Nonparametric Identification and Estimation of Truncated Regression Models

Songnian Chen*
Department of Economics
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
Email: snchen@ust.hk

First Version, December 2002; this version, November 2005.

Keywords: Nonparametric identification, truncation, cross-section, panel data

1 Introduction

Traditionally truncated regression models have been identified and estimated with parametrically specified regression functions and error distributions. As is well known (see, for example, Powell (1994)), misspecification of the regression function or the error distribution in limited dependent variable models leads in general to inconsistent estimates, misleading inferences and erroneous predictions. Recent developments in the semiparametric and nonparametric estimation literature are largely motivated by sensitivity to misspecfications. For the identification and estimation of truncated regression models, Honoré and Powell (1994) and Lee (1994), among others, allowed for general error distribution in the cross-sectional case, and Honoré (1992) in the panel data case with fixed effects, but these studies imposed parametric specifications on the latent regression functions. Levy (2000), on the other hand, considered nonparametric identification and estimation of the latent regression function in the cross-sectional case, but he imposed normality on the error distribution. Consequently, the identification and estimation results from these earlier methods are sensitive to misspecification of the latent regression function or the error distribution. In this paper, we provide a comprehensive treatment of nonparametric identification of truncated

*I would like to thank Xiaohong Chen, Bo Honoré, Lung-Fei Lee, Oliver Linton, Arthur Lewbel, Jim Powell, the Co-Editor and three anonymous referees, and participants at the Triangle econometrics workshop, the 2003 North American Econometric Society Summer meetings and the Conference on Semi-parametric Methods for Survival and Longitudinal Data at NUS for many helpful comments. Lewbel kindly made available the Gauss code for Lewbel and Linton (2002).

regression models for both the cross-sectional and panel data settings and further propose nonparametric estimation procedures based on the identification results. Our results are therefore robust to misspecification of parametric structures of the model.

There has been some recent progress in nonparametric identification and estimation of the truncated regression model in the cross-sectional case. By employing a novel two-step identification procedure, Lewbel and Linton (2002) proposed a two-step nonparametric estimator, which can be implemented through two nonparametric regressions (and their derivatives) and a univariate integral. However, there are two major drawbacks associated with their approach. First, they require a strict monotonicity assumption in order for the mean regression based on truncated sampling to be invertible, which, in turn, is equivalent to a log-concavity assumption on the error distribution[1]. This strong shape restriction could potentially restrict the applicability of their estimator in empirical applications. Heckman and Honoré (1990) provided some discussions on situations when this concavity assumption is reversed and further provided some counter examples[2]. Another major drawback of the approach by Lewbel and Linton (2002) is the requirement of the presence of a continuous regressor, thus ruling out the case when all the regressors are discrete, which is common in empirical applications.

In this paper, we first present an identification result for the underlying regression function in the cross-sectional case. The main insight behind this identification result is the observation that the basic shape of the error distribution is preserved under truncation beyond any truncation point, subject to rescaling; thus, the truncated distributions based on the same underlying error distribution with different truncation points belong to an equivalence class, subject to a location shift and rescaling. Equivalently, the truncated hazard functions corresponding to different truncation points are linked through a location shift. We establish an identification result for the latent regression function by exploiting this equivalence result under a non-periodicity[3] assumption, without imposing any strong shape restriction on the error distribution or requiring the presence of a continuous regressor, thus avoiding the forementioned two main drawbacks of Lewbel and Linton (2002). Heuristically, the strength of our approach in comparison with Lewbel and Linton (2002) is that we achieve identification by exploiting the full knowledge of the truncated distributions[4], whereas Lewbel and Linton (2002) only rely on the knowledge of the first two truncated moments. We further propose a consistent estimator for the underlying regression function based on this identification result.

While our identification result does not require a strong shape restriction on the underlying error distribution or the presence of a continuous regressor by exploiting the mapping between

[1] Their log-concavity assumption is somewhat weaker than that of Honoré and Powell (1994). Still, it imposes a strong shape restriction on the underlying error distribution; namely, they require strict concavity of $\int_t^\infty (1-F(v))dv$, where F is the CDF of the error term. See Abrevaya (2000), Heckman and Honoré (1990) and Levy (2000) for some detailed discussions.

[2] It is straightforward to construct counter-examples to this log-concavity assumption; in particular, this assumption fails when any portion of the right tail of the error distribution violates this log-concavity assumption.

[3] Horowitz (1998) and Ichimura (1993) noted the usefulness of non-periodicity in identification in the context of single-index models.

[4] In a way, this is equivalent to making use of an infinite number of moments.

truncated distributions at different truncation points through some location shift and rescaling, or equivalently, by exploiting the mapping between truncated hazard functions at different truncation points through some location shift, there is still a possibility of non-uniqueness of such mappings when the error distribution possesses a particular periodicity property. It is possible, however, to overcome such non-uniqueness to achieve identification, when the above-mentioned non-periodicity fails, by exploiting a smoothness property when all the regressors are continuous, or by making use of both the smoothness property and a large support condition on the underlying regression function[5] when both continuous and discrete regressors are present.

For the panel data model with fixed effects, Porter (1997) considered nonparametric estimation of the regression function when there is no censoring or truncation. Abrevaya (2000) and Honoré (1992) considered semiparametric estimation of the fixed-effect panel data regression model under truncation with a linear specification for the underlying regression function. To the best of our knowledge, nonparametric identification and estimation of the truncated regression model with panel data has not been considered in the literature. Similar to the cross-sectional case, we show that the underlying regression function can be identified under a non-periodicity condition on the error distribution and that identification is still possible in the presence of a continuous regressor even when this non-periodicity condition fails. In addition, we further propose a nonparametric estimation procedure for the underlying regression function. Our approach in the panel data case could be viewed as a bivariate extension of our results for the cross-sectional case.

The paper is organized as follows. Section 2 presents our identification results. Section 3 introduces our estimators and analyzes their large sample properties. Section 4 reports some simulation results. Section 5 concludes the paper. The appendix contains some mathematical proofs.

2 Identification

In this section, we describe the models and provide identification results. These results are useful for motivating our estimation procedures proposed in the next section.

We first consider the cross-sectional case. Suppose the observed data are generated based on the latent response variable, Y^*, defined by

$$Y^* = m_0(X) + \varepsilon \tag{2.1}$$

where X is a d-dimensional column vector of regressors, $m_0(\cdot)$ is the unknown regression function, and ε is the unobservable error term independent of X. For the truncated model, our observations consist of a random sample $\{(Y_i, X_i)\colon\ i = 1, 2, ..., n\}$ from the conditional distribution of (Y^*, X) given the event $\{Y^* > 0\}$.

[5] Horowitz (1998) and Ichimura (1993) noted that less-stringent conditions are needed for identification in single-index models when all regressors are continuously distributed. Furthermore, Horowitz (1998), Ichimura (1993) and Manski (1988) noted the usefulness of the presence of a continuous regressor and a large support of the regression function for identification in the context of single-index and binary choice models when discrete regressors are present.

To motivate our identification approaches, define

$$S^*(t,x) = E^*[1\{Y \geq t\}|X = x] = \frac{S(t - m_0(x))}{S(-m_0(x))} \tag{2.2}$$

for $x \in \Omega$, where Ω is the support of X, E^* denotes the expectation operator based on the truncated distribution, and S is the survival functions of ε. Assume[6] that $\Pr(Y^* > 0|X = x) > 0$ for each $x \in \Omega$. We focus on the identification and estimation of $m_0(x)$ up to a location since no location restriction is imposed on the error distribution[7]. Set $m_0(x_0) = 0$, for some $x_0 \in \Omega$. The main insight behind our identification results is that the basic shape of the error distribution beyond the truncation point is preserved under truncation[8]; namely, the error survival functions at two different truncation points are linked through a rescaling factor in their overlapping domain, which, in turn, implies the truncated conditional survival functions of Y given $X = x$ and $X = x_0$, $S^*(t,x)$ and $S^*(t,x_0)$ are linked through a rescaling factor together with a location shift. Suppose $m_0(x) \geq 0$; then, $S^*(t,x)$ and $S^*(t,x_0)$ are linked through the following location-scale relationship: $S^*(t,x_0) = \beta_0 S^*(t+\beta_1, x)$ for $t \geq 0$, where $\beta_0 = 1/S^*(m_0(x), x)$ and $\beta_1 = m_0(x)$, namely,

$$S^*(t,x_0) = \frac{S^*(t + m_0(x), x)}{S^*(m_0(x), x)}. \tag{2.3}$$

Similarly, for $m_0(x) < 0$, we obtain $S^*(t,x) = \beta_{01} S^*(t+\beta_{11}, x_0)$ for $t \geq 0$ with $\beta_{01} = 1/S^*(-m_0(x), x_0)$ and $\beta_{11} = -m_0(x)$, or

$$S^*(t,x) = \frac{S^*(t - m_0(x), x_0)}{S^*(-m_0(x), x_0)}. \tag{2.4}$$

Equations (2.3) and (2.4) form the basis for our identification and estimation results. Define

$$T(m) = T_1(m)1\{m \geq 0\} + T_2(m)1\{m < 0\} \tag{2.5}$$

where

$$T_1(m) = \int_0^\infty \left[S^*(t,x_0) - \frac{S^*(t+m,x)}{S^*(m,x)} \right]^2 w_1(t)dt$$

and

$$T_2(m) = \int_0^\infty \left[S^*(t,x) - \frac{S^*(t-m,x_0)}{S^*(-m,x_0)} \right]^2 w_1(t)dt$$

and $w_1(t)$ is a positive integrable function on $[0,\infty)$. Clearly, $T(m_0(x)) = 0$. Our first identification result states that $T(m)$ has a unique minimizer under a non-periodicity assumption on the error distribution.

[6] Note that, given the nonparametric treatment in the paper, extrapolation on $m(x)$ is not permitted for x with $P(Y^* > 0|X = x) = 0$.

[7] With truncated sampling, quantile restrictions are not useful in identifying $m_0(x)$ as the scale factor $1/S(-m_0(x))$ cannot be identified. Under the conditional zero mean restriction, the location factor can be identified only by relying on the tail distribution of the error term based on identification at infinity.

[8] Note that, conditional on $X = x$, ε is truncated at $-m_0(x)$.

Lemma 1: For some $x \in \Omega$, if the hazard function of ε, $\lambda(t) = -d \ln S(t)/dt$, is not periodic on $(\max\{0, -m_0(x)\}, \infty)$, then $m_0(x)$ is identified.

Proof: Suppose that $m_0(x) \geq 0$. We will show that $m_0(x)$ is the unique solution to $T(m) = 0$. Suppose that there is an $\tilde{m}_x \neq m_0(x)$ such that $T(\tilde{m}_x) = 0$. Further suppose that $\tilde{m}_x \geq 0$. Then, it follows that $S^*(t, x_0) = \frac{S^*(t+\tilde{m}_x, x)}{S^*(\tilde{m}_x, x)}$ for all $t \geq 0$. Some simple algebra shows that $\lambda(t) = \lambda(t + \tilde{m}_x - m_0(x))$, which implies that λ must be a periodic function on $(0, \infty)$, leading to a contradiction. We could use similar arguments to derive a similar contradiction if $\tilde{m}_x < 0$. The case with $m_0(x) < 0$ is treated analogously.

