EFFICIENT ESTIMATION OF LINEAR AND TYPE I CENSORED REGRESSION MODELS UNDER CONDITIONAL QUANTILE RESTRICTIONS

Whitney K. Newey
Princeton University

James L. Powell
University of Wisconsin at Madison

Econometric Research Program
Research Memorandum No. 331

June 1987

Abstract

We consider the linear regression model with censored dependent variable, where the disturbance terms are restricted only to have zero conditional median (or other prespecified quantile) given the covariates and censoring point. For this model, a lower bound for the asymptotic covariance matrix for locally-regular estimators of the regression coefficients is derived. We also show how an estimator which attains this lower bound can be constructed. As a special case, our results apply to the (uncensored) linear model.

Key words and phrases: semiparametric efficiency; kernel estimation; nearest neighbor estimation.

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, NJ 08544, USA

1. Introduction

We consider efficient estimation of the linear regression model with Type I censoring. The observed dependent variable y_t satisfies

$$(1.1) \qquad y_t = \min\{x_t'\beta_0 + \varepsilon_t,\ u_t\}, \quad t = 1, 2, \ldots,$$

where x_t is a $p \times 1$ vector of observed regressors, β_0 is a $p \times 1$ vector of parameters, ε_t is an unobserved disturbance term, u_t is an observed censoring point, and $z_t = (y_t, x_t', u_t)$ is i.i.d.. Heuristically, the uncensored linear regression model can be obtained as a special case where $u_t = +\infty$ for all t.

Attainable asymptotic efficiency for estimates of β_0 when ε_t is independent of x_t and u_t has previously been considered by Cosslett (1984) and Ritov (1984). Each has derived the semiparametric efficiency bound for estimates of β in this environment. Also, Ritov (1984) constructed an efficient estimator when $p = 1$ under Type II censoring, that is the case where the censoring point is not always observed.

We depart from this previous work by weakening the hypothesis of independence of ε_t and $\tilde{x}_t = (x_t', u_t)'$. We will consider efficient estimation of β_0 under the restriction that ε_t has median zero conditional on the regressors and the censoring point. This environment allows for dependence of the conditional distribution of ε_t on $\tilde{x}_t$, and in particular for heteroskedasticity, which can be an important phenomenon in practice.

Powell's (1984) censored least absolute deviations CLAD estimator is $\sqrt{n}$-consistent in this environment, under regularity conditions given below. For a sample of size n the CLAD estimator $\tilde{\beta}$ is defined as

$$(1.2) \qquad \tilde{\beta} = \operatorname{argmin}_{\beta\in B}\Sigma_{t=1}^{n} |y_t - \min\{x_t'\beta,\ u_t\}|,$$

for some set B of possible values for β_0. The CLAD estimator will also solve the asymptotic first-order condition

(1.3) $$\Sigma_{t=1}^{n} 1(x_t'\tilde{\beta} < u_t)x_t \mathrm{sgn}(y_t - x_t'\tilde{\beta}) = o_p(\sqrt{n}),$$

where $1(A)$ denotes the indicator function for the event A, $o_p(\circ)$ and $O_p(\circ)$ denote the usual order in probability relations, and $\mathrm{sgn}(\varepsilon)$ is the sign function which is equal to 1 for $\varepsilon > 0$, 0 for ε equal to zero, and -1 for $\varepsilon < 0$. The asymptotic covariance matrix of the CLAD estimator is given by (see Powell (1984))

(1.4) $$V_1 = \{E[2f_t 1_t x_t x_t']\}^{-1} E[1_t x_t x_t'] \{E[2f_t 1_t x_t x_t']\}^{-1},$$

where $f_t = f(0|\tilde{x}_t)$ is the conditional density function of ε_t at $\varepsilon = 0$ and $1_t = 1(x_t'\beta_0 < u_t)$.

