Rank estimation of a generalized fixed-effects regression model

Jason Abrevaya*

Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637, USA

Received 1 July 1997; received in revised form 1 April 1999; accepted 1 April 1999

Abstract

This paper considers estimation of a fixed-effects version of the generalized regression model of Han (1987, Journal of Econometrics 35, 303–316). The model allows for censoring, places no parametric assumptions on the error disturbances, and allows the fixed effects to be correlated with the covariates. We introduce a class of rank estimators that consistently estimate the coefficients in the generalized fixed-effects regression model. The maximum score estimator for the binary choice fixed-effects model is part of this class. Like the maximum score estimator, the class of rank estimators converge at less than the $\sqrt{n}$ rate. Smoothed versions of these estimators, however, converge at rates approaching the $\sqrt{n}$ rate. In a version of the model that allows for truncated data, a sufficient condition for consistency of the estimators is that the error disturbances have an increasing hazard function. © 2000 Elsevier Science S.A. All rights reserved.

JEL classification: C23; C14

Keywords: Fixed-effects models; Generalized regression model; Panel data; Rank estimators; Maximum score estimator

1. Introduction

Much has been written about the difficulties in consistently estimating the parameters of non-linear fixed-effects panel data models. The standard
first-differencing trick which eliminates the fixed effect from a linear model extends to only certain non-linear models, including the conditional logit model for binary data (Chamberlain, 1980), the Poisson model for count data (Hausman et al., 1984), and certain parametric models for duration data (Chamberlain, 1985). Each of these models share an exponential form which allows for cancellation of the fixed effect akin to first differencing in the linear panel model. Semiparametric methods, which do not require any parametric assumptions on the error term, exist for consistent estimation of the binary choice model (Manski, 1987), the linear censored and truncated models (Honore, 1992), the selection model (Kyriazidou, 1997), and the dependent-variable transformation model (Abrevaya, 1999).

In this paper, we introduce a class of rank estimators for consistent estimation of a fixed-effects version of the generalized regression model of Han (1987). Like its cross-sectional counterpart, the model allows for unspecified non-linearity and general forms of censoring. The model allows for correlation between the fixed effects and covariates and does not restrict the fixed effects to enter additively. The maximum score (MS) estimator for the binary choice fixed-effects model (Manski, 1987) is part of the proposed class of estimators. Like the MS estimator, the class of rank estimators converge at less than the $\sqrt{n}$ rate. Smoothed versions of these estimators, however, converge at rates approaching the $\sqrt{n}$ rate.

In Section 2, we introduce the generalized fixed-effects regression model. In Section 3, we propose a class of rank estimators that consistently estimate the coefficients in the model. Some estimators in the class require i.i.d. error disturbances over time for each cross-sectional unit, whereas other estimators in the class require only a stationarity assumption on the error disturbances. All of the estimators allow for heteroskedasticity across observational units. Smoothed versions of the estimators are also proposed and, applying the results of Horowitz (1992), have convergence rates approaching the $\sqrt{n}$ rate. In Section 4, we discuss estimation when the dependent variable is subject to truncation. For left-truncated data, the assumption that the error disturbances have an increasing hazard function (or, equivalently, a log-concave survival function) is sufficient for consistency. In Section 5, we present Monte Carlo evidence on the small-sample performance of the estimators. Section 6 concludes and discusses some possible extensions.

2. The generalized fixed-effects regression model

The generalized regression model of Han (1987) is

$$
\begin{align*}
  y_{it}^* &= F(x_{it}'\beta, \varepsilon_{it}) \\
  y_i &= D(y_{it}^*) \\
  (i = 1, \ldots, n),
\end{align*}
$$

(1)
where $x_i$ and $\beta$ are $q \times 1$ vectors and $\varepsilon_i$ is a i.i.d. error disturbance. The first equation describes the ‘latent’ dependent variable $y^*_i$, where $F: \mathbb{R}^2 \rightarrow \mathbb{R}$ is strictly increasing in both arguments. $F$ is an unknown function that allows for unspecified non-linearity in the model. The second equation describes the observed dependent variable $y_i$, where $D: \mathbb{R} \rightarrow \mathbb{Y}$ is weakly increasing and non-degenerate. The set $\mathbb{Y}$ is a subset of the real line that describes the set of possible dependent variable values. The binary choice model, censored model, and proportional hazards model are examples of (1); see Han (1987) for a discussion. For instance, the binary choice model has $F(x_i' \beta, \varepsilon_i) = x_i' \beta + \varepsilon_i$, $D(y^*_i) = 1$ ($y^*_i > 0$), and $\mathbb{Y} = \{0, 1\}$.

The natural extension of the generalized regression model to a fixed-effects setting is

$$
\begin{align*}
y^*_it & = F(x^*_it \beta, \alpha_i, \varepsilon_{it}) \\
y_{it} & = D(y^*_it) \quad (i = 1, \ldots, n; \ t = 1, 2),
\end{align*}
$$

(2)

where $x^*_it$ and $\beta$ are $q \times 1$ vectors, $\alpha_i$ is the fixed effect (fixed for a given cross-sectional unit over time), and $\varepsilon_{it}$ is an error disturbance. The number of cross-sectional units, $n$, is large, whereas the number of time periods is small. For simplicity, we restrict attention to the case of two time periods ($t = 1, 2$). The results extend easily to the case of more time periods and unbalanced panels.\(^1\) The fixed effect $\alpha_i$ may be correlated with the error disturbances, which makes estimation of $\beta$ more difficult than in the cross-sectional model. As with other fixed-effects estimation techniques, we can only identify the coefficients of time-varying covariates. Thus, $x^*_it$ consists of only time-varying covariates (i.e., those covariates which change with probability greater than zero). Any time-invariant covariates are considered as part of the time-invariant fixed-effect $\alpha_i$.

The first equation in (2) describes the ‘latent’ dependent variable $y^*_it$, where $F$ is strictly increasing in its first and third arguments. In order to simplify the proof of consistency, we also assume that $F$ is continuous with respect to its third argument. The only difference from the model in (1) is that $F$ has an additional argument, the fixed effect $\alpha_i$. As opposed to most fixed-effects models in the econometrics literature, the model in (2) does not restrict the fixed effect to be additive. In fact, one could have a vector of fixed effects and complicated interactions with the fixed effects, e.g.,

$$
y^*_it = \alpha_{1i}(x^*_it \beta)^{\alpha_{2i}} + \alpha_{3i} + \varepsilon_{it},
$$

(3)

where $\alpha_{1i} > 0$ and $\alpha_{2i} > 0$ (to ensure that $F$ is increasing in its first argument).

\(^1\) See Charlier et al. (1995) for an extension of Manski (1987) to more than two time periods. The same extension applies to the class of estimators described in this section.
The model implicitly assumes that \( D \) and \( F \) do not vary over time. The restriction on \( D \) is not substantive, though, since the econometrician can censor the data so that the function \( D \) does not vary over time. The time invariance of \( F \) is crucial for the ranking techniques described below to be applicable.