Remark 1: In contrast to the monotonicity assumption of the conditional regression based on truncated sampling by Lewbel and Linton (2002), the assumption that $\lambda(\cdot)$ is not a periodic function on $(\max\{-m_0(x), 0\}, \infty)$ does not impose a strong shape restriction on the underlying error distribution. In particular, their monotonicity assumption is equivalent to the condition that $\int_t^\infty S(v)dv$ is log-concave, or, equivalently, $S(t)/\int_t^\infty S(v)dv$ is monotonically decreasing, which, in turn, implies that $S(t)/\int_t^\infty S(v)dv$ cannot be periodic, or equivalently, that $\lambda(\cdot)$ cannot be periodic on $(\max\{-m_0(x), -m_0(x_0)\}, \infty)$. Thus, the monotonicity assumption by Lewbel and Linton (2002) is sufficient for our above identification result.

Remark 2: Let $\lambda(t|x) = -d \ln S^*(t, x)/dt = \lambda(t - m_0(x))$ denote the conditional hazard function of Y given $X = x$. In the proof of Lemma 1, we have shown that the location-scale restriction on the conditional survival functions based on truncated sampling is equivalent to a location shift restriction on the conditional hazard functions. Intuitively, the above non-periodicity assumption implies the uniqueness of the location shift parameter for the two conditional hazard functions. If the error hazard function, λ, is indeed periodic on $(\max\{0, -m_0(x)\}, \infty)$, then the arguments used in the proof of the above lemma actually show that $m_0(x)$, in addition to the location normalization, are identified up to an integer multiple of the corresponding period. In the next lemma, we show that this multiplicity problem does not arise in the continuous case; in other words, identification is still possible even if the error hazard function is periodic in its right tail.

Lemma 2: Assume that X is continuously distributed with a convex support Ω and that $m_0(\cdot)$ is a continuous function. Then, $m_0(x)$ can be identified for $x \in \Omega$ if error hazard function $\lambda(\cdot)$ is not a constant on $(\max\{0, -m_0(x)\}, \infty))$, that is, if $\lambda(\cdot)$ does not have an exponential right tail.

Proof: We provide only the details for the case with $m_0(x) \geq 0$, as the case with $m_0(x) < 0$ involves similar arguments. If $\lambda(\cdot)$ is not a periodic function on $(0, \infty)$, then the identification of $m_0(x)$ follows from Lemma 1. Now suppose that $\lambda(\cdot)$ is indeed a periodic function on $(0, \infty)$ with a positive period γ_{c0}. Define a functional, $T^f(m) = T_1^f(m) + T_2^f(m)$, of a function $m(\cdot)$ (with a slight abuse of notation here), where

$$T_1^f(m) = \int \int_0^\infty \left[S^*(t, x_0) - \frac{S^*(t + m(\tilde{x}), \tilde{x})}{S^*(m(\tilde{x}), \tilde{x})} \right]^2 1\{m(\tilde{x}) \geq 0\} w_2(\tilde{x}) w_1(t) dt d\tilde{x}$$

and

$$T_2^f(m) = \int \int_0^\infty \left[S^*(t, \tilde{x}) - \frac{S^*(t - m(\tilde{x}), x_0)}{S^*(-m(\tilde{x}), x_0)} \right]^2 1\{m(\tilde{x}) < 0\} w_2(\tilde{x}) w_1(t) dt d\tilde{x}$$

and w_1 is defined as above and w_2 is a positive weight function such that $\int w_2(\tilde{x})d\tilde{x} < \infty$. Clearly $T(m_0) = 0$. Now suppose that there exists another continuous function, $\tilde{m}(\cdot)$, such that $T^f(\tilde{m}) = 0$. Define $x(u) = (1-u)x_0 + ux$, for $u \in [0,1]$; thus $x(0) = x_0$ and $x(1) = x$. By following the arguments in the proof of Lemma 1 and noting Remark 2, we can show that $\lambda(t) = \lambda(t + \tilde{m}(x(u)) - m_0(x(u))$ and $\tilde{m}(x(u)) = m_0(x(u)) + \eta(u)\gamma_{c0}$ for any $u \in [0,1]$, where $\eta(u)$ takes on integer values, which can be different for different $u \in [0,1]$. Due to the continuity of $\tilde{m}(\cdot)$ and $m_0(\cdot)$, however, we can deduce that $\eta(u)$ must be a constant term, say, η_0. Furthermore, we have $\eta_0\gamma_{c0} = 0$ since $\tilde{m}(x_0) = m_0(x_0) = 0$; thus, $\eta_0 = 0$. Hence, $\tilde{m}(x(u)) = m_0(x(u))$ for any $u \in [0,1]$. Therefore, $m_0(x)$ is identified.

Remark 3: Note that $\lambda(t|x) = \lambda(t - m_0(x))$ can be identified based on the knowledge of $S^*(t,x)$ with truncated sampling. If $\lambda(\cdot)$ and $m_0(\cdot)$ are differentiable, then, similar to the arguments in Horowitz (1996), we can show that $\frac{\partial m_0(x)}{\partial x_k} = -\int \frac{\partial\lambda(t|x)}{\partial x_k}\frac{\partial\lambda(t|x)}{\partial t}w_1(t)dt / \int [\frac{\partial\lambda(t|x)}{\partial t}]^2 w_1(t)dt$, provided that $\int [\frac{\partial\lambda(t|x)}{\partial t}]^2 w_1(t)dt > 0$, where x_k is the k-th component of x. Thus, $\frac{\partial m_0(x)}{\partial x_k}$ can be identified as long as $\lambda(\cdot)$ does not have an exponential tail on $(-m_0(x), \infty)$, which is a weaker identification condition than that in Lewbel and Linton (2002) for identifying the partial derivatives of $m_0(\cdot)$.

Remark 4: An exponential tail for the error distribution can also be problematic for truncated regression models even in the semiparametric case in which m_0 has a linear specification. Suppose that $\lambda(t) = \lambda_0$, a constant, when $t \in (t_0, \infty)$ for some t_0 and $m_0(x) = x'\beta$. Then, under truncated sampling, we have $S^*(t,x) = e^{-\lambda_0 t}$ if $-x'\beta > t_0$. Therefore, the observations with $-X'\beta > t_0$ and $t > t_0$ are not informative about the finite dimensional parameter, β, under truncated sampling.

Finally, we examine the usefulness of continuous regressors in identification when discrete regressors are present. In particular, we show that identification is still possible even if the error hazard function is periodic in its right tail when a continuous regressor is present and the support of the underlying regression function is large enough.

Lemma 3: Let $X = (X_c, X_d)$ with its support $\Omega = \Omega_c \times \Omega_d$, where X_c and X_d denote the continuous and discrete components, respectively, Ω_c is a convex set and Ω_d is finite. Suppose that there exists an $x_c^* \in \Omega_c$, such that $\lambda(\cdot)$ is not a periodic function[9] on $(-m_0(x_c^*, x_d), \infty)$ and $m_0(x_c, x_d)$ is a continuous function of x_c for any $x_d \in \Omega_d$. For any $x_0 = (x_{c0}, x_{c0})$ and $x = (x_c, x_d)$ in Ω, if $\lambda(\cdot)$ is not a constant on $(\max\{-m_0(x), -m_0(x_0)\}, \infty)$, then $m_0(x) - m_0(x_0)$ is identified[10].

Proof: Note that $m_0(x_c^*, x_d) - m_0(x_c^*, x_{d0})$ is identified by Lemma 1 since $\lambda(\cdot)$ is not periodic on $(\max\{-m_0(x_c^*, x_d), -m_0(x_c^*, x_{d0})\}, \infty)$. As $\lambda(\cdot)$ is not a constant on $(\max\{-m_0(x_c, x_d), -m_0(x_c^*, x_d)\}, \infty)$ or $(\max\{-m_0(x_{c0}, x_{d0}), -m_0(x_c^*, x_{d0})\}, \infty)$, $m_0(x_c, x_d) - m_0(x_c^*, x_d)$ and $m_0(x_{c0}, x_{d0}) - m(x_c^*, x_{d0})$ are identified by Lemma 2. Consequently, we can identify $m(x_c, x_d) - m(x_{c0}, x_{d0})$ by combining these three terms.

Now, we turn to the panel data model with fixed effects. In this case the observed data are

[9] It is straightforward to show that there exists a constant, t_0, such that λ cannot be a periodic function on $[-t_0, \infty)$ if the support of ε is the real line. Therefore, we could pick an x_c^* such that $m_0(x_c^*, x_d) > t_0$, which is possible if $\sup_{x_c} m_0(x_c, x_d)$ is sufficiently large for every $x_d \in \Omega_d$.

[10] Note that no location normalization is imposed for this lemma.

generated based on the latent variables, Y_t^*, given by

$$Y_t^* = \alpha + m_0(X_t) + \varepsilon_t \qquad \text{for} \quad t = 1, 2, \tag{2.6}$$

where X_1 and X_2 are d-dimensional column vectors of regressors, $m_0(\cdot)$ is an unknown regression function, α is the fixed effect, and ε_1 and ε_2 are the unobservable error terms. In this case, our observations $\{(Y_{it}, X_{it})\colon\ i = 1, 2, ..., n,\ t = 1, 2\}$ consist of a random sample generated from the conditional distribution of $\{Y_t^*, X_t,\ t = 1, 2\}$ given the event $\{Y_1^* > 0 \text{ and } Y_2^* > 0\}$. We focus on the case with two time periods, but our results could be extended to more general cases.

We now consider identification under the conditional pairwise exchangeability condition (e.g., Honoré and Kyriazidou (2000)) that $(\varepsilon_1, \varepsilon_2)$ and $(\varepsilon_2, \varepsilon_1)$, or equivalently, $(\varepsilon_1^*, \varepsilon_2^*)$ and $(\varepsilon_2^*, \varepsilon_1^*)$ with $\varepsilon_1^* = \alpha + \varepsilon_1$ and $\varepsilon_2^* = \alpha + \varepsilon_2$, are identically distributed conditional on (α, X_1, X_2). Let Ω_p denote the support of $X = (X_1, X_2)$ and $P(Y_1^* > 0, Y_2^* > 0 | X_1 = x_1, X_2 = x_2) > 0$ for any $(x_1, x_2) \in \Omega_p$. Similar to Honoré (1992), the pairwise exchangeability implies that the conditional distribution of (Y_1^*, Y_2^*) given $(X_1 = x_1, X_2 = x_2)$ is symmetric around the 45^0 line through $(0, m_0(x_2) - m_0(x_1))$. The main insight for our identification results is the observation that such conditional symmetry[11] is preserved under truncation. As will become clear later, the identification results for the panel data case could be viewed as an extension of those for the cross-sectional case.

