Two-stage rank estimation of quantile index models

Shakeeb Khan*

Department of Economics, University of Rochester, Rochester, NY 14627, USA

Received 1 October 1998; accepted 25 June 2000

Abstract

This paper estimates a class of models which satisfy a monotonicity condition on the conditional quantile function of the response variable. This class includes as a special case the monotonic transformation model with the error term satisfying a conditional quantile restriction, thus allowing for very general forms of conditional heteroscedasticity. A two-stage approach is adopted to estimate the relevant parameters. In the first stage the conditional quantile function is estimated nonparametrically by the local polynomial estimator discussed in Chaudhuri (Journal of Multivariate Analysis 39 (1991a) 246–269; Annals of Statistics 19 (1991b) 760–777) and Cavanagh (1996, Preprint). In the second stage, the monotonicity of the quantile function is exploited to estimate the parameters of interest by maximizing a rank-based objective function. The proposed estimator is shown to have desirable asymptotic properties and can then also be used for dimensionality reduction or to estimate the unknown structural function in the context of a transformation model. © 2001 Elsevier Science S.A. All rights reserved.

JEL classification: C13; C14; C24; C41

Keywords: Monotonic transformation; Nonparametric quantile regression; Heteroscedasticity; Rank regression

* Corresponding author. Tel.: + 1-716-275-2180; fax: + 1-716-256-2309.
E-mail address: skhan@troi.cc.rochester.edu (S. Khan).

0304-4076/01/$ - see front matter © 2001 Elsevier Science S.A. All rights reserved.
Pii: S0304-4076(00)00040-3
1. Introduction and motivation

The monotonic transformation model is usually expressed as:

\[ y_i = g(x_i'\beta_0, \epsilon_i) \]  

(1.1)

where \( y_i \) is a scalar dependent variable, \( x_i \) is a \( d \)-dimensional vector of covariates, \( \epsilon_i \) is an unobservable error term, and \( \beta_0 \) are the \( d \)-dimensional parameters of interest. The function \( g \) is unknown, but is assumed to be monotonic in its arguments.

This model has received a great deal of attention in recent econometric literature. Its attractiveness stems from the fact that by leaving the transformation function \( g(\cdot, \cdot) \) unspecified, the effect of covariates on response variables can be studied in a much more general setting than known or parametric transformation models. This is desirable because economic theory rarely specifies a functional form for such relationships. Furthermore, the above specification does not impose additive separability between the regressors and the error term, an assumption imposed by more traditional econometric methods, even though in some cases it contradicts economic theory.

Several estimators have been proposed for the transformation model, each based on different assumptions on the unobserved error term. Han (1987) introduced the Maximum Rank Correlation (MRC) estimator, which maximized the objective function:

\[ R_n(\beta) = \left( \frac{n}{2} \right)^{-1} \sum_{i < j} I[y_i > y_j]I[x_i'\beta > x_j'\beta] + I[y_i < y_j]I[x_i'\beta < x_j'\beta] \]

where \( I[\cdot] \) denotes the usual indicator function. Han also gave regularity conditions for this estimator to be consistent. Under similar conditions, Sherman (1993) showed this estimator to be root-\( n \) consistent and asymptotically normal.

An alternative estimator based on a monotonicity condition, though intended to estimate a wider class of models than characterized by (1.1), is the Monotone Rank Estimator (MRE) proposed by Cavanagh and Sherman (1998), defined as the maximizer of the function:

\[ S_n(\beta) = \sum_{i \neq j} M(y_i)I[x_i'\beta > x_j'\beta] \]

\( ^1 \epsilon_i \) is usually assumed to be distributed independent of \( x_i \).

\( ^2 \) For example, in the analysis of money transfers from parents to children, Altonji et al. (1997) find that nonseparability is a generic property of transfer equations which are based on a consumer choice framework with interdependent preferences.
This condition imposes a weak monotonicity condition on the index. In contrast, Ahn et al. (1996) consider models which satisfy a strict monotonicity condition, and provide computational motivations for imposing the stronger condition.

In Buchinsky’s analysis, all the explanatory variables were discretely distributed, enabling him to use a ‘minimum distance’ estimator. In contrast, the estimator introduced in this paper requires that one of the regressors be continuously distributed, as is typically required for identification in index models. It should be noted, however, that in Buchinsky’s one-group model, an explanatory variable such as experience (which is measured in years) may take sufficiently many different values for the continuity assumption to serve as an adequate approximation. Also, the rank procedure used in the second stage of the estimator in this paper has been extended for the all discrete regressor case in Cavanagh and Sherman (1998). It seems plausible that a similar extension could be applied to the results here, but it is not pursued in this paper.

where \( M(\cdot) \) is some known monotonic function. The condition driving the consistency of the MRE is that

\[
E[M(y_i)|x_i] \text{ is monotonic in } x_i\beta_0. \tag{1.2}
\]

Other estimators in the literature which do not require monotonicity exploit the relationship:

\[
E[y_i|x_i] = G(x_i\beta_0)
\]

where \( G(\cdot) \) is an unknown, sufficiently smooth function. Examples of estimators based on this relationship include Powell et al. (1989), Ichimura (1993), and Ai (1997). While this condition is quite general as well, it too restricts the behavior of the error term, requiring that it depend on the regressors through the index \( x_i\beta_0 \). A recent paper by Chaudhuri et al. (1997), which aims to relax this restriction, considers a smoothness condition on a conditional quantile function, instead of a conditional mean function.

The purpose of this paper is to introduce and estimate a new class of models which are based on a monotonicity condition of the conditional quantiles of the response variable. Specifically, let \( x \in (0, 1) \) and denote the \( x \)th conditional quantile of the response variable by \( Q^x(y_i|x_i) \). This paper will focus on estimating models which satisfy the condition:

\[
Q^x(y_i|x_i) \text{ is monotonic in } x_i\beta^{(x)}_0 \tag{1.3}
\]

There are essentially two motivations for estimating models based on this quantile condition. First, it is often the case in areas such as labor economics that the tails of the conditional distribution of the response variable (such as wage, income) are of greater interest to the researcher than the mean. Furthermore, by allowing the slope coefficients \( \beta_0 \) to depend on the quantile \( x \) considered, additional properties of the conditional distribution of the response variable may be explored by comparing slope coefficients across different quantiles, as was done in Buchinsky (1994) in his analysis of the U.S. wage structure. The second motivation for considering the class of models based on

---

3 This condition imposes a weak monotonicity condition on the index. In contrast, Ahn et al. (1996) consider models which satisfy a strict monotonicity condition, and provide computational motivations for imposing the stronger condition.

4 In Buchinsky’s analysis, all the explanatory variables were discretely distributed, enabling him to use a ‘minimum distance’ estimator. In contrast, the estimator introduced in this paper requires that one of the regressors be continuously distributed, as is typically required for identification in index models. It should be noted, however, that in Buchinsky’s one-group model, an explanatory variable such as experience (which is measured in years) may take sufficiently many different values for the continuity assumption to serve as an adequate approximation. Also, the rank procedure used in the second stage of the estimator in this paper has been extended for the all discrete regressor case in Cavanagh and Sherman (1998). It seems plausible that a similar extension could be applied to the results here, but it is not pursued in this paper.
Typically, a quantile condition is to allow for general forms of conditional heteroskedasticity. In this context, one would assume that \( \beta_0 \) is independent of \( z \), and that (1.3) holds for a particular, known quantile \( z \). The condition in (1.3) will be satisfied for transformation models where the error terms satisfy a conditional quantile restriction, and will thus allow for many forms of conditional heteroscedasticity. This restriction provides a generalization of the transformation model in Eq. (1.1) with the usual assumption of independence between the errors and regressors. As established in Proposition 2.1 of Cavanagh (1996), the transformation model with independent errors implies the conditional quantile functions are monotonic in \( x_i \beta_0 \) for all \( z \in (0,1) \).

This quantile condition may also be used to estimate parameters in heteroscedastic models for which the estimators based on conditional mean conditions are not consistent. For example, in the transformation model with unknown transformation function, the conditions upon which these estimators are based will not be satisfied unless the errors are additively separable, or are depend on the regressors only through the index \( x_i \beta_0 \). To illustrate why these requirements for consistency may be restrictive, we consider the following transformation model exhibiting conditional heteroscedasticity:

\[
y_i = g(x_i \beta_0, \sigma(x_i) e_i)
\]

where \( g(\cdot, \cdot) \) is monotonic in its arguments and \( \sigma(\cdot) \) is a function of the regressors affecting the scale of the error term. Neither the MRC, MRE nor single index estimators provide consistent estimators for \( \beta_0 \) in this class of models without further assumptions on \( g(\cdot) \) and/or \( \sigma(\cdot) \).\(^6\) However, consistency of an estimator based on the quantile condition only requires that the conditional quantile of \( e_i \) be 0.

Finally, we point out one other advantage of the condition in (1.3) when compared with the MRE condition in (1.2). While the latter assumes monotonicity of the conditional mean of some known monotone transformation of the response variable, the equivariance property of quantiles implies monotonicity of the index in (1.3) for any monotonic transformation of the response variable, since

\[
Q^*(M(y_i|x_i)) = M(Q^*(y_i|x_i)) \text{ is monotonic in } x_i \beta_0^{(\alpha)}
\]

Thus a monotonicity condition of quantiles, in addition to allowing for the forms of heteroscedasticity in (1.4), may be more flexible in the sense that the

\(^5\)Typically, \( z = 0.5 \) in practice. As discussed in Powell (1986), the median is the most sensible location restriction in order to estimate the ‘typical response’ value in the presence of conditional heteroscedasticity.

\(^6\)The MRE will be consistent if \( M(\cdot) \) is set to \( g^{-1}(\cdot) \) in this case, but this requires that the econometrician know the transformation function.
consistency of an estimator based on this condition will not depend on the choice of $M(\cdot)^{7}$.

The rest of the paper is organized as follows. The next section describes a two-stage estimation procedure for models which satisfy the monotonicity condition in Eq. (1.3). Section 3 discusses the asymptotic properties of the estimator, specifying conditions for consistency and asymptotic normality, as well as proposing an estimator for the limiting covariance matrix. Section 4 considers two extensions of the estimation procedure. The first is a modification of the estimation procedure to accommodate censored data. The second explains how the proposed estimator can be used in follow up stages for dimensionality reduction and structural function estimation. Section 5 explores the finite sample performance of the estimator through a simulation study. Section 6 concludes by summarizing and discussing areas for future research. An appendix provides the details of the proofs of the main theorems.

2. Description of the estimation procedure

To illustrate how to estimate the parameters of interest $\beta_0$ under the condition of (1.3), it will be useful to first assume that the quantile function $Q^q(y|x_i) \equiv q^q(x_i)$ is observed by the econometrician at each data point. In this context, one could easily recover the parameters of interest by modifying the objective function of the MRE such that the response variable is replaced by its conditional quantile function:

$$S_n(\beta) = \sum_{i \neq j} q^q(x_i) I[x_i' \beta > x_j' \beta]. \quad (2.1)$$

Of course the values of the quantile function are unknown, making the proposed estimator infeasible. However, they can be estimated nonparametrically in a preliminary stage, and then ‘plugged in’ the objective function, which can then be maximized in the second stage. Thus we propose a ‘semiparametric 2-step’ estimator, defined as the value which maximizes:

$$\tilde{S}_n(\beta) = \sum_{i \neq j} \hat{q}^q(x_i) I[x_i' \beta > x_j' \beta] \quad (2.2)$$

where $\hat{q}^q(\cdot)$ is some nonparametric estimator of $q^q(\cdot)$.