Since \( F \) is unspecified in the generalized regression model, the scale and location of \( \beta \) are not identified. As a result, a suitable normalization on \( \beta \) is needed for estimation. To fix the scale of \( \beta \), we normalize the parameter vector so that \( |\beta_1| = 1 \) (i.e., the magnitude of the first component of the parameter vector is equal to one).\(^2\) Corresponding to the normalization on \( \beta \), define the parameter space of interest, \( \mathcal{B} \), as a compact subset of \( \{ \beta \in \mathbb{R}^q : |\beta_1| = 1 \} \).

We use standard notation to denote first differences:

\[
\Delta y_i \equiv y_{i2} - y_{i1}, \quad \Delta x_i \equiv x_{i2} - x_{i1}, \quad \Delta \varepsilon_i \equiv \varepsilon_{i2} - \varepsilon_{i1}.
\]

The first difference of the independent variables, \( \Delta x_i \), is also a \( q \times 1 \) vector. In a model without a time trend, the constant component can be omitted from the \( \Delta x_i \) vector.

When no parametric assumptions are made on the error disturbances \( \varepsilon_{it} \), there exist \( \sqrt{n} \)-consistent estimators in the literature for only three special cases of the model in (2):

1. Linear model

\[
F(x_{it}' \beta, x_{it}, \varepsilon_{it}) = x_{it}' \beta + x_{it} + \varepsilon_{it}, \quad D(y_{it}') = y_{it}'.
\]

Ordinary least-squares regression of \( \Delta y_i \) on \( \Delta x_i \) yields a \( \sqrt{n} \)-consistent estimate of \( \beta \) (including scale and location) if \( \Delta \varepsilon_i \) is uncorrelated with \( \Delta x_i \).

2. Censored linear model

\[
F(x_{it}' \beta, x_{it}, \varepsilon_{it}) = x_{it}' \beta + x_{it} + \varepsilon_{it}, \quad D(y_{it}') = y_{it}' \cdot 1(y_{it}' > 0).
\]

Honore (1992) develops \( \sqrt{n} \)-consistent estimators for \( \beta \) (including scale and location). The weakest assumption needed for consistency is that the error disturbances are stationary over time for each cross-sectional unit (i.e., \( \varepsilon_{i1} \) has the same marginal distribution as \( \varepsilon_{i2} \)), though this common marginal distribution may differ over cross-sectional units.

3. Transformation model

\[
F(x_{it}' \beta, x_{it}, \varepsilon_{it}) = h(x_{it}' \beta + x_{it} + \varepsilon_{it}), \quad D(y_{it}') = y_{it}'.
\]

\(^2\) Other normalizations (e.g. \( \| \beta \| = 1 \)) can be used. The normalization chosen here is the one used by Horowitz (1992).
where $h$ is an unknown, strictly increasing function. Abrevaya (1999) develops $\sqrt{n}$-consistent estimators of $\beta$ (up-to-scale), under the same stationarity assumption described above.

In the first two models, both $D$ and $F$ are known functions, which allows for identification of the location and scale of $\beta$. In the last model, $F$ is not fully specified – additivity of the fixed effects and error disturbances is assumed, but $h$ is left unknown. If $D$ and/or $F$ are specified incorrectly in the models above, the estimators for $\beta$ will be inconsistent. In this situation, if the data obeys the more flexible model given in (2), the rank estimators of this paper are consistent (up-to-scale). At the very least, these estimators can be used as a specification check when one of the three models above is considered. More importantly, however, these estimators are applicable to the more general model in (2).

The basic idea of rank estimation was first utilized by Manski (1987) for the semiparametric binary choice model with fixed effects:

$$F(x'_{1i}\beta, x_i, \varepsilon_i) = x'_{1i}\beta + x_i + \varepsilon_i,$$

$$D(y^*_i) = 1(y^*_i > 0).$$

The maximum score (MS) estimator of $\beta$ maximizes

$$S^{\text{ms}}_n(b) = \frac{1}{n} \sum_{i=1}^{n} \{1(\Delta y_i > 0) \cdot 1(\Delta x'_i b > 0) + 1(\Delta y_i < 0) \cdot 1(\Delta x'_i b < 0)\},$$

based on the idea that it is more likely to see $y_{i2} = 1$ than $y_{i1} = 1$ if the index value $x'_{12}\beta$ is greater than the index value $x'_{11}\beta$ (and vice versa for $x'_{12}\beta < x'_{11}\beta$).

The MS estimator is consistent under a stationarity assumption on the error disturbances ($\varepsilon_i$ and $\varepsilon_i$ having the same marginal distribution for each cross-sectional unit). Kim and Pollard (1990) have shown that the cross-sectional MS estimator (Manski, 1975, 1985) converges at the rate of $n^{-1/3}$ and has a complicated limiting distribution. Due to the similarity between the cross-sectional and fixed-effects MS estimators, it is quite likely that the proof of Kim and Pollard (1990) can be modified to show that the fixed-effects MS estimator also converges at rate $n^{-1/3}$. A ‘smoothed’ MS estimator developed by Horowitz (1992), and applied to the binary choice fixed-effects model by Kyriazidou (1997) and Charlier et al. (1995), is asymptotically normal with a faster convergence rate. The convergence rate of the smoothed estimator is at least $n^{-2/5}$ and can be made arbitrarily close to $n^{-1/2}$, depending on the strength of smoothness assumptions.

3. A class of rank estimators

The class of rank estimators for estimating $\beta$ in (2) is based on the simple idea that the observations corresponding to a given cross-sectional unit can be
ranked against each other over time. If \( x_i' \beta > x_i' \beta \), for instance, then the monotonicity of \( D \) and \( F \) implies that it will be more likely to see higher values for \( y_{i2} \) than for \( y_{i1} \) (under suitable assumptions on the error disturbances). These comparisons will be valid since each observation corresponding to a given cross-sectional unit has the same fixed effect. The estimators will use only within-unit (rather than between-unit) information.

The general rank estimator \( b_n \) maximizes the objective function

\[
S_n(b) = \frac{1}{n} \sum_{i=1}^{n} \left\{ H(y_{i2}, y_{i1}) \cdot 1(\Delta x_i' b > 0) + H(y_{i1}, y_{i2}) \cdot 1(\Delta x_i' b < 0) \right\}
\]  

over the parameter space \( \mathcal{B} \), where the non-degenerate function \( H: \mathcal{Y} \times \mathcal{Y} \to \mathcal{R} \) is weakly increasing in its first argument and weakly decreasing in its second argument:

\[
\begin{align*}
  u_1 > u_2 & \Rightarrow H(u_1, v) \geq H(u_2, v) \quad \forall v, \\
v_1 > v_2 & \Rightarrow H(u, v_1) \leq H(u, v_2) \quad \forall u.
\end{align*}
\]

The key conditions underlying consistency of \( b_n \) are:

\[
\Delta x' \beta > 0 \Rightarrow E[H(y_{i2}, y_{i1})|x_1, x_2, x] \geq E[H(y_{i1}, y_{i2})|x_1, x_2, x]
\]

and

\[
\Delta x' \beta < 0 \Rightarrow E[H(y_{i2}, y_{i1})|x_1, x_2, x] \leq E[H(y_{i1}, y_{i2})|x_1, x_2, x].
\]

These conditions ensure that \( \beta \) maximizes the limiting objective function.

