Econometric Theory, **24**, 2008, 1149–1173. Printed in the United States of America.
doi:10.1017/S0266466608080468

SEMIPARAMETRIC ESTIMATION OF NONSTATIONARY CENSORED PANEL DATA MODELS WITH TIME VARYING FACTOR LOADS

SONGNIAN CHEN
Hong Kong University of Science and Technology
and
National University of Singapore

SHAKEEB KHAN
Duke University

We propose an estimation procedure for a semiparametric panel data censored regression model in which the error terms may be subject to general forms of nonstationarity. Specifically, we allow for heteroskedasticity over time and a time varying factor load on the individual specific effect. Empirically, estimation of this model would be of interest to explore how returns to unobserved skills change over time—see, e.g., Chay (1995, manuscript, Princeton University) and Chay and Honoré (1998, *Journal of Human Resources* 33, 4–38). We adopt a two-stage procedure based on nonparametric median regression, and the proposed estimator is shown to be $\sqrt{n}$-consistent and asymptotically normal. The estimation procedure is also useful in the group effect setting, where estimation of the factor load would be empirically relevant in the study of the intergenerational correlation in income, explored in Solon (1992, *American Economic Review* 82, 393–408; 1999, *Handbook of Labor Economics*, vol. 3, 1761–1800) and Zimmerman (1992, *American Economic Review* 82, 409–429).

1. INTRODUCTION AND MOTIVATION

Censored regression models have received a great deal of attention in both the theoretical and applied econometric literature since the seminal work of Tobin (1958). The popularity of these models stems from the fact that many observed economic variables are censored from above or below for various reasons, such as nonnegativity constraints and/or top coding.[1] Paralleling the increase in popularity of estimating censored regression models is the estimation of panel data models. The increased availability of longitudinal panel data sets has presented

This paper was presented at the 2004 European Summer meetings of the Econometric Society in Madrid. We are grateful to the co-editor J. Wooldridge for helpful comments. Khan acknowledges support from NSF grant SES-0213621. Address correspondence to Songnian Chen, Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong; e-mail: snchen@ust.hk.

© 2008 Cambridge University Press 0266-4666/08 $15.00

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

new opportunities for econometricians to control for individual unobserved heterogeneity across agents. Panel data are related to grouped data sets, where in this context group specific heterogeneity can be controlled for.

In linear panel data models, unobserved additive individual specific heterogeneity, if assumed constant over time or group member (i.e., "fixed effects"), can be controlled for when estimating the slope parameters by first differencing the observations. Unfortunately, the first differencing approach will typically be inapplicable for estimating censored regression models because of the nonlinearity introduced by censoring. Thus alternative methods are necessary when the data are censored.

Heckman and MaCurdy (1980) treat each individual specific effect as parameters to be estimated. Their approach requires parametric specification of the error term and also an increasing number of observations per individual/group for consistent estimation of the slope parameters. In addition to its sensitivity to distributional misspecification, their approach is inapplicable to most longitudinal panel data sets, which are typically characterized by a large number of individual observations observed in just a few time periods.

Alternatively, Honoré (1992) proposed estimators for censored regression models with individual specific fixed effects that do not parametrically specify the error term's distribution and are asymptotically valid as the number of individuals increases, whereas the number of observations per individual remains constant. Although this is a significant improvement over previous approaches, a crucial assumption needed for consistency of Honoré's estimator is the time stationarity of the error terms.[2] This assumption has proved suspect in many applications. For example in the study of black-white earnings differentials Chay (1995) and Chay and Lee (2000) reject the stationarity assumption on the error distribution in their respective studies.[3] The use of Honoré's estimator when stationarity is violated may provide misleading results.

There are several ways in which the stationarity assumption may be overly restrictive. One way is that it does not allow for time varying heteroskedasticity, thus ruling out any relationship between the error terms and time varying explanatory variables. Another way this assumption may be too restrictive is that it does not allow for time varying factors loading onto the individual specific effects. As discussed in Chay (1995) and Chay and Honoré (1998), this can be especially problematic when estimating wage/earnings equations, where the individual specific effect usually serves as a proxy for unobserved components of productivity. In this case the stationarity assumption would rule out the possibility of unobserved skill premiums changing over time.

To allow for time varying factor loads Chay (1995) estimated a censored random effects model. A well-known drawback of the random effects approach is that the individual specific effect is assumed to be independent of observed covariates, which seems unlikely to be satisfied in most applied settings. Honoré (1998) estimated a fixed effects censored regression model where he assumed that the error terms were distributed multivariate normal. Although his fixed

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

effect approach allows for arbitrary correlation between individual specific effects and covariates, the normality assumption implies that the estimator is not robust to distributional misspecification.

In this paper we attempt to address these issues by proposing an estimation procedure for censored regression panel data models that is robust to nonstationary errors in the longitudinal data context and group member specific errors in the group data context. Specifically, we aim to expand upon the results of Honoré (1992) by identifying the time or group member varying factor load but without imposing the normality assumption in Honoré (1998). We do so by imposing (nonparametric) structure on the individual or group specific effect, in contrast to Honoré (1992), who left this term completely unspecified.

However, the structure we impose on the individual or group specific effect is weak. As in Newey (1994) and Chen (1998) we leave its correlation structure with the observed explanatory variables completely unspecified. However, in contrast to Newey (1994), we impose neither normality nor conditional homoskedasticity on the error term, and in contrast to Chen (1998) we allow for time or group varying factor loads. Making these generalizations is important, as the nonlinear nature of our model implies that consistency of regression coefficient estimators could be lost if these assumptions (i.e, normality, homoskedasticity, stationarity, common factor loads) are not true.

The paper is organized as follows. The next section introduces the model we propose to estimate and describes the estimation procedure. Section 3 outlines the necessary regularity conditions and investigates the asymptotic properties of the estimator. Section 4 assesses the finite-sample performance of the proposed estimators through a small-scale simulation study. Section 5 concludes by summarizing results and discussing extensions for future research. An Appendix fills in the details of the proofs of the asymptotic results.

2. MODEL AND ESTIMATION PROCEDURE

Longitudinal panel data are typically characterized by a large number of agents observed over a small number of time periods. To simplify the description of our estimation procedure, we will focus on the case with two periods.

The model we wish to estimate can be characterized by the equation

$$y_{it} = \max(\theta_t \alpha_i + x_{it}'\beta_0 + \epsilon_{it}, 0) \tag{2.1}$$

for $i = 1, \ldots, n$ and $t = 1, 2$. Here β_0 is a d_x-dimensional vector parameter of interest, x_{it} represents the (possibly) time varying[4] d_x-dimensional vector of observed explanatory variables, ϵ_{it} is an error term whose distribution will be allowed to vary across time, α_i denotes the unobservable time invariant individual specific effect, and θ_t denotes the time varying factor load.

The principal aim of our approach is the identification and estimation of β_0 and θ_t, as estimators based on stationarity assumptions, such as Honoré (1992),

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

necessarily require that $\theta_t \equiv 1$. We wish to relax this because models that allow for time varying factor loads can prove useful in a variety of empirical settings. As alluded to in the previous section, if estimating an earnings equation, this would allow for nonstationary returns on both unobserved skills and observed skills such as education. Examples in the literature that allow for time specific coefficients on the individual specific effect in addition to the observed regressors include the work of Heckman (1981), who studied a discrete choice model, Holtz-Eakin, Newey, and Rosen (1988), who examined a linear autoregressive model, and Chay (1995) and Honoré (1998), who investigated a censored regression model.