For $(x_0, x) \in \Omega_p$, define the joint conditional survival function of (Y_1, Y_2) under truncation,

$$G(s_1, s_2, x_0, x) = E^*[1\{Y_1 > s_1, Y_2 > s_2\} | X_1 = x_0, X_2 = x] = \frac{S_p(s_1 - m_0(x_0), s_2 - m_0(x), x_0, x)}{S_p(m_0(x_0), -m_0(x), x_0, x)},$$

where $S_p(s_1, s_2, x_0, x) = E(\varepsilon_1^* > s_1, \varepsilon_2^* > s_2 | X_1 = x_0, X_2 = x)$. Under the conditional pairwise exchangeability, we have

$$G(s_1, s_2 + m_0(x) - m_0(x_0), x_0, x) = G(s_2, s_1 + m_0(x) - m_0(x_0), x_0, x) \tag{2.7}$$

for $m_0(x) - m_0(x_0) \geq 0$ and

$$G(s_1 - (m_0(x) - m_0(x_0)), s_2, x_0, x) = G(s_2 - (m_0(x) - m_0(x_0)), s_1, x_0, x) \tag{2.8}$$

for $m_0(x) - m_0(x_0) < 0$ for any $s_1 > 0$ and $s_2 > 0$. Define location-shifted versions of G as $G_m^1(s_1, s_2, x_0, x) = G(s_1, s_2 + m, x_0, x)$ and $G_m^2(s_1, s_2, x_0, x) = G(s_1 - m, s_2, x_0, x)$. It is easy to show that $G^1_{m_0(x) - m_0(x_0)}(s_1, s_2, x_0, x)$ is pairwise exchangeable in (s_1, s_2) if $m_0(x) - m_0(x_0) \geq 0$, and
$G^2_{m_0(x) - m_0(x_0)}(s_1, s_2, x_0, x)$ is pairwise exchangeable in (s_1, s_2) if $m_0(x) - m_0(x_0) < 0$. This location shift restriction can be thought of as a bivariate extension of the location shift restriction utilized for identification in the cross-sectional case.

Similar to the identification approach adopted for the cross-sectional case, define

$$T_p(m) = T_{p1}(m) 1\{m \geq 0\} + T_{p2}(m) 1\{m < 0\} \tag{2.9}$$

[11] Note that Honoré (1992) exploited only the moment conditions related to the first two moments based on the symmetry, whereas we exploit the entire symmtry property of the conditional distribution based on pairwise exchangeability.

where

$$T_{p1}(m) = \int_0^\infty \int_0^\infty [G_m^1(s_1, s_2, x_0, x) - G_m^1(s_2, s_1, x_0, x)]^2 w_1(s_1) w_1(s_2) ds_1 ds_2$$

and

$$T_{p2}(m) = \int_0^\infty \int_0^\infty [G_m^2(s_1, s_2, x_0, x) - G_m^2(s_2, s_1, x_0, x)]^2 w_1(s_1) w_1(s_2) ds_1 ds_2$$

As noted above, the pairwise exchangeability implies that the joint distribution of $(\varepsilon_1, \varepsilon_2)$ is symmetric around the 45^0 line through the origin. Our next lemma rules out other 45^0 lines about which the joint conditional error distribution is symmetric.

Lemma 4: For $(x_0, x) \in \Omega_p$, if the 45^0 line through the origin is the only 45^0 line about which $(\varepsilon_1^*, \varepsilon_2^*)$ is symmetric, given $(X_1, X_2) = (x_0, x)$ in the region $(-m_0(x_0), \infty) \times (-m_0(x), \infty)$, then $m_0(x) - m_0(x_0)$ is identified.

Proof: Clearly, $T_p(m_0(x) - m_0(x_0)) = 0$. Suppose that there exists an $\tilde{m}^p_{x_0x} \neq m_0(x) - m_0(x_0)$ such that $T_p(\tilde{m}^p_{x_0x}) = 0$. Assuming that $m_0(x) - m_0(x_0) \geq 0$ and $\tilde{m}^p_{x_0x} > m_0(x) - m_0(x_0)$ (other cases involve similar arguments, thus omitted), then we can obtain

$$S_p(t_1 - m_0(x_0), t_2 + \tilde{m}^p_{x_0x} - m_0(x), x_0, x) = S_p(t_2 - m_0(x_0), t_1 + \tilde{m}^p_{x_0x} - m_0(x), x_0, x)$$

for $t_1 \geq 0, t_2 \geq 0$, or $S_p(t_1, t_2 + \delta_{x_0x}, x_0, x) = S_p(t_2, t_1 + \delta_{x_0x}, x_0, x)$ for $\delta_{x_0x} = \tilde{m}^p_{x_0x} - (m_0(x) - m_0(x_0))$, $t_1 \geq -m_0(x_0)$ and $t_2 \geq -m_0(x_0)$, or, equivalently, $S_p(t_1, t_2, x_0, x) = S_p(t_2 - \delta_{x_0x}, t_1 + \delta_{x_0x}, x_0, x)$ for $t_1 \geq -m_0(x_0)$, $t_2 \geq \delta_{x_0x} - m_0(x_0)$. Therefore, $S_p(\cdot, \cdot, x_0, x)$ is also symmetric around the 45^0 line through $(0, \delta_{x_0x})$ in the region $(-m_0(x_0), \infty) \times (-m_0(x), \infty)$, leading to a contradiction. Thus, $m_0(x) - m_0(x_0)$ is identified.

Remark 5: Similar to Lemma 1 for the cross-sectional case, Lemma 4 implies that the regression function is identified under a non-periodicity condition. Specifically, both lemmas require some form of non-periodicity on the underlying error hazard function or density (distribution) function. Indeed, we can establish a relationship between univariate periodic functions and bivariate functions symmetric around 45^0 lines. Suppose that $\xi(e_1, e_2)$ is symmetric around the two 45^0 lines through $(0, 0)$ and $(0, e_2^*)$, respectively, for some $e_2^* \neq 0$, namely,

$$\xi(e_1, e_2) = \xi(e_2, e_1) \text{ and } \xi(e_1, e_2) = \xi(e_2 - e_2^*, e_1 + e_2^*).$$

For any fixed point, (e_{10}, e_{20}), define $g(u) = \xi(e_{10} + u, e_{20} - u)$, thus $g(u)$ takes on the values of $\xi(\cdot, \cdot)$ on the 135^0 line through (e_{10}, e_{20}) when u changes. By the symmetry property of $\xi(\cdot, \cdot)$, we have

$$g(u) = \xi(e_{20} - u, e_{10} + u) = \xi(e_{10} + u - e_2^*, e_{20} - u + e_2^*) = g(u - e_2^*).$$

Thus, $g(u)$ is a periodic function; in other words, $\xi(e_1, e_2)$ is a periodic function along the 135^0 lines. Therefore, identification based on Lemma 4 requires a non-periodicity property on the conditional error distribution.

Remark 6: If $\Omega_p = \Omega_{p1} \times \Omega_{p2}$, then for any $(x_1, x_2) \in \Omega_p$, we can use Lemma 4 to identify $m_0(x_2) - m_0(x_1)$. For any $(x_1, x_2) \in \Omega_{p1} \times \Omega_{p1}$, we thus can pick an $x_0 \in \Omega_{p2}$ and use Lemma

4 to identify $m_0(x_2) - m_0(x_0)$ and $m_0(x_1) - m_0(x_0)$, which implies that $m_0(x_1) - m_0(x_2)$ is also identified. Therefore, it is not necessary for Ω_{p1} and Ω_{p2} to have a non-empty overlapping set for identification based on Lemma 4.

Remark 7: If ε_1 and ε_2 are independent and identically distributed conditional on (X_1, X_2, α), Honoré (1992) showed that, with a linear specification for $m_0(x)$, the linear regression coefficients are identified if the conditional marginal density of ε_1 and ε_2 is strictly log-concave. This is, indeed, sufficient for identifying $m_0(x) - m_0(x_0)$ nonparametrically. Let $S_{\varepsilon_1,\varepsilon_2}(s, \alpha, x_0, x)$ denote the marginal conditional survival function of ε_1 and ε_2 given α, $X_1 = x_0$, and $X_2 = x$. Note that

$$G(s_1, s_2 + m, x_0, x,) - G(s_2, s_1 + m, x_0, x_1) = \frac{E_{\alpha|x_0,x}[S^1(s_1, s_2, m, \alpha, x_0, x)S^2(s_1, s_2, m, \alpha, x_0, x)]}{S_p(s_1 - m_0(x_0), s_2 - m(x), x_0, x)},$$

where

$$S^1(s_1, s_2, m, \alpha, x_0, x) = S_{\varepsilon_1,\varepsilon_2}(s_2 - \alpha - m_0(x_0), \alpha, x_0, x)S_{\varepsilon_1,\varepsilon_2}(s_2 + m - \alpha - m_0(x), \alpha, x_0, x)$$

and

$$S^2(s_1, s_2, m, \alpha, x_0, x) = [\frac{S_{\varepsilon_1,\varepsilon_2}(s_1 - \alpha - m_0(x_0), \alpha, x_0, x)}{S_{\varepsilon_1,\varepsilon_2}(s_2 - \alpha - m_0(x_0), \alpha, x_0, x)} - \frac{S_{\varepsilon_1,\varepsilon_2}(s_1 + m - \alpha - m_0(x), \alpha, x_0, x)}{S_{\varepsilon_1,\varepsilon_2}(s_2 + m - \alpha - m_0(x), \alpha, x_0, x)}].$$

We can show that for given (s_1, s_2, m, x_0, x), $S^2(s_1, s_2, m, \alpha, x_0, x)$, as a function of α, does not change signs if $S_{\varepsilon_1,\varepsilon_2}(\cdot, \alpha, x_0, x)$ is log-concave. Therefore, $m_0(x) - m_0(x_0)$ must be the unique minimizer of $T_p(m)$ and thus is identified under the log-concavity of $S_{\varepsilon_1,\varepsilon_2}(\cdot, \alpha, x_0, x)$, which is implied by the log-concavity of the conditional marginal density of ε_1 and ε_2 (e.g., Heckman and Honoré (1990)). Furthermore, in the special case when the fixed effect, α, is degenerate at α_0, it is straightforward to deduce that $m_0(x) - m_0(x_0)$ is identified if the conditional hazard function of ε_1 and ε_2 (independent of each other) given $(X_1, X_2) = (x_0, x)$ is not periodic in $(-\alpha_0 + \max\{-m_0(x), -m_0(x_0)\}, \infty)$, which coincides with the condition in Lemma 1 for identification in the cross-sectional case.

Similar to the cross-sectional case, we now examine the usefulness of continuous regressors for identification purposes. Corresponding to Lemma 2, we first consider the case in which all the regressors are continuous, and we show that identification is possible even if the non-periodicity discussed in Remark 5 fails.

Lemma 5: Assume that (X_1, X_2) is continuously distributed with a convex support, $\Omega_p = \Omega_{p1} \times \Omega_{p2}$, $S_p(e_1, e_2|x_1, x_2)$ is continuous in (e_1, e_2, x_1, x_2) for $(x_1, x_2) \in \Omega_p$, and $m_0(\cdot)$ is a continuous function. For some $x_0 \in \Omega_{p1} \cap \Omega_{p2}$ and $x \in \Omega_{p2}$, assume that there does not exist any $u \in [0, 1]$ such that $S(e_1, e_2|x_0, x_u)$ is constant along all the 135^0 lines[12] in the region

[12] It is interesting to note a close connection between a bivariate density function constant along the 135^0 lines and the univariate density function with an exponential tail referred to in Lemmas 2 and 5. Specifically, if a bivariate density function, $\xi(\cdot, \cdot)$, satisfies $\xi(e_1, e_2) = \xi_0(e_1 + e_2) = \xi_1(e_1)\xi_2(e_2)$ in the region $(-\xi_{10}, \infty) \times (-\xi_{20}, \infty)$ for some univariate functions $\xi_0(\cdot)$, $\xi_1(\cdot)$ and $\xi_2(\cdot)$ and constant terms ξ_{10} and ξ_{20}; namely, $\xi(\cdot, \cdot)$ is constant along the 135^0 lines and is the product of the two marginal densities in the region, then it is straightforward to show through some calculus manipulation that $\xi_1(e_1) = \lambda_1 \exp(-\lambda_0 e_1)$ for $e_1 > (-\xi_{10}, \infty)$ and $\xi_2(e_2) = \lambda_2 \exp(-\lambda_0 e_2)$ for $e_2 > (-\xi_{20}, \infty)$ and some constant terms, λ_0, λ_1 and λ_2.