Before describing in more detail the two stages of the proposed estimator, it is worth comparing the approach taken here with similar approaches taken for

---

7 It should be noted that consistency may be lost in the quantile case for very specific choices of $M(\cdot)$. For example if $Q^q(y|x_i)$ has ‘flats’, for $M(\cdot)$ with ‘flats’ in the right areas, $M(Q^q(y|x_i))$ would be constant across all values of the index. Note this will not be a problem if either of these two functions is strictly monotonic.
other estimators in the semiparametric literature. The adopted approach in some sense resembles the work of Nawata (1990,1992), who was estimating, respectively, the censored regression and binary choice models under quantile restrictions. He also nonparametrically estimated the conditional quantile in the first stage. The reduced form values were then treated as raw data in a second stage fit by maximizing the likelihood function of the model in question using standard normal errors. Of course, the second stage of his approach cannot be applied under the general conditions of Eq. (1.3).

More recent work resembling the estimation approach taken here is the paper of Chaudhuri et al. (1997), who proposed average quantile derivative estimators for the slope parameters in transformation models. Their approach is based on a smoothness condition of the quantile function and yields a simple closed-form estimator, whose second stage is simpler to compute than the estimator introduced here. However, as will be discussed throughout the paper, there are certain disadvantages to adopting a smoothness approach in this context. One disadvantage is that their approach does not immediately apply to estimating coefficients of discretely distributed regressors. Another disadvantage of their derivative based methods is they generally require the selection of additional smoothing parameters and will also require stronger smoothness conditions on the regressors and the error terms than the monotonicity approach adopted here. These points will be explained in greater detail later in the paper.

2.1. First-stage of the estimator

The first stage involves nonparametrically estimating the conditional quantile of the dependent variable given the regressors \(x_i\). Several estimators have been recently proposed in the statistics and econometrics literature, notably Stute (1986), Bhattacharya and Gangopadhyay (1990) and Chaudhuri (1991a,b) and Cavanagh (1996). For the proof of the asymptotic properties of the proposed estimator, the local polynomial estimator discussed in Chaudhuri (1991a,b) and Cavanagh (1996) is used in the first stage. 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 in Chaudhuri (1991b).

First, we assume that the regressor vector \(x_i\), whose distribution function we denote by \(F_X(\cdot)\), can be partitioned as \((x_{i1}, x_{i2})\), where the \(d_1\)-dimensional vector \(x_{i1}\) is discretely distributed, and the \(d_2\)-dimensional vector \(x_{i2}\) is continuously distributed.

We let \(C_n(x_i)\) denote the cell of observation \(x_i\) and let \(h_n\) denote the sequence of bandwidths which govern the size of the cell. For some observation \(x_j, j \neq i\), we let \(x_j \in C_n(x_i)\) denote that \(x_{j1} = x_{i1}\) and \(x_{j2}\) lies in the \(d_2\)-dimensional cube centered at \(x_{i2}\) with side length \(h_n\).

Next, we let \(k\) denote the assumed order of differentiability of the quantile functions with respect to \(x_{i2}\). We let \(A\) denote the set of all \(d_2\)-dimensional
vectors whose components are nonnegative integers that sum to a number less than or equal to \( k \), and let \( s(A) \) denote the number of vectors in \( A \). We denote the individual vectors in \( A \) by \( \{b_j\}_{j=1}^{s(A)} \), assumed to be ordered such that the first vector, \( b_1 \), has all components which are 0. For each of the vectors \( b_i \), we let \( [b_i] \) denote the sum of its components.

For any \( s(A) \)-dimensional vector \( \boldsymbol{\Psi} \), we let \( \Psi_{(l)} \) denote its \( l \)th component, and for any two \( d_c \)-dimensional vectors \( a, b \), we let \( a^b \) denote the product of each component of \( a \) raised to the corresponding component of \( b \). Finally, we let \( I[\cdot] \) be an indicator function, taking the value 1 if its argument is true, and 0 otherwise. The local polynomial estimator of the conditional \( \alpha \)th quantile function at an observation \( x_i \) involves \( \alpha \)-quantile regression (Koenker and Bassett, 1978) on observations which lie in the defined cells of \( x_i \). Specifically, let the vector

\[
\hat{\Psi} = (\hat{\Psi}_{(1)}, \hat{\Psi}_{(2)}, \ldots, \hat{\Psi}_{(s(A))})
\]

be the solution to

\[
\arg\min_{\Psi} \sum_{j=1}^{n} I[x_j \in C_n(x_i)] \rho_{\alpha} \left( y_j - \sum_{i=1}^{s(A)} \Psi_{(l)} \cdot (x_j^{(i)} - x_i^{(i)})^b \right)
\]  

(2.3)

where \( \rho_{\alpha}(\cdot) \equiv \left| z \right| \cdot \left( 2\alpha - 1 \right)(\cdot) I[\cdot < 0] \). The conditional quantile estimator which will be used in the first stage will be the value \( \hat{\Psi}_{(1)} \).

A computational advantage of using this estimator is that its evaluation can be carried out by linear programming techniques. Efficient algorithms, such as that proposed by Barrodale and Roberts (1973), converge to a solution in a finite number of simplex iterations. Since 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 objective function is to achieve bias reduction of the nonparametric estimator, analogous to the bias reduction achieved in local polynomial regression estimators (Fan and Gijbels, 1996). Thus the additional parameters estimated in the above minimization problem are purely ‘nuisance’ parameters which are estimated only to improve the asymptotic properties of the second stage. If the required order of smoothness of the quantile function and the dimension of \( x_i \) are both large, the number of nuisance parameters introduced becomes enormous, making the above minimization problem computationally expensive.

This point can be illustrated with a simple example. If the quantile function is assumed to be twice differentiable, and \( x_i \) consists of two continuous components, then \( A \) would correspond to the set of vectors \((0,0),(1,0),(0,1),(1,1),(2,0),(0,2)\), so \( s(A) = 6 \). On the other hand, if the function is assumed to be three times continuously differentiable, \( A \) now includes the additional vectors \((3,0),(0,3),(1,2),(2,1)\), so \( s(A) = 10 \). Thus, from a computational point of view, it will be desirable to impose as weak a smoothness condition on the conditional quantile function as possible.
Table 1
Computation times

<table>
<thead>
<tr>
<th>n</th>
<th>d = 5</th>
<th>d = 10</th>
<th>d = 15</th>
<th>d = 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>21.652 (0.139)</td>
<td>39.666 (0.091)</td>
<td>64.568 (0.074)</td>
<td>200.886 (0.032)</td>
</tr>
<tr>
<td>1000</td>
<td>57.946 (0.271)</td>
<td>104.444 (0.185)</td>
<td>155.690 (0.145)</td>
<td>493.804 (0.061)</td>
</tr>
<tr>
<td>1500</td>
<td>118.366 (0.384)</td>
<td>205.114 (0.252)</td>
<td>289.512 (0.199)</td>
<td>888.300 (0.086)</td>
</tr>
<tr>
<td>2000</td>
<td>204.040 (0.466)</td>
<td>324.436 (0.307)</td>
<td>467.220 (0.227)</td>
<td>1403.784 (0.099)</td>
</tr>
</tbody>
</table>

2.2. Second stage of the estimator

With the quantile function estimated in the first stage, the values obtained can then be ‘plugged in’ to a rank based objective function in the second stage. The second stage estimator will therefore be defined as the maximizer of a 2nd order U-process

$$\hat{\beta} = \arg \sup_{\beta \in \mathcal{B}} \frac{1}{(n-1)} \sum_{i \neq j} \tau(x_i)q'(x_i)I[x'_i\beta > x'_j\beta]$$

(2.4)

where $\mathcal{B}$ is some parameter space. The function $\tau(\cdot)$ included in the above objective function is an ‘exogenous’ trimming function, selected by the econometrician. It serves to trim away observations for which the quantile function estimator is known to be imprecise.

We conclude this section with a brief discussion on the computation time of the proposed two-step procedure. Since the second stage objective function is not continuous, gradient-based methods cannot be employed. Nonetheless, simplex methods, such as the Nelder Meade algorithm, can be used effectively to determine local maxima. Unfortunately, this algorithm can converge quite slowly as the number of regressors get large, and in practice, hundreds of different starting values should be used to ensure a global maxima is found.

The first stage can also incur significant computation costs. Even though an algorithm such as Barrodale–Roberts will converge more rapidly than Nelder–Meade, and that a global minimum will be found, the first stage involves estimating a far greater number of parameters, and the minimization problem which must be carried out not once, but $n$ times.

Table 1 gives some indication of both the total and relative computation times of estimation procedure. Computation times are evaluated for designs with a various number of regressors at different sample sizes. A basic homoscedastic linear model was simulated with the number of regressors ranging from $d = 5$ to $d = 20$, and sample sizes ranging from $n = 500$ to $n = 2000$. A local linear regression was fit in the first stage, and 100 starting values were used in the
second stage. This approach will underestimate computation times for two reasons. First, as will be explained later in this paper, a higher-order polynomial will be required in the first stage to ensure convergence at the parametric rate. Second, a far greater number of starting values will be required in the second stage to ensure that a global maximum is found. Nonetheless, the table will still provide some insight to the relative computation times of the two stages. Furthermore, the values in the table indicate by how much computation times increase as new parameters are added, so computation times can be calculated for situations when the correct order of polynomial and more starting values are used.

The reported computation times (in seconds) are based on a simulation study performed in GAUSS on a Pentium II 400 MHz PC. As the results indicate, the second stage rank procedure is responsible for most of the computation time. Only in situations where the number of regressors is relatively small, and the sample size is relatively large, are the computation times of the two stages similar. In fact, in the case where \( d = 5 \) and \( n = 2000 \), it may be the case that the first-stage procedure is more computationally expensive than the second stage, as the correct order of polynomial (to ensure parametric rate convergence) is 8, which would introduce hundreds of additional parameters. Other than this situation, it seems that the second stage is more burdensome, as the rate of increase in computation time as parameters increase is much larger for the Nelder Meade algorithm than it is for the Barrodale–Roberts algorithm.

3. Asymptotic properties

Before specifying sufficient conditions for consistency and asymptotic normality of the proposed estimator, an appropriate normalization must be adopted to account for the fact that the parameter of interest \( \beta_0 \) is only identified up to scale. We adopt the normalization used in Sherman (1993) and Cavanagh and Sherman (1998), and set the \( d \)th component of \( \beta_0 \) to 1, expressing \( \beta_0 \) as \((\theta_0, 1)\). Furthermore, each value of \( \beta \) in the parameter space will be expressed as \( \beta(\hat{\theta}) = (\theta, 1) \). Thus this section will focus on specifying conditions for which \( \hat{\theta} \) converges to \( \theta_0 \).

3.1. Consistency

We begin by specifying conditions for consistency of the proposed estimator. Our consistency result is based on the following assumptions:

Assumptions on the parameter space

A. \( \theta_0 \) lies in the interior of \( \Theta \), a compact subset of \( \mathbb{R}^{d-1} \).
Assumptions on the quantile function

Q1. The $x$-quantile function is monotonic in the index. We express this as

$$Q^*(y_i|x_i) \equiv q^*(x_i) = H(x_i\beta_0)$$

where $H(\cdot)$ is a monotonic function.

Q2. For any value $x^{(d)}$ in the support of $x^{(d)}$, $q^*(\cdot)$ is $k$ times differentiable in $x^{(c)}$.

Letting $\nabla_k q^2(x^{(c)},x^{(d)})$ denote the vector of $k$th order derivatives of $q^*(\cdot)$ in $x^{(c)}$, we assume the following Lipschitz condition:

$$\|\nabla_k q^2(x_1^{(c)},x^{(d)}) - \nabla_k q^2(x_2^{(c)},x^{(d)})\| \leq \mathcal{K}\|x_1^{(c)} - x_2^{(c)}\|^\gamma$$

for all values $x^{(c)},x^{(d)}$ in the support of $x^{(c)}$, where $\|\cdot\|$ denotes the Euclidean norm, $\gamma \in (0,1]$, and $\mathcal{K}$ is some positive constant. In the theorems to follow, we will let $p = k + \gamma$ denote the order of smoothness of the quantile function.

Assumptions on the trimming function

T. The trimming function $\tau: \mathbb{R}^d \rightarrow \mathbb{R}^+$ is continuous, bounded, and bounded away from zero on its support, denoted by $\mathcal{X}$, a compact subset of $\mathbb{R}^d$.