In many panel data models the individual specific effect, α_i, is left unspecified. It can be simply "differenced out" in the linear panel data model. For nonlinear models, such as the binary choice, censored regression, and sample selection panel data models, estimators that have left α_i unspecified have required the assumption of stationary error terms (see, e.g., Manski, 1987; Honoré, 1992; Kyriazidou, 1997). In contrast to these approaches, we allow for nonstationary errors and time varying factor loads by imposing a weak structural form on the individual specific effect:

$$\alpha_i = \phi(x_i) + v_i, \tag{2.2}$$

where $\phi(\cdot)$ is an unknown function of the regressors $x_i = (x_{i1}', x_{i2}')'$ and v_i is an unobservable error term. This generalizes many previous specifications of the individual effect found in the literature.[5] For example, assuming that $\phi(\cdot) \equiv 0$ and v_i is independent of x_i yields the classical random effects model. Chamberlain (1984), Nijman and Verbeek (1992), and Zabel (1992) allow for possible correlation between x_i and α_i by imposing a linear specification on $\phi(\cdot)$ and assuming normality of the remaining error terms. Unfortunately, estimators based on these restrictions generally will be inconsistent if the linear structure on $\phi(\cdot)$ and/or the normality assumption is incorrect. Newey (1994) and Chen (1998) thus generalized this approach by leaving $\phi(\cdot)$ unspecified. Here, we only assume that the "composite" error term, $\mu_{it} \equiv v_i + \epsilon_{it}$, satisfies a weak conditional median restriction:[6]

$$P(\mu_{it} \leq 0 | x_i) = \frac{1}{2} \quad \text{a.s.} \qquad t = 1, 2. \tag{2.3}$$

The stochastic restriction that the sum of two error terms satisfy a conditional median restriction is not new to the literature. For example, Weiss (1993) exploited a similar restriction to estimate a censored regression model with measurement error. One particular way to interpret this restriction is to assume that for $t = 1, 2$, conditional on x_i, the vector (v_i, ϵ_{it}) is symmetrically distributed. This type of bivariate error restriction was used in Honoré, Kyriazidou, and Udry (1997) to estimate a semiparametric version of the Type III Tobit model and by Chen (1998) to estimate a sample selection model.

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

In the context of the model at hand, the preceding restriction is general enough to allow for error terms that are both nonstationary and conditionally heteroskedastic. We emphasize that the individual specific effect structure is still more restrictive[7] than in Honoré (1998), where θ_t was identified and α_i was left unspecified. However, unlike in Honoré (1998), no parametric assumptions on ϵ_{it} are required for identification in this paper.

The conditional median restriction can be exploited to estimate the parameters of interest in the panel data censored regression model in a way that is similar to the method introduced in Powell (1984, 1986), Khan and Powell (2001), and Honoré, Khan, and Powell (2002) for the cross-sectional censored regression model. Specifically, the equivariance property of medians implies that the conditional median of the observed response variable of individual i in time period t, denoted here by q_{it}, is of the form

$$q_{it} = \max(x_{it}'\beta_0 + \theta_t \phi(x_i), 0) \qquad t = 1, 2. \tag{2.4}$$

This result will be used to identify and estimate β_0 and θ_t. To illustrate how, we first note that θ_t is only identified up to scale, and we adopt the normalization $\theta_1 \equiv 1$ and drop the time subscript on θ_2. To illustrate our approach, suppose for now that the values of the conditional median function across all individuals and time periods are known to the econometrician. For observations where median function values exceed the censoring point we have

$$q_{i1} = x_{i1}'\beta_0 + \phi(x_i),$$

$$q_{i2} = x_{i2}'\beta_0 + \theta\phi(x_i),$$

where $q_t(\cdot)$ denotes the conditional median function in time period t.

Thus we have the following relationship between the two median function values:[8]

$$q_{i2} - q_{i1} = (q_{i1} + q_{i2})\frac{\theta - 1}{\theta + 1} - (x_{i1} + x_{i2})\frac{\theta - 1}{\theta + 1}\beta_0 + (x_{i2} - x_{i1})'\beta_0. \tag{2.5}$$

It can immediately be seen that "reduced form" coefficients, $\gamma_0 = ((\theta - 1)/(\theta + 1), (\theta - 1)/(\theta + 1)\beta_0', \beta_0')'$, can be recovered by a linear regression of $q_{2i} - q_{1i}$ on the "regressors" $z_i = (q_{1i} + q_{2i}, x_{i1}' + x_{i2}', x_{i2}' - x_{i1}')'$. Furthermore, the parameters of interest, (β_0, θ), can be easily recovered from γ_0.

Of course an estimator based on these simple relations is not immediately feasible, as the median functions are unknown to the econometrician. However, this approach does suggest a feasible two-stage estimation procedure. Though the values of the conditional median function are unknown, they can be estimated using nonparametric methods in a preliminary stage. The second stage would be as simple as least squares, regressing the first-stage estimated value of q_{2i} on the first-stage estimated value of z_i.

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

Remark 2.1. We have outlined our estimation procedure in the context of a longitudinal panel data model with individual specific effects. This procedure also has much empirical relevance in a group/family effect model. In this context our procedure would be able to estimate factor loads on group/family specific effects that vary across group/family members. This generalization would be more consistent with some of the empirical findings on the intergenerational correlation in earnings (see, e.g., Solon, 1992, 1999; Zimmerman, 1992; Ashenfelter and Zimmerman, 1997).

The remainder of this paper is devoted to analyzing the asymptotic and finite-sample properties of this estimator. Before proceeding, however, we will discuss each of the estimation procedures in greater detail. Specifically, we discuss the nonparametric estimation procedure adopted in the first stage and some technical complications that require modification of the second-stage least squares procedure.

2.1. First Stage of the Estimator

The first stage involves nonparametrically estimating the conditional median of the observed dependent variables y_{it}, $t = 1, 2$, given the regressors x_i. Several such estimators have been recently proposed in the statistics and econometrics literature (see Stute, 1986; Troung, 1989; Bhattacharya and Gangopadhyay, 1990; Chaudhuri, 1991a, 1991b; Koenker, Portnoy, and Ng, 1992; Koenker, Ng, and Portnoy, 1994). We use the local polynomial estimator introduced in Chaudhuri (1991a, 1991b) for two main reasons. This estimation procedure is computationally simple (it involves minimization of a globally convex objective function that can be handled using the simplex method), and it allows for simple control of the order of the bias by selecting the appropriate order of the polynomial.

Its description is facilitated by introducing new notation, and the notation adopted has been chosen deliberately to be as close as possible to that introduced[9] in Chaudhuri (1991a, 1991b), with the main difference stemming from the fact that here we are focusing on the median, whereas Chaudhuri's approach applies to any quantile between 0 and 1.

For a vector a let $[a]$ denote the sum of its components. For any two vectors of the same dimension a, b, let a^b denote $\Pi a_{(i)}^{b_{(i)}}$ where the subscripts in parentheses denote components. For any vector a and any nonnegative integer k, let a^k denote the vector of elements $\{a^{b_i}\}$ where b_i is a vector of nonnegative integers such that $[b_i] = k$.

Now assuming that the conditional median functions $q_t(\cdot)$, $t = 1, 2$, have order of differentiability k, let $s(A)$ denote the dimension of the set of all vectors of nonnegative integers $\{b_i\}$ where $[b_i] \leq k$. Let $I(\cdot)$ be an indicator function and let δ_n denote a "bandwidth" sequence used to smooth the data. The local polynomial estimator of the conditional median function at a point x involves least

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

absolute deviations (LAD) regression (on observations that are "close" to x, Koenker and Bassett, 1978). Specifically, let the vectors

$$(\hat{\Xi}_0, \hat{\Xi}_1, \ldots, \hat{\Xi}_k)$$

minimize the kernel weighted objective function:[10]

$$\sum_{i=1}^{n} I[x_i \in C_n(x)] | y_{it} - \Xi_0 - \Xi_1'(x_i - x) - \ldots - \Xi_k'(x_i - x)^k | \tag{2.6}$$

where C_n is a sequence of cubes centered at x whose sides are of length $2\delta_n$, and so

$$x_i \in C_n(x) \Leftrightarrow |x_i^{(j)} - x^{(j)}| \leq \delta_n \qquad j = 1, 2, \ldots, 2d_x,$$

where the superscripts in parentheses denote the component of the vector. The conditional median estimator that will be used in the first stage will be the value $\hat{\Xi}_0$, and we denote estimated median function values by $\hat{q}_{i1}, \hat{q}_{i2}$. A computational advantage of using this estimator is that its evaluation can be carried out by linear programming techniques, and because the objective function is globally convex, the solution found is guaranteed to be a global minimizer.

The motivation for including a higher order polynomial in the absolute value function and estimating the nuisance parameters $(\hat{\Xi}_1, \ldots, \hat{\Xi}_k)$ is to achieve bias reduction of the nonparametric estimator, analogous to the bias reduction achieved in local polynomial regression estimators (see Fan and Gijbels, 1996).