The rank estimator \( b_n \) is a conditional estimator since only observations with \( H(y_{i2}, y_{i1}) \neq H(y_{i1}, y_{i2}) \) can affect which parameter vector maximizes (5). When \( H(u, v) = 1 \ (u > v) \), the objective function in (5) is equivalent to the maximum score objective function in (4). Other examples for \( H(u, v) \) include:

- \( H(u, v) = M(u), M \) increasing.
- \( H(u, v) = u/v \) if \( \mathcal{Y} \subset \mathcal{R}^+ \).
- \( H(u, v) = 1(u > 2v) \).
- \( H(u, v) = 1(u > v) \cdot (u - v)^2 \).
- \( H(u, v) = 1(u > v) \cdot |u - v| \).

The objective function in (5) is easy to compute and requires only \( O(n) \) calculations. Since the objective function is discontinuous in \( b \), non-gradient search methods (e.g., the Nelder–Mead simplex algorithm) need to be used in order to maximize the objective function.

The following concept of ‘separability’ helps to classify the rank estimators described by (5):
Definition. \( H \) is separable if \( \exists G_1, G_2: \mathcal{Y} \rightarrow \mathcal{R} \) s.t. \( H(u, v) = G_1(u) + G_2(v) \) \( \forall u, v \in \mathcal{Y} \).

Separability is related to the assumption on the error disturbances needed for consistency (i.e., for (6) and (7) to hold). A rank estimator with separable \( H \) requires only stationarity over time of the disturbances for each cross-sectional unit. The maximum score estimator applied to the binary choice model is consistent under stationarity since \( \mathcal{Y} = \{0, 1\} \) yields separability. For general models, though, stationarity is not sufficient for consistency of the maximum score estimator. For such models, rank estimators with separable \( H \) are consistent under stationarity whereas rank estimators with non-separable \( H \) are consistent under the stronger i.i.d. assumption (i.i.d. over time for each cross-sectional unit, but possibly heteroskedastic across units).

We show that condition (6) holds for general \( H \) with i.i.d. errors or for separable \( H \) with stationary errors. By symmetry, condition (7) will also hold. If \( \varepsilon_1 \) and \( \varepsilon_2 \) are i.i.d. and \( \Delta x' \beta > 0 \), then

\[
E[H(y_2, y_1) | x_1, x_2, \varepsilon] = E[H(D \cdot F(x_2' \beta, \varepsilon, \varepsilon_2), D \cdot F(x_1' \beta, \varepsilon, \varepsilon_1)) | x_1, x_2, \varepsilon]
= E[H(D \cdot F(x_2' \beta, \varepsilon, \varepsilon_1), D \cdot F(x_1' \beta, \varepsilon, \varepsilon_2)) | x_1, x_2, \varepsilon]
\geq E[H(D \cdot F(x_1' \beta, \varepsilon, \varepsilon_1), D \cdot F(x_2' \beta, \varepsilon, \varepsilon_2)) | x_1, x_2, \varepsilon]
= E[H(y_1, y_2) | x_1, x_2, \varepsilon].
\]

The i.i.d. assumption gives the second equality. The weak inequality follows since \( F \) is strictly increasing in its first argument. If \( \varepsilon_1 \) and \( \varepsilon_2 \) are stationary and \( H \) is separable, then

\[
E[H(y_2, y_1) | x_1, x_2, \varepsilon] = E[H(D \cdot F(x_2' \beta, \varepsilon, \varepsilon_2), D \cdot F(x_1' \beta, \varepsilon, \varepsilon_1)) | x_1, x_2, \varepsilon]
= E[G_1(D \cdot F(x_2' \beta, \varepsilon, \varepsilon_2)) + G_2(D \cdot F(x_1' \beta, \varepsilon, \varepsilon_1)) | x_1, x_2, \varepsilon]
= E[G_1(D \cdot F(x_2' \beta, \varepsilon, \varepsilon_1)) + G_2(D \cdot F(x_1' \beta, \varepsilon, \varepsilon_2)) | x_1, x_2, \varepsilon]
\geq E[G_1(D \cdot F(x_1' \beta, \varepsilon, \varepsilon_1)) + G_2(D \cdot F(x_2' \beta, \varepsilon, \varepsilon_2)) | x_1, x_2, \varepsilon]
= E[H(y_1, y_2) | x_1, x_2, \varepsilon].
\]

With separability, only the marginal distributions of \( \varepsilon_1 \) and \( \varepsilon_2 \) matter above since the expectation is a linear operator. Thus, as in the i.i.d. case, \( \varepsilon_1 \) and \( \varepsilon_2 \) can be switched to yield the third equality.
In Section 3.1, the regularity conditions for strong consistency are presented. In Section 3.2, a smoothed version of the objective function is introduced to allow for faster rates of convergence. Where possible, the notation from Manski (1987) and Horowitz (1992) is used.

3.1. Consistency

In this section, strong consistency of the class of rank estimators defined by (5) is shown. The following assumptions are sufficient to satisfy the conditions of the consistency proof.

The first assumption is a standard i.i.d. sampling assumption:

**Assumption 1.** An i.i.d. sample \(\{(x_{i1}, x_{i2}, z_i, e_{i1}, e_{i2}); i = 1, \ldots, n\}\) is drawn from the population. The observed sample is \(\{y_{i1}, y_{i2}, x_{i1}, x_{i2}; i = 1, \ldots, n\}\), where \(y_i\) is generated according to the model in (2). The observed sample has \(x_1, x_2 \in \mathcal{R}^q (q > 1)\) and \(y_1, y_2 \in \mathcal{R}\).

The following two assumptions apply to the error disturbances of the model. The first, a stationarity assumption, is used to show consistency when \(H\) is separable. The second, an i.i.d. assumption, is used to show consistency for any \(H\).

**Assumption E.1.** \(e_1\) and \(e_2\) are stationary conditional on \((x_1, x_2, z)\), with positive density almost everywhere along \(\mathcal{R}\). Denote the common marginal p.d.f. and c.d.f. by \(g(\cdot | x_1, x_2, z)\) and \(G(\cdot | x_1, x_2, z)\), respectively.