Assumptions on the regressors

B1. The sequence of $d+1$ dimensional vectors $(y_i,x_i)$ are independent and identically distributed.

B2. The regressor vector $x_i$ has support which is a subset of $\mathbb{R}^d$.

We order the components of $x_i$ so it can be written as $x_i = (x_i^{(d)},x_i^{(c)})'$. Let $d_c$ denote dim($x_i^{(c)}$). Assume that $1 \leq d_c \leq d$ and that the support $x_i^{(c)}$ is a convex subset of $\mathbb{R}^d$ and has nonempty interior. Assume that the support of $x_i^{(d)}$ is a finite number of points lying in $\mathbb{R}^{d-d_c}$. We will let $f_x(x)$ denote the product of the conditional (Lebesgue) density of $x_i^{(c)}$ given $x_i^{(d)}$ (denoted by $f_{X|X^{(d)}}(x=x^{(c)})$) and the marginal probability mass function of $X^{(d)}$ (denoted by $f_{X^{(d)}}(x^{(d)})$).

B3. $f_{X|X^{(d)}}(x^{(c)})$ is continuous and bounded on the support of $x_i^{(c)}$.

B4. Assume that $\mathcal{X} = \mathcal{X}_{d-1} \times \mathcal{X}_d$ where $\mathcal{X}_{d-1}$ and $\mathcal{X}_d$ are compact subsets with nonempty interiors of the supports of the first $d-1$ components, and the $d$th component of $x_i$, respectively. For each $x \in \mathcal{X}$, denote its first $d-1$ components by $x_{(d-1)}$, $\mathcal{X}$ will be assumed to have the following properties:

B4.1. $\mathcal{X}$ is not contained in any proper linear subspace of $\mathbb{R}^d$.

B4.2. $f_x(x) \geq \varepsilon_0 > 0 \ \forall x \in \mathcal{X}$, for some constant $\varepsilon_0$.

B4.3. let $t_0$ satisfy $H(t) < H(t_0)$ if $t < t_0$, and let

$$T = \max_{(x_{(d-1)},\theta) \in \mathcal{X}_{(d-1)} \times \Theta} |x_{(d-1)}| < \infty$$

we assume $\mathcal{X}_d$ contains the interval $[t_0 - 3T, t_0 + 3T]$. 
Assumptions on the quantile residual term

D1. Let \( u_i = y_i - q^{\varphi}(x_i); \) in a neighborhood of 0, \( u_i \) has a conditional (Lebesgue) density, denoted by \( f_{u|x_i} = x(\cdot) \) which is continuous, and bounded away from 0 and infinity for all values of \( x \in \mathcal{X}. \) As a function of \( x, f_{u|x_i} = x \) is Lipschitz continuous for all values of \( u_i \) in a neighborhood of 0.

These regularity conditions are sufficient for consistency of the proposed estimator, as established in the following theorem. The proof, found in the appendix, is based on standard consistency theorems for compact parameter spaces.

**Theorem 3.1.** Assume that \( p > 0, \) and in the first stage \( 0 \leq k \leq \text{int}(p), \) where \( \text{int}(\cdot) \) is the function taking the integer of its argument. If the bandwidth satisfies \( h_n \to 0 \) and \( nh_n^d \to \infty, \) then under Assumptions A, Q, T, B, D,

\[
\hat{\theta}^p \to \theta_0.
\]  

(3.1)

Before proceeding to the discussion regarding the rate of convergence and limiting distribution of the proposed estimator, some comments are in order regarding the conditions used in the consistency theorem:

**Remark 3.1.** Assumption B2 allows for both continuous and categorical explanatory variables. This is in contrast to the average derivative approach taken in Chaudhuri et al. (1997). For their estimators, all regressors are required to have a continuous distribution, making them initially inapplicable for most econometric models, where the effects of race, gender, etc. on response variables are often of interest. While it seems plausible that coefficients of discrete regressors could be estimated in a follow up stage by following the approach developed in Horowitz and Härdle (1996), this will have the disadvantage of requiring the selection of an additional smoothing parameter.

**Remark 3.2.** Assumption B4.3 is analogous to the identification assumption found in Cavanagh and Sherman (1991), where the support of the \( d \)th regressor was assumed to be \( \mathfrak{R}. \) It ensures identification is not lost by restricting ourselves to the compact support of the trimming function. Intuitively, it requires that the size of the support of the \( d \)th regressor is sufficiently ‘large’ with respect to the other regressors.

3.2. Root-n consistency and asymptotic normality

We now establish conditions which enable the consistency result to be extended to root-\( n \) consistency and asymptotic normality. In addition to the
assumptions detailed in the previous section, we impose the following additional assumptions on the distribution of the regressors. Essentially, these additional conditions impose smoothness on distribution of the index $x_i' \beta(\theta)$ and on the trimming and quantile functions. We state these conditions in a format similar to that used in Sherman (1993) and Cavanagh and Sherman (1998). Let

$$q_1(x, h) = P_{q_1}(x) a(x) I[x \in \Pi_x(u)] dF_X(u)$$

and let $q_2(x, h) = P_{q_2}(x) a(x) I[x \in \Pi_x(u)] dF_X(u)$

and let $N$ be a neighborhood of the $d - 1$ dimensional vector $\theta_0$. Then we impose the following additional assumptions:

E1. For each $x$ in the support of $x_i$, $\tau_1(x, \cdot)$ is differentiable of order 2, with Lipschitz continuous second derivative on $N$.

E2. $E[\nabla_2 \tau_1(\cdot, \theta_0)]$ is negative definite.

E3. For each $x$ in the support of $x_i$, $\tau_2(x, \cdot)$ is continuously differentiable on $N$.

E4. $E[\|\nabla_1 \tau_2(\cdot, \theta_0)\|^2] < \infty$.

The following theorem establishes that these additional assumptions, along with a stronger smoothness condition on the quantile function and further restrictions on the bandwidth sequence, are sufficient for root-$n$ consistency and asymptotic normality of the proposed estimator:

**Theorem 3.2.** Assume that $p > 3d_c/2$, and that in the first stage, $k$ is set to $\text{int}(p)$ and the bandwidth satisfies $\sqrt{n h_n^2} \rightarrow 0$ and $\log n \sqrt{n^{-1} h_n^{-3d_c}} \rightarrow 0$. Define

$$\delta(y_i, x_i) = \tau(x_i) f_{y_i|x_i}(0) [I[y_i \leq q(x_i)] - \alpha] \nabla_1 \tau_2(x_i, \theta_0)$$

then under Assumptions A, B, Q, T, E,

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, V^{-1} \Delta V^{-1})$$

where $\Delta = E[\delta(y_i, x_i) \delta(y_i, x_i)']$ and $V = \frac{1}{2} E[\nabla_1 \tau_1(x_i, \theta_0)]$.

**Remark 3.3.** It is worth comparing the smoothness assumptions here with those required for the derivative based estimators discussed in Chaudhuri et al. (1997). Their average derivative estimator requires that $p > 3d_c/2 + 3$. This stronger condition can have significant computational implications, as it requires that a higher-order polynomial be used in the first stage estimation procedure. This
can lead to a significant number of additional nuisance parameters. For example, if two continuous regressors are present, the stronger smoothness condition translates into 20 additional parameters and if the model has three continuous regressors, their condition introduces over 100 additional parameters.

Although the density weighted average derivative approach they suggest allows for the same smoothness condition on the quantile function we impose, this comes at the expense of far stronger smoothness assumptions on the conditional density of the error term, and the marginal density of the regressors. Furthermore, this approach requires estimation (at each data point) of both the regressor density function and its derivative. In addition to the computational cost this incurs, it also introduces the problem of selecting two additional smoothing parameters.

Remark 3.4. While the bandwidth restrictions stated in the theorem allow for a range of bandwidth sequences, they rule out the sequence which yields the optimal rate of convergence of the first step estimator as discussed in Chaudhuri (1991a). As is often the case with two step estimators, ‘undersmoothing’ is required for root-$n$ consistency of the second-stage estimator.

3.3. Covariance matrix estimation

For purposes of inference, we consider estimation of the limiting covariance matrix derived in the previous theorem. As is usually the case with estimators which maximize nonsmooth objective functions, consistent estimation of the covariance matrix requires the selection of additional smoothing parameters, and we first consider an analog to the numerical derivative approach considered in Cavanagh and Sherman (1998). This involves separate estimators for the two components, $V$ and $\Delta$.

To estimate $V$, we define

$$\hat{\tau}_1(x, \theta) = \frac{1}{n} \sum_{i=1}^{n} \tau(x)q(x)I[x'\beta(\theta) > x'_i\beta(\theta)] + \tau(x_i)q(x_i)I[x'_i\beta(\theta) > x'\beta(\theta)]$$

(3.3)

and let

$$\hat{\gamma}_{kl}(x, \theta) = \varepsilon_n^{-2}[\hat{\tau}_1(x, \theta + \varepsilon_n e_k + e_l)]$$

$$- \hat{\tau}_1(x, \theta + \varepsilon_n e_l) - \hat{\tau}_1(x, \theta + \varepsilon_n e_k) + \hat{\tau}_1(x, \theta)]$$

(3.4)

where $e_i$ denotes a $d-1$ dimensional vector whose $i$th element is 1 and remaining elements are 0, and $\varepsilon_n$ is a ‘smoothing’ parameter that satisfies $\varepsilon_n \to 0$.
and \( n^{1/4}v_n \to \infty \). We propose estimating the \( k, l \)th element of \( V \) by

\[
\hat{V}_{ij} = \frac{1}{2n} \hat{\gamma}_{kl}(x_i, \hat{\theta}).
\]  

(3.5)

Estimation of \( \Delta \) is a little more complicated. We first require an estimator of the conditional density of the error term \( u_i \) at 0. For this we propose a simple Nadaraya–Watson type estimator of the form

\[
\hat{f}_{u_i|x_i}(0) = \frac{1}{\hat{h}_n} \sum_{i \neq j} K_1 \left( \frac{\hat{u}_j}{\hat{h}_n} \right) K_2 \left( \frac{x_j - x_i}{\hat{h}_n} \right) I[x_j = x_i] \]

(3.6)

where \( \hat{u}_i = y_i - \hat{q}(x_i) \); \( K_1(\cdot) : \mathbb{R} \to \mathbb{R} \) is a continuously differentiable ‘kernel’ function which integrates to 1 and is symmetric around zero; \( K_2(\cdot) : \mathbb{R} \to \mathbb{R}^d \) is a multivariate kernel function which integrates to 1 and is symmetric around the 0 vector; \( h_n \) is a bandwidth which satisfies \( h_n \to 0 \) and \( n^{1/8}h_n \to \infty \).