To motivate the efficiency results discussed below, it is useful to embed the CLAD estimator within a larger class of estimators, and discuss efficiency within this class. A class of estimators that includes the CLAD estimator as a special case is the class of weighted CLAD estimators, with weights depending on $\tilde{x}_t$. A weighted CLAD estimator $\tilde{\beta}_w$ can be defined as

(1.5) $$\tilde{\beta}_w = \mathrm{argmin}_{\beta\in B}\ \Sigma_{t=1}^{n} w_t |y_t - \min\{x_t'\beta,\ u_t\}|,$$

where w_t is some nonnegative function $w(\tilde{x}_t)$ of the regressors and the censoring point. This estimator has a corresponding asymptotic first order condition

(1.6) $$\Sigma_{t=1}^{n} 1(x_t'\tilde{\beta}_w < u_t) w_t x_t \mathrm{sgn}(y_t - x_t'\tilde{\beta}_w) = o_p(\sqrt{n}).$$

This equation is the same as equation (1.3) except that the variables that appear with the CLAD "residual" $1(x_t'\beta<u_t)\mathrm{sgn}(y_t-x_t'\beta)$ are the weighted regressors $w_t x_t$ rather than x_t. The resulting asymptotic covariance matrix of $\tilde{\beta}_w$ is

(1.7) $$V_w = \{E[2f_t 1_t w_t x_t x_t']\}^{-1} E[1_t w_t^2 x_t x_t'] \{E[2f_t 1_t w_t x_t x_t']\}^{-1}.$$

It is straightforward to show that the choice of w_t that minimizes this covariance matrix (in the positive semi-definite sense) is $w_t = 2f_t = 2f(0|\tilde{x}_t)$. The minimized covariance matrix is given by

(1.8) $$V^* = \{E[1_t 4 f_t^2 x_t x_t']\}^{-1}.$$

There is potential efficiency gain from a weighted CLAD estimator versus the unweighted estimator when the conditional density at zero, $f(0|\tilde{x}_t)$, is related to $\tilde{x}_t$ for $x_t'\beta_0 < u_t$.

In Section 2 it will be shown that, not only is V^* the lower bound for the asymptotic covariance matrices of weighted CLAD estimators, but also it is the semiparametric efficiency bound when only the conditional median restriction is imposed. An estimator that attains this bound (with $f(0|\tilde{x}_t)$ of unknown form) will be constructed in Section 3. The efficient estimator makes use of a nonparametric estimate of $f(0|\tilde{x}_t)$. Section 4 discusses some immediate extensions of the results, and the proofs are collected in Section 5.

2. The Semiparametric Efficiency Bound

We will consider semiparametric efficiency in the sense of Stein (1956), as developed by Koshevnik and Levitt (1976), Pfanzagl (1983), Begun, et al. (1983), Bickel, et al. (1987), and Chamberlain (1987).

Let $\mathcal{P}$ be a set of distributions of a single observation $z_t = (y_t, x_t', u_t)$ satisfying equation (1.1) such the ε_t has a conditional median of zero. A regular parametric submodel is a subset $\mathcal{P}_0$ of $\mathcal{P}$ parametrized by $\theta = (\beta', \eta')$ contained in an open set Θ, where η is a Euclidean parameter for the distribution of $(\varepsilon_t, x_t', u_t)$, such that the likelihood $f(z|\theta)$ is continuously Hellinger differentiable and has a nonsingular information matrix $I(\mathcal{P}_0, \theta)$. An estimator $\hat{\beta}$ of β is (locally) regular for a regular parametric submodel if, when $\{z_t\}_{t=1}^n$ is i.i.d. with density $f(z|\theta_n)$ for each n with $\theta_n = \theta +$

$O(1/\sqrt{n})$, $\sqrt{n}(\hat{\beta}-\beta_n)$ has a limiting distribution that does not depend on the particular sequence $\{\theta_n\}_{n\geq 1}$. Similarly, an estimator $\hat{\beta}$ is said to be regular for a class Q of regular parametric submodels when $\hat{\beta}$ is regular for all $\mathcal{P}_0 \in Q$. An efficiency bound for estimators that are regular for Q is given by the supremum $V_\beta = \sup_{\mathcal{P}_0 \in Q} I^{\beta\beta}(P_R, \theta)$ (in the positive semi-definite matrix sense) of the block of the inverse information matrix corresponding to β. A formal statement is that the asymptotic distribution of any estimator that is regular for Q is equal to the distribution of $\mathcal{Y} + \mathcal{U}$, where $\mathcal{Y} \sim N(0, V_\beta)$ and $\mathcal{U}$ is independent of $\mathcal{Y}$, as in Hajek (1970).

