case this holds trivially.

[2]The methods in that paper do not apply to either the independent or dependent censoring case.

Before describing our estimator, we review some related models and methods. Consistency of a partial likelihood estimator of β_0 in a duration model with additive fixed effects (e.g., Ridder and Tunali (1999)) is contingent on independent censoring and cannot be attained under the censoring scheme in (2.3). Other estimators for panel hazard or transformation models are either derived exclusively under assumptions about independent censoring (Kalbfleisch and Prentice (1980), Chamberlain (1985), Han and Hausman (1990)) or consider heterogeneity in the absence of censoring (e.g., Abrevaya (1999)). If one specifies $(\varepsilon_1, \varepsilon_2)$ conditionally i.i.d with known distribution, estimators of β_0 can be formulated using a preliminary estimate of the joint conditional survival function of durations (Horowitz and Lee (2003)) or of the survival function of the censoring variate (Lee (2004)). These methods do not extend to the model setting of this paper under Assumption A.

2.1 Estimation under Proportional Weighting

The key idea of our approach is to derive an unbiased estimating equation which eliminates the nuisance parameter α despite censoring in the duration variable. Due to the possible non-additivity of α in (1.1), the usual "first-difference" on a pair of durations will not eliminate the fixed effects. However, it follows by Assumption A that the differencing approach can be extended to (1.1) to eliminate the fixed effects in expectation. That is, while $\{h(\alpha, \varepsilon_2) - h(\alpha, \varepsilon_1)\}$ is identically zero when (α, ε_j) is linear in its arguments, under the conditional exchangeability assumption, $E[\{h(\alpha, \varepsilon_2) - h(\alpha, \varepsilon_1)\} | x_1, x_2] = 0$. The censoring, however, creates a bias when the differencing alone is used to generate an estimating equation. We illustrate this problem below, and propose a scheme of proportional weighting under which the estimating equations handle the censoring and the bias simultaneously.

To describe our approach, we use the following notation: $\bar{g} = g^{-1}$, $\Delta x = x_2 - x_1$ and $a^{\otimes 2} = aa'$ for a vector a. Under Assumption A, it follows immediately that $g(y_2) - g(y_1) - \Delta x'\beta_0 = h(\alpha, \varepsilon_2) - h(\alpha, \varepsilon_1)$ has probability distribution that is symmetric about zero. Therefore, for any odd function ψ, $\psi(g(y_2) - g(y_1) - \Delta x'\beta_0)$ has zero expectation. Use of such ψ is common in M-estimation, where well-known examples of ψ include the identity function $\psi(u) = u$, associated with the least squares normal equation, and sign function $\psi(u) = sgn|u|$, associated with least absolute deviations estimation (Huber, 1967).

Unfortunately, the preceding approach in generating an unbiased estimating equation is not directly applicable in this model, since y_1 and y_2 may not be always observed due to censoring. To see this, suppose only those pairs where both y_1 and y_2 are observed are used to construct an estimating equation. Then,

$$\begin{aligned} & \psi(g(y_2) - g(y_1) - \Delta x' \beta_0)\ I\{y_1 + y_2 < c\} \qquad (2.4) \\ = \;& \psi\left(h\left(\alpha, \varepsilon_2\right) - h\left(\alpha, \varepsilon_1\right)\right)\ I\{\bar{g}(x_1' \beta_0 + h\left(\alpha, \varepsilon_1\right)) + \bar{g}(x_2' \beta_0 + h\left(\alpha, \varepsilon_2\right)) < c\}. \end{aligned}$$

Note, under the dependent censoring mechanism, observing both y_1 and y_2 is equivalent to the condition $y_1 + y_2 < c$. Now, unless g is a linear function (and thus, so is $\bar{g}$), the form of the constraint $I\{y_1 + y_2 < c\}$ implies that (2.4) will not have zero expectation, as it imposes disproportionate weights on the errors ε_1 and ε_2 under the inequality constraint. This disproportionate weighting arises because the censoring mechanism acts on $y_1 + y_2$, and not $g(y_1) + g\left(y_2\right)$. More precisely, whereas ε_1 and ε_2 are conditionally exchangeable under Assumption A, their marginals are distorted under (2.4) in a manner where exchangeability cannot hold unless either $\Delta x' \beta_0 = 0$ (where the observed pair does not contribute to the loss), or g is linear, where the censoring scheme would bear proportionately on the pair of errors.

However, the symmetric form of the censoring $I\{y_1 + y_2 < c\}$, as well as the form of the resulting asymmetric constraints on ε_1 and ε_2 suggest that it would be possible to modify the naive approach to obtain an unbiased estimating equation. Note that $g\left(y_1\right) = x_1' \beta_0 + h\left(\alpha, \varepsilon_1\right)$ is bounded above by

$$x_1 \beta_0 + h\left(\alpha, \varepsilon_1\right) + \left(\Delta x' \beta_0\right)^+ = h\left(\alpha, \varepsilon_1\right) + \max(x_1' \beta_0, x_2' \beta_0),$$

while $g\left(y_2\right) = x_2 \beta_0 + h\left(\alpha, \varepsilon_2\right)$ is bounded by

$$x_2 \beta_0 + h\left(\alpha, \varepsilon_2\right) + \left(-\Delta x' \beta_0\right)^+ = h\left(\alpha, \varepsilon_2\right) + \max(x_1' \beta_0, x_2' \beta_0).$$

The key insight in our approach is to use the above bounds to proportionately weight the errors in deriving an unbiased estimating equation. This idea is related to the symmetric trimming idea proposed in Powell (1986), similarly used in Honore (1992) and Honore and Powell (1994). Whereas asymmetric censoring points of the errors in Powell (1986) and Honore (1992) suggest artificially trimming the error distributions to restore symmetry and derive orthogonality conditions, in model (2.2) the nonlinearity of g as well as the constraint which depends on $y_1 + y_2$ (rather than $g(y_1) + g\left(y_2\right)$) precludes trimming the error distribution directly.[3] However, as $I\{y_1 + y_2 < c\}$ may be interpreted as a function that asymmetrically weights the errors in (2.4), a related idea can be used here to symmetrize the weighting by trimming out a subset of the observed pairs. Thus, the underlying idea in our approach is to use the above bounds on

[3] Trimming the error distributions would be possible if g were linear; however, under linearity, no artificial trimming would be required to derive an unbiased estimating equation for this model.

$g(y_1)$ and $g(y_2)$ to proportionally weight the errors in (2.4). Consider the following modified estimating function

$$
\begin{aligned}
& \psi\left(g(y_2) - g(y_1) - \Delta x'\beta_0\right)\ I\{\ \bar{g}\left[g(y_1) + \left(\Delta x'\beta\right)^+\right] + \bar{g}\left[g(y_2) + \left(-\Delta x'\beta\right)^+\right]\ < c\} \\
= \ & \psi\left(h(\alpha, \varepsilon_2 - h(\alpha, \varepsilon_1)\right)\ I\{\ \bar{g}\left[g(y_1) + \left(\Delta x'\beta\right)^+\right] + \bar{g}\left[g(y_2) + \left(-\Delta x'\beta\right)^+\right]\ < c\}.
\end{aligned}
$$

Under Assumption A, the above function has zero expectation at the true parameter value β_0, conditionally on (x_1, x_2). It follows that the product of the above function with a function of x, in particular Δx, also has zero expectation. These results are formalized in the following proposition. First, introduce notation $\gamma(\beta) = \bar{g}\left[g(y_1) + (\Delta x'\beta)^+\right] + \bar{g}\left[g(y_2) + (-\Delta x'\beta)^+\right]$. Obviously, $\gamma(\beta) < c$ implies that $y_1 + y_2 < c$. Thus y_k $(k = 1, 2)$ in $\gamma(\beta)$ can be replaced by z_k $(k = 1, 2)$ and $\gamma(\beta)$ can be constructed from data.

Proposition 1 *Suppose y_1 and y_2 are generated by model (2.2), where y_1 is censored by c, y_2 is censored by $(c - y_1)^+$ and $(h(\alpha, \varepsilon_2), h(\alpha, \varepsilon_1))$ satisfy Assumption A. Then at the true parameter value β_0,*

$$E\left[\psi\left(g(y_2) - g(y_1) - \Delta x'\beta_0\right)\ I\{\gamma(\beta_0) < c\}\ \Delta x | x_1, x_2\right] = 0. \tag{2.5}$$

The zero expectation of (2.5) in Proposition 1 follows trivially as a consequence of the conditional exchangeability of ε_2 and ε_1, and yields an estimating equation under which an estimator of β_0 can be obtained. Note that the form of the trimming function $\gamma(\beta) = \bar{g}\left[g(y_1) + (\Delta x'\beta)^+\right] + \bar{g}\left[g(y_2) + (-\Delta x'\beta)^+\right]$ implies that asymptotically a positive fraction of the observed duration pairs will be excluded from estimation, except in the limiting case of "censoring at infinity".

The proposed estimator $\hat{\beta}$ of β_0 is defined as a root of the sample analog of the estimating equation or "score" function (2.5),

$$U_n(\beta) = n^{-1}\sum_{i=1}^{n} \Delta x_i\ \psi\left(g(z_{i2}) - g(z_{i1}) - \Delta x_i'\beta\right)\ I\{\gamma_i(\beta) < c\}, \tag{2.6}$$

where, as noted before, $z_{i2} = y_{i2}$ and $z_{i1} = y_{i1}$ under the constraint $I\{\gamma_i(\beta) < c_i\}$. Alternatively, we can define $\hat{\beta}$ as a minimizer of the following loss function

$$
\begin{aligned}
L_n(\beta) \quad = \quad & n^{-1}\sum_{i=1}^{n}\{\rho\left(g(z_{i2}) - g(z_{i1}) - (\Delta x_i'\beta)^+ \wedge [g(c_i - z_{i2}) - g(z_{i1})]\right) \\
& + \rho\left(g(z_{i2}) - g(z_{i1}) + (-\Delta x_i'\beta)^+ \vee [g(z_{i2}) - g(c_i - z_{i1})]\right)\},
\end{aligned} \tag{2.7}
$$

where ρ denotes the "loss" function that corresponds to ψ, i.e., $\rho' = \psi$. Straightforward calculation shows that $U_n(\beta) = -\partial L_n(\beta)/\partial\beta$. Two leading choices for ψ are $\psi(u) = u$ $(\rho(u) = u^2)$, in which case $U_n(\beta)$ is analogous to the least squares normal equation with a trimming modification for the dependent censoring, and $\psi(u) = sgn(u)$ $(\rho(u) = |u|)$ which leads to the loss function associated with LAD estimation.

The discontinuity of $U_n(\beta)$ in β due to the trimming function implies that the estimating equation may not have a unique root. Likewise, the loss function L_n may not have a unique minimizer. Such phenomena also occur in the estimating equations and loss functions in Powell (1986) and Powell and Honore (1994). Our regularity conditions are specified such that the estimator of β_0 derived as a solution to $U_n(\beta) = 0$ or as a minimizer of L_n is "essentially unique" in the sense that any two estimators so defined are asymptotically equivalent. These and other results are further discussed below.