$(\max\{-m_0(x_u), -m_0(x_0)\}, \infty) \times [\max\{-m_0(x_u), -m_0(x_0)\}, \infty)$, where $x_u = (1-u)x_0 + ux$. Then $m_0(x) - m_0(x)$ is identified[13].

Proof: Define a functional:

$$
\begin{aligned}
T_p^f(m) \quad = \quad & \int \int_0^\infty \int_0^\infty \{[G^1_{m(\tilde{x})}(s_1, s_2, x_0, \tilde{x}) - G^1_{m(\tilde{x})}(s_2, s_1, x_0, \tilde{x})]^2 1\{m(\tilde{x}) \geq 0\} \\
& + [G^2_{m(\tilde{x})}(s_1, s_2, x_0, \tilde{x}) - G^2_{m(\tilde{x})}(s_2, s_1, x_0, \tilde{x})]^2 1\{m(\tilde{x}) < 0\}\} w_1(s_1) w_1(s_2) w_2(\tilde{x}) ds_1 ds_2 d\tilde{x}.
\end{aligned}
$$

Note that $T_p^f(m_0) = 0$. Suppose now that there exists a continuous function, $\tilde{m}(\cdot)$, such that $T_p^f(\tilde{m}) = 0$, but $\tilde{m}(x) \neq m_0(x) - m_0(x_0)$. Suppose that $S_p(\cdot, \cdot | x_0, x_u)$ is also symmetric[14] about the 45^0 line through $(0, \gamma(u))$, in addition to the 45^0 line through the origin, $(-m_0(x_0), \infty) \times (-m_0(x_u), \infty)$ for $u \in \mathcal{U}_0$, where $\mathcal{U}_0$ is a subset of $[0,1]$. Then, based on the arguments in the proof of Lemma 4, we can show that $T_p^f(\tilde{m}) = 0$ implies $\tilde{m}(x_u) = m_0(x_u) - m_0(x_0) + \eta(u)\gamma(u)$ for any $u \in \mathcal{U}_0$, where $\eta(u)$ takes on integer values, which can be different for different $u \in [0,1]$, but we have $\tilde{m}(x_u) = m_0(x_u) - m_0(x_0)$ for any $u \in [0,1] \backslash \mathcal{U}_0$. Define $\gamma_0 = \inf_{u \in \mathcal{U}_0} \gamma(u)$. If $\gamma_0 > 0$, then $\tilde{m}(x_u) = m_0(x_u) - m_0(x_0)$ for $u \in [0,1]$ because of the fact that $\tilde{m}(x_0) = 0$ and the continuity of $\tilde{m}(\cdot)$ imply that $\eta(u) = 0$; thus, the identification follows immediately. Now, suppose that $\gamma_0 = 0$. Then, there exists a sequence $\{u_l\}$, such that $\lim_{u \to \infty} u_l = u_0$ and $\lim_{u \to \infty} \gamma(u_l) = 0$. Now we pick a point (t_1, t_2) such that $t_1, t_2 > \max\{-m_0(x_0), -m_0(x_{u_0})\}$. Without loss of generality, let $t_2 > t_1$. Define $t_{12} = (t_1 + t_2)/2$, $\bar{t}_{12} = (t_2 - t_1)/2$, and $n_l = [\bar{t}_{12}/\gamma(u_l)]$, the integer part of $\bar{t}_{12}/\gamma(u_l)$; thus, we can write $\bar{t}_{12} = n_l \gamma(u_l) + \bar{\varepsilon}_l$, with $\bar{\varepsilon}_l \in [0, \gamma(u_l))$. Then, for any $\zeta > 0$, there exists an L large enough such that for any $l > L$, we have $t_1, t_2 > \max\{-m_0(x_0), -m_0(x_{u_l})\}$, $|S_p(t_1, t_2, x_0, x_{u_0}) - S_p(t_1, t_2, x_0, x_{u_l})| < \zeta/3$ or

$$
|S_p(t_{12} - \bar{t}_{12}, t_{12} + \bar{t}_{12}, x_0, x_{u_0}) - S_p(t_{12} - \bar{t}_{12}, t_{12} + \bar{t}_{12}, x_0, x_{u_l})| < \zeta/3
$$

and

$$
|S_p(t_{12} - \bar{t}_{12}, t_{12} + \bar{t}_{12}, x_0, x_{u_l}) - S_p(t_{12} - n_l\gamma(u_l), t_{12} + n_l\gamma(u_l), x_0, x_{u_l})| < \zeta/3
$$

$$
S_p(t_{12} - n_l\gamma(u_l), t_{12} + n_l\gamma(u_l), x_0, x_{u_l}) = S_p(t_{12}, t_{12}, x_0, x_{u_l})
$$

by continuity and the symmetry property; in addition, we have

$$
|S_p(t_{12}, t_{12}, x_0, x_{u_0}) - S_p(t_{12}, t_{12}, x_0, x_{u_l})| < \zeta/3
$$

by continuity. Consequently, we obtain $|S_p(t_1, t_2, x_0, x_{u_0}) - S_p(t_{12}, t_{12}, x_0, x_{u_0})| < \zeta$ for any $\zeta > 0$. Therefore, $S_p(t_1, t_2, x_0, x_{u_0}) = S_p(t_{12}, t_{12}, x_0, x_{u_0})$; in other words, $S_p(\cdot, \cdot, x_0, x_{u_0})$ must be constant along all the 135^0 lines in the region $(\max\{-m_0(x_0), -m_0(x_{u_0})\}, \infty) \times (\max\{-m_0(x_0), -m_0(x_{u_0})\}, \infty)$, contradicting the assumptions in the Lemma. Thus, Lemma 5 has been established.

[13] Note that here we require at least partial overlapping supports for the regressors in two periods.

[14] Note that if a bivariate function is symmetric about the 45^0 line through $(0, \gamma(u))$ for a given u, then it is also symmetric about the 45^0 line through $(0, n\gamma(u))$ for any integer, n. In addition, if $\gamma(u) < 0$, then we can show the bivariate function is also symmetric about the 45^0 line through $(0, -\gamma(u))$; thus, without loss of generality, we can assume that $\gamma(u) > 0$.

We now examine the usefulness of continuous regressors in identification when both continuous and discrete regressors are present. Again, we show that identification is possible when a continuous regressor is present and the supports of the underlying regression functions are large enough even if the non-periodicity mentioned in Remark 5 fails in some upper-right region.

Lemma 6: Let $X_1 = (X_{1c}, X_{1d})$ and $X_2 = (X_{2c}, X_{2d})$, where (X_{1c}, X_{2c}) and (X_{1d}, X_{2d}) denote the continuous and discrete components, respectively. (i) Suppose that $\Omega_p = \Omega_{p1} \times \Omega_{p2}$ with $\Omega_{p1} = \Omega_{pc1} \times \Omega_{pd1}$, $\Omega_{p2} = \Omega_{pc2} \times \Omega_{pd2}$, where $\Omega_{pd1} = \Omega_{pd2} = \Omega_{pd0}$, Ω_{pc1} and Ω_{pc2} are convex sets and Ω_{pd0} is finite. (ii) There exists a constant, $\alpha^0 > 0$, such that $S_p(\cdot, \cdot, x_1, x_2)$ is only symmetric about the 45^0 line through the origin in the region $(-\alpha^0, \infty) \times (-\alpha^0, \infty)$ for $(x_1, x_2) \in \Omega_p$. (iii) There exists an $x_c^* \in \Omega_{pc1} \cap \Omega_{pc2}$ such that $m_0(x_c^*, x_d) > \alpha^0$ for any $x_d \in \Omega_{pd0}$. (iv) For some $x_0 = (x_{c0}, x_{d0}) \in \Omega_{p1}$, $x = (x_c, x_d) \in \Omega_{p2}$, there is no $u \in [0, 1]$ such that $S_p(\cdot, \cdot, x_0, x_u)$ is constant along all the 135^0 lines in the region $(-m_0(x_c, x_d), \infty) \times [-m_0(x_{cu}^*, x_d), \infty)$ or $(-m_0(x_c, x_{d0}), \infty) \times [-m_0(x_{cu}^*, x_{d0}), \infty)$ where $x_{cu}^* = (1-u)x_c^* + ux_c$. Then, $m_0(x_c, x_d) - m(x_{c0}, x_{d0})$ is identified.

Proof: Similar to the arguments in the proof of Lemma 3, we can show that $m_0(x_c^*, x_{d0}) - m_0(x_c^*, x_d)$ is identified by Lemma 4. Furthermore, Lemma 5 implies that $m_0(x_{c0}, x_{d0}) - m_0(x_c^*, x_{d0})$ and $m_0(x_{c0}, x_d) - m_0(x_c^*, x_d)$ are identified as well. Consequently, $m_0(x_c, x_d) - m(x_{c0}, x_{d0})$ is identified.

In summary, it is worth reiterating the similarities between the identification results in the cross-sectional and panel data cases; in particular, Lemmas 1 and 4 indicate that identification is possible when the underlying error hazard function or the conditional distribution satisfies some non-periodicity condition. In Lemmas 2 and, 5 we show that identification is still possible when all the regressors are continuous even if the non-periodicity fails. Moreover, in Lemmas 3 and 6, we show that in the case when there are discrete regressors, we can still establish identification without the above-mentioned non-periodicity if a continuous regressor is present and the support of the underlying regression is large enough.

3 Estimation

In this section, we propose our estimators for the regression functions in truncated regression models based on the identification results in the previous section, and we also investigate their large sample properties.

We first consider the cross-sectional case. The particular identification result we exploit here for estimation is that $m_0(x)$ is the unique minimizer of $T(m)$ defined in Eq. (2.5) under conditions set out in Lemma 1. Thus, the proposed estimator will be applicable to the general case in which both continuous and discrete regressors are allowed. The objective function is constructed by replacing various elements in the expression of $T(m)$ by some consistent estimates. Specifically, our estimator for $m_0(x)$, $\hat{m}(x)$, is defined as a minimizer of

$$T_n(m) = T_{n1}(m)w(m) + T_{n2}(m)w^c(m) \tag{3.1}$$

over a compact set $\mathcal{M}$, where $w(m)$ is a smoothed version of the indicator function, $1\{m \geq 0\}$, $w^c(x) = 1 - w(m)$, and $T_{n1}(m)$ and $T_{n2}(m)$ are sample analogues[15] of $T_1(m)$ and $T_2(m)$, except for some minor adjustments for technical reasons,

$$T_{n1}(m) = \int_0^\infty \left[\frac{S_n^*(t,x_0)}{S_n^*(\varepsilon_0,x_0)} - \frac{S_n^*(t+m,x)}{S_n^*(\varepsilon_0+m,x)}\right]^2 w_1(t)dt$$

and

$$T_{n2}(m) = \int_0^\infty \left[\frac{S_n^*(t,x)}{S_n^*(\varepsilon_0,x)} - \frac{S_n^*(t-m,x_0)}{S_n^*(\varepsilon_0-m,x_0)}\right]^2 w_1(t)dt,$$

where ε_0 is a small, positive constant, $S_n^*(t,x) = \sum_{i=1}^n K_y(\frac{Y_i-t}{h_y})k(\frac{X_i-x}{h})/\sum_{i=1}^n k(\frac{X_i-x}{h})$ is a smoothed version of the nonparametric estimator of $S^*(t,x)$ based on a random sample $\{Y_i, X_i$: $i = 1,...,n\}$, k is a kernel function with h as its bandwidth, and $K_y(t) = \int_{-\infty}^t k_y(v)dv$ is a smoothed step function, an integral of a kernel function, $k_y(\cdot)$, with h_y as the smoothing parameter, similar to Horowitz (1992). Note that, for simplicity, we assume that all the regressors are continuous. Discrete regressors can be easily accommodated. We make the following assumptions.