We have found it convenient to work with parametric submodels of a particular form. Let $f(\varepsilon|\tilde{x})$ denote the conditional density of ε_t given $\tilde{x}_t = (x_t', u_t)$. The parametric submodels we work with are those corresponding to conditional densities of the form

$$(2.1) \qquad f(\varepsilon|\tilde{x},\eta) = f(\varepsilon|\tilde{x})\{1 + [v(x)'\eta]q(\varepsilon,\tilde{x})\},$$

where η and $v(x)$ are p x 1 vectors and $v(x)$ and $q(\varepsilon,\tilde{x})$ satisfy

Assumption 2.1: $v(x)$ is bounded and $q(\varepsilon,\tilde{x})$ is bounded and continuously differentiable in ε, with bounded derivative $q_\varepsilon(\varepsilon,\tilde{x})$, such that

$$(2.2) \qquad \int q(\varepsilon,\tilde{x})f(\varepsilon|\tilde{x})d\varepsilon = \int \mathrm{sgn}(\varepsilon)q(\varepsilon,\tilde{x})f(\varepsilon|\tilde{x})d\varepsilon = 0.$$

Equation (2.2) imposes the condition that $f(\varepsilon|\tilde{x},\eta)$ integrates to one and has median zero for each η. Boundedness of $v(x)$ and $q(\varepsilon,\tilde{x})$ implies that $f(\varepsilon|\tilde{x},\eta)$ will be nonnegative for all η close enough to zero. Assumption 2.1 also imposes a smoothness condition on q.

Note that because the marginal distribution of $\tilde{x}_t$ does not depend on β, the efficiency bound is not affected by restricting attention to parametric submodels for the conditional distribution of ε_t. Also, the class of estimators that are regular for parametric submodels of the form given in equation (2.1) may be larger than the class of estimators that are regular for the class of all regular parametric submodels. Thus, the result that V^* is the efficiency bound for the class of parametric submodels that take the form given in equation (2.1) may be a stronger efficiency result than the result that V^* is the efficiency bound for all regular parametric submodels. However, the class of estimators that are regular for parametric submodels of this form may not be as satisfactory on uniformity grounds as the class of estimators that are regular for all parametric submodels.

In our work we have found that the convenience of the parametric submodels of equation (2.1) outweighs uniformity concerns. Checking that V^* is the efficiency bound for these parametric submodels is quite straightforward. Also, we do not know whether our candidate for an efficient estimator (discussed in the next section) is regular for all regular parametric submodels, while it is easy to show that it is regular for the restricted class of parametric submodels of equation (2.1). Further, the class of densities satisfying equation (2.1) appears to be quite flexible. We expect, but have not verified, that the tangent space (e.g. Bickel, et al. (1987)) for this class equals the tangent space for the class of all parametric submodels.

To guarantee that parametric submodels corresponding to conditional densities of the form given in equation (2.1) are regular, and that the efficiency bound exists, we will impose the following regularity conditions:

Assumption 2.2: $(\varepsilon_t, x_t', u_t)$ is i.i.d. with distribution that is absolutely continuous with respect to $\mathcal{L} \times \mu_x$, where $\mathcal{L}$ is Lebesgue measure on $\mathbb{R}$ and μ_x is the probability measure of $\tilde{x}_t$. The conditional density $f(\varepsilon|\tilde{x})$ satisfies $\int sgn(\varepsilon)f(\varepsilon|\tilde{x})d\varepsilon = 0$ a.s. μ_x. Also, $f(\varepsilon|\tilde{x})$ is absolutely continuous in ε, a.s. μ_x, with Radon-Nikodym derivative $f_\varepsilon(\varepsilon|\tilde{x})$ satisfying $\int(1+\|x\|^2)[f_\varepsilon(\varepsilon|\tilde{x})^2/f(\varepsilon|\tilde{x})]d\varepsilon d\mu_x$ finite.