2.2. Second Stage of the Estimator

The second stage involves a simple least squares procedure using first-stage estimated values as dependent and independent variables. As mentioned previously, this will involve searching for agents whose estimated values of the median functions exceed the censoring points in both time periods. Let $\hat{q}_i^+, \hat{q}_i^-$ denote $\hat{q}_{i1} + \hat{q}_{i2}$ and $\hat{q}_{i2} - \hat{q}_{i1}$, respectively. Our proposed estimator, denoted by $\hat{\gamma}$, of γ_0 can be obtained by minimizing

$$\frac{1}{n} \sum_{i=1}^{n} \tilde{t}_i \hat{\omega}_i (\hat{q}_i^- - \hat{z}_i' \gamma)^2,$$

where $\hat{z}_i$ denotes the (first-stage) estimated "regressor" vector $(\hat{q}_i^+, x_{i1}' + x_{i2}', x_{i2}' - x_{i1}')'$, $\hat{\omega}_i = \omega(\hat{q}_{1i}, \hat{q}_{2i})$ is a weighting function that only keeps observations for which the first-stage estimated values exceed the censoring value, and $\tilde{t}_i \equiv \tilde{t}(x_i)$ is a trimming function used primarily for technical purposes. Note that the estimator has the closed form expression

$$\hat{\gamma} = \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{t}_i \hat{\omega}_i \hat{z}_i \hat{z}_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \tilde{t}_i \hat{\omega}_i \hat{z}_i \hat{q}_i^- .$$

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

As discussed in the next section, under appropriate regularity conditions, this estimator will converge at the parametric ($\sqrt{n}$) rate, with a limiting normal distribution, and we derive the expression for its limiting covariance matrix for purposes of inference.

3. ASYMPTOTIC PROPERTIES OF THE ESTIMATOR

Regularity conditions will first be outlined in detail before we proceed to the limiting distribution theory. Specific assumptions are imposed on the parameter space, the error terms and the regressors, the order of smoothness of the individual effect function $\phi(\cdot)$, and the kernel regularity conditions needed for the first stage, as follows.

3.1. Assumptions

3.1.1. Assumptions on the Parameter Space.

A1. The true parameter value β_0 lies in the parameter space $\mathcal{B}$, a subset of $\mathbf{R}^{d_x}$, where d_x denotes the dimension of the regressor vector in each period.

3.1.2. Assumptions on Error Terms and Regressors.

B1. The sequence of $2d_x + 3$-dimensional vectors $(\epsilon_{i1}, \epsilon_{i2}, v_i, x_i)$ is independent and identically distributed (i.i.d.).

B2. The regressor vector x_i has support that is a convex subset of $\mathbf{R}^{2d_x}$ with nonempty interior and has a continuous density with respect to Lebesgue measure, denoted by $f_X(\cdot)$ that is uniformly bounded away from infinity and is bounded away from 0 except on the boundary of its support.

B3. Let V_x denote the support of the trimming function, with V_x assumed to have the following properties:

B3.1. V_x is a compact subset of the support of x_i with nonempty interior.
B3.2. $f_X(x) \geq \epsilon_0 > 0 \ \forall x \in V_x$, for some constant ϵ_0.
B3.3. (Identification) The $(2d_x + 1) \times (2d_x + 1)$ matrix

$$\mathrm{E}[\omega_i z_i z_i' I[x_i \in V_x]]$$

is of full rank.

B4. $\mu_{it}|x_i$ has conditional median = 0 (uniquely) a.s. for $t = 1, 2$.

B5. The conditional distribution of the latent error terms μ_{it}, $t = 1, 2$, given the regressors has a density with respect to Lebesgue measure, $f_{\mu_{it}|x_i}(e|x)$, that is positive and continuously differentiable for e in a neighborhood of 0 for all x in the support of x_i. As a function of x, $f_{\mu_{it}|x_i}(e|x)$ is uni-

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

formly Hölder continuous with exponent $\gamma \in (0,1]$ with respect to each component of the vector x for all x in the support of x_i and all e in a neighborhood of 0.

3.1.3. Assumptions on the Latent Regression Function.

C1. The latent regression functions $q_t^*(x_i) = x_{it}'\beta_0 + \phi(x_i)$, $t = 1,2$, are assumed to be differentiable up to order k on V_x with derivatives of order k being uniformly continuous with exponent $\gamma \in (0,1]$. Letting $p = k + \gamma$ denote the order of smoothness of these functions, it will be assumed that $p > 3d_x$.

3.1.4. Assumptions on the Weighting and Trimming Functions.

D1. The weighting function, $\omega(\cdot): \mathbf{R} \to \mathbf{R}^+$, has the following properties:

D1.1. $\omega \equiv 0$ if its argument is less than c, a small positive constant. Bounding the support of ω away from 0 enables us to avoid technical complications, as was done in Buchinsky and Hahn (1998).

D1.2. $\omega(\cdot)$ is twice continuously differentiable.

D1.3. $\omega(\cdot)$ is increasing on the interval (c,∞).

D2. The trimming function $\tilde{t}(\cdot): \mathbf{R}^{2d_x} \to \mathbf{R}^+$ is positive and bounded on V_x.

3.1.5. First Step Kernel Regularity Conditions.

E1. The kernel function in the first step estimator is the product of $2d_x$ uniform kernels. Specifically, let $C_n(x_i)$ denote the cube in $\mathbf{R}^{2d_x}$, centered at x_i, with side length $2\delta_n$. Then the other observations in the sample that are used to estimate the median function at x_i are indexed by the set

$$S_n(x_i) = \{j : 1 \le j \le n, j \ne i, x_j \in C_n(x_i)\}.$$

E2. The first-stage bandwidth sequence, denoted δ_n, is of the form

$$\delta_n = c_1 n^{-\xi},$$

where c_1 is some positive constant and $\xi \in (\frac{1}{2}[p], \frac{1}{6}d_x)$ where $[\,]$ denotes the integer argument.

Under these conditions, we are able to derive the limiting distribution of $\hat{\gamma}$, as characterized in the following theorem. The proof is left to the Appendix.

THEOREM 3.1. *Let J^* and Ω^* denote*

$$\mathrm{E}[\tilde{t}_i \omega_i z_i z_i']$$

and

$$\mathrm{E}\left[\frac{4}{(\theta+1)^2}(\delta_2^*(y_{i2},x_i) - \delta_1^*(y_{i1},x_i))(\delta_2^*(y_{i2},x_i) - \delta_1^*(y_{i1},x_i))'\right],$$

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

respectively, where

$$\delta_2^*(y_{i2}, x_i) = \tilde{t}_i \omega_i f^{-1}_{\mu_{i2}|x_i}(0|x_i)\left(\frac{1}{2} - I[y_{i2} \leq q_{i2}]\right) z_i$$

and

$$\delta_1^*(y_{i1}, x_i) = \tilde{t}_i \omega_i \theta f^{-1}_{\mu_{i1}|x_i}(0|x_i)\left(\frac{1}{2} - I[y_{i1} \leq q_{i1}]\right) z_i;$$

then

$$\sqrt{n}(\hat{\gamma} - \gamma_0) \Rightarrow N(0, (J^*)^{-1}\Omega^*(J^*)^{-1}).$$

Note that the limiting distribution of the parameters of interest $(\beta_0', \theta)'$ follows directly from the delta method.

Inference on the parameters of interest can be performed by consistently estimating the matrix $(J^*)^{-1}\Omega^*(J^*)^{-1}$, which will require additional smoothing parameters to estimate the density of the median residuals at 0. An estimator for the density function in a variance matrix that arises from using nonparametric quantile regression was proposed in Khan (2001), where conditions on the additional smoothing parameters are provided.

4. MONTE CARLO RESULTS

In this section, the finite-sample properties of the proposed estimation procedure are examined through the results of a small-scale simulation study.[11] We compare the performance of our estimator to Honoré's (1998) fixed effects factor loading (FEFL) estimator.

For the study, we simulated censored regression models of the form

$$y_{it} = \max(x_{it}\beta_0 + \theta_t \alpha_i + \epsilon_{it}, 0) \qquad t = 1, 2.$$

The regressor values x_{i1}, x_{i2} were drawn in the following way: x_{i1} was drawn from a standard normal, and x_{i2} was set to the average of the value of x_{i1} and the value of an independently drawn standard normal random variable; β_0 was set to 1. The individual specific effect α_i was modeled as

$$\alpha_i = a + x_{i1} + x_{i2} + x_{i1}x_{i2} + \eta_i.$$

The constant a was set to 0.5, and the random term η_i was drawn from a standard normal distribution. The time varying factor loads were $\theta_1 = 1$ and $\theta_2 = 2$. The distribution of the error terms ϵ_{it} was altered for the various designs considered.