Assumption 1: The random vector, (Y^*, X), $X \in R^d$, satisfies (2.1) and $\{Y_i, X_i$: $i = 1,...,n\}$ is a random sample from the distribution of (Y^*, X) conditional on $Y^* > 0$ with $P_0 = P(Y^* > 0) > 0$.

Assumption 2 (i) x and x_0 are two interior points of Ω; (ii) the density function of X, $p(\cdot)$, and the regression function, m_0, are continuously differentiable up to order q in some neighborhoods of x_0 and x with both $p(x)$ and $p(x_0)$ being positive, and the cumulative distribution function of ε, $F(\cdot)$, is continuously differentiable up to order $q^* = \max\{q, q_y\}$ for q_y to be specified below, and these derivatives are uniformly bounded. (iii) The weight function, $w(\cdot)$, is a nonnegative twice continuously differentiable function such that $w(m) \geq 0$, $w(m) = 1$ for $m \geq \varepsilon_0/2$, $w(m) = 0$ for $m \leq -\varepsilon_0/2$; $w_1(\cdot)$ is a continuous non-negative integrable weight function such that $w_1(t) = 0$ for $t \in [0, \varepsilon_0]$ and $w_1(t) > 0$ for $t > \varepsilon_0$.

Assumption 3: The parameter space, $\mathcal{M}$, is a compact set and $m_0(x)$ is an interior point in $\mathcal{M}$.

Assumption 4: For the kernel functions, k is continuously differentiable and k_y is twice continuously differentiable with bound supports, and they are q- and q_y-order kernel functions, respectively[16]: $\int k(v)dv = 1$, $\int v^j k(v)dv = 0$ if $1 \leq |j| < q$, and $\int v^j k(v)dv \neq 0$ if $|j| = q$; $\int k_y(v)dv = 1$, $\int v^j k_y(v)dv = 0$ if $1 \leq j < q_y$, and $\int v^j k_y(v)dv \neq 0$ if $|j| = q_y$.

Assumption 5: The sequence of bandwidths satisfy $nh^d h_y^{3/2}/\ln n \to \infty$, $h_y^{q_y} = o(h^q)$, $h^{2q}/h_y = o(1)$ and $nh^{d+2q} = O(1)$ as $n \to \infty$.

Assumption 6: The hazard function of the error distribution, $\lambda(\cdot)$, is not a periodic function on $(\max\{-m_0(x), 0\} + \varepsilon_0, \infty)$.

[15] Here $m_0(x)$ is viewed as an unknown parameter. Pinske and Robinson (1995) and Pendakur (1999) used a similar strategy in estimation with regression functions with a similar shape.

[16] For a d-dimensional vector, $v = (v_1, ..., v_d)'$ and a corresponding vector of intergers, $j = (j_1, ..., j_d)'$, v^j denotes $v_1^{j_1} \cdots v_d^{j_d}$. Also, we adopt the convention $\partial^j \omega(v)/\partial v^j = \partial^{|j|}\omega(v)/\partial^{j_1} v_1 \cdots \partial^{j_d} v_d$ for a differentiable function, ω, with $|j| = \sum_{i=1}^d j_i$.

Assumption 1 describes the data generation process for the truncated regression model in the cross-sectional case. Assumption 2 states some smoothness and boundedness conditions. Assumption 3 is common for optimization-based estimation procedures. Assumptions 4 and 5 provide conditions on the kernel functions and the rates of convergence of the bandwidths. For example, if $d = 1$ and we choose second-order kernels with $q = q_y = 2$, then we can choose $h \propto n^{-1/5}$ and $h_y \propto n^{-1/3}$, in which case, h_y can go to zero more rapidly than typical bandwidths in nonparametric estimation. Assumption 6 is an identification condition, which is slightly stronger than that in Lemma 1 due to the minor adjustment in the estimation procedure.

Theorem 1 *Under Assumptions 1-6, $\hat{m}(x)$ is consistent for $m_0(x)$ and asymptotically normal,*

$$\sqrt{nh^d}((\hat{m}(x) - m_0(x)) - h^q b(x)) \xrightarrow{d} N(0, \sigma(x))$$

where $b(x))$ and $\sigma(x)$ are defined in the appendix.

Now, we consider the panel data model with fixed effects. Similar to the cross-sectional case, we make use of the identification result in Lemma 4 for the purpose of estimation, where the vector of regressors can have both discrete and continuous components. We again focus on the continuous case. Suppose that (x_0, x) is an interior point of Ω_p. For a random sample, $\{(Y_{it}, X_{it})$, $i = 1, 2, ..., n$, $t = 1, 2\}$, we define our estimator $\hat{m}_p(x)$, for $m_0(x) - m_0(x_0)$, as a minimizer of the objective function[17]:

$$T_{pn}(m) = T_{pn1}(m)w(m) + T_{pn2}(m)w^c(m),$$

where

$$T_{pn1}(m) = \int_0^\infty \int_0^\infty [1 - \frac{G_n(s_1, s_2 + m, x_0, x)}{G_n(s_2, s_1 + m, x_0, x)}]^2 w_1(s_1, s_2) ds_1 ds_2,$$

$$T_{pn2}(m) = \int_0^\infty \int_0^\infty \int [1 - \frac{G_n(s_1 - m, s_2, x_0, x)}{G_n(s_2 - m, s_1, x_0, x)}]^2 w_1(s_1, s_2) ds_1 ds_2$$

with $w_1(s_1, s_2) = w_1(s_1) w_1(s_2)$ and

$$G_n(s_1, s_2, x_0, x) = \frac{1}{n} \sum_{i=1}^n K_y(\frac{Y_{i1} - s_1}{h_y}) K_y(\frac{Y_{i2} - s_2}{h_y}) \frac{1}{h^{2d}} k(\frac{X_{i1} - x_0}{h}) k(\frac{X_{i2} - x}{h}).$$

We make the following assumptions for the panel data case.

Assumption 7: The random sample, (Y_{it}^*, X_{it}), $i = 1, 2, ...n$, $t = 1, 2$, satisfies (2.5) with $Y_{it} = Y_{it}^*$ conditional on the event $\{Y_{i1}^* > 0$ and $Y_{i2}^* > 0\}$ where it is assumed that $P\{Y_1^* > 0, Y_2^* > 0\} > 0$.

Assumption 8: The density function of (X_1, X_2) at (x_1, x_2), $p(x_1, x_2)$, is bounded away from zero and continuously differentiable up to order q in some neighborhood of (x_0, x); $S_p(e_1, e_2, x_1, x_2)$ is continuously differentiable with respect to each argument up to order $q^* = \max\{q, q_y\}$; $m_0(\cdot)$

[17] We could have used a sample analogue of $T_p(m)$ in (2.9), say $T_{np}^*(m)$, by directly replacing G by its corresponding nonparametric estimates. One drawback, however, is that both $T_p(m)$ and $T_{np}^*(m)$ would go to zero when m increases, thus leading to difficulty in estimation in practice.

is continuously differentiable up to order q in some neighborhoods of x_0 and x; in addition, these derivatives are uniformly bounded.

Assumption 9: The sequence of bandwidths satisfy $nh^{2d}h_y^{3/2} \to \infty$, $h_y^{q_y} = o(h^q)$, $h^{2q}/h_y = o_p(1)$ and $nh^{2(d+q)} = O(1)$ as $n \to \infty$.

Assumption 10: The 45^0 line through the origin is the only 45^0 line about which $(\varepsilon_1^*, \varepsilon_2^*)$ is symmetric in the region $(-m_0(x_0) + \varepsilon_0,\infty) \times (-m_0(x) + \varepsilon_0,\infty)$, conditionally on (X_1, X_2).

These assumptions are similar to those in the cross-sectional case. Assumption 7 describes the data-generating process for the truncated regression model in the panel data setting. Assumption 8 states some smoothness and boundedness conditions. Assumption 9 provides conditions on the rates of convergence of the bandwidths. Finally, Assumption 10 is an identification condition that corresponds to Lemma 4. Set $m_0(x_0)$ to 0 for notational simplicity.

Theorem 2 *Under Assumptions 2(iii),3-4 and 7-10, $\hat{m}_p(x)$ is consistent for $m_0(x)$ and asymptotically normal*

$$\sqrt{nh^{2d}}((\hat{m}_p(x) - m_0(x)) - h^q b_p(x)) \xrightarrow{d} N(0, \sigma_p(x))$$

where $b_p(x)$ and $\sigma_p(x)$ are specified in the appendix.

4 A Simulation Study

In this section, we examine some of the finite sample properties of our estimator, in comparison with that of Lewbel and Linton (2002) in the cross-sectional case. The data are generated according to the following model

$$Y^* = 0.8 + X + \varepsilon$$

where X is uniformly distributed on $(-1, 1)$, the error term ε, with different distributions for three different designs, is generated given X, and the observation is kept if the resulting Y^* is positive. The sample size was set to 400 and 1000 replications were carried out for each design.

In the simulation study, we set $m_0(0) = 0$ with $x_0 = 0$ as the location normalization point, and we report the mean value (Mean), the standard deviation (SD), and the root mean square error (RMSE) for our estimator (Chen) and the estimator of Lewbel and Linton (LL) (2002) for ten equally spaced points in the interval $[-0.9, 0.9]$. In implementing our estimation approach, we used the non-smoothed version of the conditional survival function, ε_0 was set to 0, the weight functions were chosen as $w(m) = 1\{m \geq 0\}$ and $w_1(t) = 1\{0 \leq t < y_{0.975}\}$, where $y_{0.975}$ is the 0.975th quantile of Y in the sample[18]. We adopt the standard normal density function as the kernel function, k, and the bandwidth is set to ch_s, where h_s was chosen according to Silverman's (1986) rule-of-thumb for both our estimation approach and each of the two steps in Lewbel and Linton (2002). We experimented with $c = 0.6, 0.8$ and 1.0 which produce quite stable results, and it appears that our approach is less sensitive to the choice of bandwidth than is that of Lewbel and Linton (2002).

[18] The minor adjustments made in defining the objective function in (3.1) do not produce a noticable difference in our experiments.

Table 1 reports the simulation results for the first design in which ε is a standard normal variable. In this case, both estimators are consistent and perform reasonably well, but our estimator outperforms that of Lewbel and Linton (2002) for all values of c. Note that the estimator of Lewbel and Linton (2002) is more sensitive to the choice of c, especially in the region where the truncation rate was high.

In the second design for which the results are reported in Table 2, ε is chosen from the logistic distribution standardized to have unit variance. As pointed out by Levy (2000), the monotonicity assumption of the truncated regression in Lewbel and Linton (2002) is not as strong as in the standard normal case used in the first design. As a result, the performance of the estimator by Lewbel and Linton (2002) deteriorates noticeably; in contrast, the overall performance of our approach is still comparable to the performance in the standard normal case.