Throughout, ratios are taken equal to zero when their denominators are zero. Besides imposing an i.i.d. hypothesis and the conditional median restriction this assumption specifies that the conditional density of ε_t given $\tilde{x}_t$ is regular in the sense of Hajek and Sidak (1967).

The next two assumptions impose further regularity conditions on the conditional density of ε_t. Let $r(\varepsilon,\tilde{x}) = f(\varepsilon|\tilde{x})^{1/2}$, $r_\varepsilon(\varepsilon,\tilde{x}) = f_\varepsilon(\varepsilon|\tilde{x})/[2f(\varepsilon|\tilde{x})^{1/2}]$, $F(m|\tilde{x}) = \int_{-\infty}^{m} f(\varepsilon|\tilde{x})d\varepsilon$, $R(m,\tilde{x}) = [1-F(m|\tilde{x})]^{1/2}$, and $\alpha = (\beta',-1)'$.

Assumption 2.3: There is a neighborhood $N \subset \mathbb{R}^p$ of 0 and a function $\gamma(\varepsilon,\tilde{x})$ such that $\sup_{\delta\in N} r(\varepsilon+x'\delta,\tilde{x}) \le \gamma(\varepsilon,\tilde{x})$ and $\int(1+\|x\|^2)\gamma(\varepsilon,\tilde{x})^2 d\varepsilon d\mathcal{P}_x$ is finite. Also, at all δ in N, $r(\varepsilon+x'\delta|\tilde{x})$ is Hellinger differentiable in $\mathcal{L} \times \mathcal{P}_x$ with derivative $r_\varepsilon(\varepsilon+x'\delta,\tilde{x})x$. Furthermore, $r_\varepsilon(\varepsilon,\tilde{x})$ is continuous in ε a.s. $\mathcal{L} \times \mathcal{P}_x$ and $\sup_{\delta\in N}|r_\varepsilon(\varepsilon+x'\delta,\tilde{x})| \le \gamma(\varepsilon,\tilde{x})$.

Assumption 2.4: $R(m,\tilde{x})$ is continuously differentiable in m a.s. $\mathcal{P}_x$. Also, there is a neighborhood $N_0 \subset \mathbb{R}^p$ of β_0 and a function $\gamma(\tilde{x})$ such that $\int(1+\|x\|^2)\gamma(\tilde{x})^2 d\mathcal{P}_x$ is finite and $R(m,\tilde{x})$ and its derivative $R_m(m,\tilde{x})$ satisfy $\sup_{\beta\in N_0}|R(-\tilde{x}'\alpha,\tilde{x})| \le \gamma(\tilde{x})$, $\sup_{\beta\in N_0}|R_m(-\tilde{x}'\alpha,\tilde{x})| \le \gamma(\tilde{x})$.

The following assumption restricts somewhat the allowed heterogeneity for the conditional distribution of ε_t. In particular, note that when $f(\varepsilon|\tilde{x}) = f(\varepsilon/\sigma(\tilde{x}))/\sigma(\tilde{x})$ for some median zero density function $f(\varepsilon)$ that is continuous and positive at $\varepsilon = 0$ and a positive function $\sigma(\tilde{x})$, then this assumption implies that $\sigma(\tilde{x})$ is bounded and bounded away from zero.

Assumption 2.5: For some $\nu > 0$, $\nu^{-1} \ge f(\varepsilon|\tilde{x}) \ge \nu$ for all ε with $|\varepsilon| \le \nu$, a.s. $\mathcal{P}_x$.

The following Assumption imposes a local identification condition for the regression parameters.

Assumption 2.6: $E[1(x_t'\beta_0 < u_t)x_t x_t']$ is nonsingular and $Prob(x_t'\beta_0 = u_t) = 0$.