Table 1 reports the simulation results for the first design where the error terms $\epsilon_{i1}, \epsilon_{i2}$ were each drawn independently from a standard normal distribution. The table reports results for both Honoré's FEFL estimator and the two-stage median (2SM) estimator introduced here. The statistics used to compare performance

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

TABLE 1. Homoskedastic normal errors

	FEFL		2SM	
	β_0	θ_2/θ_1	β_0	θ_2/θ_1
$n = 100$				
Mean bias	0.0249	−0.0775	−0.2829	−0.1461
Median bias	−0.0696	−0.0297	−0.2916	−0.2423
RMSE	0.7150	0.4744	0.3759	0.6031
MAD	0.3615	0.2258	0.3024	0.3475
$n = 200$				
Mean bias	−0.0028	−0.0257	−0.1937	−0.1561
Median bias	−0.0352	−0.0212	−0.1975	−0.2181
RMSE	0.5330	0.3461	0.2545	0.3696
MAD	0.2634	0.1643	0.2055	0.2548
$n = 400$				
Mean bias	−0.0691	−0.0353	−0.1492	−0.1408
Median bias	−0.0651	−0.0111	−0.1549	−0.1816
RMSE	0.3540	0.2849	0.1897	0.2946
MAD	0.2110	0.1187	0.1576	0.2188

were mean bias, median bias, root mean squared error (RMSE), and median absolute deviation (MAD). We considered sample sizes of 100, 200, and 400, replicating the design 401 times.

The FEFL is based on a conditional moment condition, and so for its computation we used generalized method of moments (GMM) with 20 (unconditional) moment conditions. As the objective function is highly nonlinear and nonconvex in the parameters, it was minimized using the Nelder–Meade simplex algorithm, with 40 different starting values.

For the 2SM, the main implementation difficulty is the selection of the bandwidth sequence, and we adopted the plug-in method described in Chen and Khan (2000), which itself is based on the procedure introduced in Fan and Gijbels (1996).

Table 1 reports results from a design where the error terms were assumed to be i.i.d. standard normal. As the table indicates, the FEFL estimator performs quite well for this design. This is to be expected, as all the assumptions it is based on are satisfied.

The 2SM estimator also performs reasonably satisfactorily relative to the FEFL estimator for this design, though it exhibits a larger mean and median bias for smaller sample sizes than the FEFL. Nonetheless, its performance in terms of the four statistics compares well for sample sizes of 200 or greater. At 200 observations the two estimators had similar values for RMSE, but the FEFL

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

had a smaller bias, indicating that the 2SM is more precise but centered around the incorrect parameter value. Somewhat surprisingly, for larger sample sizes the 2SM outperformed the FEFL in terms of RMSE, even though the FEFL is based on the normality assumption.

Table 2 reports results from a design without normal errors. Specifically, here we assume that $\epsilon_{i1}, \epsilon_{i2}$ are i.i.d. normal mixtures, where the two normals in the mixture are each of unit variance and centered around -2 and 2. Under this design the FEFL is inconsistent, and this is reflected in Table 2. Here the values of the bias and RMSE are large for all sample sizes and even increase as the sample size increases from 100 to 200. In contrast, the 2SM performs satisfactorily in the simulation. As indicated in Table 2, it appears to exhibit a large bias in small sample sizes, but the RMSE shrinks with the sample size at the parametric (root-n) rate.

Table 3 reports results from a design with normal but conditionally heteroskedastic errors. Specifically, each error term was multiplied by a function of the regressor values that changed across time periods. The first-period error term was multiplied by $|x_{i1}x_{i2}|$, and the second-period error term was multiplied by $2|x_{2i}|$. The FEFL is inconsistent for this design also. Table 3 shows that it performs even worse for this design than the design in Table 2. The 2SM performs better for this design in terms of bias and RMSE. As with the other designs root-n consistency is reflected in the rate at which the RMSE decreases with the sample size.

TABLE 2. Mixed normal errors

	FEFL		2SM	
	β_0	θ_2/θ_1	β_0	θ_2/θ_1
$n = 100$				
Mean bias	−0.1143	−0.2731	−0.3285	−0.2954
Median bias	−0.4031	−0.2145	−0.3523	−0.3681
RMSE	1.3721	0.8842	0.4444	0.5666
MAD	0.9349	0.5840	0.3600	0.4234
$n = 200$				
Mean bias	−0.1983	−0.1573	−0.2403	−0.2443
Median bias	−0.2905	−0.1370	−0.2459	−0.2758
RMSE	0.9158	0.6923	0.3255	0.4390
MAD	0.6403	0.4130	0.2536	0.3219
$n = 400$				
Mean bias	−0.3129	−0.3118	−0.1776	−0.2148
Median bias	−0.2882	−0.1722	−0.1750	−0.2398
RMSE	0.9253	0.9171	0.2433	0.3618
MAD	0.6333	0.4989	0.1874	0.2692

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

TABLE 3. Heteroskedastic normal errors

	FEFL		2SM	
	β_0	θ_2/θ_1	β_0	θ_2/θ_1
$n = 100$				
Mean bias	0.4749	−0.4443	−0.2213	0.3503
Median bias	0.0595	−0.3461	−0.2199	0.2259
RMSE	1.9586	1.3018	0.3066	0.7836
MAD	1.2827	0.9113	0.2205	0.3454
$n = 200$				
Mean bias	0.0809	−0.4182	−0.1495	0.2546
Median bias	−0.0480	−0.2674	−0.1586	0.1363
RMSE	1.7220	1.2381	0.2163	0.6134
MAD	1.0659	0.8430	0.1692	0.2534
$n = 400$				
Mean bias	−0.4851	−0.8723	−0.1223	0.1633
Median bias	−0.3440	−0.4054	−0.1318	0.0842
RMSE	1.4511	1.6203	0.1712	0.4508
MAD	1.0738	1.1261	0.1407	0.2223

Table 4 reports results from a design where the assumptions of both the FEFL and 2SM are violated. Specifically, we assume that the error terms have a chi-squared distribution, violating the FEFL normality assumption and the 2SM symmetry assumption. In this design the FEFL again performs poorly, with bias increasing with the sample size, though it is quite small for the slope coefficient. The 2SM does exhibit a larger bias for the slope coefficient but exhibits a smaller bias for the factor load and much smaller RMSEs that decrease with the sample size. This design indicates that the 2SM may not perform too poorly even when its assumptions are violated.

Overall, the results from our simulation study are consistent with theoretical results. The FEFL can be inconsistent if the error terms are not normal and/or heteroskedastic, and an attractive feature of the 2SM introduced in this paper is its robustness to these forms of misspecification. The 2SM performed quite adequately in a design when the symmetry condition it required was not satisfied. Nonetheless, the finite-sample performance of both estimators leaves room for improvement.

5. SUMMARY AND CONCLUDING REMARKS

This paper introduces a two-stage procedure for estimating panel data censored regression models with nonstationary and conditionally heteroskedastic error

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

Table 4. Chi-squared errors

	FEFL		2SM	
	β_0	θ_2/θ_1	β_0	θ_2/θ_1
$n = 100$				
Mean bias	0.0174	−0.2144	−0.3090	0.0471
Median bias	−0.2536	−0.0515	0.3245	−0.0861
RMSE	1.5202	0.7123	0.3887	0.7822
MAD	1.1230	0.4316	0.3267	0.3413
$n = 200$				
Mean bias	0.0250	−0.2105	−0.2614	0.0278
Median bias	−0.2067	−0.0442	−0.2685	−0.0319
RMSE	1.3925	0.6833	0.3067	0.5813
MAD	1.0193	0.3905	0.2684	0.3093
$n = 400$				
Mean bias	−0.1312	−0.1156	−0.2192	−0.0238
Median bias	−0.2639	−0.0139	−0.2261	−0.0729
RMSE	1.1033	0.5617	0.2461	0.3517
MAD	0.7889	0.2759	0.2259	0.2250

terms. The procedure is also able to estimate time varying factors loading onto the individual specific effect. If it is applied to a grouped data set, group member varying factors can be estimated. The estimator is shown to have favorable asymptotic properties and perform relatively well in comparison with FEFL in finite samples for the designs we considered.