We next define

\[
\hat{\tau}_2(x, \hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} I[x_i\beta(\theta) > x_i\beta(\hat{\theta})]
\]

(3.7)

and its \( d - 1 \)-dimensional numerical derivative, \( \nabla \hat{\tau}_2(x, \hat{\theta}) \), whose \( k \)th element is

\[
v_n^{-1} \left[ \hat{\tau}_2(x, \hat{\theta} + v_ne_k) - \hat{\tau}_2(x, \hat{\theta}) \right]
\]

where \( v_n \) is a smoothing parameter that satisfies \( v_n \to 0 \) and \( \sqrt{n}v_n \to \infty \). We propose an estimator of \( \Delta \) of the form

\[
\hat{\Delta} = \frac{1}{n} \sum_{i=1}^{n} \hat{\delta}_i \hat{\delta}_i
\]

(3.8)

where

\[
\hat{\delta}_i = \tau(x_i) \sqrt{\tau(1-\tau)} \hat{f}_{u_i|x_i}(0) \nabla \hat{\tau}_2(x_i, \hat{\theta}).
\]

The following theorem, whose proof can be found in the appendix, establishes the consistency of our proposed estimator:

**Theorem 3.3.** Under the conditions for root-\( n \) consistency of \( \hat{\beta} \), and the restrictions on the smoothing parameters \( h_n, v_n, \hat{\theta}_n \),

\[
\hat{V}^{-1} \hat{\Delta} \hat{V}^{-1} \overset{p}{\to} V^{-1} \Delta V^{-1}.
\]

The presence of additional smoothing parameters required for the above estimator is somewhat problematic as there are very few guidelines in the literature on how they should be selected in finite samples.
An alternative approach is to reformulate the limiting covariance matrix in terms of the model primitives, to which more standard nonparametric estimation methods may be applied. Such an approach was considered in Cavanagh and Sherman (1998). Let $x_{(d-1)i}$ denote the first $d-1$ components of $x_i$, and let $\bar{x}_{(d-1)i}$ denote $E[x_{(d-1)i}|x'_i\beta_0]$. Also, we let $H'(\cdot)$ denote the derivative of the function $H(\cdot)$, and let $g_0(\cdot)$ denote the density function of the index $x'_i\beta_0$. Using identical arguments as used in Cavanagh and Sherman (1998), we have the following representations\(^8\) for the components of the limiting covariance matrix:

$$
A = E[(x_{(d-1)i} - \bar{x}_{(d-1)i})(x_{(d-1)i} - \bar{x}_{(d-1)i})'z(1 - z)f_{d|x}(0|x_i)^{-2}g_0(x'_i\beta_0)^2],
$$

(3.9)

$$
V = E[(x_{(d-1)i} - \bar{x}_{(d-1)i})(x_{(d-1)i} - \bar{x}_{(d-1)i})'H'(x'_i\beta_0)g_0(x'_i\beta_0)].
$$

(3.10)

Thus nonparametric methods can be used to estimate the separate density, expectation, and derivative functions at each observation point, and then $A$ and $V$ can be estimated by averaging values.

A third approach to estimating the limiting variance matrix would be to bootstrap confidence intervals. It should be noted, however, that bootstrapped intervals have yet to be proven to be consistent in this context, though we have little reason to believe they would not be. Also, for this particular estimator, bootstrapping will be extremely costly from a computational standpoint.

4. Extensions of the estimation procedure

In this section, we discuss how the estimation procedure may be modified in certain situations. First, we discuss how the procedure needs to be altered to accommodate data which are censored. Second, we consider estimation of the function $H(\cdot)$.

4.1. Estimation with censored data

We first consider estimating a model where the response variable is censored at a particular point, say 0, and the conditional quantile function, still assumed to be monotonic in the index $x'_i\beta_0$, does not fall below this censoring point. As an illustrative example of such a model assume a latent transformation model of the form:

$$
y^*_i = g(x'_i\beta_0, \varepsilon_i)
$$

(4.1)

\(^8\)The trimming functions are suppressed here for simplicity.
where \( g(\cdot, \cdot) \) is monotonic in its first argument and strictly monotonic in its second argument, and the econometrician observes the \( d + 1 \) dimensional vector of variables \((y_i, x_i')\), where

\[
y_i = \max(y_i^*, 0). \tag{4.2}
\]

While the example implies the data is censored at 0, we do not require the censoring point to be at 0, nor constant along observations, as long as its value is known to the econometrician even for uncensored observations.

Let \( c > 0 \) be an arbitrarily small number set by the econometrician; for this censored monotone index model we propose an estimator which minimizes:

\[
\frac{1}{n(n-1)} \sum_{i \neq j} \tau(x_i)w(q^*(x_i))I[x_i'\beta(\theta) > x_j'\beta(\theta)] \tag{4.3}
\]

where \( w(\cdot) \) has the following properties:

(i) \( w(\cdot) \) is nonnegative and twice continuously differentiable with bounded second derivative.

(ii) \( w(\cdot) = 0 \) if its argument is less than \( c \); \( w(\cdot) \) is strictly increasing on \([c, \infty]\).

Note the new objective function only differs by the fact that the quantile function estimator is now transformed by a smooth weighting function, whose support is \([c, \infty]\). Bounding the quantile function estimator away from 0 is necessary to avoid technical complications. With this minor adjustment, the same arguments may be used to prove the following theorem establishing the asymptotic normality of the estimator under censored data. The proof is left to the appendix.

**Theorem 4.1.** Let \( \hat{\theta}_c \) denote the estimator defined by minimizing (4.3). Let \( \mathcal{X}^c \) denote the set \( \{ x \in \mathcal{X} : q^*(x) \geq c \} \). Then under Assumptions A, Q, T and replacing \( \mathcal{X} \) with \( \mathcal{X}^c \) in Assumptions B, D and E,

\[
\sqrt{n}(\hat{\theta}_c - \theta_0) \Rightarrow \mathcal{N}(0, V_c^{-1}\Delta_c V_c^{-1}) \tag{4.4}
\]

where

\[
\Delta_c = \mathbb{E}[\omega'(q^*(x_i))^2\delta(y_i, x_i)\delta(y_i, x_i)']
\]

and

\[
V_c = \frac{1}{2}\mathbb{E}[\nabla^2 \tau_{1c}(x_i, \theta_0)]
\]

with

\[
\tau_{1c}(x_i'\theta_0) = \int \tau(x)\omega(q^*(x))I[x'\beta(\theta) > u'\beta(\theta)]dF_X(u)
\]

\[
+ \int \tau(u)\omega(q^*(u))I[u'\beta(\theta) > x'\beta(\theta)]dF_X(u).
\]
4.2. Dimensionality reduction and transformation function estimation

With a root-$n$ consistent estimator for the parameter $\beta_0$ now available, the quantile function can be estimated as a function of the single index. Specifically, noting that

$$Q^a(y|x) = q^a(x) = H(x'\beta_0)$$

the function $H(\cdot)$ can be estimated through nonparametric quantile regression of the dependent variable $y_i$ on the one-dimensional 'generated' regressor, $\xi_i = x_i'\hat{\beta}$. As will be shown below, the estimator of $H(\cdot)$ will converge at the usual one-dimensional nonparametric rate, thus circumventing the usual curse of dimensionality encountered in nonparametric estimation.

In the special context of additive transformation models of the form

$$y_i = g(x_i'\beta_0 + \epsilon_i)$$

this will provide an estimator of the transformation function $g(\cdot)$ under a quantile restriction on the error term $\epsilon_i$. Estimating the transformation function is especially useful in certain duration models (see, for example, Ridder, 1990). Estimators for the transformation function under the assumption of independence between errors and regressors have been proposed by Horowitz (1996), Ye and Duan (1997), and Gorgens and Horowitz (1999).

Without loss of generality, we consider estimating $H(\cdot)$ at 0 and denote the parameter of interest by $Y_0 \equiv H(0)$. This estimator is defined as the minimizer of

$$\frac{1}{nh_n} \sum_{i=1}^{n} I[|x_i'\hat{\beta}| \leq \tau_n/2] \rho_2(y_i - Y)$$

(4.5)

where $\tau_n$ denotes the bandwidth sequence.

The following theorem establishes that the rate of convergence and limiting distribution of this estimator is the same as the infeasible estimator which uses $\xi_i = x_i'\beta_0$ as the regressor; this is to be expected, as the estimator $\hat{\theta}$ converges at the parametric rate. A proof is available from the author upon request.

**Theorem 4.2.** Assume that $\tau_n = cn^{-1/5}$ where $c$ is some constant. Let $f_{\xi_i}$ denote the density function of $\xi_i$, and let $F_{u_i|\xi_i}(\cdot)$ and $f_{u_i|\xi_i}(\cdot)$ denote the conditional distribution and density functions of $u_i$ given $\xi_i$. Then

$$n^{2/5}(\hat{Y} - Y_0) \Rightarrow N(B, \alpha(1 - \alpha) f_{u_i|\xi_i}^{-2}(0) f_{\xi_i}^{-1}(0))$$

where

$$B = \frac{c^{3/2}}{24} \frac{d^2}{d\xi^2} (f_{\xi}(\xi) (x - F_{u_i|\xi_i}=\xi(Y_0 - H(\xi)))|_{\xi=0}.$$ 

For constructing confidence intervals or testing hypothesis regarding the structural function, the components of the bias and variance can be estimated
nonparametrically and ‘plugged in’. As is the case with nonparametric kernel estimation of the conditional mean, one can avoid estimating the bias term by ‘undersmoothing’ (selecting a suboptimal bandwidth sequence such that the bias shrinks at a quicker rate than the variance) at the cost of the optimal rate of convergence.

Note that since the proposed estimator does not involve isotonic regression, the estimated function will not necessarily be monotonic in finite samples. Alternatively, if an estimator of the structural parameters which does not require monotonicity is used in the first step (see Chaudhuri et al., 1997) a test for monotonicity of the proposed estimator would be a useful specification test regarding the properties of the structural function and/or the error term.

5. Monte Carlo results

In this section, the finite sample properties of the proposed estimator, referred to in this section as the quantile rank estimator (QRE), are examined through the results of a Monte Carlo simulation study. For comparison, the performance of Cavanagh and Sherman’s MRE, with \( M(\cdot) \) set to the identity function, and that of the quantile average derivative (QAD) estimator in Chaudhuri et al. (1997) are also examined. For the study, two models were simulated. The first was a linear regression model of the form:

\[
y_i = \alpha + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + \epsilon_i.
\]

The other was a log-linear regression:

\[
\log y_i = \alpha + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + \epsilon_i.
\]

The covariates \( x_{1i}, x_{2i}, x_{3i} \) were mutually independent; \( x_{2i} \) and \( x_{3i} \) were each drawn from a standard normal distribution, and \( x_{3i} \) was a binary variable, taking the values 0 and 1 with probability 0.5. The parameter values, \( \alpha, \beta_1, \beta_2, \beta_3 \) were 0, 1, 1 and 1 respectively. For the error term, \( \epsilon_i \), multiplicative heteroscedasticity was adopted:

\[
\epsilon_i = f(x_{1i}, x_{2i}, x_{3i})u_i
\]

where \( f \equiv 1 \) in the case of homoscedasticity, and

\[
f(x_{1i}, x_{2i}) = \exp(-0.5(x_{1i} + x_{2i} + x_{3i}))
\]

for the heteroscedastic case. The term \( u_i \) was drawn from a standard normal distribution.

Each Monte Carlo experiment involved 401 replications for sample sizes of 100, 200 and 400. Tables 2–5 reports four statistics, mean, median, mean square error, and mean absolute deviation for the various designs of the three estimators. The parameters estimated were \( \beta_2/\beta_1 = \beta_3/\beta_3 = 1 \) for the QRE and
Table 2
Monte Carlo simulation – slope coefficients, linear regression – homoscedastic design

<table>
<thead>
<tr>
<th></th>
<th>MRE</th>
<th>QRE</th>
<th>QAD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_2/\beta_1$</td>
<td>$\beta_3/\beta_1$</td>
<td>$\beta_2/\beta_1$</td>
</tr>
<tr>
<td>$n = 100$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean bias</td>
<td>0.0065</td>
<td>0.0095</td>
<td>0.0196</td>
</tr>
<tr>
<td>Med. bias</td>
<td>−0.0190</td>
<td>−0.0050</td>
<td>−0.0010</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.2023</td>
<td>0.2740</td>
<td>0.1984</td>
</tr>
<tr>
<td>MAD</td>
<td>0.1570</td>
<td>0.2110</td>
<td>0.1553</td>
</tr>
</tbody>
</table>

| $n = 200$ |     |     |     |     |     |     |
| Mean bias | 0.0049 | 0.0085 | 0.0017 | 0.0026 | 0.0099 | — |
| Med. bias | −0.0004 | −0.0026 | −0.0053 | −0.0110 | −0.0050 | — |
| RMSE    | 0.1183 | 0.1896 | 0.1199 | 0.1907 | 0.1208 | — |
| MAD     | 0.0923 | 0.1535 | 0.0913 | 0.1520 | 0.0957 | — |

| $n = 400$ |     |     |     |     |     |     |
| Mean bias | −0.0015 | 0.0049 | 0.0120 | 0.0772 | 0.0012 | — |
| Med. bias | −0.0010 | 0.0030 | 0.0120 | 0.0698 | −0.0037 | — |
| RMSE    | 0.0826 | 0.1293 | 0.0901 | 0.1394 | 0.0909 | — |
| MAD     | 0.0659 | 0.1052 | 0.0712 | 0.1119 | 0.0723 | — |