**Assumption E.2.** \(e_1\) and \(e_2\) are i.i.d. conditional on \((x_1, x_2, z)\), with positive density almost everywhere along \(\mathcal{R}\). Denote the common marginal p.d.f. and c.d.f. by \(g(\cdot | x_1, x_2, z)\) and \(G(\cdot | x_1, x_2, z)\), respectively.

The function \(H\) must satisfy some regularity conditions:

**Assumption 2.** The function \(H\) satisfies:

(a) \(u_1 > u_2 \Rightarrow H(u_1, v) \geq H(u_2, v) \forall v\) and \(v_1 > v_2 \Rightarrow H(u, v_1) \leq H(u, v_2) \forall u;\)

(b) \(E[|H(y_2, y_1) - H(y_1, y_2)|^{2+\eta}] < \infty\) for some \(\eta > 0;\)

(c) For any \(u \in \mathcal{R}\) and conditional on \((x_1, x_2, z)\), either \(H(y_1, u)\) or \(H(u, y_2)\) is a non-constant random variable.

Part (a) re-states the monotonicity property introduced previously. Part (b) can be weakened slightly \((\eta = 0)\) to yield weak consistency; see the appendix for details. Part (c) is used to prove uniqueness of the limiting objective function’s maximum. When error disturbances are additive (and satisfy either Assumptions E.1 or E.2), this assumption is satisfied if \(H\) is non-degenerate. For the
maximum score objective function, this assumption ensures that both 
Pr(\Delta y > 0) and Pr(\Delta y < 0) occur with positive probability.
The following assumption on the covariates is needed for identification of \( \beta \):

**Assumption 3.** (a) The support of the distribution of \( \Delta x \) is not contained in any
proper linear subspace of \( \mathbb{R}^q \);
(b) \( \beta_1 \neq 0 \), and for almost every \( \Delta \tilde{x} \equiv (\Delta x_2, \ldots, \Delta x_q)' \), the distribution of \( \Delta x_1 \)
conditional on \( \Delta \tilde{x} \) has everywhere positive density with respect to Lebesgue
measure.

Assumption 3(a) is the usual full-rank condition. Assumption 3(b) is a conti-
uinity assumption frequently made in the semiparametric literature.
Finally, we normalize the parameter vector and assume compactness of the
parameter space:

**Assumption 4.** \( |\beta_1| = 1 \), and \( \tilde{\beta} \equiv (\beta_2, \ldots, \beta_q)' \) is contained in a compact subset
\( \mathcal{B} \) of \( \mathbb{R}^{q-1} \).

For notational purposes, let \( \tilde{b} \equiv (b_2, \ldots, b_q)' \) denote the last \( q - 1 \) compo-
nents of any parameter vector \( b \in \mathbb{R}^q \). Also, define the parameter space of interest
as \( \mathcal{B} \equiv \{ b : |b_1| = 1, b \in \mathcal{B} \} \).

The strong-consistency result is given in the following theorem:

**Theorem 1.** Let Assumptions 1–4 hold. Let \( b_n \) be a solution to \( \max_{b \in \mathcal{B}} S_n(b) \). If (i)
Assumption E.2 holds or (ii) \( H \) is separable and Assumption E.1 holds, then
\( \lim_{n \to \infty} b_n = \beta \) almost surely.

The complete proof of Theorem 1 is given in the appendix. The main
technicality involves using Assumption 2(c) to strengthen the weak inequalities
in conditions (6) and (7) to strict inequalities so that \( \beta \) is the unique maximizer of
the limiting objective function.

3.2. Smoothed estimators

Note that the objective function in (5) is maximized by the same parameter
value as

\[
1 \sum_{i=1}^n \{ H(y_{i2}, y_{i1}) - H(y_{i1}, y_{i2}) \} I(\Delta x'_i b > 0).
\]

(8)

To smooth this objective function, we follow Horowitz (1992) and replace
\( I(\Delta x'_i b > 0) \) by a smooth function of \( \Delta x'_i b \). In particular, the smoothed estimator
maximizes the objective function

\[
S_n(b; \sigma_n) = \frac{1}{n} \sum_{i=1}^{n} \left\{ H(y_{i2}, y_{i1}) - H(y_{i1}, y_{i2}) \right\} K\left( \frac{\Delta x_i'b}{\sigma_n} \right) \]

over the parameter space \( \mathcal{B} \), where the function \( K: \mathbb{R} \to \mathbb{R} \) and the sequence \( \{ \sigma_n \} \) satisfy the following two assumptions:

**Assumption 5.** (a) \( K \) is a continuous function such that \( |K(v)| < M \forall v \) for some finite \( M \); (b) \( \lim_{v \to -\infty} K(v) = 0 \) and \( \lim_{v \to \infty} K(v) = 1 \).

**Assumption 6.** \( \{ \sigma_n \} \) is a positive sequence with \( \lim_{n \to \infty} \sigma_n = 0 \).

Strong consistency is easy to show since the smoothed objective function is a weighted version of the original objective function. Using Assumptions 5 and 6, the proof follows that of Theorem 1 with the additional technicalities handled as in the proof of Lemma 4 of Horowitz (1992). We formally state the strong consistency result in the following theorem:

**Theorem 2.** Let Assumptions 1–6 hold. Let \( b_n \) be a solution to \( \max_{b \in \mathcal{B}} S_n(b; \sigma_n) \). If (i) Assumption E.2 holds or (ii) \( H \) is separable and Assumption E.1 holds, then \( \lim_{n \to \infty} b_n = \beta \) almost surely.

Extension of the asymptotic normality results of Horowitz (1992) (see also Kyriazidou, 1997; Charlier et al., 1995) is rather straightforward. The key difference is that smoothness assumptions on the c.d.f. of the error disturbance (see Assumption 9 of Horowitz, 1992) are replaced by smoothness assumptions on the conditional expectation of \( H(y_2, y_1) - H(y_1, y_2) \) near \( \Delta x'\beta = 0 \). Rather than repeating the relevant assumptions and results from Horowitz (1992), we summarize the necessary changes in the appendix.

Horowitz (1992) discusses how to choose the bandwidth for estimation and how to correct the asymptotic bias. The same techniques can be used here as well. Presumably, use of the bootstrap can improve upon asymptotic approximations in finite samples, as is the case for the smoothed MS estimator of the binary choice model (see Horowitz, 1996).