2.1.1 Iterative Algorithm

In describing an algorithm for estimating β_0 as the root of the estimating equation (2.6), a primary concern is that the trimming function $I\{\gamma(\beta) < c\}$ is a step function. As a consequence, $U_n(\beta)$ will not be continuous in β, which introduces difficulties in using the usual root-finding Newton-Raphson method of finding a minimizer. In fact, there may not exist an exact solution at all due to the discontinuity in the estimating equation. We propose an iterative algorithm to find an approximate solution to $U_n(\beta)$. However, as the discontinuity in $U_n(\beta)$ emerges principally due to the presence of the unknown parameter β in the trimming function, the problem can be circumvented by fixing the unknown β with an initial estimator. More precisely, let $\widehat{\beta}_{(0)}$ denote an initial choice for β. For example, under the usual squared loss, an initial estimate may be taken as

$$\widehat{\beta}_{(0)} = \arg\min \sum_{i:\ \delta_{i2}=1} \left[\, g(z_{i2}) - g(z_{i1}) - \Delta x_i'\beta\right]^2, \tag{2.8}$$

and iteratively one can employ the estimate from the $(k-1)$th minimization in the kth minimization, such as

$$\widehat{\beta}_{(k)} = \arg\min \sum_{i:\ I\left\{\gamma_i\left(\widehat{\beta}_{(k-1)}\right) \le c_i\right\}=1} [g(z_{i2}) - g(z_{i1}) - \Delta x_i'\beta]^2, \tag{2.9}$$

until the estimator converges to a stable limit, that we denote $\widehat{\beta}$. The limiting minimizer of (2.9) is the proportionately weighted, or "proportionately trimmed" estimator of the model. Our numerical studies indicate that this method works efficiently, with rapid convergence to a minimizer even for large samples. Other estimators can be formulated by altering the function ψ. One useful alternative to the squared loss is the L_1 loss function, which is well known to be less susceptible to outliers in the data (see, e.g., Powell (1984)). An iterative algorithm can be similarly carried out for this case. Here, the initial estimate $\widehat{\beta}_{(0)}$ could be obtained as

$$\widehat{\beta}_{(0)} = \arg\min \sum_{i:\ \delta_{i2}=1} \left|\ g(z_{i2}) - g(z_{i1}) - \Delta x_i'\beta \right|. \tag{2.10}$$

Then, as above, by treating $\widehat{\beta}_{(0)}$ as the true value of the parameter, a new estimator $\widehat{\beta}_{(1)}$ of β can be obtained by minimizing the above objective function in the set that satisfies $i : I\{\gamma_i(\widehat{\beta}_{(0)}) \leq c_i\} = 1$. Repeating this iteratively until the estimator stabilizes in the kth iteration, where the $(k-1)$th minimizer is used to determine the set $I\{\gamma_i(\widehat{\beta}_{(k-1)}) \leq c_i\} = 1$ used in the kth minimization, leads to the proportionately weighted estimator $\widehat{\beta}$. It is clear the above approach is applicable to a general loss function ρ. Extensions of this method to the multi-spell case are straightforward, and described as a generalization below. In the following section, a large sample result for the proposed estimator and construction of an estimator of the limiting covariance matrix is considered.

3 Variance Estimation and Asymptotic Theory

3.1 Variance Estimation

The nonsmooth form of the objective function $U_n(\beta)$ indicates that covariance matrix estimation may involve practical difficulties, a well known problem when the minimand is not twice continuously differentiable and involves density estimation in the semiparametric setting (Pakes and Pollard (1989)). Density functions are notably difficult to compute by conventional methods (e.g., Efron and Tibshirani, 1993) and would generally depend on subjective smoothing choices. In addition, straightforward but tedious derivation shows that the Jacobian is given by

$$\begin{aligned} A = \frac{\partial}{\partial\beta} EU_n(\beta)\Big|_{\beta=\beta_0} = & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (3.11)\\ & - E\left[(\Delta x)^{\otimes 2} I(\gamma(\beta_0) < c)\psi'(h(\alpha,\epsilon_2) - h(\alpha,\epsilon_1))\right] - E\Big\{(\Delta x)^{\otimes 2}\, I\left(\gamma\left(\beta_0\right) < c\right) \\ & \times \Big[I\left(\Delta x'\beta_0 \geq 0\right) \int \psi\left(u - g\left(c - \overline{g}\left(u\right)\right)\right) \tilde{f}\left(g\left(c - \overline{g}\left(u\right)\right) - \Delta x'\beta_0, u | x_1, x_2\right) du \\ & - \ I\left(\Delta x'\beta_0 < 0\right) \int \psi\left(g\left(c - \overline{g}\left(u\right)\right) - u\right)\ \tilde{f}\left(u, g\left(c - \overline{g}\left(u\right)\right) + \Delta x'\beta_0 | x_1, x_2\right) du \Big] \Big\} \end{aligned}$$

where $\tilde{f}$ is the conditional joint conditional density of $(h(\alpha,\epsilon_1), h(\alpha,\epsilon_2))$ given x_1 and x_2, indicating that direct estimation of A would be difficult. Variance estimation in the current problem introduces an additional difficulty since the conditioning event $\gamma(\beta)$ cannot be made asymptotically orthogonal to the summand $[\psi(g(z_2) - g(z_1) - \Delta x'\beta)]$, which leads the theoretical variance of the estimator to depend on the unknown density of ε; however, the estimation scheme can lead only to estimation of the convolution of $\varepsilon_2 - \varepsilon_1$.

We address the covariance matrix estimation problem by circumventing density estimation and integration altogether. Variance estimation is instead approached with a resampling method originating in Jin, Ying and Wei (2001). Here, we consider randomly perturbing the original score with externally generated weights that are i.i.d and independent of data. Repeated perturbation permits the construction of an empirical distribution and consequently a variance estimator. Under the conditions below, this empirical distribution approximates the sampling distribution of the estimate when suitably centered. As with related methods (bootstrap, jackknife, etc.), this variance estimator is not reliant on a preliminary subjective nonparametric estimate. One useful additional virtue is that the method will apply for any choice of loss ρ $(\psi = \rho')$, so that the theoretical results do not have to be derived separately for each case. Another virtue is that the explanatory variables need not be identically distributed. In other words, the results remain valid as long as certain stability assumptions are satisfied.

To be precise, let $\{w_i\}$ denote a known and non-negative $n-$vector of an i.i.d variable $\{w_i : i = 1, ..., n\}$ that satisfies $E(w) = Var(w) = 1$. Consider a modified estimating equation,

$$U_n^*(\beta) = n^{-1} \sum_{i=1}^{n} w_i \Delta x_i \psi\left(g(z_{i2}) - g(z_{i1}) - \Delta x_i'\beta\right) I\{\gamma_i(\beta) \le c\}, \tag{3.12}$$

or the corresponding loss function

$$\begin{aligned} L_n^*(\beta) \quad = \quad & n^{-1} \sum_{i=1}^{n} w_i \left\{\rho\left(g(z_{i2}) - g(z_{i1}) - \left[(\Delta x_i'\beta) \vee 0\right] \wedge [g(c_i - z_{i2}) - g(z_{i1})]\right)\right. \quad (3.13) \\ & \left. + \rho\left(g(z_{i2}) - g(z_{i1}) - \left[(\Delta x_i'\beta) \wedge 0\right] \vee [g(z_{i2}) - g(c_i - z_{i1})]\right)\right\}, \end{aligned}$$

and let $\hat{\beta}^*$ denote the minimizer of $L_n^*(\beta)$. Our theoretical results show that under mild regularity conditions, the distribution of $\sqrt{n}(\hat{\beta}^* - \hat{\beta})$ is approximately that of $\sqrt{n}(\hat{\beta} - \beta_0)$. Thus, inferences about β_0 may be derived simply by using a large sequence of estimators $\{\hat{\beta}^*\}$ from repeatedly minimizing the perturbed minimand (3.12). Key ingredients in the making the distributional approximation appropriate are: (1) $\partial E(U_n(\beta)) / \partial\beta|_{\beta=\beta_0} = \partial E(U_n^*(\beta)) / \partial\beta|_{\beta=\beta_0}$; (2) the asymptotic variance of $U_n(\beta_0)$ and

$$U_n^*(\hat{\beta}) \approx n^{-1} \sum_{i=1}^{n} (w_i - 1) \Delta x_i \psi\left(g(z_{i2}) - g(z_{i1}) - \Delta x_i'\hat{\beta}\right) I\left\{\gamma_i(\hat{\beta}) \le c\right\}.$$

are identical.

3.2 Large Sample Properties

We next derive the main large sample properties for the proposed estimation procedure. In addition to the conditional exchangeability Assumption A of Section 2, the following assumptions are needed in the derivation.

Assumption B *The data are a random sample* $\{y_{1i}, y_{2i}, x_{1i}, x_{2i}, \alpha_i, \varepsilon_{1i}, \varepsilon_{2i}) : i = 1, ..., n\}$ *generated by* (2.2), *the observed data are* $\{z_1, z_2, x_1, x_2, \delta_1, \delta_2\}$ *and* x_1 *and* x_2 *have compact supports, i.e.,* $\Pr(||x_j|| < C) = 1$ for $j = 1, 2$ *and some constant* C.

Assumption C *The conditional joint density of* $(h(\alpha, \varepsilon_1), h(\alpha, \varepsilon_2))$ *given* x_1 *and* x_2 *exists and is bounded.*

Assumption D *The expectation*

$$E\left[\Delta x\ \psi\left(\ h(\alpha, \varepsilon_2) - h(\alpha, \varepsilon_1) - \Delta x'(\beta - \beta_0)\right)\ I\{\gamma(\beta) < c\}\right],$$

is differentiable with respect to β *and the derivative matrix is negative definite.*

Assumption B is the sampling assumption and an additional condition is that all covariates are bounded. This is a standard assumption in modern empirical process theory. However, it can be relaxed to include the case where the variates have finite moment generating functions. In addition, the i.i.d. assumption may be relaxed to include covariates, i.e., the x_{ij} are nonrandom, provided certain stability assumptions are met. Assumption C holds trivially when errors ε_1 and ε_2 are mutually independent and independent of α. It is also satisfied when ε_1 and ε_2 are exchangeable and independent of α. Thus there are two directions of relaxation of the standard assumption on the fixed effects model : possible dependence between the two errors and possible dependence between the error and the fixed effect. Under Assumptions A and C, the estimating function, which is non-smooth, can have an asymptotically linear expansion. Finally, Assumption D requires that the slope matrix of the aforementioned linear expansion is invertible. This condition appears to be quite mild as it would be trivially satisfied under the usual identification assumption of non-singularity of the expected second moment matrix of Δx in the absence of the trimming function $I\{\gamma(\beta) < c\}$.

The first theorem gives consistency and asymptotic normality for the proposed estimator.

