

University of Toronto
Department of Economics



Working Paper 374

Nonstandard Estimation of Inverse Conditional
Density-Weighted Expectations

By Chuan Goh

September 30, 2009

Nonstandard Estimation of Inverse Conditional Density-Weighted Expectations

Chuan Goh*

First draft: 27 January 2009

This version: 3 April 2009

*This paper was inspired by a conversation of the author with Arthur Lewbel and was completed during an academic visit to Indiana University. The author is grateful to his hosts in the Department of Economics, Indiana University, for their hospitality and support, and also to Herman Bierens, Sung Jae Jun and Joris Pinkse for their helpful comments. The usual disclaimer applies. Address correspondence to: Department of Economics, University of Toronto, Max Gluskin House, 150 St. George St., Toronto, ON, Canada, M5S 3G7. goh@economics.utoronto.ca, <http://www.chuangoh.org>.

Abstract

This paper is concerned with the semiparametric estimation of function means that are scaled by an unknown conditional density function. Parameters of this form arise naturally in the consideration of models where interest is focused on the expected value of an integral of a conditional expectation with respect to a continuously distributed “special regressor” with unbounded support. In particular, a consistent and asymptotically normal estimator of an inverse conditional density-weighted average is proposed whose validity does not require data-dependent trimming or the subjective choice of smoothing parameters. The asymptotic normality result is also rate adaptive in the sense that it allows for the formulation of the usual Wald-type inference procedures without knowledge of the estimator’s actual rate of convergence, which depends in general on the tail behaviour of the conditional density weight. The theory developed in this paper exploits recent results of Goh and Knight (2008) concerning the behaviour of estimated regression-quantile residuals. Simulation experiments illustrating the applicability of the procedure proposed here to a semiparametric binary-choice model are suggestive of good small-sample performance.

JEL Classification: C14, C21, C24, C25

KEYWORDS: Semiparametric, identification at infinity, special regressor, rate-adaptive, regression quantile

1 Introduction

This paper is concerned with statistical inference regarding a parameter of interest taking the form of an inverse conditional density-weighted expectation. In particular, interest is assumed to be focused on an object of interest given by

$$\boldsymbol{\theta}_0 \equiv E \left[\frac{1}{f(v_i | \mathbf{x}_i)} \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0) \right], \quad (1)$$

where

$$\mathbf{Z}_i \equiv (y_i, v_i, \mathbf{x}_i^\top)^\top, \quad i = 1, \dots, n$$

denotes a sample from the corresponding population of random variates $\mathbf{Z} \equiv (y, v, \mathbf{x}^\top)^\top$, where for each i , y_i and v_i are scalar-valued and \mathbf{x}_i is d -variate. The quantity $f(v_i | \mathbf{x}_i)$ appearing in (1) denotes the density function corresponding to the conditional distribution of v_i given \mathbf{x}_i , which is taken to be absolutely continuous with respect to Lebesgue measure on \mathbb{R} . It is assumed that $f(v | \mathbf{x})$ is positive for all $v \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^d$. The object $\mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0)$ also appearing in (1) is taken to be a realization of a known \mathbb{R}^l -valued measurable function of \mathbf{Z}_i , while $\boldsymbol{\kappa}_0$ denotes an unknown m -dimensional nuisance parameter.

The estimation of parameters of the form given in (1) is generally relevant in the consideration of models where interest is directed at the expectation of an integral of a conditional mean function with respect to a continuously distributed “special regressor” with unbounded support. In particular, it is possible to equate the parameter of interest given in (1) to the expectation of the integral of the conditional mean function $E[\mathbf{w}(\mathbf{Z}, \boldsymbol{\kappa}_0) | v, \mathbf{x}] = E[\mathbf{w}(y, v, \mathbf{x}, \boldsymbol{\kappa}_0) | v, \mathbf{x}]$ with respect to v over the real line.¹ In particular, we have

$$\boldsymbol{\theta}_0 = E \left[\int_{-\infty}^{\infty} E[\mathbf{w}(y, v, \mathbf{x}, \boldsymbol{\kappa}_0) | v, \mathbf{x}] dv \right].$$

¹To see this, note that

$$\begin{aligned} E \left[\frac{1}{f(v | \mathbf{x})} \mathbf{w}(y, v, \mathbf{x}, \boldsymbol{\kappa}_0) \right] &= E \left[E \left[\frac{1}{f(v | \mathbf{x})} \mathbf{w}(y, v, \mathbf{x}, \boldsymbol{\kappa}_0) \middle| \mathbf{x} \right] \right] \\ &= E \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{f(v | \mathbf{x})} \mathbf{w}(y, v, \mathbf{x}, \boldsymbol{\kappa}_0) f(y, v | \mathbf{x}) dy dv \right] \\ &= E \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{w}(y, v, \mathbf{x}, \boldsymbol{\kappa}_0) f(y | v, \mathbf{x}) dy dv \right] \\ &= E \left[\int_{-\infty}^{\infty} E[\mathbf{w}(y, v, \mathbf{x}, \boldsymbol{\kappa}_0) | v, \mathbf{x}] dv \right]. \end{aligned}$$

As such, statistical inference regarding parameters taking the form of an inverse conditional density-weighted average as given by θ_0 above is relevant in the consideration of a large number of important models in econometrics. Notable examples include semiparametric models of qualitative choice² as well as density-weighted least squares,³ average derivatives,⁴ entropy measures of dependence⁵ and semiparametric models of willingness to pay⁶ and of consumer surplus.⁷

The form of the generic parameter of interest in (1) gives rise to at least two nontrivial considerations from the point of view of formulating suitable estimation and test procedures. The first and more fundamental is the fact that θ_0 is essentially point-identified by those points in the support of $\mathbf{Z} \equiv (y, v, \mathbf{x}^\top)^\top$ such that the inverse conditional density weight $\frac{1}{f(v|\mathbf{x})}$ is arbitrarily large. Given the positivity of $f(v|\mathbf{x})$ for all $(v, \mathbf{x}^\top)^\top \in \mathbb{R}^{1+d}$, it follows that θ_0 is determined by those points in the support of \mathbf{Z} with v taking values at the extremities of the support of its conditional distribution given \mathbf{x} . As such, parameters of the type θ_0 as given above in (1) can be generically labelled as “identified at infinity”.⁸ Parameters in this class have the generic feature of not being estimable at a parametric rate.⁹ In particular, the actual rate of convergence of any estimator of a parameter taking the form in (1) can be shown to depend strongly on the tail behaviour of the conditional density $f(v|\mathbf{x})$,¹⁰ and is as such unknown from the standpoint of empirical practice.

The second consideration for the development of estimation and test procedures regarding an inverse conditional density-weighted average is perhaps more immediately apparent—namely, it is how to deal with the unknown conditional density of v given \mathbf{x} . In this connection, it is natural to embed a suitable nonparametric kernel density estimator in a semiparametric analogue estimator having the form

$$\frac{1}{n} \sum_{i=1}^n \frac{\tau_{ni}}{\hat{f}_{in}(v_i)} \mathbf{w}(\mathbf{Z}_i, \hat{\boldsymbol{\kappa}}_n), \quad (2)$$

where τ_{ni} is a data-dependent trimming function, $\hat{\boldsymbol{\kappa}}_n$ is an estimator of the finite-

²Cf. Lewbel (1998, 2000, 2007).

³Cf. Newey and Ruud (2005).

⁴Cf. Härdle and Stoker (1989).

⁵Cf. Hong and White (2005).

⁶Cf. Lewbel (1997); McFadden (1999).

⁷Cf. Hausman and Newey (1995); Newey (1997).

⁸Cf. Chamberlain (1986); Heckman (1990).

⁹Cf. e.g., Andrews and Schafgans (1998).