### 4. Truncated data

In this section, we discuss applicability of the class of rank estimators when the dependent variable may be truncated. We use the notion of truncation in
Honore (1992), where the data for a cross-sectional unit \( i \) are observed only if both \( y_{i1} \) and \( y_{i2} \) are positive (i.e., left truncation at zero).\(^3\)

We consider a version of the generalized fixed-effects regression model in (2) with additive error disturbances:\(^4\)

\[
y_{it}^p = F(x'_{it} \beta, \alpha_i) + \varepsilon_{it} \\
y_{it} = D(y_{it}^p),
\]

(10)

where \((y_{i1}, y_{i2}, x_{i1}, x_{i2})\) is observed if and only if \( y_{i1} > 0 \) and \( y_{i2} > 0 \). Since \( D \) is an increasing function, there exists some \( \lambda \in \mathbb{R} \) such that \( y_{it} > 0 \) if and only if \( y_{it}^p > \lambda \). To formalize the truncated sampling, we modify Assumption 1 as follows:

**Assumption 1’**. An i.i.d. sample \( \{(x_{i1}, x_{i2}, \alpha_i, \varepsilon_{i1}, \varepsilon_{i2}); i = 1, \ldots, n\} \) is drawn from the population, conditional on the event \( y_{i1} > 0 \) and \( y_{i2} > 0 \) (where \( y_{it} \) is generated according to the model in (10)). The observed sample is \( \{(y_{i1}, x_{i1}, y_{i2}, x_{i2}); i = 1, \ldots, n\} \), with \( x_{1}, x_{2} \in \mathbb{R}^q (q > 1) \) and \( y_{1}, y_{2} \in \mathbb{R} \).

Implicit in Assumption 1’ is that the unconditional probability \( \Pr(y_{1} > 0, y_{2} > 0) \) is non-zero.

As before, consistency requires the monotonicity conditions (6) and (7) to hold. Truncation complicates matters since the marginal distributions of \( \varepsilon_1 \) and \( \varepsilon_2 \) are no longer the same after conditioning on observability. If \( y_1 \) and \( y_2 \) are observed, it must be the case that \( \varepsilon_1 > L - F(x'_{1} \beta, \alpha) \) and \( \varepsilon_2 > L - F(x'_{2} \beta, \alpha) \). For \( \Delta x' \beta > 0 \), \( \varepsilon_2 \) can take on smaller values than \( \varepsilon_1 \), meaning that some sort of shape restriction on the error distribution is needed in order to still ‘expect’ \( y_{2}^p \) to be larger than \( y_{1}^p \).

Recall that \( G(\cdot \mid x_1, x_2, \alpha) \) and \( g(\cdot \mid x_1, x_2, \alpha) \) are the common marginal c.d.f. and p.d.f. for the error disturbances. The conditioning arguments will be suppressed in what follows. With left-truncated data, a sufficient condition for consistency (in addition to the assumptions from the previous section) is that \( (1 - G) \) is strictly log-concave (i.e., the logarithm of \( (1 - G) \) is strictly concave). This condition is equivalent to an increasing hazard function \( g/(1 - G) \), since the derivative of \( \log(1 - G) \) is \( -g/(1 - G) \). The consistency result is formally stated in the following theorem:

---

\(^3\)The results of this section can be easily generalized to other forms of truncation. For right-truncated data, a sufficient condition for consistency in Theorem 3 is that the common c.d.f. of the error disturbance is strictly log-concave.

\(^4\)The assumption of additive errors is not as restrictive as it might seem. For instance, a model with multiplicative errors can be written as a model within the framework of (10) by applying a logarithmic transformation.
Theorem 3. Let Assumptions 1’, 2–4, and E.2 hold. Let $b_n$ be a solution to $\max_{b \in \mathbb{R}} S_n(b)$. Assume that $G$ is differentiable. If $(1 - G)$ is strictly log-concave, then $\lim_{n \to \infty} b_n = \beta$ almost surely.

The proof of Theorem 3 is in the appendix. The basic idea is that for $x'_2 \beta > x'_1 \beta$ and $x$, the log-concavity assumption is sufficient to show that the random variable $F(x'_2 \beta, x) + \epsilon_2$ stochastically dominates the random variable $F(x'_1 \beta, x) + \epsilon_1$ in the first-order sense, conditional on both random variables being observed. (When there is no truncation, the stochastic dominance is trivial since $\epsilon_1$ and $\epsilon_2$ have the same marginal distribution and $F$ is strictly increasing in its first argument.)

It is interesting to compare the log-concavity condition of Theorem 3 to the assumption used by Honoré (1992) for the linear truncated model with fixed-effects. Honoré (1992) assumes log-concavity of the common p.d.f. $g(\cdot | x_1, x_2, x)$ of the error disturbances (rather than log-concavity of the common c.d.f. of the error disturbances). The consistency proof in Honoré (1992) is based on the symmetry and unimodality of $\Delta e$ conditional on $(\epsilon_1 + \epsilon_2)$, for which log-concavity of the p.d.f. is needed (see Lemma 1 of Honoré, 1992). Whether log-concavity of the p.d.f. can be relaxed is unknown, but the unimodality condition seems essential for identification (see also Powell, 1986, where unimodality is discussed in the cross-sectional version of the estimator). The class of rank estimators, on the other hand, do not require unimodality of $\Delta e$. The log-concavity assumption of Theorem 3 (i.e., log-concavity of the common survival function) is weaker than the assumption of Honoré (1992) since

$$g \text{ strictly log-concave } \Rightarrow G \text{ and } (1 - G) \text{ strictly log-concave},$$

a well-known result in statistics (see Pratt, 1981 or Dharmadhikari and Joag-dev, 1988). The converse is not true. The class of distributions with log-concave $g$ corresponds to the class of strongly unimodal p.d.f.s, whereas the class of distributions with log-concave $(1 - G)$ includes multimodal p.d.f.s. It is straightforward to construct distributions having multimodal p.d.f.s and log-concave c.d.f.s. The difference in log-concavity assumptions is mainly of theoretical interest. As a practical matter, simple simulations by the author have found that the estimators of Honoré (1992) seem to work fine when $(1 - G)$ is log-concave and $g$ is not.

The results on smoothed estimation from the previous section extend immediately to truncated data once the conditions for consistency are satisfied.

---

5 In the special case where $D$ is the identity function and $H(u, v) = u$, one can show that a weaker condition (strict log-concavity of $\int (1 - G(v)) \, dv$) is sufficient for strong consistency. See, for instance, Eq. (B - 7) in Heckman and Honoré (1990).
5. Monte Carlo simulations

In this section, we present some Monte Carlo evidence on the performance of the estimators (both unsmoothed and smoothed versions) for sample sizes of 250 and 1000. The design considered is

\[ y_t = \alpha(x_{1t} + x_{2t}) + \epsilon_t, \quad t = 1, 2, \tag{11} \]

where

\[ x_{1t} \sim \text{N}(0, 4), \quad x_{2t} = \begin{cases} 0 & \text{if } t = 1, \\ 1 & \text{if } t = 2, \end{cases} \quad \alpha \sim \text{U}[0, 1], \text{ and } \epsilon_t \sim \text{N}(0, 1). \]

Each of the random variables are independent of each other. The second covariate is a time trend. In some sense, the design considered is the simplest possible for the covariates since one component is continuous (needed for identification) and the other has no variation at all. The fixed effect is multiplicative. (In a model of firm output, for instance, \( \alpha \) could be thought of as an unobserved measure of firm efficiency that remains fixed over time.)