The result that the semiparametric efficiency bound for estimation of β_0 is $V^* = \{E[1(x_t'\beta_0<u_t)4f(0|\tilde{x}_t)^2 x_t x_t']\}^{-1}$ can be motivated in the following way. Note that V^* is the inverse covariance matrix of the "efficient score" $s(z) = 2f(0|\tilde{x})1(x'\beta_0<u)sgn(y-x'\beta_0)x$. In order to understand the form of the efficient score (and the efficiency bound) there are two important cases to consider, depending on the value of the indicator $1(x'\beta_0<u)$.

When $x'\beta_0 \geq u$, $s(z)$ is zero. Because the conditional median restriction does not restrict the shape of the lower half of the distribution of y, but only the relationship of the upper half to the lower half, no "local" information relevant to asymptotic efficiency is available when the censoring point is to the left of the median. (However, as discussed by Powell (1986b), all observations have "global" information which is incorporated in the minimization problem for the CLAD estimator.) In the mathematics this intuition translates to a choice of $q(\varepsilon,\tilde{x})$ and $v(x)$ such that for $x'\beta_0 \geq u$ the score (i.e. derivative of the log-likelihood) for η is identical to the score for β, i.e. so that the effect on the conditional distribution of y_t of a shift in β cannot be distinguished from the effect of a change in η. A choice which qualifies on these grounds, as well as satisfying the essential integral condition of equation (2.2) is

$$(2.3a) \qquad q(\varepsilon,\tilde{x}) = -1(\varepsilon < -\tilde{x}'\alpha_0)f_\varepsilon(\varepsilon|\tilde{x})/f(\varepsilon|\tilde{x})$$

$$+ 1(-\tilde{x}'\alpha_0 \leq \varepsilon \leq 0)f(-\tilde{x}'\alpha_0|\tilde{x})/[F(0|\tilde{x})-F(-\tilde{x}'\alpha_0|\tilde{x})],$$

$$v(x) = x, \qquad x'\beta_0 \geq u,$$

where $f_\varepsilon(\varepsilon|\tilde{x}) = \partial f(\varepsilon|\tilde{x})/\partial\varepsilon$.

When $x'\beta_0 < u$, $s(z) = 2f(0|\tilde{x})sgn(y-x'\beta_0)x$. Note that this is the natural generalization to the *uncensored* (conditional median zero) regression model of the efficient score for the location (median) model, which was given by Begun, et al. (1983). Since the median of y_t is invariant to censoring for observations where the censoring point occurs to the right of the median, the efficient score for such observations is expected to be similar to the efficient score for the uncensored regression model. Furthermore, the natural generalization to regression models of the parametric submodel that attains the efficiency bound, given by Begun, et al. (1983) for the location model, should work for the censored regression model. This natural generalization is

(2.3b) $q(\varepsilon,\tilde{x}) = -f_\varepsilon(\varepsilon|\tilde{x})/f(\varepsilon|\tilde{x}) - 2f(0|\tilde{x})sgn(\varepsilon), \quad v(x) = x, \quad x'\beta_0 < u.$

Although the $v(x)$ and $q(\varepsilon,\tilde{x})$ taken from equation (2.3) will not satisfy the boundedness and smoothness assumptions of Assumption 2.1, they can be approximated arbitrarily closely (in an appropriate sense) by functions that do satisfy Assumption 2.1. This approximation of the candidate for $q(\varepsilon,\tilde{x})$ given in equation (2.3) will give $V_\beta = V^*$.

Theorem 2.1: Let Q denote the class of regular parametric submodels with conditional density for ε_t of the form given in equation (2.1). satisfying Assumption 2.1. Suppose that Assumptions 2.2 - 2.6 are satisfied. Then

(2.4) $\sup_{\varphi_0\in Q} I^{\beta\beta}(\varphi_R,P_0) = \{E[1(x_t'\beta_0<u_t)4f(0|\tilde{x}_t)^2x_tx_t']\}^{-1} = V^*.$

3. An Efficient Estimator

An efficient estimator will take the form

(3.1) $\hat{\beta} = \beta_0 + V^*\Sigma_{t=1}^n s(z_t)/n + o_p(1/\sqrt{n}),$

where $s(z) = 2f(0|\tilde{x})1(x'\beta_0<u)sgn(y-x'\beta_0)x$ is the efficient score discussed in Section 3 and $V^* = \{E[s(z_t)s(z_t)']\}^{-1}$. To construct such an estimator, we follow the approach proposed by Schick (1986). In this approach, the sample is divided into two subsamples of (approximately) equal size. In each of these subsamples an estimate of the efficient score is formed. The estimated score is then averaged over the other subsample, and used in the construction of a one-step estimator based on an initial $\sqrt{n}$-consistent estimator of β_0.