¹⁰Cf. Khan and Tamer (2009).

dimensional nuisance parameter κ_0 and where

$$\hat{f}_{in}(v_i) \equiv \frac{\hat{f}_{v_i \mathbf{x}_i, n}(v_i, \mathbf{x}_i)}{\hat{f}_{\mathbf{x}_i, n}(\mathbf{x}_i)}$$

for kernel density estimators $\hat{f}_{v_i \mathbf{x}_i, n}(\cdot, \cdot)$ and $\hat{f}_{\mathbf{x}_i, n}(\cdot)$ of the joint density $f_{v_i \mathbf{x}_i}$ and marginal density $f_{\mathbf{x}_i}$, respectively.¹¹ The implementation and analysis of the large-sample behaviour of estimators having the form (2) are clearly complicated by the effects of any “rules of thumb” used to implement the kernel estimate \hat{f}_{in} and the trimming function τ_{ni} , tasks that are in turn further complicated by the fact that the rate of convergence of the estimator given in (2) is generally unknown.

This paper proposes a semiparametric estimator for an inverse conditional density-weighted average taking the general form given in (1). In addition to being consistent and asymptotically normal, the proposed estimator also affords practitioners the convenience of avoiding the need to select smoothing parameters or implement trimming functions. The asymptotic normality result presented below is rate adaptive in the sense that it permits the formulation of natural Wald-type inference procedures without the need to specify the exact rate of convergence of the estimator.¹²

The remainder of this paper is organized as follows. The next section defines the estimator and describes the procedure used to circumvent the need to embed an explicit estimator of the conditional density of $f(v|\mathbf{x})$ in the overall estimation procedure. The consistency and asymptotic normality of a general inverse conditional-density weighted average are established formally in Section 3. Section 4 provides some numerical evidence on the finite-sample behaviour of the proposed estimation procedure in the context of a semiparametric latent-variable model of binary choice. Section 5 concludes.

¹¹Examples of estimators having the form in (2) can be found in Lewbel (1998), Lewbel (2000), Honoré and Lewbel (2002), Khan and Lewbel (2007) as well as in Lewbel (2007).

¹²The estimator proposed in this paper can also be viewed as a generalization of the procedure proposed by Lewbel and Schennach (2007), which focused on the case where the parameter of interest has the form $\theta_0 = E\left[\frac{w}{f(v)}\right]$, where both w and v are scalar valued, and where the marginal density $f(\cdot)$ of v is such that θ_0 is estimable at a parametric rate. In a manner similar to the procedure proposed in this paper, the estimator of Lewbel and Schennach (2007) also does not involve any requirement to select smoothing parameters or implement trimming functions.

2 The Estimator

We again assume the existence of a random sample

$$\{\mathbf{Z}_i \equiv (y_i, v_i, \mathbf{x}_i^\top)^\top : i = 1, \dots, n\},$$

where y_i and v_i are scalar-valued and \mathbf{x}_i is d -variate.¹³ The random variables v_1, \dots, v_n are assumed to have unbounded support. It is assumed that for each i , the conditional distribution

$$F_i \equiv F(v_i | \mathbf{x}_i)$$

of v_i given \mathbf{x}_i is absolutely continuous with respect to Lebesgue measure on \mathbb{R} . For f_i denoting the density function corresponding to F_i , the parameter of interest is given by

$$\boldsymbol{\theta}_0 \equiv E \left[\frac{1}{f_i} \mathbf{w}_i \right], \quad (3)$$

where

$$\mathbf{w}_i \equiv \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0), \quad (4)$$

for some known measurable function $\mathbf{w}(\cdot, \boldsymbol{\kappa}_0) : \mathbb{R}^k \rightarrow \mathbb{R}^l$ and some unknown m -dimensional nuisance parameter $\boldsymbol{\kappa}_0$. It is proposed in this paper to estimate $\boldsymbol{\theta}_0$ using an estimator of the form

$$\hat{\boldsymbol{\theta}}_n \equiv \frac{1}{n} \sum_{i=1}^n s_{n,\tau}^*(v_i | \mathbf{x}_i) \mathbf{w}(\mathbf{Z}_i, \hat{\boldsymbol{\kappa}}_n) \quad (5)$$

where $\hat{\boldsymbol{\kappa}}_n$ is a \sqrt{n} -consistent estimator of $\boldsymbol{\kappa}_0$ and where $s_{n,\tau}^*(v_i | \mathbf{x}_i)$ denotes a ‘‘pseudo-estimator’’ such that $s_{n,\tau}^*(v_i | \mathbf{x}_i) \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa})$ converges in distribution to a random variable with mean $E \left[\frac{1}{f(v_i | \mathbf{x}_i)} \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}) \right]$ for all $\boldsymbol{\kappa}$ in an open neighbourhood of $\boldsymbol{\kappa}_0$. Details concerning the construction and asymptotic behaviour of $s_{n,\tau}^*(v_i | \mathbf{x}_i)$ are given as follows.

2.1 Inconsistent ‘‘estimation’’ of the inverse conditional density function

The estimator $\hat{\boldsymbol{\theta}}_n$ of the interest parameter $\boldsymbol{\theta}_0$ given in (1) involves a non-standard treatment of the unknown conditional density function of v_i given \mathbf{x}_i . This non-standard handling of the conditional density is bound up in the quantity denoted

¹³The sequence $\{\mathbf{x}_i\}_{i=1}^n$ is permitted to be deterministic; see Remark 2 on p. 10 below.

by $s_{n,\tau}^*(v_i|\mathbf{x}_i)$ in the expression for the estimator $\hat{\theta}_n$ given above in (5). This section of the paper describes $s_{n,\tau}^*(v_i|\mathbf{x}_i)$ and its large-sample properties.

The quantity $s_{n,\tau}^*(v_i|\mathbf{x}_i)$ exploits a number of the large-sample properties of the regression-quantile optimization problem considered by Goh and Knight (2008). In this connection, for any constant $\alpha \in (0, 1)$, let $F_{v_i|\mathbf{x}_i}^{-1}(\alpha)$ denote the conditional α -quantile of v_i given \mathbf{x}_i . When $F_{v_i|\mathbf{x}_i}^{-1}(\alpha)$ is linear, i.e., when $F_{v_i|\mathbf{x}_i}^{-1}(\alpha) = \mathbf{x}_i^\top \boldsymbol{\beta}(\alpha)$ for some d -vector $\boldsymbol{\beta}(\alpha)$, then it is customary to estimate the parameter $\boldsymbol{\beta}(\alpha)$ using the regression α -quantile estimator of Koenker and Bassett (1978). In particular, the regression α -quantile $\hat{\boldsymbol{\beta}}_n(\alpha)$ is defined to solve

$$\min_{\mathbf{b} \in \mathbb{R}^d} \sum_{i=1}^n \rho_\alpha(v_i - \mathbf{x}_i^\top \mathbf{b}), \quad (6)$$

where

$$\rho_\alpha(u) \equiv u [\alpha - 1 \{u < 0\}].$$

In general, provided certain regularity conditions are met, the regression α -quantile will be consistent for the quantity $\boldsymbol{\beta}(\alpha)$, which is defined to be the solution of

$$\min_{\mathbf{b} \in \mathbb{R}^d} \left\| \int \int \psi_\alpha(v - \mathbf{x}^\top \boldsymbol{\beta}(\alpha)) \mathbf{x} f(v|\mathbf{x}) \mu(d\mathbf{x}) dv \right\|, \quad (7)$$

where

$$\psi_\alpha(u) \equiv \alpha - 1 \{u < 0\}, \quad (8)$$

and μ is a probability measure having a non-lattice component. This is a general result that holds even when $F_{v_i|\mathbf{x}_i}^{-1}(\alpha) \neq \mathbf{x}_i^\top \boldsymbol{\beta}(\alpha)$ for any non-zero value of $\boldsymbol{\beta}(\alpha) \in \mathbb{R}^d$.¹⁴

It is well known that the regression α -quantile $\hat{\boldsymbol{\beta}}_n(\alpha)$ may be easily computed by solving a linear program.¹⁵ In particular, under certain regularity conditions on $\{(v_i, \mathbf{x}_i^\top)^\top : i = 1, \dots, n\}$,¹⁶ it is well-known that

$$v_i = \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha)$$

for exactly d elements of the set $\{(v_i, \mathbf{x}_i^\top)^\top : i = 1, \dots, n\}$. In other words, the precise value of $\hat{\boldsymbol{\beta}}_n(\alpha)$ is determined only by those observations with indices i

¹⁴Cf. Angrist et al. (2006).