Table 3
Monte Carlo simulation – slope coefficients, linear regression – heteroscedastic design

<table>
<thead>
<tr>
<th></th>
<th>MRE</th>
<th>QRE</th>
<th>QAD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_2/\beta_1$</td>
<td>$\beta_3/\beta_1$</td>
<td>$\beta_2/\beta_1$</td>
</tr>
<tr>
<td>$n = 100$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean bias</td>
<td>0.0144</td>
<td>0.0142</td>
<td>0.0342</td>
</tr>
<tr>
<td>Med. bias</td>
<td>−0.0118</td>
<td>−0.0025</td>
<td>−0.0082</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1987</td>
<td>0.2775</td>
<td>0.2317</td>
</tr>
<tr>
<td>MAD</td>
<td>0.1526</td>
<td>0.2104</td>
<td>0.1697</td>
</tr>
</tbody>
</table>

| $n = 200$ |     |     |     |     |     |     |
| Mean bias | 0.0016 | 0.0072 | 0.0018 | 0.0097 | 0.0096 | — |
| Med. bias | −0.0110 | −0.0078 | −0.0101 | 0.0059 | 0.0075 | — |
| RMSE    | 0.1131 | 0.1735 | 0.1303 | 0.1941 | 0.1151 | — |
| MAD     | 0.0872 | 0.1375 | 0.0969 | 0.1515 | 0.0866 | — |

| $n = 400$ |     |     |     |     |     |     |
| Mean bias | −0.0003 | 0.0087 | 0.0035 | 0.0114 | 0.0028 | — |
| Med. bias | −0.0100 | 0.0002 | −0.0043 | −0.0013 | −0.0016 | — |
| RMSE    | 0.0835 | 0.1254 | 0.0885 | 0.1435 | 0.0845 | — |
| MAD     | 0.0638 | 0.0989 | 0.0699 | 0.1092 | 0.0671 | — |
Table 4
Monte Carlo simulation – slope coefficients, log-linear regression – homoscedastic design

<table>
<thead>
<tr>
<th></th>
<th>MRE</th>
<th>QRE</th>
<th>QAD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_2/\beta_1$</td>
<td>$\beta_3/\beta_1$</td>
<td>$\beta_2/\beta_1$</td>
</tr>
<tr>
<td>$n = 100$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean bias</td>
<td>0.0068</td>
<td>0.0356</td>
<td>0.0422</td>
</tr>
<tr>
<td>Med. bias</td>
<td>— 0.0418</td>
<td>0.0087</td>
<td>— 0.0011</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.2754</td>
<td>0.3765</td>
<td>0.3129</td>
</tr>
<tr>
<td>MAD</td>
<td>0.2150</td>
<td>0.2916</td>
<td>0.2209</td>
</tr>
<tr>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean bias</td>
<td>0.0176</td>
<td>0.0032</td>
<td>0.0404</td>
</tr>
<tr>
<td>Med. bias</td>
<td>0.0087</td>
<td>0.0148</td>
<td>0.0236</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1907</td>
<td>0.2762</td>
<td>0.2104</td>
</tr>
<tr>
<td>MAD</td>
<td>0.1489</td>
<td>0.2197</td>
<td>0.1608</td>
</tr>
<tr>
<td>$n = 400$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean bias</td>
<td>0.0074</td>
<td>0.0161</td>
<td>0.0095</td>
</tr>
<tr>
<td>Med. bias</td>
<td>0.0003</td>
<td>0.0096</td>
<td>0.0015</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1318</td>
<td>0.2051</td>
<td>0.1361</td>
</tr>
<tr>
<td>MAD</td>
<td>0.1049</td>
<td>0.1620</td>
<td>0.1061</td>
</tr>
</tbody>
</table>

Table 5
Monte Carlo simulation – slope coefficients, log-linear regression – homoscedastic design

<table>
<thead>
<tr>
<th></th>
<th>MRE</th>
<th>QRE</th>
<th>QAD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_2/\beta_1$</td>
<td>$\beta_3/\beta_1$</td>
<td>$\beta_2/\beta_1$</td>
</tr>
<tr>
<td>$n = 100$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean bias</td>
<td>0.3223</td>
<td>— 0.0532</td>
<td>— 0.0180</td>
</tr>
<tr>
<td>Med. bias</td>
<td>— 0.0206</td>
<td>— 0.0210</td>
<td>— 0.0152</td>
</tr>
<tr>
<td>RMSE</td>
<td>7.8618</td>
<td>16.6937</td>
<td>0.6349</td>
</tr>
<tr>
<td>MAD</td>
<td>1.4732</td>
<td>3.0838</td>
<td>0.1959</td>
</tr>
<tr>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean bias</td>
<td>1.6835</td>
<td>— 2.0966</td>
<td>0.0327</td>
</tr>
<tr>
<td>Med. bias</td>
<td>0.0100</td>
<td>— 0.0003</td>
<td>0.0052</td>
</tr>
<tr>
<td>RMSE</td>
<td>17.3916</td>
<td>24.5651</td>
<td>0.3655</td>
</tr>
<tr>
<td>MAD</td>
<td>2.5508</td>
<td>4.2113</td>
<td>0.0992</td>
</tr>
<tr>
<td>$n = 400$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean bias</td>
<td>16.6481</td>
<td>— 19.3289</td>
<td>0.0249</td>
</tr>
<tr>
<td>Med. bias</td>
<td>0.0171</td>
<td>— 0.0285</td>
<td>0.0041</td>
</tr>
<tr>
<td>RMSE</td>
<td>209.6651</td>
<td>185.3507</td>
<td>0.2711</td>
</tr>
<tr>
<td>MAD</td>
<td>19.8315</td>
<td>23.6785</td>
<td>0.0711</td>
</tr>
</tbody>
</table>
Table 6
Monte Carlo simulation – variances, linear regression

<table>
<thead>
<tr>
<th></th>
<th>Homosced.</th>
<th>Heterosced.</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_2/\beta_1$</td>
<td>$\beta_3/\beta_1$</td>
<td>$\beta_2/\beta_1$</td>
<td>$\beta_3/\beta_1$</td>
</tr>
<tr>
<td>$n = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean bias</td>
<td>$-0.2842$</td>
<td>$-0.2272$</td>
<td>$-0.1604$</td>
<td>$-0.1405$</td>
</tr>
<tr>
<td>Med. bias</td>
<td>$-0.2986$</td>
<td>$-0.2124$</td>
<td>$-0.1114$</td>
<td>$-0.1252$</td>
</tr>
<tr>
<td>RMSE</td>
<td>$0.4431$</td>
<td>$0.4243$</td>
<td>$0.5454$</td>
<td>$0.5284$</td>
</tr>
<tr>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean bias</td>
<td>$-0.2549$</td>
<td>$-0.2050$</td>
<td>$-0.1388$</td>
<td>$-0.0918$</td>
</tr>
<tr>
<td>Med. bias</td>
<td>$-0.2459$</td>
<td>$-0.1507$</td>
<td>$-0.2146$</td>
<td>$-0.1840$</td>
</tr>
<tr>
<td>RMSE</td>
<td>$0.3586$</td>
<td>$0.3371$</td>
<td>$0.4392$</td>
<td>$0.3914$</td>
</tr>
<tr>
<td>$n = 400$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean bias</td>
<td>$-0.1504$</td>
<td>$-0.1276$</td>
<td>$-0.1773$</td>
<td>$-0.1111$</td>
</tr>
<tr>
<td>Med. bias</td>
<td>$-0.1057$</td>
<td>$-0.0756$</td>
<td>$-0.1438$</td>
<td>$-0.1082$</td>
</tr>
<tr>
<td>RMSE</td>
<td>$0.2514$</td>
<td>$0.2217$</td>
<td>$0.3016$</td>
<td>$0.2515$</td>
</tr>
</tbody>
</table>

MRE, and only $\beta_2/\beta_1 = 1$ for the QAD estimator, since this approach does not apply to discrete covariates.

Tables 6 and 7 report mean bias, median bias and RMSE\(^9\) for the ‘model primitives’ estimator of the diagonal elements of the limiting covariance matrix of the QRE for each of the four designs.

The simulation study was performed in GAUSS. For the QRE and MRE, the objective function was maximized using the Nelder–Meade simplex algorithm to determine local maxima. Rescaled least squares and least absolute deviations estimators were used as endpoints of a diagonal to construct a rectangle on which a grid of 100 starting values were used. The estimators were selected as the global maxima among the local maxima obtained by the different starting values.

For the QRE and QAD estimators, a local linear model was fit in the first stage with $\alpha = 0.5$ using the Barrodale Roberts algorithm. The bandwidth was selected by a procedure similar to the rule of thumb approach discussed on p. 202 of Fan and Gijbels (1996). These were of the form $c_{\text{qre}} n^{-1/6.5}$ and $c_{\text{quad}} n^{-1/9.5}$, where the rates are consistent with Theorem 3.2 for the QRE and the assumptions in Chaudhuri et al. (1997) for the QAD. The constant was proportional to the asymptotic variance of the local linear estimator. For the QRE this

\(^9\) Results reported are standardized in the sense that statistics are divided by the true values of the variance.
was of the form $1.35(\alpha(1 - \alpha)f_{K_0}(0)^{-2})^{1/6.5}$, where 1.35 is the constant based on functions of moments of the uniform kernel function listed on p. 67 of Fan and Gijbels (1996), and $f_{K_0}(0)$ is a kernel estimator of the residual density function (evaluated at 0) obtained from a preliminary global quadratic fit.

For estimation of the limiting covariance matrix, the conditional expectation and index density functions were estimated using the Nadaraya-Watson estimator with bandwidth set to $n^{-1/5}$. $H(\cdot)$ was estimated using a local linear fit and the bandwidth $n^{-1/5}$. The conditional density of the residual was estimated using the approach detailed in Section 6.2.3 of Fan and Gijbels (1996).

Tables 2 and 3 report the results for the linear regression model. As the table indicates, for the homoscedastic design, the monotone rank estimator performs extremely well. Even for small samples, it yields a small median and mean bias, and has a small mean square estimator. The QRE also performs quite well, though the QRE exhibits larger biases than in the homoscedastic case.

When heteroscedasticity is introduced, the performance of the MRE still performs well, as the conditional mean is still monotonic in the index. The QRE and QAD also perform quite well, though the QRE exhibits larger biases than in the homoscedastic case.
Tables 4 and 5 list the results for the log-linear regression model. The qualitative results for the homoscedastic case do not change. For the heteroscedastic design, the performance of the MRE rapidly deteriorates, as predicted by the theory, since the conditional mean is no longer monotonic in the index. As the sample size increases, the MRE gets further away from the truth, and the root mean square error increases with the sample size. In contrast, the QRE performs quite well, as the conditional median is still increasing in the index (despite the fact that the scale function is decreasing in the index). Its RMSEs appear to be decreasing at the parametric rate. The QAD also performs well in both homoscedastic and heteroscedastic designs, and outperforms the QRE in terms of RMSE in the heteroscedastic design.

Tables 6 and 7 report results for the estimated variances of the two coefficients. The results are not nearly as favorable as for the slope coefficients. Large biases are exhibited for sample sizes of 100 and 200, and it is not always the case that the bias shrinks with the sample size. Nonetheless, the fact that the RMSE’s are shrinking with the sample size for all designs is somewhat encouraging, though more work is needed in deriving a data driven approach to the selection of smoothing parameters which works better in practice.

Overall, the simulation results indicate that the QRE and QAD performs better than the MRE for heteroscedastic models where there is no additive separability between the error term and the conditional mean function, yet a penalty (in terms of RMSE) is incurred otherwise.