Table 3 reports the results for the third design in which ε was set to $(\varepsilon^* - 0.2)/1.865$ to have zero mean and unit variance, where $\varepsilon^* = 0.5 * \varepsilon_a^{1.5} + 0.5 * \left[-|\varepsilon_b|^{0.8}\right]$, a mixture distribution with ε_a and ε_b generated from the standard exponential distribution. In this case, the monotonicity in Lewbel and Linton (2002) no longer holds. Consequently, the procedure by Lewbel and Linton (2002) incurs substantial biases and large variances. Our procedure still performs quite satisfactorily, though there is noticeable bias in the region where the truncation rate is very high.

5 Conclusion

In this paper, we have provided a comprehensive treatment of nonparametric identification of truncated regression models in both the cross-sectional and panel data settings and have proposed nonparametric estimators based on the identification results. Our nonparametric approach for the cross-sectional case overcomes some major drawbacks associated with the approach of Lewbel and Linton (2002), and simulation results indicate the superior performance of our estimator. Moreover, we have provided the first systematic treatment of nonparametric identification and estimation of the truncated panel data model with fixed effects, and the results could be viewed as bivariate extensions of those for the cross sectional case. The proposed estimators are shown to be consistent and asymptotically normal.

Appendix

In the appendix, we only provide sketches of the proofs of Theorems 1 and 2.

Sketch Proof of Theorem 1: Under Assumptions 1–2 and 4-5, with some standard arguments (e.g., Newey (1994), Pollard (1995)), it is straightforward to establish that $\sup_t |S_n^*(t,\tilde{x}) - S^*(t,\tilde{x})| = O_p(\delta_n)$, $\sup_t |\frac{\partial S_n^*(t,\tilde{x})}{\partial t} - \frac{\partial S^*(t,\tilde{x})}{\partial t}| = O_p(\delta_{n1})$, and $\sup_t |\frac{\partial^2 S_n^*(t,\tilde{x})}{\partial t^2} - \frac{\partial^2 S^*(t,\tilde{x})}{\partial t^2}| = O_p(\delta_{n2})$ for $\tilde{x}$ uniformly in some neighborhoods of x and x_0, where $\delta_n = (\ln n/nh^d)^{1/2} + h^q + h_y^{q_y}$, $\delta_{n1} = (\ln n/nh^d h_y)^{1/2} + h^q + h_y^{q_y}$, $\delta_{n2} = (\ln n/nh^d h_y^3)^{1/2} + h^q + h_y^{q_y}$. Then, we can show that $T_n(m)$ converges to

$$T^0(m) = \int_0^\infty \left\{ \left[\frac{S^*(t,x_0)}{S^*(\varepsilon_0,x_0)} - \frac{S^*(t+m,x)}{S^*(\varepsilon_0+m,x)} \right]^2 w(m) + \left[\frac{S^*(t,x)}{S^*(\varepsilon_0,x)} - \frac{S^*(t-m,x_0)}{S^*(\varepsilon_0-m,x_0)} \right]^2 w^c(m) \right\} w_1(t)dt$$

uniformly in $m \in \mathcal{M}$. Moreover, $T^0(m)$ achieves the unique minimum at $m_0(x)$ under Assumption 6. Consistency then follows from the standard arguments in Amemiya (1985).

To establish the asymptotic normality, we first establish an initial rate for $\hat{m}(x)$. We can show that

$$T_n(m) = O_p(\delta_n^2) + O_p(\delta_n)(m - m_0(x)) + C(m - m_0(x))^2$$

for some $C > 0$ uniform in m in a $o_p(1)$ neighborhood of $m_0(x)$. Then, standard arguments for extreme estimators (e.g., Sherman (1993)) yield $\hat{m}(x) - m_0(x) = \delta_n$. Then a Taylor expansion of the first-order condition, together with the above preliminary results, yields

$$\hat{m}(x) - m_0(x) = V(x)^{-1} U_n(x) + o_p((nh^d)^{-1/2}) \tag{A.1}$$

where

$$V(x) = \int \left[\lambda_1^*(t,x)\right]^2 w_1(t) dt w(m_0(x)) + \int \left[\lambda_2^*(t,x)\right]^2 w_1(t) dt w^c(m_0(x)),$$

with $\lambda_1^*(t,x) = \frac{S(t)}{S(\varepsilon_0)}(\lambda(\varepsilon_0) - \lambda(t))$ and $\lambda_2^*(t,x) = \frac{S(t-m_0(x))}{S(\varepsilon_0 - m_0(x))}(\lambda(\varepsilon_0 - m_0(x)) - \lambda(t - m_0(x)))$, and $U_n(x) = U_{n1}(x) w(m_0(x)) + U_{n2}(x) w^c(m_0(x))$, where

$$U_{n1}(x) = \int [\frac{S_n^*(t,x_0)}{S_n^*(\varepsilon_0,x_0)} - \frac{S_n^*(t+m_0(x),x)}{S_n^*(\varepsilon_0 + m_0(x),x)}] \lambda_1^*(t,x) w_1(t) dt$$

and

$$U_{n2}(x) = \int [\frac{S_n^*(t,x)}{S_n^*(\varepsilon_0,x)} - \frac{S_n^*(t-m_0(x),x_0)}{S_n^*(\varepsilon_0 - m_0(x),x_0)}] \lambda_2^*(t,x) w_1(t) dt.$$

Through linearization, we can establish that

$$\begin{aligned} U_{n1}(x) &= \frac{1}{n} \sum_{i=1}^n \left[\phi_{11}(Y_i,X_i) + \frac{1\{Y_i > \varepsilon_0\} P_0}{S(\varepsilon_0 - m_0(X_i))} B_{11}(X_i,x_0,x) \right] \frac{1}{h^d} k\left(\frac{X_i - x_0}{h}\right) \\ &\quad - \frac{1}{n} \sum_{i=1}^n [\phi_{12}(Y_i,X_i) + \frac{1\{Y_i > \varepsilon_0 + m_0(x)\} P_0}{S(\varepsilon_0 + m_0(x) - m_0(X_i))} B_{12}(X_i,x_0,x)] \frac{1}{h^d} k\left(\frac{X_i - x}{h}\right)] + o_p((nh^d)^{-1/2}) \end{aligned}$$

if $w(m_0(x)) \neq 0$, and

$$\begin{aligned} U_{n2}(x) &= \frac{1}{n} \sum_{i=1}^n \left[\phi_{21}(Y_i,X_i) + \frac{1\{Y_i > \varepsilon_0\} P_0}{S(\varepsilon_0 - m_0(X_i))} B_{21}(X_i,x_0,x) \right] \frac{1}{h^d} k\left(\frac{X_i - x}{h}\right) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \left[\phi_{22}(Y_i,X_i) + \frac{1\{Y_i > \varepsilon_0 - m_0(x)\} P_0}{S(\varepsilon_0 - m_0(X_i))} B_{22}(X_i,x_0,x) \right] \frac{1}{h^d} k\left(\frac{X_i - x_0}{h}\right)] + o_p((nh^d)^{-1/2}) \end{aligned}$$

if $w^c(m_0(x)) \neq 0$, where

$$\phi_{11}(Y_i,X_i) = \frac{P_0}{S(\varepsilon_0) p(x_0)} \int \left[1\{Y_i > t\} - 1\{Y_i > \varepsilon_0\} \frac{S(t - m_0(X_i))}{S(\varepsilon_0 - m_0(X_i))} \right] \lambda_1^*(t,x) w_1(t) dt$$

$$\phi_{12}(Y_i,X_i) = \frac{P_0}{S(\varepsilon_0) p(x)} \int [1\{Y_i > t + m_0(x)\} - 1\{Y_i > \varepsilon_0 + m_0(x)\} \frac{S(t + m_0(x) - m_0(X_i))}{S(\varepsilon_0 + m_0(x) - m_0(X_i))}] \lambda_1^*(t,x) w_1(t) dt$$

$$\phi_{21}(Y_i,X_i) = \frac{P_0}{S(\varepsilon_0 - m_0(x)) p(x)} \int \left[1\{Y_i > t\} - 1\{Y_i > \varepsilon_0\} \frac{S(t - m_0(X_i))}{S(\varepsilon_0 - m_0(X_i))} \right] \lambda_2^*(t,x) w_1(t) dt$$

$$\phi_{22}(Y_i, X_i) = \frac{P_0}{S(\varepsilon_0 - m_0(x))p(x_0)} \int [1\{Y_i > t - m_0(x)\} - 1\{Y_i > \varepsilon_0 - m_0(x)\} \frac{S(t - m_0(X_i))}{S(\varepsilon_0 - m_0(X_i))}] \lambda_2^*(t, x) w_1(t) dt$$

$$B_{11}(X_i, x_0, x) = \frac{S(\varepsilon_0 - m_0(X_i))}{S(\varepsilon_0)p(x_0)} \int \left[\frac{S(t - m_0(X_i))}{S(\varepsilon_0 - m_0(X_i))} - \frac{S(t)}{S(\varepsilon_0)} \right] \lambda_1^*(t, x) w_1(t) dt$$

$$B_{12}(X_i, x_0, x) = \frac{S(\varepsilon_0 + m_0(x) - m_0(X_i))}{S(\varepsilon_0)p(x)} \int \left[\frac{S(t + m_0(x) - m_0(X_i))}{S(\varepsilon_0 + m_0(x) - m_0(X_i))} - \frac{S(t)}{S(\varepsilon_0)} \right] \lambda_1^*(t, x) w_1(t) dt$$

$$B_{21}(X_i, x_0, x) = \frac{S(\varepsilon_0 - m_0(X_i))}{S(\varepsilon_0 - m_0(x))p(x)} \int \left[\frac{S(t - m_0(X_i))}{S(\varepsilon_0 - m_0(X_i))} - \frac{S(t - m_0(x))}{S(\varepsilon_0 - m_0(x))} \right] \lambda_2^*(t, x) w_1(t) dt$$

and

$$B_{22}(X_i, x_0, x) = \frac{S(\varepsilon_0 - m_0(X_i))}{S(\varepsilon_0 - m_0(x))p(x_0)} \int \left[\frac{S(t - m_0(X_i))}{S(\varepsilon_0 - m_0(X_i))} - \frac{S(t - m_0(x))}{S(\varepsilon_0 - m_0(x))} \right] \lambda_2^*(t, x) w_1(t) dt.$$