A drawback of the estimators we introduce in this paper is the nonparametric estimation required in the first stage. The well-known "curse of dimensionality" encountered in nonparametric estimation could possibly have a profound effect on the second-order properties of our second-stage estimator. This is especially problematic in the context of a longitudinal panel data model, as the number of variables in the structure of correlation between the fixed effect and the regressors increases with the number of time periods in the data set. It is less of a problem in the grouped data setting, especially in the context of intergenerational earnings where typically only a cross-section of fathers, sons, and brothers is available.

There are two suggestions we have for estimating these models if there are several time periods. One way is to assume that the regressors x_i are only correlated with the individual specific effects through their average value across time periods. This assumption has been commonly used in practice (Heckman and MaCurdy, 1980; Hsiao, 1986; Verbeek and Nijman, 1992) and prevents the dimension from increasing with the number of time periods in the context of our model.

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

A different approach to this problem would be to assume additive separability of the structure of correlation between the individual specific effect and the regressors. This would imply additive separability of the conditional median functions in their ranges exceeding the censoring value. It would seem likely that this additional structure could be exploited to improve efficiency and convergence rates of the first stage of our estimation procedure (see Stone, 1985; Andrews and Whang, 1990; Andrews, 1994), as the dimensionality of the problem is invariant to the number of time periods. This in turn could have a significant positive impact on the finite-sample properties of the second-stage estimators. A further useful approach to consider for improved finite-sample performance would be to combine the proposed quantile regression method with the Chamberlain (1982) specification of the individual effects.

NOTES

1. Throughout the paper we will not distinguish between these two causes of censoring. Because our parameters of interest are slope coefficients our proposed procedure is more useful for true data censoring.

2. Actually, Honoré (1992) imposed the even stronger assumption of exchangeability, which was relaxed to stationarity in Honoré and Kyriazidou (2000).

3. The Chay (1995) and Chay and Lee (2000) test was based on estimating a model with nonstationary errors, violating the error term behavior assumed in Honoré (1992).

4. One important example of such a regressor would be a time specific dummy variable, allowing for the estimation of time specific intercept terms. When dealing with wage data, this permits changes resulting from inflation or productivity, that is separate from the return to ability changing over time. We are grateful to J. Wooldridge for pointing this out to us.

5. Honoré and Lewbel (2002) impose a structural form on the individual effect that is neither stronger nor weaker than that imposed here. They require an exclusion condition that prevents α_i from depending on all the covariates, and a strong support condition on one of the covariates, but impose less structure in the sense of having the disturbance terms v_i satisfy the conditions imposed here.

6. Setting this median equal to 0 is simply a normalization to identify the difference in values of the time specific effects.

7. One particular disadvantage of the structure we impose is that an inconsistency can arise in the sense that $\phi(\cdot)$ will change once the time periods change and so it will be misspecified. Therefore, as mentioned in Arellano and Honoré (2001), approaches like ours implicitly place strong assumptions on the regressor distributions and parameters for internal consistency to hold. However, Arellano and Honoré (2001), acknowledge that one might still be willing to proceed along these lines because it can be a useful approximation of the relationship between the individual specific effect and the regressors, which, unlike the fixed effect approach, still allows one to estimate all the parameters needed for calculating the effect of the covariates on a location parameter (e.g., mean or, as in our case, median) of the latent dependent variable.

8. This is one of various relationships that exist. We illustrate this one because it is analogous to the classical relationship for the linear model, where we have a first difference of dependent variables on the left-hand side and a first difference of covariates on the right-hand side.

9. In particular the notation we adopt will be based on the assumption of the covariates being continuously distributed as was assumed in Chaudhuri (1991a, 1991b). We emphasize, however, that this assumption is not required for the asymptotic theory to hold and is only maintained in this paper for notational simplicity. In fact, root-n consistency will also be obtained if all the covariates are discretely distributed, in which case we perform LAD on observations whose regressor values

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

match up exactly with the point of interest. The mixed discrete-continuous case can also be accommodated. Specifically, one could use a product of kernel functions—using the kernel function for the continuous regressors multiplied by the product of indicator functions for each of the discrete components.

10. For technical reasons used in proving asymptotic properties of the second-stage estimator, we actually require that this objective function be minimized over a compact subset of $\mathbf{R}^{s(A)}$.

11. The simulation study was performed with GAUSS.

REFERENCES

Ahn, H. & J.L. Powell (1993) Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics* 58, 3–29.

Andrews, D.W.K. (1994) Empirical process methods in econometrics. In R.F. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2247–2294. North-Holland.

Andrews, D.W.K. & Y.J. Whang (1990) Additive interactive regression models: Circumvention of the curse of dimensionality. *Econometric Theory* 6, 466–479.

Arellano, M. & B. Honoré (2001) Panel data models: Some recent developments. In J.J. Heckman & E. Leamer (eds.), *Handbook of Econometrics*, vol. 5, pp. 3229–3296. North-Holland.

Ashenfelter, O. & D.J. Zimmerman (1997) Estimates of returns to schooling from sibling data. *Review of Economics and Statistics* 79, 1–9.

Bhattacharya, P.K. & A.K. Gangopadhyay (1990) Kernel and nearest neighbor estimation of a conditional quantile. *Annals of Statistics* 18, 1400–1415.

Buchinsky, M. & J. Hahn (1998) An alternative estimator for the censored quantile regression model. *Econometrica* 66, 653–672.

Chamberlain, G. (1982) Multivariate regression models for panel data. *Journal of Econometrics* 18, 5–46.

Chamberlain, G. (1984) Panel data. In Z. Griliches & M. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, pp. 1247–1318. North-Holland.

Chaudhuri, P. (1991a) Nonparametric quantile regression. *Annals of Statistics* 19, 760–777.

Chaudhuri, P. (1991b) Global nonparametric estimation of conditional quantiles and their derivatives. *Journal of Multivariate Analysis* 39, 246–269.

Chaudhuri, P., K. Doksum, & A. Samarov (1997) On average derivative quantile regression. *Annals of Statistics* 25, 715–744.

Chay, K.Y. (1995) Evaluating the Impact of the 1964 Civil Rights Act on the Economic Status of Black Men Using Censored Longitudinal Earnings Data. Manuscript, Princeton University.

Chay, K.Y. & B.E. Honoré (1998) Estimation of semiparametric censored regression models. *Journal of Human Resources* 33, 4–38.

Chay, K.Y. & D.S. Lee (2000) Changes in relative wages in the 1980s: Returns to observed and unobserved skills and black-white wage differentials. *Journal of Econometrics* 99, 1–38.

Chen, S. (1998) Root-*n* Consistent Estimation of a Panel Data Sample Selection Model. Manuscript, Hong Kong University of Science and Technology.

Chen, S. & S. Khan (2000) Estimation of censored regression models in the presence of nonparametric multiplicative heteroskedasticity. *Journal of Econometrics* 98, 283–316.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

Heckman, J.J. (1981) Statistical models for discrete panel data. In D. McFadden & C. Manski (eds.), *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press.

Heckman, J.J. & T.E. MaCurdy (1980) A life cycle model of female labor supply. *Review of Economic Studies* 47, 47–74.

Holtz-Eakin, D., W.K. Newey, & H.S. Rosen (1988) Estimating vector autoregressions with panel data. *Econometrica* 56, 1371–1395.

Honoré, B.E. (1992) Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica* 60, 533–565.