Theorem 1 *Suppose Assumptions A, B, C and D are satisfied. Then, there exists a neighborhood of* β_0, *within which any minimizer of* $L_n(\beta)$, $\hat{\beta}$, *is consistent and asymptotically normal,*

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) \xrightarrow{d} N\left(0, A^{-1}VA^{-1}\right)$$

where A is defined by (3.11) and $V = E\left[\ \Delta x^{\otimes 2}\psi^2\left(h\left(\alpha,\varepsilon_2\right)-h\left(\alpha,\varepsilon_1\right)\right)\ I\{\gamma\left(\beta_0\right)<c\}\right]$. Suppose further that β_0 is the unique minimizer of L_n in a compact region $Q \subset \mathcal{R}^p$ that contains it as an interior point. Then $\hat{\beta}$ as a minimizer of L_n in Q is consistent.

The above theorem essentially contains two components. The first part deals with the local behavior near β_0. It asserts that within a small neighborhood, the estimator as a minimizer is always consistent and asymptotically normal. The second part asserts that under an additional assumption on the global uniqueness of the expected loss $EL_n(\beta)$, the consistency can be extended to any compact region. This additional assumption is necessary since the loss function is not convex. The proof of Theorem 1 is given in the Appendix.

On the other hand, there are alternative ways to construct globally consistent estimators without assuming uniqueness of the minimizer for expected loss function. Below we outline an approach based on reducing the construction to a simple one-sample problem.

Suppose that $x_{i1} \equiv 0$ and $x_{i2} \equiv 1$. Then, this is essentially a one-sample problem ($(x_{i1}, x_{i2}) \equiv (0,1)$ for all i) with a crossover of x from 0 to 1. There is no fixed effects here since the covariates are constant across i and the regression parameter β is simply the location shift between the two periods. The distribution of the first period, y_1, can be estimated nonparametrically by the usual product-limit estimator. The distribution of the second period, y_2, can also be estimated using the approach of Lin, Sun and Ying (1999). By taking the difference of a specific quantile, such as the median, of the two estimators, we get a consistent estimator of β_0.

We next argue that the general case can be reduced to the above one-sample problem. Since A is nonsingular, we can always find p pairs of p-vectors, $(a_k,\ b_k)$, $k = 1, \ldots, p$, that are in the support of the distribution of (x_1, x_2), and that $b_k - a_k$, $k = 1, \ldots, p$, are linearly independent. Suppose we can consistently estimate $(b_k - a_k)'\beta_0$, $k = 1, \ldots, p$. Then a consistent estimator of β_0 can be obtained by solving a set of linear equations. Now, for each k, the estimation of $(b_k - a_k)'\beta_0$ is the one-sample problem provided we include only those i such that (x_{i1}, x_{i2}) are close to (a_1, b_1) and regarding them as equal to (a_1, b_1).

To see the distributional approximation arising from $\hat{\beta}^*$, it is useful to give an outline for the derivation of the asymptotic normality of $\hat{\beta}$. The modern empirical process theory ensures the following asymptotic approximation and linearity:

$$U_n\left(\beta\right) - U_n\left(\beta_0\right) = EU_n\left(\beta\right) - EU_n\left(\beta_0\right) + o\left(n^{-1/2}\left(\beta-\beta_0\right) + n^{-1/2}\right),$$

and

$$EU_n\left(\beta\right) - EU_n\left(\beta_0\right) = A\left(\beta-\beta_0\right) + o(||\beta-\beta_0||).$$

Now, as $EU_n(\beta_0) = 0$, we obtain

$$0 \approx U_n(\hat{\beta}) \approx U_n\left(\beta_0\right) + A(\hat{\beta}-\beta_0),$$

or

$$\sqrt{n}(\hat{\beta}-\beta_0)\approx-\sqrt{n}A^{-1}U_n(\beta_0)\xrightarrow{d}N\left(0,A^{-1}VA^{-1}\right).$$

In the preceding subsection, we proposed using repeatedly generated $\hat{\beta}^*$ for the approximation of the distribution of $\hat{\beta}$. The theoretical justification is given by the following theorem.

Theorem 2 *Under the same assumptions as in Theorem 1, the conditional distribution of $\sqrt{n}(\hat{\beta}^*-\hat{\beta})$ given data converges to the limiting distribution of $\sqrt{n}(\hat{\beta}-\beta_0)$. In particular, the variance of $\hat{\beta}$ is approximated consistently by the conditional variance of $\hat{\beta}^*$.*

The proof of Theorem 2 is given in the Appendix. The validity is seen intuitively from the following expansion

$$U_n^*\left(\hat{\beta}^*\right)\approx U_n^*\left(\hat{\beta}\right)+A\left(\hat{\beta}^*-\hat{\beta}\right),$$

where A is the same slope matrix as in (3.11), since U_n^* has the same expectation as that of U_n. Since the (conditional) expectation of $U_n^*(\hat{\beta})$ is $U_n(\hat{\beta})\approx 0$, and (conditional) variance of $U_n^*(\hat{\beta})$ is approximately the same as the variance of $U_n(\beta_0)$, we can see that the distributional approximation given by Theorem 2 should hold.

Before we turn to numerical studies of the proposed estimator and of the specification tests, we describe a generalization of our methods to the multi-spell case.

3.3 Extension to the Multi-spell Model

There is a straightforward extension of the method proposed in (2.6) to the multiple-duration ($J>2$) case. For this case, the sampling assumption is that the observed data are the i.i.d sample $\{(z_{i1},..,z_{iJ},x_{i1},..,x_{iJ},\delta_{i1},..,\delta_{iJ},c_i),\ i=1,...,n\}$ where

$$z_j=c\wedge\left(\textstyle\sum_{k=1}^{j}y_j\right),\quad \delta_j=I\{y_j\le c-\left(\textstyle\sum_{k=1}^{j-1}y_k\right)\}\qquad(j=1,...,J),$$

and y_j is censored by $c-(\sum_{k=1}^{j-1}y_k)$. A generalization of the conditional exchangeability in Assumption A is also required, and is given by the following pairwise exchangeability condition.

Assumption A′ *For every pair $(j',l;\ j'<l)$, conditional on $(x_1,..,x_l,\alpha,\varepsilon_{k,k<l,k\neq j'})$, the distribution of $(\varepsilon_{j'},\varepsilon_l,c)$ is identical to the distribution of $(\varepsilon_l,\varepsilon_{j'},c)$.*

Now, because the fixed-effects enter equivalently in all spells, any pair of within-subject differences will eliminate the fixed-effect, while the conditioning argument $\gamma(\beta)$ will need to be modified in accordance with the dependence between a specific pair of spells. Specifically, for durations y_j and y_l, $(1 \le j < l \le J)$, define $\Delta x^{jl} = x_l - x_j$, and

$$\gamma_i^{jl}(\beta) = \overline{g}\left[g(z_{ij}) + (\Delta x_i^{jl\prime}\beta)^+\right] + \overline{g}\left[g(z_{il}) + (-\Delta x_i^{jl\prime}\beta)^+\right] + \sum\nolimits_{m \le l, m \ne l, m \ne j} y_{im}.$$

Then an estimating equation for the $(j, l; 1 \le j \le l \le J)$ pair is given by

$$U_n^{jl}(\beta) = n^{-1}\sum_{i=1}^{n} \Delta x_i^{jl}\ \psi\left\{g(z_{il}) - g(z_{ij}) - \Delta x_i^{jl\prime}\beta\right\}\ I\{\gamma_i^{jl}(\beta) < c_i\}. \tag{3.14}$$

An estimator of β_0 in the multi-spell case could be chosen by computing estimators separately for each $U_n^{(jl)}(\beta)$, and computing an overall estimator that is a weighted sum of all estimators. An alternative is to choose an estimator as the root of a sum of the score functions $U^{(jl)}$ $(1 \le j < l \le J)$. One way to combine these estimating functions is by a simple average, i.e.,

$$U_n(\beta) = \frac{1}{n\begin{pmatrix} J \\ 2 \end{pmatrix}}\sum_{i=1}^{n}\sum_{1 \le j < l \le J} \Delta x_i^{jl}\ \psi\left\{g(z_{il}) - g(z_{ij}) - \Delta x_i^{jl\prime}\beta\right\}\ I\{\gamma_i^{jl}(\beta) < c_i\}. \tag{3.15}$$

The corresponding loss function is then

$$\begin{aligned} L_n(\beta) = \frac{1}{n\begin{pmatrix} J \\ 2 \end{pmatrix}} \quad & \sum_{i=1}^{n}\sum_{1 \le j < l \le J}\{\rho(g(z_{il}) - g(z_{ij}) - (\Delta x_i^{jl\prime}\beta)^+ \wedge [g(c_i - \sum_{\substack{m \le l \\ m \ne j}} z_{im}) - g(z_{ij})]) \\ & + \rho(g(z_{il}) - g(z_{ij}) - (\Delta x_i^{jl\prime}\beta)^+ \wedge [g(z_{il} - g(c_i - \sum_{m \le l-1} z_{im}))])\}. \end{aligned}$$

The estimator of β_0 outlined by (2.8) and (2.9) has a simple extension in the multi-spell case. For instance, the initial estimate $\widehat{\beta}_{(0)}$ would be obtained as

$$\widehat{\beta}_{(0)} = \arg\min \sum_{i:\ \delta_{il}=1}\ \sum_{1 \le j < l \le J} \rho\left[\ g(z_{il}) - g(z_{ij}) - \Delta x_i^{jl\prime}\beta\right],$$

and one would proceed as above by using the $(k-1)$th estimate to construct the conditioning sets $I\{\gamma^{jl}(\widehat{\beta}_{(k-1)}) < c\}$ $(1 \le j < l \le J)$ for the kth minimization,

$$\widehat{\beta}_{(k)} = \arg\min \sum_{i:\ \gamma_i(\widehat{\beta}_{(k-1)}) < c_i}\ \sum_{1 \le j < l \le J} \rho\left[\ g(z_{il}) - g(z_{ij}) - \Delta x_i^{jl\prime}\beta\right].$$

For distributional approximation, we can extend the random perturbation approach to the current setting. Specifically, we can define perturbed versions as follows, with weights w_i as before,

$$U_n^*(\beta) = \frac{1}{n\begin{pmatrix} J \\ 2 \end{pmatrix}} \sum_{i=1}^{n} w_i \sum_{1\le j<l\le J} \Delta x_i^{jl}\ \psi \left\{ g\left(z_{il}\right) - g\left(z_{ij}\right) - \Delta x_i^{jl\prime}\beta \right\}\ I\{\gamma_i^{jl}\left(\beta\right) < c_i\}$$

and

$$L_n^*\left(\beta\right) = \frac{1}{n\begin{pmatrix} J \\ 2 \end{pmatrix}} \qquad \sum_{i=1}^{n} w_i \sum_{1\le j<l\le J} \{\rho(g\left(z_{il}\right) - g\left(z_{ij}\right) - (\Delta x_i^{jl\prime}\beta)^+ \wedge [g(c_i - \sum_{\substack{m\le l \\ m\ne j}} z_{im}) - g\left(z_{ij}\right)])$$

$$+\rho(g\left(z_{il}\right) - g\left(z_{ij}\right) - (\Delta x_i^{jl\prime}\beta)^+ \wedge [g(z_{il} - g(c_i - \sum_{m\le l-1} z_{im}))])\}.$$

4 Specification Tests

An important issue in the analysis of panel data is testing for the presence of fixed effects. Although the proposed proportionately weighted estimator is a consistent estimator of β_0 in the fixed effects specification (2.2), if an alternate estimator of the model exists and is relatively more efficient under the null hypothesis of random effects but inconsistent in the fixed effects setting, an important objective of empirical research is to test the failure of the null. In the linear panel model, such testing is usually achieved by the Hausman (1978) specification test, which tests whether the numerical difference of two estimators is significantly different from zero. The well known practicability of the test in the linear panel case arises from the fact that under the null, the variance of the difference is the difference of the variances.