¹⁵Cf. e.g., Koenker and d'Orey (1987); Koenker and Park (1996) and Koenker (2005, Chapter 6).

¹⁶Cf. e.g., Koenker (2005, §2.2.1).

such that $\hat{\epsilon}_i(\alpha) = 0$, where

$$\hat{\epsilon}_i(\alpha) \equiv v_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha),$$

i.e., the corresponding fitted regression α -quantile residual. In this connection, define the set

$$\mathcal{H}_n(\alpha) \equiv \{i : \hat{\epsilon}_i(\alpha) = 0\}.$$

The limiting behaviour of observations with indices i belonging to $\mathcal{H}_n(\alpha)$ is markedly different from those with indices lying in the complement of $\mathcal{H}_n(\alpha)$.¹⁷ Consider a constant $\tau \in (0, 1)$. For any $\alpha \in [\tau, 1 - \tau]$, it is shown that the asymptotic behaviour of the pseudo-estimator $s_{n,\tau}^*(v_i | \mathbf{x}_i)$ depends on the asymptotic behaviour of observations with indices belonging solely to the complement of $\mathcal{H}_n(\alpha)$.

In this connection, consider the asymptotic behaviour of the sequence

$$\{(\hat{\epsilon}_i(\alpha), \mathbf{x}_i^\top)^\top : i \notin \mathcal{H}_n(\alpha)\}.$$

This is bound up with the limiting behaviour of the point process

$$M_{n\alpha}(A \times B) \equiv \sum_{i \notin \mathcal{H}_n(\alpha)} 1 \{n\hat{\epsilon}_i(\alpha) \in A, \mathbf{x}_i \in B\}. \quad (9)$$

The asymptotic behaviour of $\{M_{n\alpha}\}$ is spelled out below in Lemma 1. Regularity conditions governing the validity of this result are first stated.

Assumption 1. *The conditional distributions $\{F_i\}$ of v_i given \mathbf{x}_i are absolutely continuous with respect to Lebesgue measure on \mathbb{R} with density functions*

$$f_i(v) \equiv f(v | \mathbf{x}_i),$$

where for each $i = 1, 2, \dots$, f_i is bounded away from zero and infinity in each $v \in \mathbb{R}$ and also uniformly continuous in $v \in \mathbb{R}$ uniformly over the support of \mathbf{x}_i .

Assumption 2. *For each $\alpha \in [\tau, 1 - \tau] \subset (0, 1)$, and fixed design sequence $\{\mathbf{x}_i\}$, the following hold for each $\mathbf{b} \in \mathbb{R}^d$ and for $\kappa_\alpha(\mathbf{x}) \equiv f(\mathbf{x}^\top \boldsymbol{\beta}(\alpha) | \mathbf{x})$:*

1.

$$\max_{1 \leq i \leq n} \left| f_i \left(F_{v_i | \mathbf{x}_i}^{-1}(\alpha) + \frac{1}{\sqrt{n}} \mathbf{x}_i^\top \mathbf{b} \right) - \kappa_\alpha(\mathbf{x}_i) \right| \rightarrow 0$$

as $n \rightarrow \infty$.

¹⁷Further details are spelled out in Goh and Knight (2008); Knight and Goh (2008) and Knight (2008).

2.

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[F_i \left(F_{v_i|\mathbf{x}_i}^{-1}(\alpha) + \frac{1}{\sqrt{n}} \mathbf{x}_i^\top \mathbf{b} \right) - F_i \left(F_{v_i|\mathbf{x}_i}^{-1}(\alpha) \right) \right] \mathbf{x}_i \\
&= \frac{1}{n} \sum_{i=1}^n \sqrt{n} \left[F_i \left(F_{v_i|\mathbf{x}_i}^{-1}(\alpha) + \frac{1}{\sqrt{n}} \mathbf{x}_i^\top \mathbf{b} \right) - F_i \left(F_{v_i|\mathbf{x}_i}^{-1}(\alpha) \right) \right] \mathbf{x}_i \\
&\rightarrow \mathbf{D}_1(\alpha) \mathbf{b}
\end{aligned}$$

as $n \rightarrow \infty$, where

$$\mathbf{D}_1(\alpha) \equiv \int \kappa_\alpha(\mathbf{x}) \mathbf{x} \mathbf{x}^\top \mu(d\mathbf{x}) < \infty$$

is positive-definite for a probability measure μ having a non-lattice component.

3. The design sequence $\{\mathbf{x}_i\}$ satisfies

$$\max_{1 \leq i \leq n} \frac{1}{\sqrt{n}} \|\mathbf{x}_i\| \rightarrow 0$$

as $n \rightarrow \infty$. In addition, there exists a constant $\delta > 0$ such that

$$\int \|\mathbf{x}\|^{2+\delta} \mu(d\mathbf{x}) < \infty,$$

where μ is the same non-lattice probability measure used in the definition of $\mathbf{D}_1(\alpha)$.

Remark 1. Parts 2 and 3 of Assumption 2 guarantee the continued convexity as $n \rightarrow \infty$ of the localized regression α -quantile objective function

$$\tilde{Z}_n(\mathbf{u}) \equiv \sum_{i=1}^n \left[\rho_\alpha \left(\epsilon_i(\alpha) - \frac{1}{\sqrt{n}} \mathbf{x}_i^\top \mathbf{u} \right) - \rho_\alpha(\epsilon_i(\alpha)) \right],$$

where $\epsilon_i(\alpha) \equiv v_i - \mathbf{x}_i^\top \boldsymbol{\beta}(\alpha)$, which in turn guarantees the uniqueness in large samples of the solution to the regression α -quantile minimization problem given above in (6) for all $\alpha \in [\tau, 1 - \tau]$.

Assumption 3. For μ denoting the same non-lattice probability measure appearing in Assumption 2,

$$\frac{1}{n} \sum_{i=1}^n 1\{\mathbf{x}_i \in B\} \rightarrow \mu(B)$$

for all sets B with $\mu(\partial B) \rightarrow 0$.

Remark 2. Assumption 3 ensures that in large samples the design behaves essentially like a random sample from a population with probability measure μ . This holds even when the design sequence $\{\mathbf{x}_i\}$ is deterministic.

Remark 3. Assumptions 1 and 3 jointly imply that the empirical measure Q_n given by

$$Q_n(A) \equiv \frac{1}{n} \sum_{i=1}^n 1\{(\mathbf{x}_i, v_i) \in A\}$$

converges weakly to a measure Q with

$$Q(d\mathbf{x} \times dv) = \mu(d\mathbf{x}) f(v|\mathbf{x}) dy.$$

Remark 4. Assumptions 1–3 jointly guarantee the convergence in large samples of $\hat{\beta}_n(\alpha)$ to the quantity $\beta(\alpha)$, where $\beta(\alpha)$ is the solution of equation (7) above. This convergence is also uniform for $\alpha \in [\tau, 1 - \tau]$, where $\tau \in (0, 1)$ may be made arbitrarily small, as shown by Angrist et al. (2006, Theorem 3).