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

Honoré, B.E. (1998) IV Estimation of Panel Data Tobit Models with Normal Errors. Working paper, Princeton University.
Honoré, B.E., S. Khan, & J.L. Powell (2002) Quantile regression under random censoring. *Journal of Econometrics* 109, 67–105.
Honoré, B.E. & E. Kyriazidou (2000) Estimation of Tobit-type models with individual specific effects. *Econometric Reviews* 19, 341–366.
Honoré, B.E., E. Kyriazidou, & C. Udry (1997) Estimation of type 3 Tobit models using symmetric trimming and pairwise comparisons. *Journal of Econometrics* 76, 107–128.
Honoré, B.E. & A. Lewbel (2002) Semiparametric binary choice panel data models without strictly exogeneous regressors. *Econometrica* 70, 2053–2063.
Hsiao, C. (1986) *Analysis of Panel Data*. Cambridge University Press.
Khan, S. (2001) Two stage rank estimation of quantile index models. *Journal of Econometrics* 100, 319–355.
Khan, S. & J.L. Powell (2001) Two-step quantile estimation of semiparametric censored regression models. *Journal of Econometrics* 103, 73–110.
Koenker, R. & G.S. Bassett Jr. (1978) Regression quantiles. *Econometrica* 46, 33–50.
Koenker, R., P. Ng, & S. Portnoy (1994) Quantile smoothing splines. *Biometrika* 81, 673–680.
Koenker, R., S. Portnoy, & P. Ng (1992) Nonparametric estimation of conditional quantile function. In Y. Dodge (ed.), *Proceedings of the Conference on L_1-Statistical Analysis and Related Methods*, pp. 217–229. Elsevier.
Kyriazidou, E. (1997) Estimation of a panel data sample selection model. *Econometrica* 65, 1335–1364.
Manski, C.F. (1987) Semiparametric analysis of random effects linear models from binary panel data. *Econometrica* 55, 357–362.
Newey, W.K. (1994) The asymptotic variance of semiparametric estimators. *Econometrica* 62, 1349–1382.
Newey, W.K. & D. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. 4. North-Holland.
Nijman, T. & M. Verbeek (1992) Nonresponse in panel data: The impact on estimates of a life cycle consumption function. *Journal of Applied Econometrics* 7, 243–257.
Powell, J.L. (1984) Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* 25, 303–325.
Powell, J.L. (1986) Censored regression quantiles. *Journal of Econometrics* 32, 143–155.
Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57, 1404–1430.
Solon, G. (1992) Intergenerational income mobility in the United States. *American Economic Review* 82(3), 393–408.
Solon, G. (1999) Intergenerational mobility in the labor market. In O. Ashenfelter & D. Card (eds.), *Handbook of Labor Economics*, vol. 3, pp. 1761–1800. North-Holland.
Stone, C.J. (1985) Additive regression and other nonparametric models. *Annals of Statistics* 13, 689–705.
Stute, W. (1986) Conditional empirical processes. *Annals of Statistics* 14, 638–647.
Tobin, J. (1958) Estimation of relationships for limited dependent variables. *Econometrica* 26, 24–36.
Troung, Y. (1989) Asymptotic properties of kernel estimates based on local medians. *Annals of Statistics* 17, 606–617.
Verbeek, M. & T. Nijman (1992) Testing for selectivity bias in panel data models. *International Economics Review* 33, 681–703.
Weiss, A.A. (1993) Some aspects of measurement error in a censored regression model. *Journal of Econometrics* 56, 169–188.
Zabel, J. (1992) Estimating fixed and random effects models with selectivity. *Economic Letters* 40, 269–272.
Zimmerman, D.J. (1992) Regression towards mediocrity in economic stature. *American Economic Review* 82, 409–429.

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

APPENDIX

As a first step in deriving the asymptotic properties of the estimator, we will establish two main results involving global properties of the local polynomial estimator used in the first stage. The first relates to global rates of convergence for the local polynomial estimator for uncensored data. Such rates, which depend on the bandwidth sequence δ_n, the number of regressors, $2d_x$, and the order of smoothness of the conditional quantile function, p, have been established in Chaudhuri (1991b) and Chaudhuri, Doksum, and Samarov (1997).

LEMMA A.1 (From Chaudhuri et al., 1997, Lem. 4.3a). *Under Assumptions B4, B5, C1, and E2 we have for* $t = 1, 2$

$$\max_{i:1\le i\le n,\, q_{it}\ge c/2,\, x_i\in V_x} |\hat{q}_{it} - q_{it}| = o_p(n^{-1/4}). \tag{A.1}$$

The second result establishes a weak form of consistency for the local polynomial estimator of the conditional median function for censored data.

LEMMA A.2. *For the small constant c used in the weighting function* $\omega(\cdot)$*, we have for* $t = 1, 2$

$$P\left(\max_{1\le i\le n} |\hat{q}_{it} - q_{it}| \ge c/2\right) \to 0 \quad \textit{at an exponential rate.}$$

Proof. Without loss of generality we restrict attention to the first-period median function. We first show that the preceding result is true for a "local constant" estimator, denoted by $q_{i1}^{\dagger}$ and defined as

$$q_{i1}^{\dagger} = \operatorname{argmin}_{\theta} \sum_{x_j\in C_{ni}} |y_{j1} - \theta|.$$

Define the random variables

$$U_n = I\left[\max_{1\le i\le n} |q_{i1}^{\dagger} - q_{i1}| \ge c/4\right],$$

$$N_{ni} = \sum_{j=1}^{n} I[x_j \in C_{ni}],$$

$$V_{n,i} = I[|q_{i1}^{\dagger} - q_{i1}| \ge c/4],$$

$$\mathrm{E}_{n,i} = I[c_1 n\delta_n^{2d_x} \le N_{ni} \le c_2 n\delta_n^{2d_x}, \qquad 1 \le i \le n] \qquad c_1, c_2 \in \mathbf{R}.$$

By Theorem 3.1 of Chaudhuri (1991a) we can choose constants c_1, c_2 s.t.

$$P(\liminf \mathrm{E}_{n,i}) = 1.$$

Thus to prove the desired result, because $P(U_n \cap \mathrm{E}_{n,i}) \le \sum_{i=1}^{n} P(V_{n,i} \cap \mathrm{E}_{n,i})$ it will suffice to show that for all $x_i \in V_x$, $P(V_{n,i} \cap \mathrm{E}_{n,i})$ has an exponential bound.

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

First note that by Assumption B5, there exists a constant $\lambda_1 > 0$ such that for all $x_i \in V_x$,

$$\min(P(0 < \mu_{i1} \leq c/8), P(-c/8 \leq \mu_{i1} < 0)) \geq \lambda_1.$$

Because $q_{i1}^{\dagger} \geq 0$ by construction, for $0 \leq q_{i1} < c/4$, the event $V_{n,i}$ implies that

$$q_{i1}^{\dagger} \geq q_{i1} + c/4.$$

Thus we have the event $V_{n,i}$ implying

$$\sum_{j: x_j \in C_{ni}} I[y_{j1} \geq q_{i1} + c/4] \geq N_{ni} \frac{1}{2}.$$

Furthermore, by Assumptions C1 and E2, there exists an integer N_0 such that for all $n \geq N_0, x_j \in C_{ni}$, we have

$$q_{j1} \in [0, q_{i1} + c/8],$$

which implies

$$P(y_{j1} \geq q_{i1} + c/4 | x_j \in C_{ni}, n \geq N_0) \leq P(\mu_{j1} \geq c/8)$$

$$\leq \frac{1}{2} - \lambda_1. \tag{A.2}$$

So it immediately follows by Hoeffding's inequality that for $0 \leq q_{i1} < c/4, n \geq N_0$,

$$P(V_{n,i} \cap \mathrm{E}_{n,i}) \leq 2 \exp(-2c_1^2/c_2 \lambda_1 n \delta_n^{2d_x}). \tag{A.3}$$

We next consider the case where $q_{i1} \geq c/4$. If $q_{i1}^{\dagger} = 0$, then it must be the case that

$$\sum_{x_j \in C_{ni}} I[y_{j1} = 0] \geq N_{ni} \frac{1}{2}; \tag{A.4}$$

also, by Assumptions C1 and E2, there exists an integer N_1 such that $q_{j1} \geq c/8$ for $n \geq N_1$, and so we have

$$P(y_{j1} = 0 | x_j \in C_{ni}, n \geq N_1) \leq P(\mu_{j1} \leq -c/8 | x_j)$$

$$\leq \frac{1}{2} - \lambda_1. \tag{A.5}$$

Therefore by Hoeffding's inequality the probability of the event (A.4) is bounded above by

$$2 \exp(-2c_1^2/c_2 \lambda_1 n \delta_n^{2d_x}).$$

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

Also, for $q_{i1} \geq c/4$, if $0 < q_{i1}^{\dagger} \leq q_{i1} - c/4$ it follows that

$$\sum_{j: x_j \in C_{ni}} I[y_{j1} \leq q_{i1} - c/4] \geq N_{ni} \frac{1}{2} \tag{A.6}$$

and for $n \geq N_2$, $q_{j1} \geq q_{i1} - c/8$, which implies that

$$P(y_{j1} \leq q_{i1} - c/2 | q_{j1} \geq q_{i1} - c/8) \leq \frac{1}{2} - \lambda_1,$$

and so once again Hoeffding's inequality can be used to obtain the same exponential bound for the probability of the event (A.6). A similar argument can be used for the case where $q_{i1} \geq c/4$, $q_{i1}^{\dagger} \geq q_{i1} + c/4$. Thus by Assumption E2 we have an exponential rate of convergence for $P(U_n)$.