In this section, we devise a specification test for the semiparametric panel duration model (2.2), along the lines of Hausman (1978). We will consider the null hypothesis

$$H_0 : \alpha \ \text{ independent of } (x_1, x_2). \tag{4.16}$$

The test we devise deviates from the usual Hausman test in that under the null (4.16), various consistent estimators of β_0 in model (2.2) can be constructed, but an efficient estimator is not available. The main obstacle in constructing an efficient estimator for β_0 is the unspecified dependency across errors of the two time periods, and the induced dependent censoring on the second period. Although the Hausman test is not predicated on the availability of an efficient estimator (see, e.g., Newey (1985)), the significant consequence of testing between two inefficient estimators is that the difference of the variances of the two inefficient estimators is no longer the variance of their difference. Nevertheless, a test can be constructed where we can control the Type 1 error at any specified level, and is described as follows.

To fix our notation, let β_1 be a component of the parameter vector β and $\hat{\beta}_1$ the corresponding component in the proposed estimator $\hat{\beta}$. Suppose that under the null hypothesis we can construct an alternative estimator of β_1, to be denoted by $\hat{\beta}_{a1}$, that is also asymptotically normal. A valid test statistic for the hypothesis (4.16) is given by

$$J = \left(\hat{\beta}_1 - \hat{\beta}_{a1}\right)' \; \left[\hat{\Sigma}\left(\hat{\beta}_1 - \hat{\beta}_{a1}\right)\right]^{-1} \left(\hat{\beta}_1 - \hat{\beta}_{a1}\right)$$

where $\hat{\Sigma}(u)$ denotes a consistent estimator of the covariance matrix of the random vector u. Note that both estimators are consistent under the null hypothesis while only $\hat{\beta}_1$ is consistent under the alternative. As usual, setting the Type 1 error at level π, the null hypothesis is rejected if $J > \chi^2_{(p,1-\pi)}$ where $\chi^2_{(p,1-\pi)}$ is the $(1-\pi)$th quantile of the χ^2 distribution with p degrees of freedom.

To describe our approach, first note that under the null hypothesis of random effects, an alternative consistent estimator of regression parameter vector in model (2.2) is given by Powell's (1984) LAD estimator (using only first period data), under regularity conditions assumed by Powell for the asymptotic normality of this estimator. As to be see in Section 5.2, not all components of β are identifiable and Powell's estimate typically includes an intercept term that is not in β. To accommodate this generality, we shall use $\tilde{x}_1$ to denote the vector of all covariates in Powell's LAD model for the first period, including the intercept. Let $\tilde{\beta}$ denote the corresponding regression parameter vector, with $\tilde{\beta}_1$ denoting the common component shared with β in model (2.2). Powell's estimator is defined as

$$\hat{\beta}_P = \arg\min \sum_{i=1}^{n} |g(z_{1i}) - \min(\tilde{x}_{1i}'\tilde{\beta}, c_i)|. \tag{4.17}$$

Let $\hat{\beta}_{P1}$ denote the component in $\hat{\beta}_P$ that estimates $\tilde{\beta}_1$. Consider the random vector $\hat{\beta}_1 - \hat{\beta}_{P1}$. As being the case with $\hat{\beta}$, the covariance matrix $\Sigma\left(\hat{\beta}_1 - \hat{\beta}_{P1}\right)$, which is required in specification testing, has a complicated analytical form involving density estimates that are difficult to compute. However, we propose a relatively straightforward estimator of the variance covariance matrix, invoking the same method of random perturbation used in approximating the limiting variance of $\hat{\beta}$. Specifically, let $\hat{\beta}^*$ be as defined following equation (3.12), and let

$$\hat{\beta}_P^* = \arg\min \sum_{i=1}^{n} w_i |g(z_{1i}) - \min(\tilde{x}_{1i}'\tilde{\beta}, c_i)| \tag{4.18}$$

where the $\{w_i\}$ is the identical sequence of weights used in generating the original perturbed equation (3.12). The following theorem is the critical theoretical foundation underlying the limiting distribution of the proposed test.

Theorem 3 *Suppose Assumptions A-D are satisfied. Then for* $\hat{\beta}_P$ *and* $\hat{\beta}_P^*$ *defined in* (4.17) *and* (4.18), *under the null hypothesis* (4.16) ,

$$\sqrt{n}\begin{pmatrix} \hat{\beta}-\beta_0 \\ \hat{\beta}_P-\tilde{\beta}_0 \end{pmatrix} \xrightarrow{d} N\left(0,\ B\right), \qquad \sqrt{n}\begin{pmatrix} \hat{\beta}^*-\hat{\beta} \\ \hat{\beta}_P^*-\hat{\beta}_P \end{pmatrix} \xrightarrow{d^*} N\left(0,\ B\right)$$

where d^* *indicates convergence of conditional distribution given data and* B *is a* $(p+\tilde{p})\times(p+\tilde{p})$ *matrix defined by equation* (A.42) *in the Appendix, and* $\tilde{p}$ *is the dimension of* $\tilde{\beta}$.

It follows from the above theorem that $\sqrt{n}\left(\hat{\beta}_1-\hat{\beta}_{P1}\right)$ is asymptotically normal with mean zero and a $p_1\times p_1$ covariance matrix denoted by B_1. Theorem 3 also implies that the limiting distribution of $\sqrt{n}\left(\hat{\beta}_1-\hat{\beta}_{P1}\right)$ is approximated by the conditional distribution of $\sqrt{n}\left(\ \hat{\beta}_1^*-\hat{\beta}_{P1}^*-(\hat{\beta}_1-\hat{\beta}_{P1})\right)$ given data. Thus we have the following corollary.

Corollary 1 *Suppose Assumptions A-D are satisfied. Then, under the null hypothesis* (4.16),

$$\sqrt{n}\left(\hat{\beta}_1-\hat{\beta}_{P1}\right) \xrightarrow{d} N\left(0, B_1\right),$$

and

$$\sqrt{n}\left(\ \hat{\beta}_1^*-\hat{\beta}_{P1}^*-\left(\hat{\beta}_1-\hat{\beta}_{P1}\right)\right) \xrightarrow{d^*} N\left(0,\ B_1\right).$$

In particular,

$$n\left(\hat{\beta}_1-\hat{\beta}_{P1}\right)'\hat{B}_1^{-1}\left(\hat{\beta}_1-\hat{\beta}_{P1}\right) \xrightarrow{d} \chi_p^2$$

where $\hat{B}_1^{-1}/n$ *is the conditional covariance matrix of* $\left(\hat{\beta}_1^*-\hat{\beta}_{P1}^*\right)$ *given data.*

Note that $\hat{B}_1^{-1}/n$ may be approximated arbitrarily well by the law of large numbers, since $\hat{\beta}_1^*$and $\hat{\beta}_{P1}^*$ can be generated as many times as required under the random perturbation approach. In particular, suppose that for some large integer L, L pairs $\left(\hat{\beta}_1^{*(1)}, \hat{\beta}_{P1}^{*(1)}, ..., \hat{\beta}_1^{*(L)}, \hat{\beta}_{P1}^{*(L)}\right)$ are obtained under (3.12) and (4.18). Let

$$\hat{B}_{1,L} = \frac{1}{L}\sum_{l=1}^{L}\left[\hat{\beta}_1^{*(l)}-\hat{\beta}_{P1}^{*(l)}-(\hat{\beta}_1-\hat{\beta}_{P1})\right]^{\otimes 2}. \tag{4.19}$$

Then, the statistic

$$\hat{J} = \left(\hat{\beta}_1-\hat{\beta}_{P1}\right)'\ \hat{B}_{1,L}^{-1}\ \left(\hat{\beta}_1-\hat{\beta}_{P1}\right) \sim \chi_p^2$$

with cutoff point at $\chi^2_{p,(1-\pi)}$ for a size π significance test. The proposed specification test is a useful complement to the confidence intervals for inference, and its usefulness is studied in numerical studies below.

5 Numerical Studies

The usefulness of the proposed methods will require illustration of efficient and reliable performance in numerical studies, that are in accordance with the theory. To investigate their finite-sample performance, two studies are conducted. The first one is a Monte Carlo simulation study. The results reveal a number of features of the finite sample properties of the proportionately weighted estimator of β_0, as well as the practicability of the variance estimator based on the suggested resampling scheme. Being reported are coverage probabilities, under various censoring designs, and for different error distributions. The second one is an application to a study of initial wage effects on job duration, among a sample of young men.

5.1 Monte Carlo Study

The data are generated for a two-period panel in all cases. We use the logarithmic transform, $g(\cdot) = \log(\cdot)$, that corresponds with the AFT model. In the first experiment the data are generated as:

$$\log(y_{i1}) = x_{i1}\beta_0 + \alpha_i + \varepsilon_{i1}, \quad \log(y_{i2}) = x_{i2}\beta_0 + \alpha_i + \varepsilon_{i2}, \tag{5.20}$$

where $\beta_0 = 3$, $x_1 \sim N(0,1)$ and $x_2 \sim N(0,1)$. As in Lee (2004), the fixed-effect is generated by $\alpha = 0.5 \times x_1 + 0.5 \times x_2$. Using this common setup, we consider two distributions for the errors. In the first case, ε_1 and ε_2 are generated from a standard normal distribution, while in the second they are generated as independent variates from the Laplace distribution with density $\frac{1}{2}e^{-|\varepsilon|}$. Finally, the censoring thresholds were generated by specifying $c = e^{uk}$, and $c_2 = I\{y_1 \leq c\}(c - y_1)$, where $u \sim U[0,1]$ and the parameter k regulates the degree of censoring. Our results include designs with low censoring ($k = 9$ censors approximately 25% of all pairs) as well as heavy censoring ($k = 2$ censors approximately 60% of all pairs). The Monte Carlo results use $n = 300, 500, 1000$ for each experiment, replicating 1000 independent copies of this design.