6. Summary and concluding remarks

This paper introduces a procedure for estimating the parameters of a general class of models. An important special case of this class is the transformation model with heteroscedastic errors. The estimator exploits the monotonicity of the conditional quantile function, which permits weaker assumptions than approaches based on the smoothness of quantile function. The procedure was shown to have desirable asymptotic properties and performed reasonably well in the simulation studies conducted.

The results of this paper suggest areas for further research. First, while the estimator was modified to allow for censored data, the asymptotic properties for the case when the dependent variable is binary or ordered remains to be developed. Second, the estimator proposed in this paper immediately suggests the construction of hypothesis tests for heteroscedasticity in the context of transformation models. It seems plausible that the tests for heteroscedasticity proposed by Koenker and Bassett (1982) and Powell (1986) could be adapted for the transformation model by using the estimator proposed in this paper, and determining the limiting distribution of such a statistic could prove useful. Finally, the computational costs of the first stage estimation when the order of
smoothness gets large, suggests methods for estimating the conditional quantile function which do not require linear programming methods to compute, such as a jackknife based approach, where optimal sorting algorithms could be employed. The asymptotic properties of such an approach would be useful for this and other semiparametric estimators requiring preliminary nonparametric estimation of the quantile function.

Acknowledgements

I am grateful to J. Powell, B. Honoré, S. Chen, two anonymous referees, an associate editor, and seminar participants at Boston College/Boston University, Johns Hopkins, Queens, Rutgers, and the Universities of British Columbia, Michigan, Rochester, Virginia, and Western Ontario for helpful comments.

Appendix

In this section, we adopt notation which serves to keep expressions in a more convenient form. Specifically, we let $q_i, q_i, \tau_i, \tau_{1i}(\cdot), \tau_{2i}(\cdot), f_u(\cdot), \tilde{f}_u(\cdot), \gamma_{kl}(\cdot)$ denote $q^a(x_i), q^b(x_i), \tau(x_i), \tau_1(x_i, \cdot), \tau_2(x_i, \cdot), f_u(x_i, \cdot), \tilde{f}_u(x_i, \cdot), \gamma_{kl}(x_i, \cdot)$, respectively. For the nonparametric estimation procedure, we let $C_n, n, \sum_{j \neq i} I[x_j \in C_n]$. As before, for any matrix $A$ with elements $a_{ij}$, we let $\|A\|_2$ denote $(\sum_{i,j} a_{ij}^2)^{1/2}$.

Proof of Theorem 3.1. Define

$$G(\theta) = E[\tau_i q_i (I[x_i^j \beta(\theta) > x_i^j \beta(\theta_0)] - I[x_i^j \beta(\theta_0) > x_i^j \beta(\theta_0)])]$$

and

$$G_n(\theta) = \frac{1}{n(n-1)} \sum_{i \neq j} \tau_i q_i (I[x_i^j \beta(\theta) > x_i^j \beta(\theta_0)] - I[x_i^j \beta_0 > x_i^j \beta_0]).$$

We note that $\hat{\theta}$ maximizes $G_n(\theta)$. Given Assumptions A, B2, T, which ensure the continuity of $G(\cdot)$, it will suffice to show, by Theorem 4.1.1 of Amemiya (1985), the following conditions are satisfied:

(i) $G_n(\theta)$ converges uniformly in probability to $G(\theta)$.
(ii) $G(\theta)$ is uniquely maximized at $\theta_0$.

To establish condition (i), we write $G_n(\theta)$ as

$$G_n(\theta) = \frac{1}{n(n-1)} \sum_{i \neq j} \tau_i q_i (I[x_i^j \beta(\theta) > x_i^j \beta(\theta_0)] - I[x_i^j \beta_0 > x_i^j \beta_0]) + R_n.$$

(A.1)
The global rates of convergence established in Chaudhuri (1991b) and Chaudhuri et al. (1997) immediately imply that
\[
\sup_{\theta \in \Theta} R_n = o_p(1).
\]
It follows from Corollary 7 of Sherman (1994b) that
\[
\sup_{\theta \in \Theta} \frac{1}{n(n-1)} \sum_{i \neq j} \tau_i \eta_i(I[x_i'\beta(\theta) > x_j'\beta(\theta)] - I[x_i'\beta_0 > x_j'\beta_0])
\]
\[
- G(\theta) = O_p(n^{-1/2})
\] (A.2)
establishing (i). To establish (ii), note that, by expressing \( q(\theta) = H(x'\beta_0) \), we can write \( G(\theta) \) as
\[
\frac{1}{2} E[\tau_i H(x_i'\beta_0)]I[x_i'\beta > x_j'\beta] + \tau_j H(x_j'\beta)I[x_i'\beta < x_j'\beta]
\]
\[
- \max(\tau_i H(x_i'\beta(\theta_0)), \tau_j H(x_j'\beta(\theta_0)))
\]
The rest of the argument is almost identical to that used in the proof of Theorem 1 in Cavanagh and Sherman (1998). The only difference is that Assumption B4.3 replaces their requirement that the \( d \)th regressor has support on \( \mathbb{R} \).

**Proof of Theorem 3.2.** For the proof, without loss of generality, we set \( \theta_0 = 0 \). Since discontinuities in the objective function prevent standard Taylor series methods from applying, our strategy for attaining these results is based on approximating the objective functions with quadratic functions. Specifically, we establish that the approximation errors converge uniformly at a sufficiently fast rate, so that following two results, discussed in Sherman (1994a), can be applied to establish root-\( n \) consistency and asymptotic normality respectively:

**Lemma A.1.** Suppose \( \hat{\theta} \) maximizes \( G_n(\theta) \) and \( 0 \) maximizes \( G(\theta) \). Let \( \{\delta_n\} \) and \( \{\epsilon_n\} \) be sequences of nonnegative numbers converging to 0; if
1. \( \|\hat{\theta}\| = O_p(\delta_n) \),
2. there exists a neighborhood \( \mathcal{N} \) of 0 and a positive constant \( \kappa \) for which
   \[
   G(\theta) \leq -\kappa \|\theta\|^2
   \]
   for all \( \theta \) in \( \mathcal{N} \),
3. uniformly over \( O_p(\delta_n) \) neighborhoods of 0,
   \[
   G_n(\theta) = G(\theta) + O_p(\|\theta\|/\sqrt{n}) + o_p(\|\theta\|^2) + O_p(\epsilon_n)
   \]
then
\[
\|\hat{\theta}\| = O_p(\max[\epsilon_n^{1/2}, 1/\sqrt{n}]).
\] (A.3)
This $\sqrt{n}$-consistency result can then be applied to the following result to get asymptotic normality:

**Lemma A.2.** Suppose $\hat{\theta}$ is $\sqrt{n}$-consistent, and 0 is an interior point of $\Theta$. If, uniformly over $O_p(1/\sqrt{n})$ neighborhoods of 0,

$$G_n(\theta) = \frac{1}{2} \theta' V \theta + \frac{1}{\sqrt{n}} \theta' W_n + o_p(1/n)$$

where $V$ is a negative definite matrix, and $W_n$ converges in distribution to a $N(0, \Delta)$ random vector. Then

$$\sqrt{n} \hat{\theta} \Rightarrow N(0, V^{-1} \Delta V^{-1}) \quad (A.4)$$

To verify the conditions for which these two results can apply we begin by decomposing the objective function as follows:

$$G_n(\theta) = \frac{1}{n(n-1)} \sum_{i \neq j} \tau_i q_i (I[x_i' \beta(\theta) > x_j' \beta(\theta)] - I[x_i' \beta_0 > x_j' \beta_0])$$

$$+ \frac{1}{n(n-1)} \sum_{i \neq j} \tau_i (\hat{q}_i - q_i) (I[x_i' \beta(\theta) > x_j' \beta(\theta)] - I[x_i' \beta_0 > x_j' \beta_0])$$

$$= \Gamma_n(\theta) + H_n(\theta). \quad (A.5)$$

The following lemma establishes an asymptotic expression for the first term in the above expression. Since this term does not involve the preliminary non-parametric estimator, the derivation follows from well-developed results for U-processes (see for example Sherman, 1993).

**Lemma A.3.** If the assumptions for consistency and Assumptions E1, E2 hold, then uniformly over $o_p(1)$ neighborhoods of 0,

$$\Gamma_n(\theta) = \frac{1}{2} \theta' V \theta + o(||\theta||^2) + o_p(1/n). \quad (A.6)$$

**Proof.** $\Gamma_n(\theta)$ can be decomposed as follows: (see Serfling, 1980):

$$\Gamma_n(\theta) = \frac{1}{2} E[\tau_{1i}(\theta)] + \frac{1}{n} \sum_{i=1}^{n} \tau_{1i}(\theta) - E[\tau_{1i}(\theta)] + h(x_i, x_j, \theta) \quad (A.7)$$
where \( h(\cdot, \cdot, \theta) \) is a second-order degenerate U-statistic. A Taylor series expansion \( E[\tau_{1i}(\theta)] \) around 0 yields:

\[
E[\tau_{1i}(\theta)] = E[\tau_{1i}(0)] + \theta' V_1 E[\tau_{1i}(0)] + \frac{1}{2} \theta' V_2 E[\tau_{1i}(0)] \theta + R_n
\]

where the first two terms vanish because of the normalization, and the fact that 0 maximizes the limiting objective function, respectively. By Assumption E1, the remainder term \( R_n \) is \( o(||\theta||^2) \) uniformly over \( o_p(1) \) neighborhoods of 0.

The second term in the above expression is a class of the average of mean 0 random variables indexed by \( \theta \). A mean value expansion and the dominated convergence theorem applied to this term yields

\[
\theta' \frac{1}{n} \sum_{i=1}^{n} V_1 \tau_{1i}(\theta^*) - E[V_1 \tau_{1i}(\theta^*)]
\]

where \( \theta^* \) denotes an intermediate value. Uniform central limit theorems from empirical process theory can be applied to the above average of mean 0 variables. From Lemma 2.17 of Pakes and Pollard (1989) it follows that the above expression is \( o_p(||\theta||/\sqrt{n}) \) uniformly over \( o_p(1) \) neighborhoods of 0. Finally, the second-order degenerate U-statistic \( h(x_i, x_j, \theta) \) is \( o_p(1/n) \) uniformly over \( o_p(1) \) neighborhoods of 0 by Corollary 6 in Sherman (1994b).

The second term is more difficult to approximate due to the presence of the nonparametric estimator. The following lemma, based on linear representations for the first stage local polynomial estimator, establishes that under sufficient smoothness conditions on the quantile function and an appropriately chosen bandwidth sequence, this term can be approximated by a sequence of random variables which are asymptotically normal.

**Lemma A.4.** Assume that \( p > 3d_c/2 \), \( k = \text{int}(p) \) and the bandwidth satisfies

(i) \( \sqrt{nh_n^p} \to 0 \),

(ii) \( \log n \sqrt{n^{-1}h_n^{-3d_c}} \to 0 \).

And the Assumptions A, B, T, Q, D, E.3, E.4 hold; then uniformly over \( o_p(1) \) neighborhoods of 0,

\[
H_n(\theta) = \theta' \frac{1}{n} \sum_{i=1}^{n} \delta(y_i, x_i) + o_p(||\theta||/\sqrt{n}) + o_p(||\theta||^2) + O_p(1/nh_n^d)
\]

where the \( O_p(1/nh_n^d) \) term has order \( o_p(1/n) \) uniformly over \( O_p(1/\sqrt{nh_n^d}) \) neighborhoods of 0.
Proof. The proof will be carried out in a series of steps. The first step involves ‘plugging in’ a local Bahadur representation for the conditional quantile estimator. Such representations have been developed in Chaudhuri (1991a), Chaudhuri et al. (1997) and Cavanagh (1996). We deliberately choose notation so it is as close as possible to that used in Chaudhuri et al. (1997).