Remark 5. Assumptions 1–3 also jointly imply the convergence

$$\sqrt{n} \left(\hat{\beta}_n(\alpha) - \beta(\alpha) \right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{D}_1^{-1}(\alpha) \mathbf{C}_\alpha \mathbf{D}_1^{-1}(\alpha)\right), \quad (10)$$

where

$$\mathbf{C}_\alpha \equiv \int \int \psi_\alpha^2(v - \mathbf{x}^\top \beta(\alpha)) \mathbf{x} \mathbf{x}^\top f(v|\mathbf{x}) \mu(d\mathbf{x}) dv.$$

The convergence in (10) is also known to hold uniformly on arbitrary closed subintervals $[\tau, 1 - \tau]$ of $(0, 1)$, where $\tau > 0$.¹⁸

The asymptotic distribution of the point process in (9) is given as follows.

¹⁸Further details are given in Angrist et al. (2006, Theorem 3).

Lemma 1 (Goh & Knight (2008, Lemma 2)). *Given the conditions of Assumptions 1–3, the point process $\{M_{n\alpha}\}$ for any $\alpha \in (0, 1)$ has a behaviour in large samples that is approximable by that of the point process*

$$\bar{M}_{n\alpha}(A \times B) \equiv \sum_{i=1}^n 1 \{n\epsilon_i(\alpha) \in A, \mathbf{x} \in B\},$$

where

$$\epsilon_i(\alpha) \equiv v_i - \mathbf{x}_i^\top \boldsymbol{\beta}(\alpha).$$

The process $\bar{M}_{n\alpha}$ converges in distribution with respect to the vague topology to a Poisson process M_α with mean measure

$$m_\alpha(d\epsilon, d\mathbf{x}) \equiv \lambda(d\epsilon)\mu(d\mathbf{x})\kappa_\alpha(\mathbf{x}),$$

where λ denotes Lebesgue measure and μ is the same non-lattice probability measure appearing in the statement of Assumption 2.

Proof. The proof appears in Appendix A.1. □

Once again, pick a constant $\tau \in (0, 1)$. For any $\alpha \in [\tau, 1 - \tau]$, it is clear that the conditional α -quantile of v_i given \mathbf{x}_i need not be a linear combination of the components of \mathbf{x}_i . This fact notwithstanding, it is nevertheless possible to generate a perfect linear fit to each v_i by adjusting α so as to make $(v_i, \mathbf{x}_i^\top)^\top$ basic for the corresponding regression α -quantile optimization problem.¹⁹ In other words, if one sets $\alpha = \alpha_{ni,\tau}^*$, where

$$\alpha_{ni,\tau}^* \equiv \sup \left\{ \alpha \in [\tau, 1 - \tau] : \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha) = v_i \right\}, \quad (11)$$

then $v_i = \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha_{ni,\tau}^*)$ with probability one.²⁰

In this connection, consider that for any $\alpha \in [\tau, 1 - \tau]$, the points of the limiting Poisson process M_α in Lemma 1 are given by $\{(\Gamma_k, \mathbf{X}_k^\top)^\top : k \neq 0\}$,

¹⁹In practice, this would require setting $\tau \in (0, 1)$ to small values in order to ensure the existence for each $i \in \{1, \dots, n\}$ of a quantile $\alpha_i \in [\tau, 1 - \tau]$ such that $(v_i, \mathbf{x}_i^\top)^\top$ belongs to a basic solution for the corresponding regression α_i -quantile $\hat{\boldsymbol{\beta}}_n(\alpha_i)$. It is generally appropriate to make τ smaller when dealing with larger sample sizes.

²⁰It should be noted that the supremum in the definition of $\alpha_{ni,\tau}^*$ in (11) is arbitrary in the sense that it can be replaced with any other mapping that induces a means of selecting a quantile α_i^* in the interval $[\tau, 1 - \tau]$ satisfying $\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha_i^*) = v_i$ for a given $(v_i, \mathbf{x}_i^\top)^\top$.

where $\{\mathbf{X}_k\}$ is an iid sequence with population measure μ and

$$\begin{aligned}\Gamma_k &= \sum_{j=1}^k E_{j,k}, \\ \Gamma_{-k} &= -\sum_{j=1}^k E_{-j,k},\end{aligned}$$

where, conditional on $\{\mathbf{X}_k\}$, the $\{E_{j,k}\}$ are independent exponential random variables with mean

$$\frac{1}{\kappa_\alpha(\mathbf{X}_k)} = \frac{1}{f(\mathbf{X}_k^\top \boldsymbol{\beta}(\alpha) | \mathbf{X}_k)}.$$

This fact motivates the definition of the pseudo-estimator $s_{n,\tau}^*(v_i | \mathbf{x}_i)$.

First consider a design point \mathbf{x}_i with $i \in \mathcal{H}_n(\alpha_{ni,\tau}^*)$. Then by construction,

$$\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha_{ni,\tau}^*) = v_i.$$

Arrange the non-zero regression $\alpha_{ni,\tau}^*$ -quantile residuals in order of magnitude. In particular, define indices $(i_1), \dots, (i_{n-d})$ such that

$$|\hat{\epsilon}_{(i_1)}(\alpha_{ni,\tau}^*)| \leq \dots \leq |\hat{\epsilon}_{(i_{n-d})}(\alpha_{ni,\tau}^*)|,$$

and consider the quantity

$$\hat{\epsilon}_{(i_{n-d})}(\alpha_{ni,\tau}^*) \equiv v_{(i_{n-d})} - \mathbf{x}_{(i_{n-d})}^\top \hat{\boldsymbol{\beta}}_n(\alpha_{ni,\tau}^*), \quad (12)$$

where $(v_{(i_{n-d})}, \mathbf{x}_{(i_{n-d})}^\top)^\top$ is the corresponding element of $\{(v_i, \mathbf{x}_i^\top)^\top : i = 1, \dots, n\}$.

By Angrist et al. (2006, Theorem 3), we have that the regression-quantile process $\alpha \rightarrow \hat{\boldsymbol{\beta}}_n(\alpha)$ is uniformly consistent on arbitrary closed subintervals $[\tau, 1 - \tau]$ of $(0, 1)$. From this it follows that for each observation $i \in \{1, \dots, n\}$ and sufficiently small $\tau > 0$, there exists an $\alpha_\tau^* \in [\tau, 1 - \tau]$, not necessarily unique for each observation, such that $\hat{\boldsymbol{\beta}}_n(\alpha_{ni,\tau}^*) \xrightarrow{P} \boldsymbol{\beta}(\alpha_\tau^*)$. In addition, we have that $n |\hat{\epsilon}_{(i_{n-d})}(\alpha_{ni,\tau}^*)|$ is asymptotically equivalent to $n |\epsilon_{(i_{n-d})}(\alpha_\tau^*)|$, where the index (i_{n-d}) denotes the same observation indicated above in (12).

By Lemma 1, however, we have that $n |\epsilon_{(i_{n-d})}(\alpha_\tau^*)|$ has a limiting distribution given by the sum $\sum_{j=1}^{n-d} E_j$, where, conditional on a sample $\{\mathbf{X}_j : j = 1, \dots, n - d\}$ drawn from μ , the corresponding sequence $\{E_j : j = 1, \dots, n - d\}$ is a sequence of independent exponential random variables with mean

$$\frac{1}{\kappa_{\alpha_\tau^*}(\mathbf{X}_j)} = \frac{1}{f(\mathbf{X}_j^\top \boldsymbol{\beta}(\alpha_\tau^*) | \mathbf{X}_j)}.$$

In this connection, define the pseudo-estimator

$$s_{n,\tau}^*(v_i | \mathbf{x}_i) \equiv \frac{n |\hat{\epsilon}_{(i_{n-d})}(\alpha_{ni,\tau}^*)|}{n-d}. \quad (13)$$