The finite-sample properties of the estimator are considered under both the squared loss as well as the absolute deviations (LAD) loss functions. For the L_1 case, the estimators are obtained by linear programming as suggested in Koenker and D'Orey (1987) (see also Koenker and Portnoy (1999)). Using a preliminary least-squares or LAD estimate from just the observed data, in each replication the proportionately trimmed estimator $\hat{\beta}$ of Section 2 is then computed as a solution to the trimmed moment conditions. By replication we accumulate 1000 estimators $(\hat{\beta}_1, .., \hat{\beta}_{1000})$, which are used to construct an empirical distribution function (e.d.f) of $\hat{\beta}$.

To implement the resampling method for computing an estimator of the asymptotic standard error, the algorithm outlined in Section 3.1 is implemented. For the rth ($r = 1, .., 1000$)

replication of the data, an i.i.d n-vector

$$\{w_i : i = 1, ..., n\}, \quad w_i \sim Gam(1,1) \tag{5.21}$$

that satisfies $E\,(w_i) = Var\,(w_i) = 1$ is generated to construct a modified score function as described in the text, equation (3.12). Using $\hat{\beta}_r$ as an initial value, an estimator $\hat{\beta}_r^*$ is computed as the minimizer of the perturbed score function. For each replication this resampling is repeated 1000 times, yielding the sequence of estimators $\{\hat{\beta}_r^{*(1)}, ..., \hat{\beta}_r^{*(1000)}\}$ by repeated minimization of $U_n^*(\beta)$. The theoretical distribution function of $\hat{\beta}_r$ is approximated using this sequence, with variance numerically computed as $\hat{V}^{*(1)} = (1/1000)\sum_{q=1}^{1000}(\hat{\beta}_r^{*(q)} - \hat{\beta}_r)^2$, and a 95% confidence interval for $\hat{\beta}_r$ constructed in the usual way.

In the results below, we report the average censoring percent, mean bias, standard error computed using the resampling method (SD (Res.)), standard deviation of the empirical distribution of $\hat{\beta}$ (SD (e.d.f.)), 95% coverage probabilities (95% ECP) using the resampling standard error to compute the intervals, and quantiles of the data using the e.d.f. (the "lower" quantile (LQ) for the 25th quantile, "upper" quantile (UQ) for the 75th quantile). The median absolute error (MAE) is also given. Note, SD (Res.) is a simple average of the 1000 estimated standard errors $(\sqrt{\hat{V}^{*(1)}}, .., \sqrt{\hat{V}^{*(1000)}})$. The results of this experiment are summarized in Table 1.

Panel (A) gives results for normal errors. In the squared loss case, there is little difference between the resampling standard error estimator or the standard deviation of the empirical distribution. The coverage probabilities for the estimator are consistent with the theoretical distribution of the estimator, even for experiments with small n and heavy censoring (60%). The quantiles and MAE are each consistent with the theory, and the magnitudes of the MAE are indicative of no unfavorable tail problems in the e.d.f of the estimator. Qualitatively similar results are observed for the L_1 case. The resampling standard error is virtually indistinguishable from that of the empirical distribution, but these are each larger than the corresponding ones for squared-loss, which is an expected finding when the error has the standard normal distribution. The quantiles indicate that the empirical distribution function is more diffused relative to that with squared loss. The coverage probabilities remain reasonably correct however, even for an extremely large censoring case such as 60%.

In Panel (B) are results for Laplace errors, for which estimators based on L_1 are well known to have better performance than estimators based on squared loss. This is generally what we find. Using RMSE, we note that while the estimator was more efficient under squared loss than L_1 loss in with normal errors (Panel A), with Laplace errors the RMSE-efficiency is reversed. It is interesting to note that the length of the 95% interval is too large for the L_1 case with $n = 300$ and 25% censoring, but reduces towards the correct length as n increases; while the interval length is too small for the squared-loss case with $n = 300$ and 60% censoring, but increases to the correct size as n increases. The remaining features are quite similar to that of Panel A, including consistency and small MAE.

There is one other feature we draw attention to. The quantiles of the e.d.f. may be used to infer the shape of the sampling distribution of the estimator. Note, for both designs the 25th and 75th quantile[4] are approximately symmetric about the median, for both loss functions considered and all sample sizes, and thus consistent with the theoretical (normal) asymptotic distribution of the estimator. In conjunction with the virtual absence of any bias in our results, this suggests that the finite sample distribution of $\hat{\beta}$ provides a correct approximation to the theoretical distribution despite the trimming involved.

This finding is in contrast with a pronounced asymmetry in the finite sample distribution of other trimmed estimators, such as Powell's (1986). In fact, this discrepancy is due to the role of trimming in the two estimators. In Powell's case the asymmetry arises because the trimming condition has a first-order effect on the number of excluded observations, i.e., induces a discrepancy of the first-order in the number of exclusions for high versus low estimates. Here, the role of the trimming function $\gamma(\beta)$ is "second-order" in that the majority of the exclusion is done on observability of a pair, i.e., $y_1 + y_2 \leq c$, whereas the effect of $\gamma(\beta)$ is in large part the removal of pairs (y_1, y_2) that are observed, but whose trimmed sum is larger than c. It is largely this "second-order" effect of trimming that leads to the absence of the asymmetry in the finite sampling distribution of $\hat{\beta}$. Because the experimental designs are not directly comparable between Powell and (5.20), a precise gauge of this difference cannot be given. However, it is useful to note that for the "typical" experiment in Powell, the trimming indicator trims out approximately 50% of observations, while in our "typical" design the trimming leads to exclusion of only 12-15% of observations.

As an illustration of the analytical results developed in Section 5, we next conduct a specification test and report its results in Table 2. For the test, we assume the same design as (5.20), but generate the data differentially. As before, x_1 and x_2 are independent standard normal variates, (η, η_1, η_2) are also independent standard normal variates and

$$\alpha_i = 1, \quad \varepsilon_{ij} = \eta_{ij} + \eta_i \ (i = 1, ...n; \ j = 1, 2). \tag{5.22}$$

Thus, in this design α is independent of (x_1, x_2) and all other variates, while the errors ε_1 and ε_2 are mutually dependent. The censoring thresholds are generated by specifying $c = e^{uk}$, and $c_2 = I\{y_1 \leq c\}(c - y_1)$, where u is uniform and k controls the degree of censoring. Our results include designs with low censoring ($k = 13$ censors approximately 25% of all pairs) as well as heavy censoring ($k = 4.5$ censors approximately 60% of all pairs).

[4] We also compare the 10th and 90th quantiles in results not reported, and reach the same conclusion.

Table 2 gives results from specification tests of the above design, comparing the proposed estimator $\hat{\beta}$ against Powell's (1984) LAD estimator $\hat{\beta}_P$. As noted previously, the proposed estimator is consistent under both the null hypothesis as well as the alternative, while Powell's LAD (using first period data alone) is consistent under only the null. For the specification test, all results correspond to the case where the proposed estimator $\hat{\beta}$ is computed using the L_1 loss function of (2.10). Tests for several sample sizes and two censoring designs are considered. For each design, an initial pair of estimates $(\hat{\beta}_1, \hat{\beta}_{P1})$ is obtained from the estimating equations (2.10) and (4.17).[5] Using the exact weights as in (5.21) above, for each n and each censoring percent considered, a sequence of 1000 pairs $\left(\hat{\beta}_1^{*(1)}, \hat{\beta}_P^{*(1)}, ..., \hat{\beta}_1^{*(1000)}, \hat{\beta}_{P1}^{*(1000)}\right)$ is generated using the perturbed score equations (3.12) and (4.18), and the covariance matrix of $\left(\hat{\beta}_1 - \hat{\beta}_{P1}\right)$ constructed using given by equation (4.19), with $L = 1000$.

Table 2 gives the results of the specification test, reporting the probability of rejecting the null hypothesis (Type 1 error), where the probabilities are computed as the fraction of the 1000 pairs which reject the correctly specified null model. The test statistic has distribution χ^2_2, and is done at the 5% error level. Our results are indicative of the high accuracy of the proposed specification test, and the practicability of the suggested covariance matrix estimator of $\left(\hat{\beta}_1 - \hat{\beta}_{P1}\right)$. Even in small samples the rejection probability computed by the proposed test is approximately 5%, with the lower censoring design more accurate than the heavy censoring design, and the accuracy improves as n increases.

In summary, our results indicate that the estimator is consistent, with reliable performance in our numerical studies, performing well for both low and heavy censoring designs even in reasonably small samples. Our simulations are also indicative of the practicability of the resampling approach for standard error estimation, and the covariance matrix estimation required for specification tests. We next use the estimator in an application with actual data.

5.2 Empirical Application

We consider the initial wage effect on job durations in a sample of young men, using data from the National Longitudinal Survey of Youth (NLSY). We use data that have been analyzed previously in Horowitz and Lee (2002), whose results using a complementary method are compared to those proposed in our work. A number of data features characterize this application, including censored durations, dependently censored spells and unobserved heterogeneity.

A large literature on job duration and match theory has emerged with a number of styl-

[5] Note that in this model α is the intercept for Powell's estimator, thus $\hat{\beta}_P = (\hat{\alpha}, \hat{\beta}_{P1})$, where $\hat{\beta}_{P1}$ is the coefficient of x_1. Thus, in this case the common component $\tilde{\beta}_1$ between Powell's estimator and the proposed estimator is the coefficient of x (i.e., $\hat{\beta} = \hat{\beta}_1$).

ized facts. Low-skill (typically measured as low-education) workers have higher incidence of unemployment and thus shorter employment spells, lower rates of job tenure and also lower on-the-job propensities for new matches (see, e.g., Pissarides and Mortensen (1994), Topel and Ward (1992)). Such patterns indicate that high initial wage-offers (made to high-skill workers) are likely to be negatively correlated with the probability of job termination, and thus associated positively with longer observed durations. These facts are consistent with an older tradition (e.g., Singer and Spilerman (1976)) where workers may differ in their unobserved propensities to leave a job. However, if "movers" are characterized by low wage prospects relative to "stayers", then regression models may erroneously estimate a negative wage effect on job duration if unobserved heterogeneity is unaccounted for; see Topel and Ward (1992).

Even among competing ideas about the appropriate duration model specification in empirical research, one commonality has been the concern for incorporating heterogeneity in individual behavior.[6] Workers may not only be heterogeneous in unobserved skill quality, but along other dimensions such as propensities to move (Singer and Spilerman (1976)), each of which affect both durations as well as observed choice variates. Thus, regression may erroneously estimate the wage effect on job duration if such heterogeneity is unaccounted for. This heterogeneity is treated as the possibly nonlinear fixed-effect α. The NLSY data we use are collected over a period of fixed length, 1979-1996, over which the job durations of a sample of men are recorded successively, consistent with the dependent censoring modelling framework in this paper. Censoring of job spells is also present, although it is small (approximately 20%; see Table 3). Due to such concerns, methods for uncensored or independently censored data will be inappropriate modelling choices for the problem at hand, but we report those results for comparison. Because our methods require that the durations occur successively (i.e., a "pure renewal" process, with no time spent in other states), the censoring variate c is measured as the last interview date minus the start date of the first job minus time spent unemployed between jobs.[7]

The original NLSY sample of 3003 young men is further subsampled to homogenize the population, as done in Horowitz and Lee (2002). This subsample is limited to young men each of whom is a high school graduate, is a full time worker (working more than thirty hours a week) and works in the private sector. Individuals not meeting these criteria, as well as those individuals with suspect, erroneous or missing data are excluded, to yield a final sample of 890 individuals for whom we have data concerning their first two jobs, including initial real wage offer (in 1996 dollars). The study focuses on the first two-jobs only, although some individuals

[6] The importance of this is underscored in Heckman and Singer (1982), who state that "...every microeconomic study in which the hypothesis of heterogeneity is subject to test, it is not rejected".