To conclude the proof of the lemma we show that for n sufficiently large,

$$|q_{i1}^{\dagger} - \hat{q}_{i1}| \leq c/4$$

for all $x_i \in V_x$. Note that on the compact set V_x, Assumption C1 implies that the derivatives of $q_1(\cdot)$ up to order k are all bounded and note that the local polynomial estimator $\hat{q}_{i1}$ is equal to a local constant estimator where the dependent variable is

$$y_{j1}^{*} = y_{j1} - \hat{\theta}_1'(x_j - x_i) - \ldots - \hat{\theta}_k'(x_j - x_i)^k,$$

where $\hat{\theta}_i$ are the local polynomial estimators of the local derivatives of order 1 through k. Thus $|y_{j1}^{*} - y_{j1}| < c/4$ for n sufficiently large by the compactness assumption on the parameter space in the local polynomial estimator. ■

These two lemmas will be used jointly when expanding the weighting function of estimated medians around the weighting function evaluated at true medians. Both lemmas are necessary to account for the problems encountered when the median functions are near the censoring point, as they are not smooth at 0. To illustrate how they will be used, consider the generic case

$$\frac{1}{n} \sum_{i=1}^{n} \tilde{t}_i \hat{\omega}_i s(y_{i1}, y_{i2}, x_i) \tag{A.7}$$

for some function $s(\cdot,\cdot,\cdot)$. Expanding $\tilde{t}_i \hat{\omega}_i$ around $\tilde{t}_i \omega_i$ we get that the preceding expression is

$$\frac{1}{n} \sum_{i=1}^{n} \tilde{t}_i \omega_i s(y_{i1}, y_{i2}, x_i) + \frac{1}{n} \sum_{i=1}^{n} \tilde{t}_i (\omega_{1i}^{*}(\hat{q}_{1i} - q_{1i}) + \omega_{2i}^{*}(\hat{q}_{2i} - q_{2i})) s(y_{i1}, y_{i2}, x_i), \tag{A.8}$$

where ω_{1i}^{*} denotes the partial of the weighting function with respect to its first argument evaluated at a value that lies between the true and estimated quantile values and ω_{2i}^{*} denotes the analogous value with the partial taken with respect to the second argument. Note that we may decompose the second term in the preceding expansion into the cases where the true median values exceed $c/2$ and cases where they do not. In the former case we may apply Lemma A.1. In the latter, we note that if the derivative of the weight-

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

ing function is positive it must be the case that estimated median function value exceeds c, in which case the difference between true and estimated values exceeds $c/2$ in absolute value, to which we can apply Lemma A.2.

In what follows, we will be replacing $\tilde{t}_i\hat{\omega}_i$ with $\tilde{t}_i\omega_i$ and deal with the remainder terms in this way, using Lemmas A.1 and A.2 as just explained. For concise exposition, we will leave out the details when making this replacement.

We can now turn attention to deriving a linear representation for the estimator. We note that

$$\hat{\gamma} - \gamma_0 = \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\hat{\omega}_i\hat{z}_i\hat{z}_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\hat{\omega}_i\hat{z}_i(\hat{q}_i^- - \hat{z}_i'\gamma_0),$$

$$\equiv \hat{J}^{-1}\hat{N}$$

where

$$\hat{J} = \frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\hat{\omega}_i\hat{z}_i\hat{z}_i'$$

and

$$\hat{N} = \frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\hat{\omega}_i\hat{z}_i(\hat{z}_i\hat{q}_i^- - \hat{z}_i'\gamma_0).$$

We first show that $\hat{J} \xrightarrow{p} J^*$. It will suffice to show by a mean value expansion of the estimated weighting function around the truth (as discussed earlier), that

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\omega_i(z_iz_i' - \hat{z}_i\hat{z}_i') = o_p(1).$$

Let $\|\cdot\|$ denote the euclidean norm. By the same argument used in Lemma 8.3 in Newey and McFadden (1994) it will suffice to show that

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\omega_i\|z_i - \hat{z}_i\|^2 = O_p\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\omega_i|\hat{q}_{i1} - q_{i1} + \hat{q}_{2i} - q_{2i}|^2\right)$$

is $o_p(1)$. This follows directly from Lemma A.1.

The rest of the proof involves establishing a linear representation for

$$\hat{N} = \frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\hat{\omega}_i(\hat{q}_i^-\hat{z}_i - \hat{z}_i\hat{z}_i'\gamma_0),$$

which after a mean value expansion around the true weighting function can be decomposed as

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\omega_i(\hat{q}_i^- - q_i^-)z_i \tag{A.9}$$

$$-\frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\omega_i\frac{\theta-1}{\theta+1}(\hat{q}_i^+ - q_i^+)z_i + R_n, \tag{A.10}$$

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

where the remainder term will involve the summation of terms involving the partial derivatives of the weighting function, and the square of the difference between estimated and true quantile function values. By Assumption D1, which ensures that the derivative of the weighting function is bounded in a neighborhood of 0, and the discussion on replacing $\tilde{t}_i\hat{\omega}_i$ with $\tilde{t}_i\omega_i$, we can conclude that the remainder term R_n is $o_p(n^{-1/2})$. We complete the proof by establishing linear representations for (A.9) and (A.10). Note that we may combine these two summations and get an expression in terms of the period specific median functions:

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\omega_i\frac{2}{\theta+1}(\hat{q}_{i2}-q_{i2}) \tag{A.11}$$

$$-\frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\omega_i\frac{2\theta}{\theta+1}(\hat{q}_{i1}-q_{i1}). \tag{A.12}$$

Here we only derive a representation for (A.12). Specifically, we will show that

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\omega_i\frac{2\theta}{\theta+1}(\hat{q}_{i1}-q_{i1})z_i=\frac{1}{n}\sum_{i=1}^{n}\frac{2\theta}{\theta+1}\delta_1^*(y_{i1},x_i)+o_p(n^{-1/2}). \tag{A.13}$$

Throughout we will drop the constant $2\theta/(\theta+1)$ for ease of exposition. First note that because the difference $\hat{q}_{1i}-q_{1i}$ is multiplied by the weighting function ω, observations for $q_{i1}\le c$ are "trimmed" in the sum in (A.13). This means that for untrimmed observations, the "residual," u_{i1}, satisfies the necessary assumptions for the local Bahadur representation developed in Chaudhuri (1991a) and Chaudhuri et al. (1997). To plug in this linear representation, we introduce the notation used in Chaudhuri et al. (1997) and define the following terms: for estimation of q_{1i} using observations $j\neq i$, we let $b(\delta_n,x_j-x_i)$ denote the $s(A)$-dimensional vector:

$$b(\delta_n,x_j-x_i)=(\{\delta_n^{-[u]}(x_j-x_i)^u,1\le[u]\le k\}')'. \tag{A.14}$$

Denote the $s(A)\times s(A)$ matrix $G_n(x_i)$ as the density weighted conditional expectation of the outer product of $b(\delta_n,x_j-x_i)$ given that $x_j\in C_{ni}$. Let $e_{(1)}$ denote an $s(A)$-dimensional vector whose first component is 1 and remaining components are 0 and denote its transpose by $e'_{(1)}$. Finally, let $q_{n1}(x_j,x_i)$ denote kth-order Taylor approximation of q_{1j} evaluated at the point x_i, so that we have that $q_{1j}-q_{n1}(x_j,x_i)=O((x_j-x_i)^p)$. The specific form of the linear representation we will plug in is

$$\hat{q}_{1i}-q_{1i}=N_n(x_i)^{-1}e'_{(1)}G_n(x_i)^{-1}\sum_{j=1,j\neq i}^{n}b(\delta_n,x_j-x_i)$$
$$\times\left(\frac{1}{2}-I[y_{j1}\le q_{n1}(x_j,x_i)]\right)I[x_j\in C_{ni}]+R_n(x_i), \tag{A.15}$$

where $R_n(x_i)$ is $o_p(n^{-1/2})$ *uniformly* in $x_i\in V_x$. By plugging in this representation into the left-hand side of (A.13), we obtain