Consider a value $\boldsymbol{\kappa}$ of the finite-dimensional nuisance parameter in some open neighbourhood containing the true value $\boldsymbol{\kappa}_0$. By Lemma 1 and the uniform consistency of the regression-quantile process $\alpha \rightarrow \hat{\boldsymbol{\beta}}_n(\alpha)$ on closed subintervals of $(0, 1)$, the limiting distribution of $s_{n,\tau}^*(v_i | \mathbf{x}_i) \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa})$ is approximable by that of

$$\frac{n |\epsilon_{(i_{n-d})}(\alpha_\tau^*)|}{n-d} \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}) = \frac{n |\epsilon_{(i_{n-d})}(\alpha_\tau^*)|}{n-d} \mathbf{w}(y_i, v_i, \mathbf{x}_i, \boldsymbol{\kappa}),$$

which for a sample $\{\mathbf{X}_j : j = 1, \dots, n-d\}$ drawn from a population with measure μ , behaves asymptotically as a form of rescaled gamma random variable with conditional mean given $\{\mathbf{X}_j : j = 1, \dots, n-d\}$ equal to

$$\frac{1}{n-d} \sum_{j=1}^{n-d} \frac{1}{f(\mathbf{X}_j^\top \boldsymbol{\beta}(\alpha_\tau^*) | \mathbf{X}_j)} \mathbf{w}(y_i, v_i, \mathbf{X}_j, \boldsymbol{\kappa}). \quad (14)$$

Note that the expression in (14) is asymptotically equivalent to

$$\frac{1}{n-d} \sum_{j=1}^{n-d} \frac{1}{f(\mathbf{X}_j^\top \hat{\boldsymbol{\beta}}_n(\alpha_{ni,\tau}^*) | \mathbf{X}_j)} \mathbf{w}(y_i, v_i, \mathbf{X}_j, \boldsymbol{\kappa}),$$

which in turn is asymptotically equivalent to

$$E \left[\frac{1}{f(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha_{ni,\tau}^*) | \mathbf{x}_i)} \mathbf{w}(y_i, v_i, \mathbf{x}_i, \boldsymbol{\kappa}) \right] = E \left[\frac{1}{f(v_i | \mathbf{x}_i)} \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}) \right]$$

for \mathbf{x}_i drawn from the limiting design measure μ . From this it would appear that subject to additional regularity conditions, the specification of the pseudo-estimator $s_{n,\tau}^*(v_i | \mathbf{x}_i)$ is sufficient to induce the estimator $\hat{\boldsymbol{\theta}}_n$ as given above in (5) to be consistent for $\boldsymbol{\theta}_0$.

The next section of this paper provides a detailed consideration of the effect of the large-sample behaviour of the pseudo-estimator $s_{n,\tau}^*(v_i | \mathbf{x}_i)$ on the asymptotic behaviour of $\hat{\boldsymbol{\theta}}_n$.

3 Consistency and Asymptotic Normality of $\hat{\theta}_n$

This section of the paper is devoted to showing that the estimator $\hat{\theta}_n$ as given in (5) above is consistent and asymptotically normal for the general estimand θ_0 whose expression is given in (1). In this connection, recall the notation

$$\mathbf{w}_i \equiv \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0),$$

and define

$$\mathbf{w}_i(\boldsymbol{\kappa}) \equiv \mathbf{w}_i(\mathbf{Z}_i, \boldsymbol{\kappa}) \tag{15}$$

to be the quantity in (4) above with an arbitrary value $\boldsymbol{\kappa}$ of the m -dimensional nuisance parameter appearing in place of the true value $\boldsymbol{\kappa}_0$.

Additional regularity conditions governing the analysis of the first-order behaviour of the estimator $\hat{\theta}_n$ are given as follows.

Assumption 4. $\mathbf{w}_i(\boldsymbol{\kappa})$ as given in (15) is continuously differentiable for all $\boldsymbol{\kappa} \in \mathcal{K}$, where $\mathcal{K} \subset \mathbb{R}^m$ is an open neighbourhood of the true value $\boldsymbol{\kappa}_0$.

Assumption 5.

$$E \left[\sup_{\boldsymbol{\kappa} \in \mathcal{K}} \left\| \frac{1}{f(v_i | \mathbf{x}_i)} \mathbf{w}_i(\boldsymbol{\kappa}) \right\| \right] < \infty.$$

Assumption 6. As a function of $\boldsymbol{\kappa}$,

$$E \left[\frac{1}{f(v_i | \mathbf{x}_i)} \mathbf{w}_i(\boldsymbol{\kappa}) \right]$$

is continuous at $\boldsymbol{\kappa}_0$, while

$$E \left[\frac{1}{f(v_i | \mathbf{x}_i)} \nabla_{\boldsymbol{\kappa}} \mathbf{w}_i(\boldsymbol{\kappa}_0) \right] < \infty,$$

where $\nabla_{\boldsymbol{\kappa}}$ denotes the partial derivative operator with respect to the m -dimensional nuisance parameter.

Assumption 7. The estimator $\hat{\boldsymbol{\kappa}}_n$ of $\boldsymbol{\kappa}_0$ is asymptotically linear and \sqrt{n} -consistent, i.e.,

$$\hat{\boldsymbol{\kappa}}_n - \boldsymbol{\kappa}_0 = \frac{1}{n} \sum_{i=1}^n \boldsymbol{\psi}_i + o_p \left(n^{-\frac{1}{2}} \right), \tag{16}$$

where $E[\boldsymbol{\psi}_i] = \mathbf{0}$ and $E[\|\boldsymbol{\psi}_i\|^2] < \infty$.

An analysis of the first-order asymptotic behaviour of $\hat{\boldsymbol{\theta}}_n$ is complicated by the unboundedness of the inverse conditional-density weight at the extremities of the support of the corresponding conditional distribution. This has the consequence of potentially inducing the asymptotic variance of the summands in the expression for $\hat{\boldsymbol{\theta}}_n$ to be infinite. In this connection, define the functions

$$\boldsymbol{\Lambda}_{\boldsymbol{\kappa}}(v_i | y_i, \boldsymbol{x}_i) \equiv \frac{1}{f(v_i | \boldsymbol{x}_i)} \boldsymbol{w}(y_i, v_i, \boldsymbol{x}_i, \boldsymbol{\kappa}),$$

and

$$S_{\boldsymbol{\kappa}}(c | y_i, \boldsymbol{x}_i) \equiv E \left[\|\boldsymbol{\Lambda}_{\boldsymbol{\kappa}}(v_i | y_i, \boldsymbol{x}_i)\|^2 \mathbf{1} \{ \|\boldsymbol{\Lambda}_{\boldsymbol{\kappa}, \gamma}(v_i | y_i, \boldsymbol{x}_i)\| \leq c \} \mid y_i, \boldsymbol{x}_i \right],$$

where $\boldsymbol{\kappa}$ is restricted to an open neighbourhood \mathcal{K} containing the true value $\boldsymbol{\kappa}_0$. The following condition is imposed to facilitate the development of the central limit theory for the proposed estimator.

Assumption 8. *For all $\boldsymbol{\kappa} \in \mathcal{K}$, where $\mathcal{K} \subset \mathbb{R}^m$ is an open neighbourhood of the true value $\boldsymbol{\kappa}_0$ of the finite-dimensional nuisance parameter, the quantity $S_{\boldsymbol{\kappa}}(c | y_i, \boldsymbol{x}_i)$, taken as a function of c , is slowly varying at infinity.*

Assumption 8 is a domain of attraction condition and essentially imposes a restriction on the tail thicknesses of the conditional densities of each component of the vector $\frac{1}{f(v_i | \boldsymbol{x}_i)} \boldsymbol{w}(y_i, v_i, \boldsymbol{x}_i, \boldsymbol{\kappa})$ given $(y_i, \boldsymbol{x}_i^\top)^\top$. In particular, the condition of Assumption 8 is equivalent to the restriction

$$\begin{aligned} & c^2 \{ P[\|\boldsymbol{\Lambda}_{\boldsymbol{\kappa}}(v_i | y_i, \boldsymbol{x}_i)\| > c | y_i, \boldsymbol{x}_i] + P[\|\boldsymbol{\Lambda}_{\boldsymbol{\kappa}}(v_i | y_i, \boldsymbol{x}_i)\| \leq -c | y_i, \boldsymbol{x}_i] \} \\ &= o(S_{\boldsymbol{\kappa}}(c | y_i, \boldsymbol{x}_i)) \end{aligned}$$

as $c \rightarrow \infty$. In particular, in cases where $\|\boldsymbol{w}(y, v, \boldsymbol{x}, \boldsymbol{\kappa})\|$ tends to a finite limit as $v \rightarrow \pm\infty$ for arbitrary fixed values of y, \boldsymbol{x} and $\boldsymbol{\kappa}$, Assumption 8 essentially rules out conditional densities $f(v_i | \boldsymbol{x}_i)$ with overly thin tails.²¹

Let

$$\boldsymbol{\Omega}_n \equiv \text{Cov} [s_{n,\tau}^*(v_i | \boldsymbol{x}_i) \boldsymbol{w}(\boldsymbol{Z}_i, \boldsymbol{\kappa}_0)].$$

The conditions of Assumptions 4–8, combined with those governing the large sample behaviour of $s_{n,\tau}^*(v_i | \boldsymbol{x}_i)$ as spelled out in Lemma 1, yield the main result of this paper.