[7] This is the same approach in Horowitz and Lee (2002), and a valid modification of the data so long as time spent outside the labor force is independent of the eventual duration of the job.

had additional job information.

Of the 890 individuals surveyed, 80% (707) have uncensored job durations in the first two jobs; see Table 3 (A). Simple correlations in the raw data suggest a weak positive correlation between wages and duration. In the entire sample, the mean of second job duration is 4.5 weeks larger than the first job, and the wage offer of the second job $.65 higher (in 1996 dollars), while amongst individuals with both completed spells, the mean job duration is 19 weeks larger, for a higher wage offer of $.61. To examine whether this correlation persists in regression, we use two covariates in our study: the initial real hourly wage and an indicator for second job. Other common covariates (e.g., race, marital status, geographic location) do not change at all or sufficiently over spells and were treated as a part of the time-invariant fixed effect.

To illustrate the presence of various features in the data, we compare results from a number of different methods and report the results in Table 3. Our results include (a) estimates of the Cox (1972) partial MLE; (b) an untrimmed version of the estimator we propose, i.e., first differences on all uncensored pairs with $g = \log$, on all uncensored pairs ("First Differences on Uncensored Pairs"); (c) we reproduce results from the complementary model where g is unspecified and ε is Type 1 (Horowitz and Lee (2002), Table 4); and (d) the proposed proportionately trimmed estimator. The simulation study has illustrated the practicability and accuracy of the resampling variance estimator, which we use to construct confidence intervals for our parameter estimates (using a series of 1000 resampling estimates). Because neither (c) nor (d) nests the other, we do not consider either to provide a natural "benchmark". However, we consider whether a score test of the null (which is a test for both models) can be rejected, using the test discussed in Section 3.

Our results are as follows (Table 3 (B)). Using conventional Wald tests, the coefficients of both initial hourly wages (β_{01}) as well as the second job indicator (β_{02}) using the proportionately trimmed estimator are positive and highly statistically significant . The direction of these effects are consistent with the Cox PMLE, Horowitz-Lee, as well as the untrimmed version of our estimator (column 3). The signs are consistent with search models that predict higher persistence in jobs with higher wages, even if there is for e.g., unobserved quality or moving propensities that are negatively correlated with initial wage offers. Because the estimates are obtained under different transformations, a more judicious comparison could be done using ratios of parameter estimates, β_{01}/β_{02}. Using ratios, the proposed method (.302) and the Cox partial likelihood (.335) are quite close, the "first-differences" estimator without trimming has ratio .243, but these are somewhat disparate from the Horowitz-Lee estimator (.045). In fact, the Horowitz-Lee estimator produces statistically insignificant parameter estimates for both covariates, which merits further analysis.

Let $\beta_0 = (\beta_{01}, \beta_{02})$. It is well known that the Wald tests and score tests are generally in agreement in detecting deviations from a null. Therefore, in addition to individual parameter significance in Table 3, we consider a score test for the null model ($\beta_0 = 0$) to compare the power of tests in competing models. In particular, under the dependent censoring scheme (2.3), consider the complementary model:

$$H(y_j) = x_j\beta_0 + \alpha + \tilde{\varepsilon}_j, \quad (j = 1, 2) \tag{5.23}$$

where $H(\cdot)$ is an unspecified strictly monotone function, possibly different from g, and $\tilde{\varepsilon}$ is a random error with specified probability distribution. An example of (5.23) are the models considered in Horowitz and Lee (2002, 2003) and Lee (2004) where $(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2)$ i.i.d conditional on (x_1, x_2, α). Although (5.23) and the model of this paper are specified under different assumptions about the data generating process, the two intersect at the null model ($\beta_0 = 0$) that can lay a foundation for evaluating the relative power of tests. This "equivalence" means that in the null model the observed durations are consistent with the data generating mechanism in either (2.2) or (5.23). To see this, note that the null model of (5.23) implies that the observed durations satisfy

$$y_j = H^{-1}(\alpha, \tilde{\varepsilon}_j)$$

for a strictly monotone function $\tilde{H}$, while the null model of (2.2) satisfies

$$g(y_j) = h(\alpha, \varepsilon_j).$$

Now, the score vector for the null hypothesis of model (2.2) has the form

$$U_n(0) = n^{-1}\sum_{\delta_{i2}=1} \Delta x_i \left[\psi\{g(z_{i2}) - g(z_{i1})\}\right] = n^{-1}\sum_{\delta_{i2}=1} \Delta x_i \left[\psi\{h_i(\alpha_i, \varepsilon_{i2}) - h_i(\alpha_i, \varepsilon_{i1})\}\right]$$

with $E(U_n(0)) = 0$ under Assumption A. Further, under the assumption that $(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2)$ i.i.d conditional on (x_1, x_2, α) used in the models of Horowitz and Lee (2002, 2003) and Lee (2004), the score vector $U_n(0)$ is equivalently a test of the null model of (5.23), since the i.i.d assumption implies that for the increasing function $g(H^{-1}(\cdot))$,

$$\begin{aligned} E(U_n(0)) &= n^{-1}E\left[\sum_{\delta_{i2}=1} \Delta x_i \left[\psi\{g(z_{i2}) - g(z_{i1})\}\right]\right] \\ &= n^{-1}E\left[\sum_{\delta_{i2}=1} \Delta x_i \left[\psi\{g(H^{-1}(\alpha_i, \tilde{\varepsilon}_{i2})) - g(H^{-1}(\alpha_i, \tilde{\varepsilon}_{i1})\}\right]\right] = 0. \end{aligned}$$

Thus, in the null model, a score test can be used to discriminate between the power of tests in the two models. Letting $V_n = Var(U_n(0))$, the score statistic for the null hypothesis is $S_n = nU_n(0)V_n^{-1}U_n(0)'$. By standard conditions for the sum of i.i.d observations, $\sqrt{n}U_n(0)$ is asymptotically normal and S_n is asymptotically χ^2 with p degrees of freedom ($p = \dim(x)$), so that

$$S_n \overset{d}{=} \chi^2_p, \qquad V_n = n^{-1} \sum_{i:\delta_{i2}=1} \Delta x_i^{\otimes 2} \left[\psi\left(g\left(z_{i2}\right) - g\left(z_{i1}\right)\right)\right]^2. \tag{5.24}$$

Using this formulation, the test statistic is found to be 13.03, while the critical $\chi^2_{(2)}$ value is 5.991. While the proportionately weighted estimator of this paper is found to strongly rejects the null model using either Wald tests or score tests, the alternate estimator in Horowitz and Lee (2003) cannot reject the null for conventional error levels. Although the rejection of the null cannot lead to "acceptance" of any alternative, this discrepancy is suggestive that hypotheses tests using variance estimators in Horowitz and Lee (2003) may have smaller power relative to the proportionately weighted estimator.[8]

As a complement to the above results, we also consider a specification test of model (5.23) under the null hypothesis that α is independent of the covariates. As in the numerical studies, the test is done by comparing the proposed estimator with Powell's LAD estimator (the LAD is computed using only first period data). Note that in this application x_2, the second period indicator, is equal to one only in the second period and zero otherwise. Using first period data in Powell's LAD implies that only β_{01} is identified, which is the coefficient of initial hourly wages. Thus, the test statistic has the χ^2_1 distribution. Using an initial pair of estimates from (2.6) and (4.17), a sequence of $L = 1000$ additional pairs from (3.12) and (4.18) and the covariance matrix estimator described in (4.19), the test statistic is 4.001 while the critical value at 5% error level is 3.841. Thus, the test rejects the null hypothesis of random effects.

6 Conclusions and Discussion

In this paper, we devise an extremely practicable estimator for a panel duration model with dependent censoring and fixed effects, including the case where the unobserved fixed-effects enter the model nonlinearly. For inference, a weighted resampling algorithm which repeatedly perturbs the minimand is suggested. Similarly to the bootstrap, this approach avoids the density estimation and numerical integration that would be required if an estimator of the theoretical variance was constructed directly. An extensive set of simulation studies, as well as an empirical indicates both the practicability and accuracy of this approach, even for reasonably small samples and high censoring.

While efficiency is not directly evaluated, we expect that the proposed estimator is generally semiparametrically inefficient. We speculate that this inefficiency is related to an asymptotically suboptimal amount of trimming in certain cases, induced by the suggested estimator. One

[8] Horowitz and Lee (2002) contain the empirical study, while it is omitted from Horowitz and Lee (2003).

example is a panel Tobit model with dependent censoring, which is a special case of (2.2) with $g(u) = u$. For this case, Assumption A implies that

$$E\ \left[\ \psi\left(y_2 - y_1 - \Delta x'\beta_0\right)\ I\left\{y_1 + y_2 < c\right\}|x_1, x_2\right] = 0. \tag{6.25}$$

Thus, if $(\varepsilon_1, \varepsilon_2, c)$ and $(\varepsilon_2, \varepsilon_1, c)$ are exchangeable conditionally on (x_1, x_2, α), an estimator that solves the sample analog of (6.25) will require no artificial trimming, whereas the estimator of Section 2.1 induces a strictly positive amount of trimming. In this sense, the estimator (2.9) makes inefficient use of the data.

There is possibly a separate source of inefficiency of the estimator that could arise in samples where no pairs of durations are subject to censoring. Conceptually, the form of the proposed estimator is such that a strictly positive fraction of observed pairs are always trimmed out of the estimation, except in the limiting case where the censoring threshold is so large (i.e., $c \to \infty$) that $\{y_1 + y_2\}$ is never censored. However, such trimming would inefficiently discard data if the censoring threshold is finite but all observed durations in the sample satisfy $\{y_1 + y_2\} < c$. This is because, for any strictly monotone transform g, under Assumption A the model satisfies

$$E\left[\ \psi\left(g(y_2) - g(y_1) - \Delta x'\beta_0\right)|x_1, x_2\right] = 0, \tag{6.26}$$

such that an estimator solving the sample analog of (6.26) requires no trimming; but the proposed estimator (2.9) involves a strictly positive amount of trimming which makes an inefficient use of the data. In actual empirical analysis, one may avoid the unnecessary trimming for each of these cases. For instance, in an empirical study with two-spells if $\delta_{1i} = \delta_{2i} = 1$ $(i = 1, ..., n)$ then one may consider an estimator of β_0 by solving the sample analog of (6.26) rather than that of (2.9). If, in addition, it is known that g is linear, then an estimator of β_0 could be obtained by solving the sample analog of (6.25). In each of these cases, the inference procedures will remain unchanged from the resampling approach developed in Section 3.