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\omega_i z_i e'_{(1)}G_n(x_i)^{-1}N_n(x_i)^{-1}\times \tag{A.16}$$

$$\sum_{j\neq i}b(\delta_n, x_j - x_i)\left(\frac{1}{2} - I[y_{j1}\leq q_{n1}(x_j, x_i)]\right)I[x_j\in C_n(x_i)] + \frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\omega_i z_i R_n(x_i) \tag{A.17}$$

We first show that the remainder term (A.17) is $o_p(n^{-1/2})$. Note that this term is bounded by

$$\max_{i:1\leq i\leq n, q_{i1}\geq c, x_i\in V_x}|R_n(x_i)|\frac{1}{n}\sum_{i=1}^{n}\|\tilde{t}_i\omega_i z_i\|.$$

It follows from Lemma 4.1 in Chaudhuri et al. (1997) and Assumption E2 that

$$\max_{i:1\leq i\leq n, q_{i1}\geq c, x_i\in V_x}|R_n(x_i)| = o_p(n^{-1/2})$$

and by the law of large numbers that

$$\frac{1}{n}\sum_{i=1}^{n}\|\tilde{t}_i\omega_i z_i\| = O_p(1),$$

which together establish that (A.17) is $o_p(n^{-1/2})$. We now show that (A.16) can be represented as a second-order U-statistic. The steps involved follow those used in Chaudhuri et al. (1997), and so we will omit some of the details. We again attempt to keep our notation as close as possible to theirs. To begin, we define the event D_{ni} as

$$D_{ni} = \{c_3 n\delta_n^{2d_x}\leq N_{ni}\leq c_4 n\delta_n^{2d_x}\}.$$

By multiplying each term in the double summation in (A.16) by $I[D_{ni}]$, the resulting remainder term is $o_p(n^{-1/2})$, because as established in Chaudhuri et al. (1997),

$$P\left(\max_{i:1\leq i\leq n, x_i\in V_x}I[D_{ni}] = 0\right)\to 0 \quad \text{at an exponential rate.}$$

The next step involves replacing the term $I[y_{j1}\leq q_{n1}(x_j, x_i)]$ with $I[y_{j1}\leq q_{j1}]$. Because $\|G_{n1}(x_i)\|$ is bounded away from 0 for all x_i s.t. $\omega > 0$, and $b_n(\delta_n, x_j - x_i)I[x_j\in C_n(x_i)]$ is bounded by 1, the remainder term of this replacement is bounded above by a constant times the U-statistic:

$$\frac{1}{n(n-1)}\sum_{i\neq j}\delta_n^{-2d_x}\tilde{t}_i\omega_i\|z_i\|I[x_j\in C_{ni}]I[|\mu_{j1}|\leq C\delta_n^p]. \tag{A.18}$$

Noting that $\mathrm{E}[I[|\mu_{i1}|\leq C\delta_n^p|x_i] = O(\delta_n^p)$ a.s., by the change of variables $u = \delta_n^{-1}(x_j - x_i)$ we get that the unconditional second moment of the norm of the kernel in the preceding U-statistic is $O(\delta_n^{p-2d_x})$, which is $o(n)$ by Assumption E2. Thus by Lemma 3.1 in Powell, Stock, and Stoker (1989), we can express this U-statistic as the sum of its projection plus a negligible term.

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

It follows directly that the projection term is $o_p(n^{-1/2})$ by the fact that

$$\mathrm{E}[I[|\mu_{i1}| \le C\delta_n^p | x_i] = O(\delta_n^p) = o(n^{-1/2}),$$

where the second equality follows from Assumption E2.

The next step is to replace the number of observations used for the first-stage nonparametric estimator, N_{ni}, by its conditional expectation, $n\rho_{ni}$ where $\rho_{ni} = P(x_j \in C_{ni} | x_i)$. We use the following two results from Chaudhuri et al. (1997):

$$\max_{1\le i\le n} \tilde{t}_i \omega_i \left(\frac{1}{N_{ni}} - \frac{1}{n\rho_{ni}} \right) I[D_{ni}] = O(n^{-3/2}\delta_n^{-4d_x}\sqrt{\log n}), \tag{A.19}$$

$$\max_{1\le i\le n} \sum_{j\ne i} b(\delta_n, x_j - x_i) \left(\frac{1}{2} - I[y_{j1} \le q_{j1}] \right) I[x_j \in C_{ni}] I[D_{ni}] = O\left(\sqrt{n\delta_n^{2d_x} \log n}\right). \tag{A.20}$$

Because the product of these two terms is $O(n^{-1}\delta_n^{-3d_x} \log n)$, which is $o(n^{-1/2})$ by Assumption E2, and because the term $\|G_{n1}^{-1}(x_i)\|$ is uniformly bounded, the desired result follows from the fact that the term $(1/n)\sum_{i=1}^n \tilde{t}_i \omega_i \|z_i\|$ is $O_p(1)$ by the law of large numbers.

Next, we represent this triangular array as a sum of i.i.d. random variables. Let $f_n(\cdot,\cdot)$ denote the kernel of the second-order U-statistic:

$$\frac{1}{n(n-1)} \sum_{i\ne j} \tilde{t}_i \omega_i z_i e'_{(1)} G_n^{-1}(x_i) \rho_{ni}^{-1} \times b(\delta_n, x_j - x_i) \left(I[y_{j1} \le q_{j1}] - \frac{1}{2} \right) I[x_j \in C_{ni}].$$

Note that

$$\mathrm{E}[\|f_n(\cdot,\cdot)\|^2] = O(\delta_n^{-3d_x}) = o(n),$$

where the first equality follows from Lemma 4.2.b in Chaudhuri et al. (1997) and the second equality follows from Assumption E2. Thus by Lemma 3.1 in Powell et al. (1989; see also Ahn and Powell, 1993) we can represent this U-statistic as

$$\hat{U}_n + o_p(n^{-1/2}),$$

where $\hat{U}_n$ denotes the U-statistic's projection. Note that by the conditional quantile restriction we have

$$\mathrm{E}[f_n(\cdot,\cdot)] = 0,$$

$$\mathrm{E}[f_n(\hbar_i,\cdot)] = 0,$$

where $\hbar_i$ denotes the vector $(y_{i1}, x_i')'$. It thus remains to be shown that

$$\frac{1}{n}\sum_{i=1}^n (\mathrm{E}[f_n(\cdot,\hbar_i)] - \delta_1^*(\hbar_i)) = o_p(n^{-1/2}).$$

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press

Note that both random variables in the summation have expected value 0, and so it will suffice to show that

$$\mathrm{E}[\|\mathrm{E}[\,f_n(\cdot,\hbar_i)] - \delta_1^*(\hbar_i)\|^2] \to 0.$$

The preceding expectation is bounded above by

$$\frac{1}{4}\,\mathrm{E}\left[\left\|\int_{V_x} \omega(q_1(x),q_2(x))ze'_{(1)}\,G_n^{-1}(x)\rho_n^{-1}(x)b(\delta_n,x_i-x)f_X(x)\,dx - \tilde{t}_i\omega_i z_i f^{-1}_{\mu_1|x_i}(0|x_i)\right\|^2\right],$$

which converges to 0 by the change of variables $u = (x_i - x)\delta_n^{-1}$ and two applications of the dominated convergence theorem, as shown in, e.g., Chen and Khan (2000). This establishes (A.13). Note that we can use the same arguments to show that (A.11) has the representation

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{t}_i\omega_i\,\frac{2}{\theta+1}\,\delta_2^*(y_{i2},x_i) + o_p(n^{-1/2}). \tag{A.21}$$

Collecting our results we have that the desired linear representation for $\hat{N}$, combined with the probability limit established for $\hat{J}$ yields the desired linear representation for the estimator $\hat{\gamma}$. The limiting distribution in the theorem then follows by applying the central limit theorem. ■

https://doi.org/10.1017/S0266466608080468 Published online by Cambridge University Press