²¹A leading example involves a binary choice model with a latent mean restriction. This example serves as the immediate context for the simulation experiments presented in Section 4.

Theorem 1. *Suppose the conditions of Assumptions 1–8 hold. Then the estimator $\hat{\boldsymbol{\theta}}_n$ given above in (5) satisfies*

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$$

and

$$\sqrt{n}\boldsymbol{\Omega}_n^{-\frac{1}{2}} \left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}).$$

Proof. The proof appears in Appendix A.2. □

Note that the central limit result in Theorem 1 is rate adaptive in the sense that it enables the construction of Wald-type inference procedures whose validity does not require any assumptions regarding the rate of convergence of the estimator.²² As such, this result is analogous to that of Andrews and Schafgans (1998, Theorem 3), who consider the limiting distribution of Heckman (1990)’s semi-parametric “identification at infinity” estimator of the sample-selection model.

4 Numerical Evidence

This section of the paper presents the results of a series of simulation experiments designed to illustrate the sampling behaviour of the proposed estimator in the case of samples of no more than moderate size. The precise context considered here is that analyzed by Lewbel (2000), namely, a qualitative-response model originating from a latent linear model with an unknown error distribution. The specific model that is the subject of the simulations presented here involves a binary dependent variable y , two scalar-valued covariates v and x and a latent disturbance term e , to wit:

$$y = 1 \{v + \beta_1 + \beta_2 x + e > 0\},$$

where the distributions of e and x satisfy the restrictions

$$\begin{aligned} E[e] &= 0, \\ E[ex] &= 0 \end{aligned}$$

and

$$0 < E[x^2] < \infty.$$

The distribution of e is explicitly taken to be unknown, while the covariate v is taken to be the “special regressor” in this context. In particular, it is shown in

²²Note that $\boldsymbol{\Omega}_n$ is asymptotically equivalent to its sample analogue.

Lewbel (2000, Theorem 1) that if the conditional distribution of v given the other regressor x is absolutely continuous with respect to Lebesgue measure and has a large support relative to that of $\beta_1 + \beta_2 x + e$, and if the disturbance term e is conditionally independent of v given x , then the parameter vector $\beta \equiv (\beta_1 \ \beta_2)^\top$ has the form

$$\beta = (E[\mathbf{x}\mathbf{x}^\top])^{-1} E\left[\mathbf{x} \frac{y - 1\{v > 0\}}{f(v|x)}\right], \quad (17)$$

where $\mathbf{x} \equiv (1 \ x)^\top$ and where $f(v|x)$ is the conditional density of v given x .

The goal of the simulations presented here is to verify the suitability of the nonstandard estimation procedure developed in this paper for

$$\theta = E\left[\mathbf{x} \frac{y - 1\{v > 0\}}{f(v|x)}\right]$$

and by extension for β as given above in (17).²³

The specific data-generating process used in the simulation experiments is as follows. The covariate x was simulated from a uniform distribution on the interval $(-1, 1)$, while the special regressor v was drawn from a $N(0, 4)$ distribution. The error term e was set to be standard normal, and the parameters of interest were set to be $\beta_1 = \beta_2 = 1$. Samples of sizes $n = 50, 100, 200$ were used, and the number of Monte Carlo replications was set to 1000. When implementing the pseudo-estimator of the inverse conditional-density weights according to the procedure described above in Section 2.1, basic solutions for the regression-quantile optimization problem were obtained by searching a grid of quantiles each separated by a distance of .01 within the interval $[\alpha_{.01}, .99]$. For a given observation indexed by $i \in \{1, \dots, n\}$, the search algorithm starts from $\alpha = .99$ and works its way downwards until a quantile $\alpha_{ni,.01}^*$ satisfying

$$\left|v_i - \mathbf{x}_i^\top \hat{\beta}_n(\alpha_{ni,.01}^*)\right| < .0001 \quad (18)$$

is found. If searching a grid of points in the unit interval with uniform separation widths of .01 doesn't lead to a quantile $\alpha_{ni,.01}^*$ satisfying (18) for a given covariate vector $(v_i, \mathbf{x}_i^\top)^\top$, then the algorithm restarts at $\alpha = .99$ and searches downward along a grid of quantiles separated by width .005. Subsequent iterations of the algorithm, if needed, involve the separation widths halving by what they were in

²³In particular, nonstandard estimates of θ of the general form given above in (5) were simply rescaled by the inverse of $\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top$ to obtain (nonstandard) estimates of the parameter vector β .

the immediately previous iteration. The search algorithm was found not to require a number of iterations greater than three for the series of simulations presented here.

For purposes of comparison, the nonstandard estimates of β were set against three other estimates generated by a series of more conventional procedures. The first competing procedure was probit maximum-likelihood, which is efficient in the case where the specific data-generating process used in these simulations is taken to be true. The second and third procedures proceed from a consideration of the representation of β given in (17). In particular, (17) suggests that β_1 and β_2 may be consistently estimated by applying ordinary least squares in a regression of

$$\tilde{y}_i \equiv \frac{y_i - 1 \{v_i > 0\}}{\hat{f}_n(v_i | x_i)} \quad (19)$$

on a constant and x_i given the availability of a random sample

$$\{(y_i, v_i, x_i)^\top : i = 1, \dots, n\}$$

drawn from the joint distribution of $(y, v, x)^\top$ and a suitable estimate $\hat{f}_n(v_i | x_i)$ of the conditional density of v_i given x_i . In this connection, one set of OLS results involved an infeasible regression using the true value of the conditional density of v_i given x_i in place of the estimated density in \tilde{y}_i , while the other involved a conditional density estimate of the form

$$\hat{f}_n(v_i | x_i) \equiv \frac{\hat{f}_{vx,n}(v_i, x_i)}{\hat{f}_{x,n}(x_i)},$$

where $\hat{f}_{vx,n}$ and $\hat{f}_{x,n}$ denote kernel density estimates implemented using Epanechnikov kernels. The bandwidth used to implement the estimate of the joint density of (v, x) was set to decay at rate $n^{-\frac{1}{5}}$, while that used to implement the estimate of the marginal density of x was set to be proportional to $n^{-\frac{1}{4}}$. Leading constants for both bandwidths were computed using the popular rule of thumb of Silverman (1986).

The behaviour of the four estimators of β across the 1000 Monte Carlo replications considered for each of the three sample sizes used is summarized in Tables 1 and 2. In both tables the rows labelled “Nonstandard” correspond to the nonstandard estimation procedure proposed in this paper, while “Probit” denotes probit maximum-likelihood and “OLS” and “SP-OLS” denote the least-squares procedure using the true and estimated conditional densities of v given x , respectively.