Three different rank estimators were used in the simulations, with objective functions:

**OBJ1**: \( H(u, v) = 1(u > v). \)

**OBJ2**: \( H(u, v) = 1(u > v) \cdot |u - v|. \)

**OBJ3**: \( H(u, v) = 1(u > v) \cdot (u - v)^2. \)

OBJ1 is the maximum-score objective function, and OBJ2 and OBJ3 weight observational units by the absolute difference in the dependent variable and the squared difference in the dependent variable, respectively. For smoothed estimation, we report results obtained by using the fourth-order smoothing function used by Horowitz (1992):\(^7\)

\[ K(v) = \begin{cases} 0 & \text{if } v < -5, \\ 0.5 + (105/64)[(v/5) - (5/3)(v/5)^3 \\ \quad + (7/5)(v/5)^5 - (3/7)(v/5)^7], & \text{if } -5 \leq v \leq 5, \\ 1 & \text{if } v > 5. \end{cases} \tag{12} \]

The results from other smoothing functions were quite similar and, thus, are not reported.

\(^6\) Other designs were also considered, including designs in which (i) there was dependence between the covariates and the fixed effect, and (ii) there was heteroskedasticity of the error disturbance. The basic results were quite similar, so the results from the simplest design are presented.

\(^7\) The optimal convergence rate for this smoothing function is \( n^{-4/9} \).

Simulations were performed for \( n = 250 \) and 1000. In all cases, 500 replications were carried out. Using the normalization from the previous sections, the coefficient of the first covariate was held fixed at one. The coefficient on the second covariate (whose true value is 1) was estimated using a grid search. The results are summarized in Table 1, with mean, standard deviation, root-mean-squared error, and mean absolute deviation reported for each estimator. The first row for each estimator corresponds to the unsmoothed estimator. The second row corresponds to the optimal smoothed estimator (in terms of RMSE), which was determined by doing 500 replications at increments of 0.05 for \( \sigma_w \).

The three unsmoothed estimators seem to perform pretty well for both sample sizes. The estimators have very little bias. Not surprisingly, the OBJ1 estimator is somewhat less efficient (RMSE’s 10–15% higher) than the OBJ2 and OBJ3 estimators. The OBJ1 estimator does not use any information about the levels of the dependent variables, only their relative rankings. However, the OBJ1 estimator does have the virtue of being more robust (than OBJ2 and OBJ3) to outliers; see Cavanagh and Sherman (1998) for a discussion. The optimal smoothed OBJ1 and OBJ2 estimators yield substantial efficiency gains over their unsmoothed counterparts. The smoothed OBJ3 estimator is almost more efficient than its unsmoothed counterpart, but the bias of the estimator reduces the efficiency gain (in terms of RMSE) to about 10% for both sample sizes.

Finally, notice that the results in Table 1 are consistent with the slower rates of convergence for the rank estimators. For \( \sqrt{n} \)-consistent estimators, one

\[8\] In cases where multiple values maximized the objective function, the estimate was chosen at random from the set of maximizers.
would expect the standard deviation or RMSE to halve when the sample size is quadrupled from $n = 250$ to 1000. The standard deviations for the unsmoothed estimators are about 62–65% lower for $n = 1000$, whereas the standard deviations for the smoothed estimators are about 55–60% lower for $n = 1000$. These numbers are consistent with the theoretical finding that the smoothed estimators converge faster than the unsmoothed estimators but slower than the $\sqrt{n}$ rate.

While the smoothed estimators are certainly more attractive from a theoretical perspective and perform better than their unsmoothed counterparts when the bandwidth is chosen optimally, Horowitz (1992) finds that the RMSE of the smoothed estimator is quite sensitive to the bandwidth choice. Our Monte Carlo results also confirm the sensitivity of the estimates to the bandwidth choice. For $n = 250$, Fig. 1 graphs the bias (in absolute value), standard deviation, and RMSE of the smoothed OBJ1 estimator for different values of the bandwidth $\sigma_n$. Fig. 2 is the same, except for $n = 1000$. From Table 1, the RMSE of the unsmoothed OBJ1 estimator is 0.5951 for $n = 250$ and 0.3825 for $n = 1000$. Comparing these values to the RMSE curves in Figs. 1 and 2, the smoothed OBJ1 estimator performs better (in terms of RMSE) for bandwidths between 1.17 and 1.56 for $n = 250$ and for bandwidths between 1.21 and 1.53 for $n = 1000$. The graphs for the OBJ2 and OBJ3 estimators look quite similar and are thus omitted. Table 2 summarizes the range of bandwidths for which the
smoothed estimators outperform the unsmoothed estimators in terms of RMSE. For each of the estimators, deviations of more than 10–20% from the optimal bandwidth choice cause the smoothed estimator to be less efficient than the unsmoothed estimator. Since there is no automatic procedure for choosing $\sigma_n$ (see Horowitz (1992) for a method akin to the plug-in method for kernel estimation), the unsmoothed estimators may have appeal for practitioners. At the very least, if smoothed estimation is used, it would make sense to report estimates for several bandwidth choices.
6. Conclusion

This paper has focused on the case of two time periods \((t = 1, 2)\). Generalizing the objective function in (5) to more time periods \((T > 2)\) is straightforward:

\[
\left[ \eta \left( \frac{T}{2} \right) \right]^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \{ H(y_{is}, y_{it}) \mathbb{1}(x_{it}'b > x_{is}'b) + H(y_{is}, y_{it}) \mathbb{1}(x_{it}'b < x_{is}'b) \}.
\]

For \(H(u, v) = 1(u > v)\) (maximum score estimation), Charlier et al. (1995) give details on the asymptotic theory for \(T > 2\). Computation is still quick since \(T\) is fixed and small.

The approach discussed in this paper suffers many of the same drawbacks as previous papers in the fixed-effects estimation literature. First, estimation relies on a strict exogeneity assumption. Dealing with lagged dependent variables in non-linear fixed-effects models is a difficult problem. For instance, Honoré (1993) has considered moment conditions for estimation of the censored and truncated models with fixed-effects and lagged dependent variables, but identification of the parameter vector is not obtained. More connected to the rank estimators of this paper is Honoré and Kyriazidou (1997), which considers use of the maximum score estimator for a binary choice fixed-effects model with lagged dependent variables. Their approach requires four time periods worth of data as well as additional restrictions on the model (e.g., no time trend between the second and third time periods).

Second, the coefficients on time-invariant covariates are not identified by the class of estimators. Again, this weakness is shared by other papers in the fixed-effects literature. In order to identify these coefficients, additional exclusion restrictions on the fixed effects and covariates (e.g., Hausman and Taylor (1981) for the linear model) need to be made. This type of approach is an interesting topic for future research on non-linear fixed-effects models.