Following some standard arguments in nonparametric regression analysis, we can show that

$$E\left\{ \left[\frac{1\{Y_i > \varepsilon_0\}P_0}{S(\varepsilon_0 - m_0(X_i))} B_{11}(X_i, x_0, x) - \frac{1\{Y_i > \varepsilon_0 - m_0(x)\}P_0}{S(\varepsilon_0 - m_0(X_i))} B_{22}(X_i, x_0, x) \right] \frac{1}{h^d} k\left(\frac{X_i - x_0}{h} \right) \right\} = h^q b_1(x) + o\left(h^q\right)$$

and

$$E\left\{ \left[\frac{1\{Y_i > \varepsilon_0\}P_0}{S(\varepsilon_0 - m_0(X_i))} B_{21}(X_i, x_0, x) - \frac{1\{Y_i > \varepsilon_0 + m_0(x)\}P_0}{S(\varepsilon_0 + m_0(x) - m_0(X_i))} B_{12}(X_i, x_0, x) \right] \frac{1}{h^d} k\left(\frac{X_i - x}{h} \right) \right\} = h^q b_2(x) + o\left(h^q\right)$$

for some $b_1(x)$ and $b_2(x)$. Let $b(x) = V^{-1}(x)[b_1(x)w(m_0(x)) + b_2(x)w^c(m_0(x))]$ and $\sigma(x) = \sigma_1(x) + \sigma_2(x)$, where

$$\sigma_1(x) = V^{-2}(x) E\left[\phi_{11}(Y_i(x_0), x_0) w(m_0(x)) - \phi_{22}(Y_i(x_0), x_0) w^c(m_0(x))\right]^2 \int k^2(u) du\, p(x_0)/P_0$$

and

$$\sigma_2(x) = V^{-2}(x) E\left[\phi_{12}(Y_i(x), x) w(m_0(x)) - \phi_{22}(Y_i(x), x) w^c(m_0(x))\right]^2 \int k^2(u) du\, p(x)/P_0,$$

with $Y_i(x_0) = \varepsilon_i$ and $Y_i(x) = m_0(x) + \varepsilon_i$. Then, we can obtain

$$\sqrt{nh^d}(\hat{m}(x) - m_0(x) - h^q b(x)) = \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n \left\{ \phi_{i1} k\left(\frac{X_i - x_0}{h} \right) + \phi_{i2} k\left(\frac{X_i - x}{h} \right) \right\} + o_p(1), \tag{A2}$$

where $\phi_{i1} = \phi_{11}(Y_i, X_i) w(m_0(x)) - \phi_{22}(Y_i, X_i) w^c(m_0(x))$ and $\phi_{i2} = \phi_{21}(Y_i, X_i) w^c(m_0(x)) - \phi_{12}(Y_i, X_i) w(m_0(x))$. Consequently, Theorem 1 follows by applying a triangular-array central limit theorem to (A.2).

Sketch Proof of Theorem 2:

Proceeding as in the proof of Theorem 1, we can show that $\hat{m}_p(x)$ is consistent and

$$\hat{m}_p(x) - m_0(x) = V_p(x)^{-1} U_{pn}(x) + o_p((nh^{2d})^{-1/2}), \tag{A.3}$$

where

$$V_p(x) = \int\int \left\{ \left[\lambda_{p1}^*(s_1, s_2, x)\right]^2 w(m_0(x)) + \left[\lambda_{p2}^*(s_1, s_2, x)\right]^2 \right\} w_1(s_1, s_2) ds_1 ds_2 w^c(m_0(x))$$

and $U_{pn}(x) = U_{pn1}(x)w(m_0(x)) + U_{pn2}(x)w^c(m_0(x))$,

$$U_{pn1}(x) = \int\int [\frac{G_n(s_2, s_1 + m_0(x), x_0, x) - G_n(s_1, s_2 + m_0(x), x_0, x)}{G(s_2, s_1 + m_0(x), x_0, x)}]\lambda_{p1}^*(s_1, s_2, x)w_1(s_1, s_2)ds_1 ds_2,$$

$$U_{pn2}(x) = \int\int [\frac{G_n(s_2 - m_0(x), s_1, x_0, x) - G_n(s_1 - m_0(x), s_2, x_0, x)}{G(s_2 - m_0(x), s_1, x_0, x)}]\lambda_{p2}^*(s_1, s_2, x)w_1(s_1, s_2)ds_1 ds_2$$

with $\lambda_{p1}^*(s_1, s_2, x) = \frac{d}{dm}\left[\frac{S_p(s_1, s_2 + m_0(x), x_0, x)}{S_p(s_2, s_1 + m_0(x), x_0, x)}\right]$ and $\lambda_{p2}^*(s_1, s_2, x) = \frac{d}{dm}\left[\frac{S_p(s_1 - m_0(x), s_2, x_0, x)}{S_p(s_2 - m_0(x), s_1, x_0, x)}\right]$.

Define $\phi_{pi} = \phi_{pi1}w(m_0(x)) + \phi_{pi2}w^c(m_0(x))$, where

$$\phi_{pi1} = \int\int \frac{[\phi_{pi11}(s_1, s_2) - \phi_{pi12}(s_1, s_2)]}{G(s_2, s_1 + m_0(x), x_0, x)}\lambda_{p1}^*(s_1, s_2, x)w_1(s_1, s_2)ds_1 ds_2$$

$$\phi_{pi2} = \int\int \frac{[\phi_{pi21}(s_1, s_2) - \phi_{pi22}(s_1, s_2)]}{G(s_2 - m_0(x), s_1, x_0, x)}\lambda_{p2}^*(s_1, s_2, x)w_1(s_1, s_2)ds_1 ds_2$$

with

$$\phi_{pi11}(s_1, s_2) = 1\{Y_{i1} > s_2, Y_{i2} > s_1 + m_0(x)\} - \frac{S_p[s_2 - m(X_{i1}), s_1 + m_0(x) - m(X_{i2}), X_{i1}, X_{i2}]}{S_p[-m(X_{i1}), -m(X_{i2}), X_{i1}, X_{i2}]}$$

$$\phi_{pi12}(s_1, s_2) = 1\{Y_{i1} > s_1, Y_{i2} > s_2 + m_0(x)\} - \frac{S_p[s_1 - m(X_{i1}), s_2 + m_0(x) - m(X_{i2}), X_{i1}, X_{i2}]}{S_2[-m(X_{i1}), -m(X_{i2}), X_{i1}, X_{i2}]}$$

$$\phi_{pi21}(s_1, s_2) = 1\{Y_{i1} > s_2 - m_0(x), Y_{i2} > s_1\} - \frac{S_p[s_2 - m_0(x) - m(X_{i1}), s_1 - m(X_{i2}), X_{i1}, X_{i2}]}{S_p[-m(X_{i1}), -m(X_{i2}), X_{i1}, X_{i2}]}$$

$$\phi_{pi22}(s_1, s_2) = 1\{Y_{i1} > s_1 - m_0(x), Y_{i2} > s_2\} - \frac{S_p[s_1 - m_0(x) - m(X_{i1}), s_2 - m(X_{i2}), X_{i1}, X_{i2}]}{S_p[-m(X_{i1}), -m(X_{i2}), X_{i1}, X_{i2}]}.$$

We can then show that

$$\sqrt{nh^{2d}}(\hat{m}_p(x) - m_0(x) - h^q b_p(x)) = \frac{1}{\sqrt{nh^{2d}}}\sum_{i=1}^{n}\phi_{pi}\frac{1}{h^{2d}}k(\frac{X_{i1} - x_0}{h})k(\frac{X_{i2} - x}{h}) + o_p(1) \qquad \text{(A4)}$$

for some $b_p(x)$. Let

$$\sigma_p(x) = V_p^{-2}(x)E\frac{[\phi(Y_{i1}(x_0), Y_{i2}(x), x_0, x)]^2}{S_p(0, -m_0(x), x_0, x)P_0}\left\{\int k^2(u)du\right\}^2 p(x_0, x)/P_0$$

where $Y_{i1}(x_0) = \varepsilon_{i1}^*$ and $Y_{i2}(x) = m_0(x) + \varepsilon_{i2}^*$. Consequently, Theorem 2 follows by applying a triangular-array central limit theorem to (A.4).

References

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

[2] Amemiya, T. (1985): *Advanced Econometrics,* Cambridge, MA: Harvard University Press.

[3] Heckman J.J. and B. Honoré (1990): "The Empirical Content of the Roy Model," *Econometrica*, 5, 1121-1149.

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

[5] B.E. Honoré and E. Kyriazidou(2000): "Estimation of Tobit-Type Models with Individual Specific Effects," *Econometric Reviews*, Volume 19, No. 3, 341-366.

[6] B.E. Honoré and J.L. Powell (1994): "Pairwise difference estimators of censored and truncated regression models," *Journal of Econometrics,* 64, 241-278.

[7] Horowitz, J.L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505-531.

[8] Horowitz, J.L. (1996):"Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable," *Econometrica* 64, 103-137.

[9] Horowitz, J.L. (1998): Semiparametric Methods in Econometrics, Springer

[10] Ichimura, H. (1993): "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models," *Journal of Econometrics,* 58, 71-120.

[11] Lee, L-F. (1992): "Semiparametric Nonlinear Least-Squares Estimation of Truncated Regression Models," Econometric Theory, 8, 52-94.

[12] Levy, A. (2000): "A Simple Consistent Non-parametric Estimator for the Regression Function in Truncated Sample," Manuscript.

[13] Lewbel, A. and O. Linton (2002): "Nonparametric Censored and Truncated Regression," *Econometrica*, 70, 765-779.

[14] Manski, C. F. (1988):"Identification of Binary Response Models," *Journal of the American Statistical Association*," 83, 729-738.

[15] Newey, W.K. (1994): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233-253.

[16] Pendakur, K. (1999): "Semiparametric Estimates and Tests of Based-independent Equivalence Scales," 88, 1-40.

[17] Pinkse, C.A.P. and P.M. Robinson (1995): "Pooling Nonparametric Estimates of Regression Functions with a Similar Shape," in *Advances in Econometrics and Quantitative Economics*, ed. by G.S. Maddala, P.C.B. Phillips and T.N. Srinivasan, Blackwell.

[18] Pollard, D. (1995): "Uniform Ratio Limit Theorems for Empirical Processes," *Scandinavian Journal of Statistics*, 22, 271-278.

[19] Porter, J. (1997): "Nonparametric Regression Estimation for a Panel Data Model with Additive Individual Effects", Manuscript.

[20] Powell, J.L. (1994): "Estimation of Semiparametric Models. Handbook of Econometrics," in *Handbook of Econometrics Vol 4*, ed. by R. Engle and D. McFadden. North-Holland: Elsevier Science, 2444-2521.

[21] Sherman, R.P., 1993, The limiting distribution of the maximum rank correlation estimator, *Econometrica* 61, 123-138.

[22] Silverman, B.W. (1986): Density estimation for statistics and data analysis, Chapman and Hall, London.