Roughly the same qualitative pattern emerges in the results summarized in both Tables 1 and 2. In particular, the nonstandard estimates dominate those obtained using the semiparametric OLS procedure in terms of bias, while the semiparametric OLS estimator has a smaller variance. The bias of the semiparametric OLS procedure appears to decay only slowly as the sample size is increased. Probit maximum-likelihood appears to be rather unstable in terms of both bias and precision at the smallest sample size, but tends to outperform the other estimators in larger sample sizes, particularly in the case of the slope estimate. The performance of the nonstandard estimator is generally closer to the infeasible OLS procedure than the infeasible OLS procedure is to its semiparametric counterpart.

The overall impression from the simulation study conducted here is that the nonstandard procedure provides a feasible and computationally convenient alternative to more established methods, particularly when a normal latent-error assumption may not be justified or when the researcher desires an estimator whose finite-sample performance is unaffected by the choice of subjective smoothing parameters.

5 Conclusion

This paper has proposed an approach to generating consistent and asymptotically normal estimates of inverse conditional density-weighted expectations that circumvents both the computation of preliminary nonparametric estimates of the inverse conditional density weights and the implementation of data-dependent trimming functions. The new approach proposed here exploits results concerning the behaviour of fitted regression-quantile residuals developed in recent work by Goh and Knight (2008). The asymptotic normality result for the proposed estimator is explicitly rate adaptive in order to facilitate the formulation of suitable Wald-type inference procedures. Simulation evidence in the context of a binary-choice model originating from a linear latent-variable model indicates that the estimation procedure proposed here provides a viable alternative to more established methods requiring the choice of bandwidths or kernel functions.

A Proofs

A.1 Proof of Lemma 1

Note that

$$M_{n\alpha}(A \times B) = \sum_{i \notin \mathcal{H}_n(\alpha)} 1 \{n\epsilon_i(\alpha) - \mathbf{x}_i^\top \mathbf{W}_n^{-1} \boldsymbol{\xi}_n \in A, \mathbf{x}_i \in B\},$$

where \mathbf{W}_n is a $d \times d$ matrix with columns \mathbf{x}_j for indices $j \in \mathcal{H}_n(\alpha)$ and $\boldsymbol{\xi}_n$ is a vector with components $n\epsilon_j(\alpha)$ for $j \in \mathcal{H}_n(\alpha)$.

Argue conditionally given $\mathbf{W}_n^{-1} \boldsymbol{\xi}_n$ and observe that $\mathbf{W}_n^{-1} \boldsymbol{\xi}_n = O_p(\sqrt{n})$ in order to deduce the desired result.

A.2 Proof of Theorem 1

Define

$$\tilde{\boldsymbol{\theta}}_n \equiv \frac{1}{n} \sum_{i=1}^n s_{n,\tau}^*(v_i | \mathbf{x}_i) \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0). \quad (20)$$

Theorem 1 follows from the following three lemmas, which are shown in Appendices A.2.1–A.2.3 to hold under the conditions of Theorem 1:

Lemma 2.

$$\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n = o_p(1).$$

Lemma 3.

$$\tilde{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0.$$

Lemma 4.

$$\sqrt{n} \boldsymbol{\Omega}_n^{-\frac{1}{2}} \left(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}).$$

A.2.1 Proof of Lemma 2

We have for some intermediate point $\bar{\boldsymbol{\kappa}}_n$ between $\hat{\boldsymbol{\kappa}}_n$ and $\boldsymbol{\kappa}_0$ that

$$\begin{aligned} & \hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n \\ &= \frac{1}{n} \sum_{i=1}^n s_{n,\tau}^*(v_i | \mathbf{x}_i) (\mathbf{w}(\mathbf{Z}_i, \hat{\boldsymbol{\kappa}}_n) - \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0)) \\ &= \left(\frac{1}{n} \sum_{i=1}^n s_{n,\tau}^*(v_i | \mathbf{x}_i) \nabla_{\boldsymbol{\kappa}} \mathbf{w}(\bar{\boldsymbol{\kappa}}_n) \right) (\hat{\boldsymbol{\kappa}}_n - \boldsymbol{\kappa}_0) \\ &\stackrel{p}{=} E \left[\frac{1}{f(v_i | \mathbf{x}_i)} \nabla_{\boldsymbol{\kappa}} \mathbf{w}(\bar{\boldsymbol{\kappa}}_n) \right] (\hat{\boldsymbol{\kappa}}_n - \boldsymbol{\kappa}_0), \end{aligned}$$

The desired result follows from Assumptions 6 and 7.

A.2.2 Proof of Lemma 3

The restriction imposed by Assumption 8 on the conditional tail thicknesses of each component of $\frac{1}{f(v|\mathbf{x})}\mathbf{w}(y, v, \mathbf{x}, \boldsymbol{\kappa}_0)$ as $v \rightarrow \pm\infty$ ensures that

$$\frac{1}{n} \sum_{i=1}^n s_{n,\tau}^*(v_i|\mathbf{x}_i) \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0) \stackrel{p}{=} E [s_{n,\tau}^*(v_i|\mathbf{x}_i) \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0)].$$

We note that $E [s_{n,\tau}^*(v_i|\mathbf{x}_i) \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0)] \rightarrow \boldsymbol{\theta}_0$ by construction.

A.2.3 Proof of Lemma 4

We have from Assumption 8 that

$$\sqrt{n}\boldsymbol{\Omega}_n^{-\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{f(v_i|\mathbf{x}_i)} \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0) - \boldsymbol{\theta}_0 \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}). \quad (21)$$

Observe that

$$\begin{aligned} & \sqrt{n}\boldsymbol{\Omega}_n^{-\frac{1}{2}} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ &= \boldsymbol{\Omega}_n^{-\frac{1}{2}} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n (s_{n,\tau}^*(v_i|\mathbf{x}_i) \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0) - \boldsymbol{\theta}_0) \right] \\ &= \boldsymbol{\Omega}_n^{-\frac{1}{2}} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{1}{f(v_i|\mathbf{x}_i)} \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0) - \boldsymbol{\theta}_0 \right) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(s_{n,\tau}^*(v_i|\mathbf{x}_i) - \frac{1}{f(v_i|\mathbf{x}_i)} \right) \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0) \right] \\ &\equiv \mathbf{A}_{n1} + \mathbf{A}_{n2}. \end{aligned}$$

We have $\mathbf{A}_{n1} \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$ by (21), so it remains to show that \mathbf{A}_{n2} is asymptotically negligible.

In this connection, note that

$$\begin{aligned} & E \left[\frac{1}{n^2} \left\| \sum_{i=1}^n \left(s_{n,\tau}^*(v_i|\mathbf{x}_i) - \frac{1}{f(v_i|\mathbf{x}_i)} \right) \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0) \right\|^2 \right] \\ &\leq \frac{1}{n} E \left[\left(s_{n,\tau}^*(v_i|\mathbf{x}_i) - \frac{1}{f(v_i|\mathbf{x}_i)} \right)^2 \|\mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0)\|^2 \right] \\ &= o(n^{-1}). \end{aligned}$$

From this it follows that

$$\left\| \frac{1}{n} \sum_{i=1}^n \left(s_{n,\tau}^*(v_i|\mathbf{x}_i) - \frac{1}{f(v_i|\mathbf{x}_i)} \right) \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0) \right\| = o_p \left(\frac{1}{\sqrt{n}} \right),$$

which in turn implies that

$$\sqrt{n} \|\boldsymbol{\Omega}_n\|^{-\frac{1}{2}} \cdot \left\| \frac{1}{n} \sum_{i=1}^n \left(s_{n,\tau}^*(v_i|\mathbf{x}_i) - \frac{1}{f(v_i|\mathbf{x}_i)} \right) \mathbf{w}(\mathbf{Z}_i, \boldsymbol{\kappa}_0) \right\|$$

$$\begin{aligned} &= \|\Omega_n\|^{-\frac{1}{2}} \cdot o_p(1) \\ &= o_p(1) \end{aligned}$$

by virtue of the fact that $\|\Omega_n\| = O(1)$.²⁴ It follows that $\mathbf{A}_{n,2}$ is asymptotically negligible.