Further research on the rank estimators described in this paper might focus on issues concerning efficiency. The choice of \(H\) affects the efficiency of the estimators but there is no theory telling the practitioner which \(H\) is “optimal”. Cavanagh and Sherman (1998) provide some “rules of thumb” for choosing this function for cross-sectional rank estimators. There may also be efficiency gains from weighted versions of the rank estimators.

Acknowledgements

An earlier version of this paper (“Maximum Score Estimation of Linear Index Panel Models”) was part of the author’s Ph.D. dissertation at MIT. Comments from Jerry Hausman, Bo Honoré, Whitney Newey, and two anonymous referees substantially improved this paper. Financial support from the National Science
Appendix

Proof of Theorem 1. Notice that the objective function $S_n(b)$ has the same maximizer as the objective function

$$R_n(b) = \frac{1}{n} \sum_{i=1}^{n} 1(\Delta x'_i b > 0)(H(y_{i2}, y_{i1}) - H(y_{i1}, y_{i2}))$$

since

$$S_n(b) = R_n(b) + \frac{1}{n} \sum_{i=1}^{n} H(y_{i1}, y_{i2}).$$

The last term is not a function of $b$. Consistency will be shown based on the objective function $R_n(b)$. Denote the limiting objective function by $R_0(b)$.

Let $z \equiv (y_1, x_1, y_2, x_2)$. Define

$$f(z, b) = 1(\Delta x'_b > 0)(H(y_2, y_1) - H(y_1, y_2)).$$

Then,

$$R_n(b) - R_0(b) = \sum_{i=1}^{n} f(z_i, b) - R_0(b) = P_n f(\cdot, b) - R_0(b),$$

where $P_n$ places mass $n^{-1}$ at each $z_i$ (adopting the notation of Pakes and Pollard, 1989; Sherman, 1994). \{P_n f(\cdot, b) - R_0(b): b \in \mathcal{B}\} is a zero-mean empirical process.

We verify the following conditions of Newey and McFadden (1994, Theorem 2.1 (almost-sure version)) needed for strong consistency: (i) $R_0(b)$ uniquely maximized at $\beta$; (ii) $\mathcal{B}$ compact; (iii) $R_0(b)$ continuous; and, (iv) $R_n(b)$ converges uniformly almost surely to $R_0(b)$ (i.e., $\sup_{b \in \mathcal{B}} |R_n(b) - R_0(b)| \xrightarrow{a.s.} 0$). The bulk of the proof involves verifying condition (i), so the other conditions are considered first.

Condition (ii) is satisfied by Assumption 4. Condition (iii) follows directly from the proof of Lemma 5 of Manski (1985). In that proof, continuity relies on the continuity of $\Delta \tilde{x}$ from Assumption 3(b), and the existence of the second moment of $|H(y_2, y_1) - H(y_1, y_2)|$ allows use of the dominated convergence theorem. For condition (iv), note that the class of functions $\mathcal{F} \equiv \{f(\cdot, b): b \in \mathcal{B}\}$ is Euclidean for the envelope $|H(y_2, y_1) - H(y_1, y_2)|$ (see Example 2.11 of Pakes.
and Pollard (1989), of which $\mathcal{F}$ is a slight variant). Then, Assumption 2(b) yields condition (iv) by Corollary 9 of Sherman (1994). (Note that weak consistency only requires existence of the second moment; see Corollary 7 of Sherman (1994).

For condition (i), recall from Section 3 that

$$\Delta x' \beta > 0 \Rightarrow E[H(y_2, y_1) - H(y_1, y_2) \mid x_1, x_2, \alpha] \geq 0$$  \hspace{1cm} (A.1)

and

$$\Delta x' \beta < 0 \Rightarrow E[H(y_2, y_1) - H(y_1, y_2) \mid x_1, x_2, \alpha] \leq 0.$$  \hspace{1cm} (A.2)

We now show that (A.1) and (A.2) hold with strict inequalities. Without loss of generality, we prove strict inequality for (A.1). Consider any $(x_1, x_2, a)$ with $\Delta x' \beta > 0$. We consider the case of i.i.d. error disturbances and arbitrary $H$. Extension to stationarity and separability is trivial. By the weak monotonicity of $D$, $F$, and $H$, we have

$$E[H(D \cdot F(x_2' \beta, \alpha, \varepsilon_1), D \cdot F(x_1' \beta, \alpha, \varepsilon_2))]$$

$$\geq E[H(D \cdot F(x_1' \beta, \alpha, \varepsilon_1), D \cdot F(x_1' \beta, \alpha, \varepsilon_2))].$$  \hspace{1cm} (A.3)

Without loss of generality, we know from Assumption 2(c) that $H(y_1, u)$ is a non-constant random variable (conditional on $(x_1, x_2, \alpha)$). (A symmetric argument works for non-constant $H(u, y_2)$.) Thus, for any given $\varepsilon_2$, there exists some $\varepsilon^*$ s.t. for all $\varepsilon'$ and $\varepsilon''$ satisfying $\varepsilon' > \varepsilon^* \geq \varepsilon''$, we have

$$\varepsilon' > \varepsilon^* \geq \varepsilon'' \Rightarrow H(D \cdot F(x_2' \beta, \alpha, \varepsilon'), D \cdot F(x_1' \beta, \alpha, \varepsilon_2))$$

$$> H(D \cdot F(x_2' \beta, \alpha, \varepsilon''), D \cdot F(x_1' \beta, \alpha, \varepsilon_2)).$$

By monotonicity of $F$ and continuity in its third argument, there exists some $\eta > 0$ s.t.

$$\varepsilon \in (\varepsilon^* - \eta, \varepsilon^*) \Rightarrow F(x_2' \beta, \alpha, \varepsilon) > F(x_1' \beta, \alpha, \varepsilon^*).$$

Combined with the previous inequality, we have

$$\varepsilon_1 \in (\varepsilon^* - \eta, \varepsilon^*) \Rightarrow H(D \cdot F(x_2' \beta, \alpha, \varepsilon_1), D \cdot F(x_1' \beta, \alpha, \varepsilon_2))$$

$$> H(D \cdot F(x_1' \beta, \alpha, \varepsilon_1), D \cdot F(x_1' \beta, \alpha, \varepsilon_2)).$$

Since $\varepsilon_1$ has positive density almost everywhere along $\mathcal{R}$, the event $\{\varepsilon_1 \in (\varepsilon^* - \eta, \varepsilon^*)\}$ occurs with positive probability. As a result, (A.3) holds with a strict inequality. Follow the link of equalities and inequalities from Section 3 to show that (A.1) holds with strict inequality.
Since both (A.1) and (A.2) hold with strict inequality for any $x$, we have

$$\Delta x'\beta > 0 \Rightarrow E[H(y_2, y_1) - H(y_1, y_2) | x_1, x_2] > 0$$  \hspace{1cm} (A.4)

and

$$\Delta x'\beta < 0 \Rightarrow E[H(y_2, y_1) - H(y_1, y_2) | x_1, x_2] < 0.$$  \hspace{1cm} (A.5)