Lemma A.5. Let \( z_i = (y_i, x_i)' \) and let

\[
f_n(z_i, z_j, z_k) = e^{s(A)}(p_{ni} G_{ni}^{n1})^{-1} \tau_i(z - I[y_k \leq q_k])b(h_n, x_k - x_i)I[x_k \in C_{ni}]
\]

\[
\times (I[x_i' \beta(\theta) > x_j' \beta(\theta)] - I[x_i' \beta_0 > x_j' \beta_0])
\]

(A.8)

where

(i) \( p_{ni} \equiv p_n(x_i) \) and \( p_n(x) \) is a real valued function equal to

\[
f_{x^{(d)}(x^{(d)})}^{d}\int_{[-1/2,1/2]^s} f_{x^{(d)}(x^{(d)})}^{d}(x^{(c)} + h_n t) \, dt
\]

(ii) \( e^{s(A)} \) is an \( s(A) \)-dimensional row vector whose first component is 1, and the rest are 0;

(iii) \( b(h_n, x_1 - x_2) \) is an \( s(A) \)-dimensional vector, whose components, indexed by vectors \( u \), s.t. \( [u] \leq k \), are of the form

\[
h_n^{-[u]}(x_1 - x_2)^u,
\]

(iv) \( G_{ni} \equiv G_n(x_i) \) where \( G_n(x) \) is an \( s(A) \times s(A) \) matrix, whose elements, indexed by the vectors \( u, v \), s.t. \( [u] \leq k, [v] \leq k \) are of the form

\[
\int_{[-1/2,1/2]^s} t^u t^v f_{u,x^{(d)}(x^{(d)})}^{d}(0,x^{(c)} + h_n t) \, dt
\]

\[
\int_{[-1/2,1/2]^s} f_{x^{(d)}(x^{(d)})}^{d}(x^{(c)} + h_n t) \, dt.
\]

Then under the bandwidth conditions in Lemma A.4, and Assumptions B,T,E, \( H_n(\theta) \) can be written as

\[
\frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} f_n(z_i, z_j, z_k) + R_n
\]

(A.9)

where

\[
R_n = o_p(||\theta||/\sqrt{n}) + o_p(1/n)
\]

(A.10)

uniformly over \( o_p(1) \) neighborhoods of 0.

Proof. It follows directly from Lemma 4.1 of Chaudhuri et al. (1997), the Bahadur representation for the local polynomial quantile estimator can be
written as the first component of the $s(A)$-dimensional vector:

$$
\tilde{q}_i - q_i = (N_{ni} \Phi_{ni})^{-1} \sum_{j=1, j \neq i}^n b(h_n, x_j - x_i) (x - I[y_j \leq \tilde{q}_{ij}]) I[x_j \in C_{ni}] + R_{ni}
$$

(A.11)

where $N_{ni} = \# \{ j : 1 \leq j \leq n, j \neq i, x_j \in C_{ni} \}$ and $\tilde{q}_{ij}$ is a $k$th-order Taylor approximation of $q_i$ around $q_j$.

The rest of the proof now parallels that in Chaudhuri et al. (1997). The only difference is that remainder terms need be shown to be $O_p(||\theta||/\sqrt{n}) + o_p(1/n)$ uniformly over $o_p(1)$ neighborhoods of 0. We will therefore just sketch the steps involved. First, let $D_{ni} = I[c_1 n h_n^d \leq N_{ni} \leq c_2 n h_n^d]$, where $c_1 < c_2$ are positive constants. By Bernstein’s inequality, and the assumptions on the bandwidth, we have for some $c_1$ and $c_2$,

$$
\max_{1 \leq i \leq n} |D_{ni}| = o_p(n^{-1/2})
$$

(A.12)

with $D_{ni}$ denoting the complement of the event $D_{ni}$. The restrictions on the bandwidth also imply that (see Lemma 4.1 in Chaudhuri et al., 1997)

$$
\max_{1 \leq i \leq n} |R_{ni}| = o_p(n^{-1/2}).
$$

(A.13)

These two results, combined with (A.11) yield

$$
\frac{1}{n(n-1)} \sum_{i \neq j} \tau_i (\tilde{q}_i - q_i) (I[x_i^g \beta(\theta) > x_j^g \beta(\theta)] - I[x_i^g \beta_0 > x_j^g \beta_0])
$$

$$
= \frac{1}{n(n-1)} \sum_{i \neq j} \tau_i I[x_j^g \beta(\theta) > x_j^g \beta(\theta)] - I[x_j^g \beta_0 > x_j^g \beta_0] N_{ni}^{-1}
$$

(A.14)

$$
\times \sum_{k=1, k \neq i, j}^n (x - I[y_k \leq \tilde{q}_{ki}]) I[x_k \in C_{ni}] I[D_{ni}]
$$

$$
+ o_p(||\theta||/\sqrt{n}) + o_p(1/n)
$$

(A.15)

uniformly over $o_p(1)$ neighborhoods of 0. The next step is to replace $\tilde{q}_{ki}$ with $q_k$ and show the remainder stays the same order of magnitude. Note that

$$
|I[y_k \leq \tilde{q}_{ki}] - I[y_k \leq q_k]| \leq |u_k| \leq |\tilde{q}_{ki} - q_k|.
$$

So if $x_k \in C_{ni}$, the difference in indicators is bounded above by

$$
I[|u_k| \leq c_3 h_n^d]
$$

for some positive constant $c_3$. By the bandwidth restrictions, Bernstein’s inequality implies that

$$
\max_{1 \leq k \leq n, k \neq i} |I[y_k \leq q_k] - I[y_k \leq \tilde{q}_{ki}]| I[x_k \in C_{ni}] = o_p(n^{-1/2}).
$$

(A.16)
So by the same argument as in the previous step, the remainder as a result of this replacement is \( o_p(||\theta||/\sqrt{n}) + o_p(1/n) \) uniformly over \( o_p(1) \) neighborhoods of 0. The final step in the proof is to replace \( N_n \) with \((n - 2)p_{ni}\) and remove \( I[D_{ni}]\). By Lemma 4.2 of Chaudhuri et al. (1997), and Bernstein’s inequality, we have

\[
\max_{1 \leq i \leq n} \left( \frac{1}{N_{ni}} - \frac{1}{(n - 2)p_{ni}} \right) I[D_{ni}] = O(n^{-3/2}h_n^{-2d} \sqrt{\log n}). \tag{A.17}
\]

and

\[
\max_{1 \leq i \leq n} \sum_{k = 1, k \neq i, j}^{n} (\alpha - I[y_k \leq q_k])I[x_k \in C_{ni}]I[D_{ni}] = O(\sqrt{nh_n^{d} \log n}). \tag{A.18}
\]

So by the bandwidth restrictions, the remainder term for these replacements is also \( o_p(||\theta||/\sqrt{n}) + o_p(1/n) \) uniformly over \( o_p(1) \) neighborhoods of 0.

The remaining steps involve working with Hoeffding decomposition of the third order U-process:

\[
\frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} f_n(z_i, z_j, z_k).
\]

Note by the quantile restriction and the law of iterated expectations, we have

\[
E[f_n(z_1, z_2, z_3)] = E[f_n(z_1, z_2, z_3)|z_1] = E[f_n(z_1, z_2, z_3)|z_2] = 0. \tag{A.19}
\]

Thus the usual Hoeffding decomposition leaves us with

\[
\frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} f_n(z_i, z_j, z_k)
\]

\[
= \frac{1}{n} \sum_{i = 1}^{n} E[f_n(z_1, z_2, z_3)|z_3 = z_i] + U_n^2 f_n^2 + U_n^3 f_n^3 \tag{A.20}
\]

where \( U_n^2 f_n^2 \) and \( U_n^3 f_n^3 \) are degenerate U processes of order 2 and 3 respectively. These two terms can be disposed with in a similar fashion as in Sherman (1994a). Following his approach, we have the following result:

**Lemma A.6.**

\[
U_n^2 f_n^2 = o_p(1/n)
\]

uniformly over \( O_p(1/\sqrt{nh_n^{d}}) \) neighborhoods of 0.

\[
U_n^3 f_n^3 = o_p(1/n)
\]

uniformly over \( o_p(1) \) neighborhoods of 0.

**Proof.** We follow the 2 step approach of Sherman (1994a); let \( \tilde{f}_n^2(\cdot, \cdot, \cdot, \theta) = h_n^2 f_n^2(\cdot, \cdot, \theta) \) and \( \tilde{F}_n = \{ \tilde{f}_n^2(\cdot, \cdot, \theta); \theta \in \Theta \} \). It can be easily established
that $\mathcal{F}_n$ is Euclidean for a constant envelope. By Theorem 3 of Sherman (1994a), we have, uniformly over $\Theta$,

$$U_n^2 f_n^2(\cdot, \cdot, \theta) = \frac{1}{h_n^d} O_p(1/n)$$  \hspace{1cm} (A.21)

Now let $\delta_n = O(1/\sqrt{nh_n^d})$, and let $\Theta_n = \{ \theta \in \Theta: ||\theta|| \leq \delta_n \}$. Note that $||\nabla_1 f_n(\cdot, \cdot, \theta)||$ is bounded by a constant. By a Taylor series expansion about 0, we get that $||\mathcal{F}_n^+(\cdot, \cdot, \theta)||$ is bounded by a multiple of $\delta_n$. Another application of Theorem 3 in Sherman (1994a) gives us that, uniformly over $O_p(1/\sqrt{nh_n^d})$ neighborhoods of 0,

$$U_n^2 f_n^2(\cdot, \cdot, \theta) = \frac{1}{h_n^d} O_p(\delta_n/n)$$  \hspace{1cm} (A.22)

and the term on the right-hand side is $o_p(1/n)$ by the restrictions on the bandwidth. Thus we have shown that $U_n^2 f_n^2 = o_p(1/n)$ uniformly over $O_p(1/\sqrt{nh_n^d})$ neighborhoods of 0 which is the desired result. Because the third-order degenerate U process converges at a (sufficiently) faster rate, this term can be handled in one step. Note that $U_n^2 f_n^3$ is bounded by $1/h_n^d$. Let

$$\tilde{f}_n^3(\cdot, \theta) = h_n^d f_n^3(\cdot, \theta)$$

and define the random process:

$$\mathcal{F}_n^3 = \{ \tilde{f}_n^3(\cdot, \theta); \theta \in \Theta \}.$$

Then by Theorem 3 of Sherman (1994a), uniformly over $\theta$,

$$U_n^2 f_n^3(\cdot, \theta) = \frac{1}{h_n^d} O_p(1/n^3/2) = o_p(1/n)$$  \hspace{1cm} (A.23)

where the last equality follows by the assumptions on the bandwidth. \hfill \Box

The final step involves showing that the first term on the right-hand side of (A.20) behaves asymptotically like the sum of i.i.d zero mean random variables. We next show the following representation:

**Lemma A.7.** Suppose Assumptions B, T, D, E3, E4 hold, and $h_n \to 0$. Then uniformly over $o_p(1)$ neighborhoods of 0, we have

$$\frac{1}{n} \sum_{i=1}^{n} E[f_n(z_1, z_2, z_3) | z_3 = z_i]$$

$$= \theta' \frac{1}{n} \sum_{i=1}^{n} \tau_i f_n(0)^{-1} (U[y_i \leq q_i] - \pi) \nabla_1 \tau_{2i}(0) + o_p(||\theta||/\sqrt{n}) + o_p(||\theta||^2).$$  \hspace{1cm} (A.24)
**Proof.** Note that \( \mathbb{E}[f_n(z_1, z_2, z_3)|z_3 = z_i] \) can be expanded as the sum of the two terms:
\[
\theta'(z - I[y_i \leq q_i])
\int e_1^{(d)}(h_n, x_i - x)(p_n(x)\mathcal{G}_n(x))^{-1} \tau(x)\nabla \tau_2(x,0)I[x_i \in C_n(x)] \, dF_X(x)
\] (A.25)
\[
\frac{1}{2} \theta'(z - I[y_i \leq q_i])
\int e_1^{(d)}(h_n, x_i - x)(p_n(x)\mathcal{G}_n(x))^{-1} \tau(x)\nabla \tau_2(x, \tilde{\theta})I[x_i \in C_n(x)] \, dF_X(x)\theta
\] (A.26)

where \( \tilde{\theta} \) denotes an intermediate value. Denote the integral in (A.25) by \( z_n \). It will be shown that
\[
\frac{1}{n} \sum_{i=1}^{n} (z_i - I[y_i \leq q_i])(z_n - \tau_{\mathcal{I}u^{-1}}(0)\nabla \tau_2(0)) = o_p(n^{-1/2}).
\] (A.27)