As discussed in the text, the model is complementary to one where the transform g is unspecified but the distribution function of ε is known, as being studied by Horowitz and Lee (2003). Another useful direction for future research is to consider a model specification where both the transform g and the distribution of ε are unspecified.

A Appendix

Proof of Theorem 1

Since L_n is a weighted sum of independent random variables (empirical process indexed with β), it follows from the uniform law of large numbers (Pollard, 1990) that L_n converges uniformly within any compact region to its expectation EL_n. Because $-\frac{\partial}{\partial\beta}EL_n(\beta) = -EU_n(\beta)$ is zero at $\beta = \beta_0$ and $-\frac{\partial}{\partial\beta} EU_n(\beta)\Big|_{\beta=\beta_0}$ is positive definite, there exists a neighborhood of β_0, inside which β_0 is the unique minimizer. Therefore, any minimizer of L_n inside the neighborhood must be strongly consistent. Moreover, if β_0 is the unique minimizer of EL_n in Q, and $\hat{\beta}$ is obtained as a minimizer of L_n in Q, then $\hat{\beta}$ must be strongly consistent. Next, under Assumptions A and C, it follows from Lemma 1 of Lai and Ying (1988) that, uniformly in β in any compact region,

$$U_n(\beta) - EU_n(\beta) = o\left(n^{-1/2+\xi}\right) \quad \text{for any } \xi > 0. \tag{A.27}$$

Unless otherwise stated, all approximations herein will be in the sense of almost sure convergence. Since $EU_n(\beta)$ is differentiable with a nonsingular slope matrix and $EU_n(\beta_0) = 0$, it follows that within a suitably small neighborhood of β_0, any minimizer $\hat{\beta}$ of $U_n(\cdot)$ must satisfy

$$||\hat{\beta} - \beta_0|| = o\left(n^{-1/2+\xi}\right) \text{ for any } \xi > 0. \tag{A.28}$$

In particular, $||\hat{\beta} - \beta_0|| = o\left(n^{-1/3}\right)$.

Next by Theorem 1 of Lai and Ying (1988), uniformly in $\beta, \tilde{\beta} \in B_n = \{b : ||b - \beta_0|| \le n^{-1/3}\}$

$$U_n(\beta) - U_n\left(\tilde{\beta}\right) = EU_n(\beta) - EU_n\left(\tilde{\beta}\right) + o\left(n^{-1/2}\right). \tag{A.29}$$

A tedious but otherwise straightforward differentiation gives

$$\begin{aligned} A = \frac{\partial}{\partial\beta} EU_n(\beta)\Big|_{\beta=\beta_0} = & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A.30)} \\ & - E\left[(\Delta x)^{\otimes 2} I(\gamma(\beta_0) < c)\psi'(h(\alpha,\epsilon_2) - h(\alpha,\epsilon_1))\right] - E\Big\{(\Delta x)^{\otimes 2} I(\gamma(\beta_0) < c) \\ & \Big[I\left(\Delta x'\beta_0 \ge 0\right) \int \psi\left(u - g\left(c - \overline{g}(u)\right)\right) \widetilde{f}\left(g\left(c - \overline{g}(u)\right) - \Delta x'\beta_0, u | x_1, x_2\right) du \\ & - I\left(\Delta x'\beta_0 \ge 0\right) \int \psi\left(g\left(c - \overline{g}(u)\right) - u\right) \widetilde{f}\left(u, g\left(c - \overline{g}(u)\right) + \Delta x'\beta_0 | x_1, x_2\right) du\Big]\Big\} \end{aligned}$$

From (A.29) and (A.30), we get the following asymptotically linear representation

$$U_n(\beta) = U_n(\beta_0) + A(\beta - \beta_0) + o\left(n^{-1/2}\right) \tag{A.31}$$

uniformly in $\beta \in B_n$. In particular,

$$U_n\left(\hat{\beta}\right) = o\left(n^{-1/2}\right) \tag{A.32}$$

and any two minimizers of $||U_n(\cdot)||$ must be asymptotically equivalent in the sense that their difference must be of order $o\left(n^{-1/2}\right)$. From (A.31) and under Assumption D we also get

$$\hat{\beta} - \beta_0 = -A^{-1} U_n(\beta_0) + o\left(n^{-1/2}\right). \tag{A.33}$$

Finally, since $U_n(\beta_0)$ is a sum of independent zero-mean random vectors under our assumptions, a standard multivariate central limit theorem implies that $n^{1/2}U_n(\beta_0)$ is asymptotically normal with mean zero and covariance matrix

$$V = E\left[(\Delta x)^{\otimes 2}\psi^2\left[h(\alpha,\varepsilon_2) - h(\alpha,\varepsilon_1)\right]I\{\gamma(\beta_0) < c\}\right].\blacksquare \tag{A.34}$$

Proof of Theorem 2

Since $E(w_i) = 1$, we have $EU_n^*(\beta) = EU_n(\beta)$. Analogously to (A.27), we get

$$U_n^*(\beta) - EU_n(\beta) = o\left(n^{-1/2+\xi}\right) \quad \text{for any } \xi > 0. \tag{A.35}$$

Therefore the same argument following (A.27) gives

$$||\hat{\beta}^* - \beta_0|| = o\left(n^{-1/2+\xi}\right) \text{ for any } \xi > 0. \tag{A.36}$$

Furthermore, uniformly in $\beta, \tilde{\beta} \in B_n$,

$$U_n^*(\beta) - U_n^*\left(\tilde{\beta}\right) = EU_n(\beta) - EU_n\left(\tilde{\beta}\right) + o\left(n^{-1/2}\right). \tag{A.37}$$

Thus, similar to (A.31),

$$U_n^*\left(\hat{\beta}^*\right) = U_n^*\left(\hat{\beta}\right) + A\left(\hat{\beta}^* - \hat{\beta}\right) + o\left(n^{-1/2}\right). \tag{A.38}$$

By their definitions, both $U_n^*\left(\hat{\beta}^*\right)$ and $U_n\left(\hat{\beta}\right)$ are of order $o\left(n^{-1/2}\right)$. Therefore, (A.38) implies

$$\begin{aligned} \hat{\beta}^* - \hat{\beta} &= -A^{-1}\left(U_n^*\left(\hat{\beta}\right) - U_n\left(\hat{\beta}\right)\right) & \text{(A.39)} \\ &= -\frac{1}{n}A^{-1}\sum_{i=1}^{n}(w_i - 1)\Delta x_i\left[\psi\left(g(z_{i2}) - g(z_{i1}) - \Delta x_i'\hat{\beta}\right)\ \right]I\left\{\gamma(\hat{\beta}) < c\right\}. \end{aligned}$$

Now, conditional on data, i.e., conditional on (x_i, z_i, c_i),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(w_i - 1)\Delta x_i\left[\psi\left(g(z_{i2}) - g(z_{i1}) - \Delta x_i'\hat{\beta}\right)\ \right]I\{\gamma\left(\hat{\beta}\right) < c\}$$

is a sum of zero-mean random vectors, with covariance matrix converging to (by the law of large numbers and consistency of $\hat{\beta}$),

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\Delta x_i)^{\otimes 2}\ \psi^2\left(g(z_{i2}) - g(z_{i1}) - \Delta x_i'\hat{\beta}\right)\ I\{\gamma\left(\hat{\beta}\right) < c_i\} = V.$$

This and (A.39) entail that $\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right)$ converges to $N\left(0, A^{-1}VA^{-1}\right).\blacksquare$

Proof of Theorem 3

By (A.33), we have

$$\hat{\beta} - \beta_0 = -A^{-1}\frac{1}{n}\sum_{i=1}^{n}\Delta x_i \psi\left(h\left(\alpha,\varepsilon_{i2}\right) - h\left(\alpha,\varepsilon_{i1}\right)\right) I\{\gamma\left(\beta_0\right) < c_i\} + o_p\left(n^{-1/2}\right) \tag{A.40}$$

which is approximately a sum of zero mean i.i.d random variables. For Powell's estimator, a similar representation holds:

$$\hat{\beta}_P - \tilde{\beta}_0 = -\tilde{A}^{-1}\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{i1}\ sgn\left(\tilde{\varepsilon}_{i1}\right) I\{\tilde{x}_{i1}'\tilde{\beta}_0 < c_i\} + o_p\left(n^{-1/2}\right) \tag{A.41}$$

where $\tilde{\varepsilon}_{i1}$ are error terms in Powell's LAD model that have zero median and $\tilde{A}$ is a $\tilde{p} \times \tilde{p}$ matrix (see Powell (1984) and Pollard (1990)). From (A.40), (A.41) and standard multivariate central limit theorem, it follows that $\sqrt{n}\left(\hat{\beta} - \beta_0,\ \hat{\beta}_P - \tilde{\beta}_0\right)$ converges jointly to a multivariate normal distribution with mean zero and covariance matrix

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \tag{A.42}$$

where $B_{11} = A^{-1}VA^{-1}$, and

$$B_{12} = B_{21}' = A^{-1}E\left[\Delta x\ \tilde{x}_1'\ \psi\left(h\left(\alpha,\varepsilon_2\right) - h\left(\alpha,\varepsilon_1\right)\right)\ \ \times sgn\left(\tilde{\varepsilon}_1\right) I\{\gamma\left(\beta_0\right) < c\} I\{\tilde{x}_1'\tilde{\beta}_0 < c\}\right]\tilde{A}^{-1}.$$

On the other hand, by (A.39)

$$\begin{aligned} \hat{\beta}^* - \hat{\beta} \quad &= \quad -A^{-1}\frac{1}{n}\sum_{i=1}^{n}\Delta x_i\ \psi\left(g\left(y_{i2}\right) - g\left(y_{i1}\right) - \Delta x_i'\hat{\beta}\right)\ \{\gamma\left(\hat{\beta}\right) < c_i\} \\ &\quad\ \times\left(w_i - 1\right) + o\left(n^{-1/2}\right). \end{aligned} \tag{A.43}$$

A similar representation can be shown as well for $\hat{\beta}_P^*$,

$$\hat{\beta}_P^* - \hat{\beta}_P = \tilde{A}^{-1}\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{i1}\ sgn\left(g\left(y_{i1}\right) - \tilde{x}_{i1}'\hat{\beta}_P\right) \times I\{\tilde{x}_1'\hat{\beta}_P < c_i\}\left(w_i - 1\right) + o\left(n^{-1/2}\right). \tag{A.44}$$

Conditional on data, both (A.43) and (A.44) are sums of zero mean random variables. By straightforward calculation of their conditional variance covariance matrix, which converges to B after suitable scaling, we conclude that $\sqrt{n}\left(\hat{\beta}^* - \beta_0,\ \hat{\beta}_P^* - \tilde{\beta}_0\right)$ converges to the same limiting normal distribution $N\left(0, B\right)$.■

References

[1] Abrevaya, J. I. (1999). Leapfrog Estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics*, 93, 203-228.