The event $\{\Delta x'\beta = 0\}$ can be ignored since it occurs with zero probability. It follows immediately that $\beta$ maximizes the limiting objective function $R_0(b)$. To show uniqueness, we need to show that for all other $b \in \mathcal{B}$ (i.e., $b \neq \beta$), $R_0(b) < R_0(\beta)$. Given (A.4) and (A.5), it suffices to show that for any $b \in \mathcal{B}$ s.t. $b \neq \beta$, the indicator functions $1(\Delta x'\beta > 0)$ and $1(\Delta x'\beta > 0)$ differ on a region of positive Lebesgue measure. This fact follows immediately from Assumption 3(b); see Manski (1985) or Theorem 2.10 of Newey and McFadden (1994) for details. Thus, condition (i) is satisfied.

**Proof of Theorem 2.** Same as Theorem 1, with $K$ handled as in Horowitz (1992, Lemma 4).

**Proof of Theorem 3.** We show that conditions (6) and (7) hold for the truncated model. Once these conditions are verified, strong consistency follows from the proof of Theorem 1. Verification of (6) and (7) is based on the following lemma:

**Lemma 1.** Suppose that

$$y^*(z) = F(z) + \varepsilon,$$

where $F$ is an increasing function and $\varepsilon$ is independent of $z$ with differentiable c.d.f. $G$. If $y^*(z)$ is observed only when $y^*(z) > L$ and $\log(1 - G)$ is strictly concave, then $y^*(z_1)$ first-order stochastically dominates $y^*(z_2)$ (in the strong sense) for $z_1 > z_2$.

**Proof of Lemma.** For ease of exposition, assume that $F$ is differentiable with respect to $z$. (The argument below can be modified easily to handle a discretized version of differentiation.) To show first-order stochastic dominance, it suffices to show that

$$\frac{d \Pr(y^*(z) < c \mid y^*(z) > L)}{dz} < 0 \quad \text{for any } c > L.$$

Note that

$$\Pr(y^*(z) < c \mid y^*(z) > L) = \frac{G(c - F(z)) - G(L - F(z))}{1 - G(L - F(z))}.$$
Differentiating with respect to \(z\) yields

\[
\begin{align*}
\frac{d\Pr(y^*(z) < c \mid y^*(z) > L)}{dz} &= -\frac{F'(z)(g(c - F(z)) - g(L - F(z)))}{1 - G(L - F(z))} \\
&\quad - \frac{F'(z)(g(L - F(z))(G(c - F(z)) - G(L - F(z)))}{(1 - G(L - F(z)))^2}.
\end{align*}
\]

We are interested in the sign of this expression. Simplifying and multiplying by 
\((1 - G(L - F(z)))/(1 - G(c - F(z)))\), which doesn’t affect the sign, yields

\[
- \frac{F'(z)}{1 - G(c - F(z))} \left\{ \frac{g(c - F(z))}{1 - G(c - F(z))} - \frac{g(L - F(z))}{1 - G(L - F(z))} \right\}.
\]

Since \(F'(z)\) is positive and the hazard function \(g/(1 - G)\) is strictly increasing (since \((1 - G)\) strictly log-concave), this expression is negative. Thus, first-order stochastic dominance in the strong sense holds.

By Lemma 1, \(\Delta x^\beta > 0\) implies that the marginal distribution of the random variable \(y^*_1\) first-order stochastically dominates the marginal distribution of the random variable \(y^*_2\). By the monotonicity of \(D, F,\) and \(H\), (6) immediately follows when \(e_1\) and \(e_2\) are i.i.d. A symmetric argument for \(\Delta x^\beta < 0\) verifies condition (7), completing the proof. \(\square\)

**Asymptotic normality of the smoothed estimators:** The results of Horowitz (1992) for smoothed maximum score estimation extend to the smoothed estimators of Section 3.2. A few modifications of the assumptions and notation are required. First, \(x\) should be replaced by \(\Delta x\) to make the results applicable to the fixed-effects, rather than cross-sectional, setting.

Then, with \(z \equiv \Delta x^\beta\), define the following conditional expectation

\[
L(z, \Delta \tilde{x}) \equiv E[H(y_2, y_1) - H(y_1, y_2) \mid z, \Delta \tilde{x}].
\]

For each positive integer \(i\), define

\[
L^{(i)}(z, \Delta \tilde{x}) \equiv \partial^i L(z, \Delta \tilde{x})/\partial z^i
\]

whenever the derivative exists. Assumption 9 of Horowitz (1992) is replaced by:

**Assumption.** For each integer \(i\) such that \(1 \leq i \leq h\), all \(z\) in a neighborhood of 0, almost every \(\Delta \tilde{x}\), and some \(M < \infty\), \(L^{(i)}(z, \Delta \tilde{x})\) exists and is a continuous function of \(z\) satisfying \(|L^{(i)}(z, \Delta \tilde{x})| < M\).

The matrices in the asymptotic covariance expressions are re-defined to cover the general class of rank estimators. For each integer \(h \geq 2\), define the \((q - 1) \times 1\)
vector $A$ and the $(q - 1) \times (q - 1)$ matrices $D$ and $Q$ by

$$A \equiv -2x_A \sum_{i=1}^{h} \{ [i!(h - i)!]^{-1} E[L^{(i)}(0, \Delta \tilde{x}) p^{(h - i)}(0|\Delta \tilde{x})\Delta \tilde{x}],
$$

$$D \equiv x_D E[ E[\{H(y_2, y_1) - H(y_1, y_2)\}^2 | z = 0, \Delta \tilde{x}] \Delta \tilde{x} \Delta \tilde{x}' p(0|\Delta \tilde{x})]$$

and

$$Q \equiv 2E[\Delta \tilde{x} \Delta \tilde{x}' L^{(1)}(0, \Delta \tilde{x}) p(0|\Delta \tilde{x})],$$

whenever these quantities exist.\(^9\)

The main asymptotic results are given by Theorems 2 and 3 of Horowitz (1992). Theorem 2 describes the asymptotic distribution of the smoothed estimators, and Theorem 3 describes how the quantities in the asymptotic distribution can be consistently estimated. In Theorem 3, the only necessary modification is to define

$$t_b(b, \sigma) \equiv (H(y_{i2}, y_{1i}) - H(y_{i1}, y_{12})(\Delta \tilde{x}_i/\sigma) K'(\Delta \tilde{x}_i/\sigma).$$

References


\(^9\) Thanks to Erwin Charlier, who pointed out an error in the expression for $D$ in an earlier version of this paper.