Formally, let $n_1 = int(n/2)$, where $int(\circ)$ denotes the largest integer less than or equal to, and let $n_2 = n - n_1$. Define the index sets $I_1 = \{1, \ldots, n_1\}$ and $I_2 = \{n_1+1, \ldots, n\}$ of two subsamples. Let $\tilde{\beta}_1$ and $\tilde{\beta}_2$ be preliminary estimators of β_0 based on the subsamples indexed by I_1 and I_2 respectively. Also, let $\hat{s}_1(z)$ and $\hat{s}_2(z)$ be score functions that are estimated using the corresponding subsamples, and let

$$(3.2) \qquad \hat{V}_1^* = [\Sigma_{t\in I_2}\hat{s}_1(z_t)\hat{s}_1(z_t)'/n_2]^{-1}, \quad \hat{V}_2^* = [\Sigma_{t\in I_1}\hat{s}_2(z_t)\hat{s}_2(z_t)'/n_1]^{-1}.$$

Our candidate for an efficient estimator will be of the form

$$(3.3) \qquad \hat{\beta} = \tfrac{1}{2}(\tilde{\beta}_1 + \tilde{\beta}_2) + [\hat{V}_2^*\circ\Sigma_{t\in I_1}\hat{s}_2(z_t) + \hat{V}_1^*\circ\Sigma_{t\in I_2}\hat{s}_1(z_t)\ ]/n.$$

Preliminary estimators $\tilde{\beta}_j$, $(j=1,2)$, can be obtained as CLAD estimators for the corresponding subsamples. That is, as

$$(3.4) \qquad \tilde{\beta}_j = \text{argmin}_{\beta\in B}\Sigma_{t\in I_j}|y_t - \min\{x_t'\beta,\ u_t\}|, \quad (j = 1,\ 2),$$

where B denotes the relevant p-dimensional parameter space.

Given preliminary estimators the only remaining unknown component of the efficient score is the conditional density function $f(0|\tilde{x})$ of ε given $\tilde{x}$, evaluated at zero. For technical reasons we consider estimation not of $f(0|\tilde{x})$, but of

$$(3.5) \qquad g(\tilde{x}) = f(0|\tilde{x})\circ 1(x'\beta_0<u)\circ\|x\|.$$

The remainder of the score function is bounded. To estimate the scalar function $g(\tilde{x})$ we use an estimation approach which combines kernel-type estimation of density functions with nearest neighbor estimation of conditional expectations. Specifically, the estimates $\hat{g}_j(\tilde{x},\beta)$ of $g(x)$ based on each subset I_j of observations and a value β for the regression parameters will take the form

(3.6) $$\hat{g}_j(\tilde{x},\beta)$$

$$= \Sigma_{t=1}^{n} W_{ntj}(\tilde{x}) \circ 1(x_t'\beta < u_t) \circ c_n^{-1} \circ 1(x_t'\beta - c_n \leq y_t \leq x_t'\beta) \circ \|x_t\|,$$

where $\{c_n\}$ is a sequence of "window widths" that tend to zero as $n \to \infty$, and $W_{ntj}(\tilde{x})$ are "k_n-nearest neighbor" (k_n-NN) weights which put nonzero weight only on those observations with $t \in I_j$ that have distance between $\tilde{x}_t$ and $\tilde{x}$ among the k_n smallest (with an appropriate tie-breaking rule). The formal definition of these weights is given in Stone (1977), and is quite general in its conditions on the "distance" between $\tilde{x}_t$ and $\tilde{x}$ and its specification of the tie-breaking rule. For our purposes it will suffice that the weights be of the uniform, triangular, or quadratic form proposed by Stone. Specifically, we will impose the following condition on the weights:

Assumption 3.1: For some $W_0 > 0$,

(3.7) $$W_{ntj}(\tilde{x}) = 0 \text{ if } t \notin I_j, \quad 0 \leq W_{ntj}(\tilde{x}) \leq W_0/k_n,$$

$$\Sigma_{t=1}^{n} W_{ntj}(\tilde{x}) = 1.$$

We will also impose the following condition on the limiting behavior of the window width c_n and the number k_n of nearest neighbors.