[2] Altonji, J and R. Matzkin (2001). Panel Data Models for Nonseparable Models with Endogenous Regressors. National Bureau of Economic Research, working paper T0267.

[3] Abbring, J. P.A. Chiappori and J. Pinquet (2003). "Moral hazard and dynamic insurance data", Journal of the European Economic Association, forthcoming.

[4] Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence, in J.J. Heckman and . Singer, editors, *Longitudinal analysis of labor market data*, Cambridge University Press, Cambridge.

[5] Cheng, S.C., Wei, L.J. and Ying, Z. (1995) Analysis of transformation models with censored data. *Biometrika* 82 835-845.

[6] Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser B* 34 187-202.

[7] Dharmalingam, A. (2002). Marriage, cohabitation and marital dissolution in New Zealand, *Journal of Population Research*, 19, 1-6.

[8] Efron, B. and R. Tibshirani (1993). *An introduction to the bootstrap.* Chapman and Hall, New York (1993)

[9] Han, A and J. A. Hausman (1990). Flexible Parametric Estimation of Duration and Competing Risk Models, *Journal of Applied Econometrics*, 5, 1-28.

[10] Hahn, J (1994). The efficiency bound of the mixed proportional hazards model. *Review of Economic Studies*, 61, 607-629.

[11] Heckman, J and B. Singer (1982). A method for minimizing the impact of distributional assumptions in econometric models of duration data. *Econometrica*, 52, 271-320.

[12] Honore, B. E (1992). Trimmed LAD and least squares estimation of censored and truncated models with fixed effects. *Econometrica*, 60, 533-565.

[13] Honore, B.E. and J.L. Powell (1994). Pairwise difference estimators for censored and truncated regression models, *Journal of Econometrics*, **64**, 247-278.

[14] Horowitz, J.L and S. Lee (2002). Semiparametric estimation of a panel data proportional hazards model with fixed effects, working paper, University of Iowa (http://cemmap.ifs.org.uk/docs/cwp2102.pdf)

[15] Horowitz, J.L and S. Lee (2003). Semiparametric estimation of a panel data proportional hazards model with fixed effects. *Journal of Econometrics*, 119, 155-198.

[16] Jin, Z., Ying, Z. and Wei, L.J. (2001). A simple resampling method by perturbing the minimand. *Biometrika,* 88, 381–390.

[17] Jovanovic, B (1979). Job Matching and the Theory of Turnover. *Journal of Political Economy*, 87, 972-990.

[18] Kalbfleisch, J.D. and Prentice, R.L. (2002) *The Statistical Analysis of Failure Time Data* (*2nd ed.*). Wiley, New York.

[19] Koenker, R. and D'Orey, V. (1987). Computing regression quantiles. *Appl. Statist.* 36, 383–93.

[20] Koenker, R. and Portnoy, S. (1997). The Gaussian Hare and the Laplacian Tortoise: computability of squared-error vs. absolute-error estimators. *Statist. Science*, 12, 279–300.

[21] Lai, T.L. and Ying, Z. (1988). Stochastic integral of empirical-type processes with applications to censored regression. *J. Multivariate Anal.* 27, 334-358.

[22] Lancaster, T. (2000). The incidental parameters problem since 1948. *Journal of Econometrics*, 95, 391-403.

[23] Lee, S (2004). Estimating Panel Data Duration Models with Censored Data. University College London working paper.

[24] Lin, D. Y. , Sun, W. and Ying, Z. (1999). Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika*, 86, 59–70.

[25] Lin, D. Y. , Wei, L.J. and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika,* 85, 605-618.

[26] Pakes, A and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57, 1027-1057.

[27] Pissarides, C and D. Mortensen (1994). Job Creation and Job Destruction in the Theory of Unemployment. *Review of Economic Studies*, 61, 397-415.

[28] Pollard, D. (1990), *Empirical Processes: Theory and Applications*, Regional Conference Series in Probability and Statistics 2, Hayward: Institute of Mathematical Statistics.

[29] Powell, J.L. (1984). Least Absolute Deviations for the Censored Regression Model. *Journal of Econometrics*, 25, 303-325.

[30] Powell, J.L (1986). Symmetrically trimmed least squares estimation for Tobit models. *Econometrica*, 54, 1435-1460.

[31] Ridder, G and I. Tunali (1999). Stratified partial likelihood estimation. *Journal of Econometrics*, 92, 93-132.

[32] Singer, B and S. Spilerman (1976). *Some Methodological Issues in the Analysis of Longitudinal Surveys.* Annals of Economics and Social Measurement, New York, Columbia University Press.

[33] Topel, R. and M. Ward (1992). Job Mobility and the Careers of Young Men. *Quarterly Journal of Economics*, 107, 439-479.

[34] Van den Berg, G (2002). Duration Models: Specification, Identification and Multiple Durations, Handbook of Econometrics, Volume 5. Elsevier. North Holland.

[35] Visser, M (1996). Nonparametric estimation of the bivariate survival function with application to vertically transmitted AIDS. *Biometrika*, 83, 503-518.

[36] Wang, W. and M. T. Wells (1998). Nonparametric estimation of successive durations under dependent censoring, *Biometrika*, 81, 561-572.

[37] Wei, L. J., D. Y. Lin and L. Weissfeld (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* 84, 1065–73.

Table 1: *Finite Sample Properties of the Proportionately Weighted Estimator*

	Cens. %	Mean Bias	SD (Res.)	SD (e.d.f)	RMSE	95% ECP	LQ	Median	UQ	MAE
				(A) Design 1 (Normal Errors), $\beta_0 = 3.00$						
Squared Loss:										
$n = 300$	25	.0034	.0816	.0850	.0816	.940	2.943	3.005	3.061	.0050
$n = 500$	25	.0007	.0607	.0606	.0607	.950	2.954	3.000	3.043	0
$n = 1000$	25	.0013	.0451	.0452	.0451	.945	2.970	3.000	3.030	0
$n = 300$	60	.0060	.1520	.1570	.1521	.939	2.904	3.000	3.099	0
$n = 500$	60	-.0032	.1194	.1200	.1194	.931	2.918	2.998	3.088	-.0020
$n = 1000$	60	-.0034	.0843	.0834	.0844	.944	2.941	2.997	3.049	-.0030
L_1 *Loss*:										
$n = 300$	25	.0001	.1069	.1050	.1069	.938	2.932	3.003	3.068	.0030
$n = 500$	25	.0026	.0784	.0850	.0784	.940	2.941	3.006	3.063	.0060
$n = 1000$	25	-.0015	.0583	.0550	.0583	.950	2.961	2.999	3.035	.0010
$n = 300$	60	.0035	.2100	.1980	.2100	.942	2.863	2.995	3.136	.0050
$n = 500$	60	-.0068	.1578	.1497	.1579	.930	2.890	2.995	3.086	.0050
$n = 1000$	60	.0011	.1100	.1070	.1100	.930	2.920	3.000	3.070	0
					(B) Design 2 (Double Exponential Errors), $\beta_0 = 3.00$					
Squared Loss:										
$n = 300$	25	.0028	.0775	.0880	.0800	.941	2.941	2.990	3.060	-.0010
$n = 500$	25	-.0008	.0604	.0627	.0604	.938	2.956	2.997	3.043	0
$n = 1000$	25	.0007	.0427	.0450	.0427	.940	2.969	2.998	3.030	-.0020
$n = 300$	60	.0059	.1600	.1730	.1601	.922	2.889	3.002	3.109	0
$n = 500$	60	.0004	.1250	.1266	.1250	.944	2.915	2.992	3.080	0
$n = 1000$	60	.0029	.0803	.0880	.0804	.944	2.941	2.990	3.060	-.0010
L_1 *Loss*:										
$n = 300$	25	-.0015	.0660	.0617	.6601	.964	2.957	2.999	3.037	-.0001
$n = 500$	25	.0003	.0490	.0463	.0490	.949	2.967	3.000	3.030	0
$n = 1000$	25	-.0028	.0334	.0337	.0335	.945	2.970	2.998	3.019	-.0002
$n = 300$	60	.0027	.1280	.1179	.1280	.949	2.928	2.999	3.077	-.0001
$n = 500$	60	-.0047	.0900	.0840	.0901	.955	2.941	2.990	3.050	-.0010
$n = 1000$	60	-.0003	.0679	.0660	.0679	.943	2.960	2.999	3.039	-.0001

Table 2: *Rejection Probabilities Computed for Specification Test*
Ho: α *independent of* x_1 x_2, H_1 : *Not* Ho
(*Nominal Level*=5%)

Sample Size	Censoring Percentage	Rejection Probability
$n = 300$	25	.048
$n = 500$	25	.051
$n = 1000$	25	.051
$n = 300$	60	.045
$n = 500$	60	.049
$n = 1000$	60	.051

Table 3: *Empirical Study of Wage Effects on Job Duration in the NLSY*[a]

(A) SUMMARY STATISTICS

	All Observations (n = 890)					Observations with Complete Spells (n = 707)				
	Mean	Median	Std. Dev	Min.	Max.	Mean	Median	Std. Dev	Min.	Max.
Censoring Time (weeks)	451.9	456	209.8	15.0	924	447.1	455.5	178	2.00	941
First job duration (weeks)	104.4	38.00	168.0	1.00	919	60.24	29.00	81.14	1.00	685
First Job Initial Hourly Wage	7.911	7.043	3.323	.798	23.7	6.907	7.742	3.153	.798	22.94
Second job duration (weeks)	108.9	43.00	163.4	1.00	906	79.87	37.00	113.1	1.00	827
Second Job Initial Hourly Wage	7.464	7.089	4.569	0	24.6	7.466	8.353	3.651	.6326	24.57

(B) EMPIRICAL RESULTS

	(1) Cox PMLE	(2) Horowitz-Lee	(3) First-Differences on Uncensored Pairs	(4) Proportionately Weighted Estimation
Initial Hourly Wage	.059	.0335	.049	.073
	(.008)	(.035)	(.014)	(.019)
Indicator for Second Job	.176	.766	.202	.236
	(.051)	(.584)	(.0511)	(.078)
Sample size in estimation	707	707	707	535

(C) SPECIFICATION TEST

	(1) Proportionately Weighted Estimator		(2) Powell LAD
Initial Hourly Wage	.0493		.1078
	(.025)		(.028)
		Test Statistic = 4.0001	
		$\chi^2_{(1)}$ =3.841 (n=832)	

[a] Wages are in 1996 dollars. Standard errors in parenthesis . The Cox PMLE and Horowitz-Lee results are reproduced from Table 4, Horowitz and Lee (2002). Regression estimates in the third column are from the proposed estimator without trimming. Fourth column standard error estimates obtained using the resampling method. Part (C) uses only first period data.