²⁴Note that $\|\Omega_n\|$ tends either to a finite positive limit or to infinity as $n \rightarrow \infty$.

Table 1: Finite-sample behaviour of slope estimators, binary choice model

| | Mean | SD | Lower quartile | Median | Upper quartile | RMSE | Mean absolute error | Median absolute error |
|----------------|--------|---------|----------------|--------|----------------|---------|---------------------|-----------------------|
| <i>n</i> = 50 | | | | | | | | |
| Nonstandard | .9322 | .3856 | .6719 | .9053 | 1.1580 | .3914 | .3126 | .2637 |
| Probit | 8.0135 | 79.5015 | .8914 | 1.1543 | 1.5457 | 79.7706 | 7.1786 | .3112 |
| OLS | .9953 | .4225 | .7068 | .9730 | 1.2478 | .4223 | .3284 | .2781 |
| SP-OLS | .7306 | .2706 | .5497 | .7280 | .8925 | .3817 | .3197 | .2934 |
| <i>n</i> = 100 | | | | | | | | |
| Nonstandard | .9957 | .3014 | .7929 | .9795 | 1.1862 | .3013 | .2378 | .2004 |
| Probit | 1.1255 | .3528 | .8992 | 1.0742 | 1.2841 | .3743 | .2570 | .1919 |
| OLS | 1.0187 | .3153 | .8034 | .9993 | 1.2129 | .3157 | .2496 | .2039 |
| SP-OLS | .8241 | .2206 | .6793 | .8113 | .9640 | .2821 | .2331 | .2129 |
| <i>n</i> = 200 | | | | | | | | |
| Nonstandard | 1.0441 | .2249 | .8872 | 1.0305 | 1.1829 | .2291 | .1806 | .1515 |
| Probit | 1.0410 | .2046 | .8925 | 1.0259 | 1.1658 | .2086 | .1629 | .1299 |
| OLS | 1.0095 | .2127 | .8612 | 1.0080 | 1.1404 | .2129 | .1679 | .1392 |
| SP-OLS | .8969 | .1590 | .7860 | .8903 | 1.0059 | .1895 | .1538 | .1338 |

Table 2: Finite-sample behaviour of intercept estimators, binary choice model

| | Mean | SD | Lower quartile | Median | Upper quartile | RMSE | Mean absolute error | Median absolute error |
|----------------|--------|---------|----------------|--------|----------------|---------|---------------------|-----------------------|
| <i>n</i> = 50 | | | | | | | | |
| Nonstandard | .9614 | .3948 | .7114 | .9258 | 1.1913 | .3964 | .3095 | .2571 |
| Probit | 7.2719 | 64.5987 | .8757 | 1.1914 | 1.6218 | 64.8703 | 6.4387 | .3151 |
| OLS | 1.0129 | .3994 | .7486 | .9888 | 1.2505 | .3994 | .3117 | .2514 |
| SP-OLS | .7623 | .2773 | .5757 | .7647 | .9351 | .3651 | .2975 | .2587 |
| <i>n</i> = 100 | | | | | | | | |
| Nonstandard | .9953 | .3049 | .7656 | .9846 | 1.1932 | .3048 | .2425 | .2101 |
| Probit | 1.1109 | .3365 | .8911 | 1.0631 | 1.2847 | .3541 | .2514 | .1910 |
| OLS | 1.0037 | .2948 | .7967 | .9884 | 1.1924 | .2947 | .2356 | .1993 |
| SP-OLS | .8307 | .2123 | .6788 | .8354 | .9765 | .2714 | .2220 | .1931 |
| <i>n</i> = 200 | | | | | | | | |
| Nonstandard | 1.0571 | .2454 | .8906 | 1.0443 | 1.2125 | .2518 | .1961 | .1621 |
| Probit | 1.0395 | .2043 | .9052 | 1.0175 | 1.1541 | .2079 | .1569 | .1247 |
| OLS | 1.0097 | .1994 | .8671 | 1.0116 | 1.1395 | .1995 | .1596 | .1366 |
| SP-OLS | .9033 | .1484 | .8088 | .9063 | 1.0025 | .1771 | .1425 | .1226 |

References

- Andrews, D. W. K., and M. M. A. Schafgans (1998) 'Semiparametric estimation of the intercept of a sample selection model.' *Review of Economic Studies* 65, 497–517
- Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006) 'Quantile regression under misspecification, with an application to the U.S. wage structure.' *Econometrica* 74, 539–563
- Chamberlain, G. (1986) 'Asymptotic efficiency in semi-parametric models with censoring.' *Journal of Econometrics* 32, 189–218
- Goh, S. C., and K. Knight (2008) 'Nonstandard quantile-regression inference.' *Econometric Theory*, forthcoming.
- Härdle, W., and T. M. Stoker (1989) 'Investigating smooth multiple regression by the method of average derivatives.' *Journal of the American Statistical Association* 84, 986–995
- Hausman, J. A., and W. K. Newey (1995) 'Nonparametric estimation of exact consumers surplus and dead-weight loss.' *Econometrica* 63, 1445–1476
- Heckman, J. J. (1990) 'Varieties of selection bias.' *American Economic Review* 80, 313–318
- Hong, Y., and H. White (2005) 'Asymptotic distribution theory for nonparametric entropy measures of serial dependence.' *Econometrica* 73, 837–901
- Honoré, B. E., and A. Lewbel (2002) 'Semiparametric binary choice panel data models without strictly exogenous regressors.' *Econometrica* 70, 2053–2063
- Khan, S., and A. Lewbel (2007) 'Weighted and two-stage least squares estimation of semiparametric truncated regression models.' *Econometric Theory* 23, 309–347
- Khan, S., and E. Tamer (2009) 'Irregular identification, support conditions, and inverse weight estimation.' Department of Economics, Duke University
- Knight, K. (2008) 'Asymptotics of the regression quantile basic solution under misspecification.' *Applications of Mathematics* 53, 223–234
- Knight, K., and C. Goh (2008) 'Asymptotics of quantile regression basic solutions with applications.' Department of Statistics, University of Toronto. An earlier version of this paper is available at <http://www.chuangoh.org/rqs.pdf>.
- Koenker, R. (2005) *Quantile Regression* (New York: Cambridge University Press)
- Koenker, R., and B. J. Park (1996) 'An interior point algorithm for nonlinear quantile regression.' *Journal of Econometrics* 71, 265–283
- Koenker, R., and G. Bassett (1978) 'Regression quantiles.' *Econometrica* 46, 33–50

- Koenker, R., and V. d'Orey (1987) 'Computing regression quantiles.' *Applied Statistics* 36, 383–393
- Lewbel, A. (1997) 'Semiparametric estimation of location and other discrete choice moments.' *Econometric Theory* 13, 32–51
- (1998) 'Semiparametric latent variable model estimation with endogenous or mismeasured regressors.' *Econometrica* 66, 105–121
- (2000) 'Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables.' *Journal of Econometrics* 97, 145–177
- (2007) 'Endogenous selection or treatment model estimation.' *Journal of Econometrics* 141, 777–806
- Lewbel, A., and S. M. Schennach (2007) 'A simple ordered data estimator for inverse density weighted expectations.' *Journal of Econometrics* 136, 189–211
- McFadden, D. L. (1999) 'Computing willingness-to-pay in random utility models.' In *Trade Theory and Econometrics: Essays in Honour of John S. Chipman*, ed. J. Moore, R. Riezman, and J. Melvin (New York: Routledge) pp. 253–274
- Newey, W. K. (1997) 'Convergence rates and asymptotic normality for series estimators.' *Journal of Econometrics* 79, 147–168
- Newey, W. K., and P. A. Ruud (2005) 'Density weighted linear least squares.' In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. D. W. K. Andrews and J. H. Stock (New York: Cambridge University Press) pp. 554–573
- Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis* (London: Chapman and Hall)