A similar argument can be used to show that (A.26) is \( o_p(\|\theta\|^2) \) uniformly over \( o_p(1) \) neighborhoods of 0. To establish (A.27), first note that each term in the sum has expected value 0 by the conditional quantile restriction; define the \( s(A) \times s(A) \) matrix \( Q \) by
\[
\int_{[-1/2,1/2]^6} b(1,t)b(1,t)' \, dt
\]
and define \( z'_n \) by replacing the term \( (p_n(x)\mathcal{G}_n(x))^{-1} \) in \( z_n \) with \( h_n^{-d_c} (f_{u,X_i'|X_i=x_i}(0,x^{(c)})Q)^{-1} \) and let
\[
r_n(x) = (p_n(x)\mathcal{G}_n(x))^{-1} - h_n^{-d_c} (f_{u,X_i'|X_i=x_i}(0,x^{(c)})Q)^{-1}
\]
with the change of variables \( t = (x_i^{(c)} - x^{(c)})/h_n \) we have
\[
z'_n = \int_{[-1/2,1/2]^6} e_1^{(d)}Q^{-1}b(1,t)\tau(x_i^{(c)} - th_n, x_i^{(d)})f_{u,X_i'|X_i=x_i}(0,x_i^{(c)} - th_n)
\times f_{X_i'|X_i=x_i}(x_i^{(c)} - th_n)\nabla \tau_2((x_i^{(c)} - th_n, x_i^{(d)}),0) \, dt
\] (A.28)

by the dominated convergence theorem and continuity of the functions \( \tau(\cdot) \) and \( \nabla \tau_2(\cdot, \cdot) \), we have
\[
\lim_{n \to \infty} z'_n = \tau_{\mathcal{I}u^{-1}}(0)\nabla \tau_2(0) \int_{[-1/2,1/2]^6} e_1^{(d)}Q^{-1}b(1,t) \, dt
\]
\[
= \tau_{\mathcal{I}u^{-1}}(0)\nabla \tau_2(0).
\] (A.29)
The law of iterated expectations and another application of the dominated convergence theorem implies
\[
\lim_{n \to \infty} \mathbb{E}[(\alpha - I[y_i \leq q_i])^2(z'_n - \tau_if_{ui}^{-1}(0)\nabla \tau_{2i}(0))^2] = 0.
\]

It thus follows by Chebyshev’s inequality that
\[
\frac{1}{n} \sum_{i=1}^{n} (\alpha - I[y_i \leq q_i])(z'_n - \tau_if_{ui}^{-1}(0)\nabla \tau_{2i}(0)) = o_p(n^{-1/2}).
\] (A.30)

Next note that by Lemma 4.2(b) in Chaudhuri et al. (1997), \( \mathbb{E}[(r_n(x_i))^2] = O(h_n^{2\gamma}) \) for some \( \gamma \in (0,1] \), so by the same change of variables and dominated convergence argument, it follows that
\[
\frac{1}{n} \sum_{i=1}^{n} (\alpha - I[y_i \leq q_i]) \int r_n(x)I[x_i \in C_n(x)]\nabla \tau_{2i}(0)b(h_n, x_i - x) dF_X(x) = o_p(n^{-1/2}).
\] (A.31)

This and (A.30) establish the desired result. \( \square \)

The approximations in Lemmas A.3 and A.4, and approximation rates characterized in Lemmas A.1 and A.2 enable us to derive the parametric rate of convergence and limiting distribution of Theorem 3.2. \( \square \)

**Proof of Theorem 3.3.** The proof of this theorem relies heavily on two results. The first is a uniformity result for averages of classes of mean zero random variables. A proof can be found in Sherman (1994b):

**Lemma A.8.** Let \( \{g(\cdot, \theta) : \theta \in \Theta \} \) be a class of mean zero functions which is Euclidean with respect to constant envelope. Then
\[
\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} g(z_i, \theta) \right| = O_p(n^{-1/2}).
\]

The second is a global rate of convergence for the local polynomial estimator, established in Chaudhuri (1991b) and Chaudhuri et al. (1997):

**Lemma A.9.** Under Assumptions B, D, T and the bandwidth conditions in Lemma A.4,
\[
\max_{1 \leq i \leq n} \tau_i |\hat{q}_i - q_i| = o_p(n^{-1/4}).
\]
These two results imply that
\[
\hat{V}_{kl} = \frac{1}{n} \sum_{i=1}^{n} \gamma_{kl}(\hat{\theta}) + o_p(\varepsilon_n^{-2} n^{-1/4}) + O_p(\varepsilon_n^{-2} n^{-1/2})
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} \gamma_{kl}(\hat{\theta}) + o_p(1).
\]

Another application of Lemma A.8, and the differentiability assumptions on \( \tau_1 \) thus imply that \( \hat{V}_{kl} \overset{p}{\to} V_{kl} \).

To establish the consistency of \( \hat{\Delta} \), we first define
\[
\delta_i = \sqrt{(1 - x) f_{ui}^{-1}(0) \nabla_1 \tau_{2i}(0)}.
\]

By the law of large numbers and the Cauchy Schwartz inequality it will suffice to show that
\[
\frac{1}{n} \sum_{i=1}^{n} \|\delta_i - \delta_i\|^2 = o_p(1). \tag{A.32}
\]

We establish the above result by showing the following:
\[
\frac{1}{n} \sum_{i=1}^{n} \tau_i \|\hat{f}_{ui}^{-1}(0) - f_{ui}^{-1}(0)\|^2 = o_p(1), \tag{A.33}
\]
\[
\frac{1}{n} \sum_{i=1}^{n} \|\nabla_1 \hat{\tau}_{2i}(\hat{\theta}) - \nabla_1 \tau_{2i}(0)\|^2 = o_p(1). \tag{A.34}
\]

To establish (A.33), we note that by replacing \( \hat{f}_{ui}^{-1}(0) \) with the infeasible kernel estimator:
\[
\hat{f}_{ui}^{-1}(0) = \frac{1}{h_n} \sum_{j \neq i} K_1 \left( \frac{u_j}{h_n} \right) K_2 \left( \frac{x_j - x_i}{h_n^d} \right)
\]
\[
\sum_{i \neq j} K_2 \left( \frac{x_j - x_i}{h_n^d} \right)
\]

the resulting remainder is \( o_p(1) \) by the differentiability of the kernel function, the restrictions on the smoothing parameter \( h_n \) and Lemma A.9. It follows immediately from uniform rates of convergence over compact sets for the Nadaraya-Watson estimator (see, for example, Bierens, 1987) that
\[
\frac{1}{n} \sum_{i=1}^{n} \|\hat{f}_{ui}^{-1}(0) - f_{ui}^{-1}(0)\|^2 = o_p(1) \tag{A.35}
\]
To establish (A.34), we first replace \( q_2 \) with \( q^2 \) in the numerical derivative, and show the remainder term is negligible. For this it will suffice to show that

\[
\sup_{\theta \in \Theta} \left| \frac{1}{n n_1} \sum_{i=1}^{n} \hat{\tau}_{2i}(\theta) - \tau_{2i}(\theta) \right| = o_p(1). \tag{A.36}
\]

Note for any \( \theta \in \Theta \), the left-hand side in the above expression is

\[
\frac{1}{n n(n-1)} \sum_{i \neq j} I[x_i^T \beta(\theta) > x_j^T \beta(\theta)] - \tau_{2i}(\theta).
\]

The term inside the summation is mean 0. Thus (A.36) holds by Corollary 7 in Sherman (1994b) and the fact assumption that \( \sqrt{n} \to \infty \).

By Lemma A.8 and the restrictions on the smoothing parameter \( v_n \) we have that

\[
\frac{1}{n} \sum_{i=1}^{n} \| v_n^{-1}(\tau_n(x_i, \hat{\theta} + 1 v_n) - \tau_n(x_i, \hat{\theta})) - \nabla_{1} \tau_{2i}(0) \|^2 = o_p(1) \tag{A.37}
\]

where 1 here denotes a \( d - 1 \) dimensional vector of 1’s. The differentiability of \( \tau_{2i}(\cdot) \), the root-\( n \) consistency of \( \hat{\theta} \), and the restrictions on \( v_n \) imply that the left-hand side of the above equation can be replaced with

\[
\frac{1}{n} \sum_{i=1}^{n} \| v_n^{-1}(\tau_{2i}(1 v_n) - \tau_{2i}(0)) - \nabla_{1} \tau_{2i}(0) \|^2.
\]

By a weak law of large numbers and the dominated convergence theorem, this term is \( o_p(1) \), establishing (A.34) and hence proving that \( \hat{\Delta} \xrightarrow{p} \Delta \). \( \square \)

**Proof of Theorem 4.1.** Noting that the linear representation and global rates of convergence established in Chaudhuri (1991a,b) do not apply to censored data, we first establish the following uniform result in a neighborhood of the censoring point. The proof is in Chen and Khan (1999).

**Lemma A.10.** Let \( \mathcal{X}_c \) denote the set

\[
\{ x_i \in \mathcal{X}, q_i < c/2 \}
\]

where \( \mathcal{X} \) denotes the support of \( \tau_i \), and let \( A_n \) denote the event

\[
\{ \hat{q}_i \geq c \text{ for all } x_i \in \mathcal{X}_c \}
\]

then under Assumptions B, T, D there exists constants \( C_1, C_2 \) such that

\[
P(A_n) \leq C_1 e^{-c_2 n h c}.
\]

We now sketch the proof for the asymptotic properties of the estimator in the presence of censored data. Consistency follows along the same lines as before to
establish asymptotic normality we take a second order Taylor series expansion of the function \( w(\cdot) \) in the objective function, yielding

\[
\frac{1}{n(n-1)} \sum_{i \neq j} \tau_i \omega_i I[x'_i \beta(\theta) > x'_j \beta(\theta)] - I[x'_i \beta_0 > x'_j \beta_0] \\
+ \frac{1}{n(n-1)} \sum_{i \neq j} \tau_i \omega'_i (\hat{q}_i - q_i) I[x'_i \beta(\theta) > x'_j \beta(\theta)] - I[x'_i \beta_0 > x'_j \beta_0] \\
+ \frac{1}{n(n-1)} \sum_{i \neq j} \tau_i \omega''_i (\hat{q}_i - q_i)^2 I[x'_i \beta(\theta) > x'_j \beta(\theta)] - I[x'_i \beta_0 > x'_j \beta_0]
\]

where \( \omega_i, \omega'_i \) denote \( \omega(q_i), \omega'(q_i) \) respectively, and \( \omega''_i \) denotes \( \omega''(\cdot) \) evaluated at an intermediate value. The first two terms can be dealt with in the same way as before. The third term can be decomposed as follows:

\[
\frac{1}{n(n-1)} \sum_{i \neq j} \omega''(\hat{q}_i - q_i)^2 I[q_i > c/2] \\
+ \frac{1}{n(n-1)} \sum_{i \neq j} \omega''(\hat{q}_i - q_i)^2 I[q_i \leq c/2].
\]

The first term only deals with values where \( q_i > c/2 \) and since the residual has a positive density in a neighborhood of 0, the global rates of convergence discussed in Chaudhuri (1991b) and Chaudhuri et al. (1997) imply that this term is \( o_p(\sqrt{n}/n) \) uniformly in \( \theta \). The second term is bounded by a term times \( I[\hat{q}_i \geq c, q_i \leq c/2] \), so by Lemma A.10, this term will be \( o_p(1/n) \) uniformly in \( \theta \). □

References