Assumption 3.2: As $n \to \infty$, the window width c_n and the number of nearest neighbors k_n satisfy

$$c_n \to \infty, \quad k_n \to \infty, \quad k_n/n \to 0, \quad n^{3/4}/(c_n k_n) \to 0.$$

The first three conditions of Assumption 3.2 are standard for kernel and nearest neighbor estimation methods. The fourth condition implies a high degree of smoothing of the corresponding estimators. For example, if $c_n = c_0 \circ n^{-\gamma}$ and $k_n = k_0 \circ n^{\delta}$ for some positive c_0 and k_0, then δ must be in the interval $(3/4, 1)$ and γ in the interval $(0, \delta - 3/4)$. Thus, this condition means that the number of nearest neighbors must be large relative to the sample size, and the window width must shrink slowly with n.

The estimates of $g(\tilde{x})$ can be used to form an estimates of the efficient score as follows. Let $\tilde{\beta}_j$, $(j=1,2)$, be the preliminary CLAD estimators defined in equation (3.4). For $\hat{g}_j(\tilde{x},\beta)$ satisfying equation (3.6) let

$$(3.8) \qquad \hat{s}_j(z) = \hat{g}_j(\tilde{x},\tilde{\beta}_j) \circ 1(x'\tilde{\beta}_j < u) \circ \text{sgn}(y - x'\tilde{\beta}_j) \circ x/\|x\|, \quad (j=1,2),$$

where $1/\|x\|$ is well defined as long as the regression includes a constant. Let $\hat{V}_j^*$, $(j=1,2)$, be obtained from equation (3.2) using the score estimates given in equation (3.8). Our candidate $\hat{\beta}$ for an efficient estimator is then as given by equation (3.3).

To characterize the asymptotic distribution of $\hat{\beta}$ it is useful to strengthen somewhat the conditions we impose on the distribution of $(\varepsilon_t, \tilde{x}_t')$.

Assumption 3.3: The conditional density $f(\varepsilon|\tilde{x})$ is bounded and Lipshitz uniformly in $\tilde{x}$. That is, there are positive constants f_0 and M_0 such that $f(\varepsilon|\tilde{x}) \le f_0$ and $|f(\varepsilon_1|\tilde{x}) - f(\varepsilon_2|\tilde{x})| \le M_0|\varepsilon_1 - \varepsilon_2|$.

With some additional notation and moment conditions the density bound f_0 and the Lipshitz constant M_0 can be allowed to vary with the regressors, but are assumed constant here for simplicity.

Assumption 3.4: For some positive constants K_0 and ν_0, $E[1(|\tilde{x}_t'\alpha| \le \|\tilde{x}_t\|\delta)(1 + \|x_t\|^2)] \le K_0\delta$ for $\|\beta - \beta_0\| \le \nu_0$, $|\delta| \le \nu_0$.

The following condition is useful in showing $\sqrt{n}$-consistency of $\tilde{\beta}_j$.

Assumption 3.5: The parameter space B is compact and contains β_0 in its interior.