Table1: Design I

	$m_0(x)$	Chen Mean	Bias	SD	RMSE	LL Mean	Bias	SD	RMSE
$c = 0.6$									
	-0.9	**-0.850**	**0.050**	**0.294**	**0.298**	**-0.783**	**0.117**	**0.380**	**0.397**
	-0.7	**-0.701**	**-0.001**	**0.282**	**0.281**	**-0.648**	**0.052**	**0.364**	**0.368**
	-0.5	**-0.522**	**-0.022**	**0.272**	**0.272**	**-0.476**	**0.024**	**0.369**	**0.369**
	-0.3	**-0.329**	**-0.029**	**0.278**	**0.279**	**-0.300**	**0.000**	**0.352**	**0.352**
	-0.1	**-0.081**	**0.019**	**0.112**	**0.113**	**-0.106**	**-0.006**	**0.255**	**0.255**
	0.1	**0.081**	**-0.019**	**0.116**	**0.118**	**0.117**	**0.017**	**0.277**	**0.277**
	0.3	**0.304**	**0.004**	**0.228**	**0.228**	**0.293**	**-0.007**	**0.343**	**0.342**
	0.5	**0.505**	**0.005**	**0.234**	**0.234**	**0.490**	**-0.010**	**0.350**	**0.350**
	0.7	**0.699**	**-0.001**	**0.235**	**0.235**	**0.690**	**-0.010**	**0.379**	**0.378**
	0.9	**0.879**	**-0.021**	**0.251**	**0.251**	**0.909**	**0.009**	**0.371**	**0.371**
$c = 0.8$									
	-0.9	**-0.824**	**0.076**	**0.287**	**0.298**	**-0.721**	**0.179**	**0.401**	**0.439**
	-0.7	**-0.680**	**0.020**	**0.269**	**0.269**	**-0.613**	**0.087**	**0.375**	**0.385**
	-0.5	**-0.514**	**-0.014**	**0.259**	**0.259**	**-0.486**	**0.014**	**0.374**	**0.374**
	-0.3	**-0.304**	**-0.004**	**0.218**	**0.218**	**-0.301**	**-0.009**	**0.371**	**0.371**
	-0.1	**-0.068**	**0.032**	**0.075**	**0.081**	**-0.107**	**-0.007**	**0.223**	**0.222**
	0.1	**0.068**	**-0.032**	**0.070**	**0.077**	**0.093**	**-0.007**	**0.213**	**0.213**
	0.3	**0.296**	**-0.004**	**0.181**	**0.181**	**0.288**	**-0.013**	**0.361**	**0.361**
	0.5	**0.502**	**0.002**	**0.199**	**0.198**	**0.499**	**-0.001**	**0.365**	**0.365**
	0.7	**0.704**	**0.004**	**0.211**	**0.211**	**0.717**	**0.017**	**0.357**	**0.357**
	0.9	**0.851**	**0.049**	**0.248**	**0.252**	**0.908**	**0.008**	**0.362**	**0.361**
$c = 1.0$									
	-0.9	**-0.836**	**0.064**	**0.287**	**0.294**	**-0.619**	**0.281**	**0.399**	**0.487**
	-0.7	**-0.709**	**-0.009**	**0.277**	**0.277**	**-0.571**	**0.129**	**0.381**	**0.401**
	-0.5	**-0.520**	**-0.021**	**0.241**	**0.242**	**-0.471**	**0.029**	**0.390**	**0.390**
	-0.3	**-0.310**	**-0.010**	**0.197**	**0.197**	**-0.305**	**-0.005**	**0.340**	**0.339**
	-0.1	**-0.065**	**0.034**	**0.074**	**0.081**	**-0.101**	**-0.001**	**0.168**	**0.168**
	0.1	**0.067**	**-0.003**	**0.061**	**0.069**	**0.098**	**-0.002**	**0.156**	**0.155**
	0.3	**0.305**	**0.005**	**0.156**	**0.156**	**0.326**	**0.026**	**0.357**	**0.357**
	0.5	**0.512**	**0.012**	**0.185**	**0.185**	**0.549**	**0.049**	**0.390**	**0.392**
	0.7	**0.696**	**-0.004**	**0.192**	**0.192**	**0.771**	**0.071**	**0.335**	**0.342**
	0.9	**0.826**	**0.074**	**0.224**	**0.236**	**0.922**	**0.022**	**0.337**	**0.337**

Table 2: Design II

	$m_0(x)$	Chen				LL			
		Mean	**Bias**	**SD**	**RMSE**	**Mean**	**Bias**	**SD**	**RMSE**
$c = 0.6$									
	-0.9	**-0.826**	**0.074**	**0.340**	**0.347**	**-0.793**	**0.107**	**0.517**	**0.528**
	-0.7	**-0.707**	**-0.007**	**0.313**	**0.313**	**-0.662**	**0.038**	**0.483**	**0.484**
	-0.5	**-0.514**	**-0.014**	**0.297**	**0.297**	**-0.468**	**0.032**	**0.433**	**0.433**
	-0.3	**-0.331**	**-0.031**	**0.282**	**0.283**	**-0.265**	**0.035**	**0.410**	**0.411**
	-0.1	**-0.086**	**0.014**	**0.110**	**0.110**	**-0.104**	**-0.004**	**0.287**	**0.287**
	0.1	**0.072**	**-0.028**	**0.085**	**0.089**	**0.103**	**0.003**	**0.285**	**0.284**
	0.3	**0.295**	**-0.005**	**0.212**	**0.212**	**0.284**	**-0.016**	**0.396**	**0.396**
	0.5	**0.515**	**0.015**	**0.221**	**0.222**	**0.490**	**-0.010**	**0.402**	**0.402**
	0.7	**0.699**	**-0.001**	**0.215**	**0.214**	**0.688**	**-0.012**	**0.431**	**0.431**
	0.9	**0.873**	**-0.027**	**0.235**	**0.237**	**0.936**	**0.036**	**0.428**	**0.429**
$c = 0.8$									
	-0.9	**-0.843**	**0.057**	**0.334**	**0.339**	**-0.700**	**0.200**	**0.539**	**0.574**
	-0.7	**-0.726**	**-0.026**	**0.299**	**0.299**	**-0.591**	**0.108**	**0.485**	**0.496**
	-0.5	**-0.521**	**-0.021**	**0.281**	**0.282**	**-0.432**	**0.067**	**0.465**	**0.468**
	-0.3	**-0.308**	**-0.008**	**0.212**	**0.212**	**-0.269**	**0.031**	**0.429**	**0.430**
	-0.1	**-0.076**	**0.024**	**0.074**	**0.077**	**-0.098**	**0.002**	**0.238**	**0.238**
	0.1	**0.075**	**-0.025**	**0.063**	**0.068**	**0.104**	**0.004**	**0.232**	**0.232**
	0.3	**0.304**	**0.004**	**0.170**	**0.170**	**0.286**	**-0.014**	**0.381**	**0.381**
	0.5	**0.509**	**0.009**	**0.177**	**0.177**	**0.510**	**0.010**	**0.415**	**0.414**
	0.7	**0.709**	**0.009**	**0.182**	**0.182**	**0.747**	**0.047**	**0.390**	**0.391**
	0.9	**0.853**	**-0.047**	**0.207**	**0.212**	**0.927**	**0.027**	**0.393**	**0.393**
$c = 1.0$									
	-0.9	**-0.834**	**0.066**	**0.301**	**0.307**	**-0.557**	**0.343**	**0.541**	**0.640**
	-0.7	**-0.706**	**-0.006**	**0.264**	**0.264**	**-0.531**	**0.169**	**0.490**	**0.518**
	-0.5	**-0.541**	**-0.041**	**0.246**	**0.249**	**-0.455**	**0.045**	**0.468**	**0.469**
	-0.3	**-0.319**	**-0.019**	**0.175**	**0.175**	**-0.315**	**-0.015**	**0.426**	**0.426**
	-0.1	**-0.079**	**0.021**	**0.064**	**0.067**	**-0.108**	**-0.008**	**0.207**	**0.207**
	0.1	**0.073**	**-0.027**	**0.054**	**0.060**	**0.094**	**-0.006**	**0.191**	**0.191**
	0.3	**0.301**	**0.001**	**0.132**	**0.132**	**0.287**	**-0.013**	**0.390**	**0.390**
	0.5	**0.505**	**0.005**	**0.157**	**0.157**	**0.529**	**0.029**	**0.419**	**0.419**
	0.7	**0.695**	**-0.005**	**0.167**	**0.167**	**0.761**	**0.061**	**0.368**	**0.373**
	0.9	**0.827**	**-0.073**	**0.187**	**0.200**	**0.919**	**0.019**	**0.399**	**0.399**

Table3: Design III

	$m_0(x)$	**Chen** Mean	**Bias**	**SD**	**RMSE**	**LL** Mean	**Bias**	**SD**	**RMSE**
$c = 0.6$									
	-0.9	**-0.514**	**0.386**	**0.698**	**0.797**	**-0.105**	**0.795**	**0.685**	**1.049**
	-0.7	**-0.618**	**0.082**	**0.541**	**0.546**	**-0.457**	**0.243**	**0.644**	**0.687**
	-0.5	**-0.498**	**0.002**	**0.523**	**0.523**	**-0.549**	**-0.049**	**0.679**	**0.680**
	-0.3	**-0.334**	**-0.034**	**0.436**	**0.437**	**-0.426**	**-0.126**	**0.616**	**0.628**
	-0.1	**-0.090**	**0.010**	**0.115**	**0.115**	**-0.151**	**-0.051**	**0.364**	**0.367**
	0.1	**0.094**	**-0.006**	**0.077**	**0.077**	**0.153**	**0.053**	**0.413**	**0.416**
	0.3	**0.314**	**0.014**	**0.201**	**0.201**	**0.557**	**0.257**	**0.646**	**0.694**
	0.5	**0.497**	**-0.003**	**0.134**	**0.134**	**0.912**	**0.412**	**0.778**	**0.880**
	0.7	**0.696**	**-0.004**	**0.142**	**0.142**	**1.169**	**0.469**	**0.786**	**0.915**
	0.9	**0.868**	**-0.032**	**0.110**	**0.114**	**1.541**	**0.641**	**0.846**	**1.060**
$c = 0.8$									
	-0.9	**-0.594**	**0.306**	**0.710**	**0.772**	**-0.042**	**0.858**	**0.613**	**1.054**
	-0.7	**-0.610**	**-0.090**	**0.553**	**0.560**	**-0.222**	**0.478**	**0.598**	**0.765**
	-0.5	**-0.515**	**-0.015**	**0.484**	**0.484**	**-0.253**	**0.247**	**0.685**	**0.727**
	-0.3	**-0.337**	**-0.037**	**0.395**	**0.397**	**-0.251**	**0.049**	**0.611**	**0.613**
	-0.1	**-0.094**	**0.006**	**0.127**	**0.127**	**-0.105**	**-0.005**	**0.378**	**0.378**
	0.1	**0.092**	**-0.008**	**0.042**	**0.043**	**0.128**	**0.028**	**0.378**	**0.379**
	0.3	**0.302**	**0.002**	**0.122**	**0.121**	**0.475**	**0.175**	**0.638**	**0.661**
	0.5	**0.507**	**0.007**	**0.089**	**0.089**	**0.748**	**0.248**	**0.803**	**0.839**
	0.7	**0.700**	**0.000**	**0.090**	**0.090**	**0.993**	**0.239**	**0.860**	**0.908**
	0.9	**0.844**	**-0.056**	**0.105**	**0.118**	**1.288**	**0.388**	**0.931**	**1.008**
$c = 1.0$									
	-0.9	**-0.674**	**0.226**	**0.603**	**0.662**	**-0.141**	**0.759**	**0.611**	**0.974**
	-0.7	**-0.624**	**0.076**	**0.531**	**0.535**	**-0.181**	**0.519**	**0.622**	**0.809**
	-0.5	**-0.494**	**0.006**	**0.511**	**0.510**	**-0.177**	**0.323**	**0.721**	**0.789**
	-0.3	**-0.343**	**-0.043**	**0.357**	**0.359**	**-0.149**	**0.151**	**0.651**	**0.668**
	-0.1	**-0.087**	**0.013**	**0.138**	**0.139**	**-0.095**	**0.005**	**0.376**	**0.376**
	0.1	**0.097**	**-0.003**	**0.112**	**0.112**	**0.111**	**0.011**	**0.414**	**0.413**
	0.3	**0.311**	**0.011**	**0.154**	**0.154**	**0.358**	**0.058**	**0.736**	**0.738**
	0.5	**0.512**	**0.012**	**0.114**	**0.114**	**0.603**	**0.103**	**0.819**	**0.825**
	0.7	**0.692**	**-0.008**	**0.093**	**0.093**	**0.875**	**0.175**	**0.886**	**0.902**
	0.9	**0.821**	**-0.078**	**0.095**	**0.123**	**1.110**	**0.210**	**0.879**	**0.902**