Theorem 3.1: Suppose that Assumptions 2.1 - 2.6 and 3.1 - 3.5 are satisfied. Then

$$(3.8) \qquad \sqrt{n}(\hat{\beta}-\beta_0) \xrightarrow{d} N(0,V^*), \quad \hat{V}^* = (\hat{V}_1^* + \hat{V}_2^*)/2 \xrightarrow{p} V^*,$$

and $\hat{\beta}$ is locally regular for the class Q of parametric submodels defined in the statement of Theorem 1

4. Extensions

Two immediate extensions of the results of Sections 2 and 3 are available. The first is the extension to the linear regression model without censoring. By ignoring terms that involve censoring (i.e. treating the censoring point u_t as if it were equal to $+\infty$) it follows immediately from the calculations done in the proof of Theorem 2.1 that the efficient score for the linear model with a conditional median zero disturbance is $s(z_t) = 2f(0|\tilde{x}_t)\circ sgn(y_t-x_t'\beta_0)\circ x_t$, with corresponding semiparametric efficiency bound $V_\beta = \{E[s(z_t)s(z_t)']\}^{-1}$. Similar to the discussion in in the introduction, this efficiency bound can be interpreted as the covariance matrix of an optimally weighted least absolute deviations (LAD) estimator, where the weighting accounts for the fact that the conditional density function $f(0|\tilde{x}_t)$ depends on x_t. Furthermore, an asymptotically efficient estimator, involving nonparametric estimation of $f(0|\tilde{x}_t)$, can be constructed by following the procedure outlined in Section 3 while omitting terms involving $1(x'\beta < u)$ in the construction of $\tilde{g}_j(x,\beta)$ and using LAD estimators as the initial estimators of β. The proof of Theorem 3.1 applies to this case.

The second extension of the results is to the case where the τ^{th} conditional quantile of ε_t is zero, i.e. $F(0|\tilde{x}_t) = \tau$, where τ is not necessarily equal to 1/2 (and there is no loss of generality in assuming a value of zero for the quantile as long as x_t includes a constant). Noting that this imposes the restriction $\int(\tau-1(\varepsilon<0))f(\varepsilon|\tilde{x})d\varepsilon = 0$, the proof of Theorem 1 can easily be modified to show that the efficient score is given by $s(z) = [\tau(1-\tau)]^{-1/2}\circ f(0|\tilde{x})\circ 1(x'\beta_0 < u)\circ[\tau - 1(y-x'\beta_0 < 0)]$ with efficiency bound $V_\beta = \{E[s(z_t)s(z_t)']\}^{-1}$. An asymptotically efficient estimator can be constructed by following the procedure outlined in Section 3 while replacing $sgn(y-x'\beta)$ with $[\tau(1-\tau)]^{-1/2}\circ[\tau - 1(y-x'\beta_0 < 0)]$ and using for initial estimates Powell's (1986a) censored regression quantiles.

5. Proofs of Theorems

We first establish some notation and terminology. For any measure μ let $L_2(\mu)$ denote the Hilbert space of functions that are square integrable with respect to μ, and let $\|\circ\|_\mu = \int(\circ)^2 d\mu$ denote its norm. H-continuity of functions of and $\theta = (\beta', \eta')'$ will mean continuity in θ for the norm $\|\circ\|_\mu$. Similarly, H-differentiability will mean Frechet differentiability in $L_2(\mu)$.

In defining the likelihood it is useful to work with the transformed dependent variable $\tilde{y} = y - u$. Then for a given parametric conditional density $f(\varepsilon|\tilde{x},\eta)$ and for each β and η the distribution of $\tilde{z} = (\tilde{y}, \tilde{x}')'$ is absolutely continuous with respect to $\tilde{\mu} = \tilde{\ell} \times p_x$, where $\tilde{\ell}$ is the sum of a point mass at zero and Lebesgue measure on $(-\infty, 0)$ and p_x is the probability measure of $\tilde{x}_t$. The density of $\tilde{z}$ is given by

$$(5.1) \qquad f(\tilde{z}|\beta,\eta) = 1(\tilde{y} = 0)[1 - F(-\tilde{x}'\alpha|\tilde{x},\eta)] + 1(\tilde{y} < 0)f(\tilde{y}-\tilde{x}'\alpha|\tilde{x},\eta),$$

where $\alpha = (\beta', -1)'$, $F(m|\tilde{x},\eta) = \int_{-\infty}^{m} f(\varepsilon|\tilde{x},\eta)d\varepsilon$.

Lemma 5.1: If Assumptions 3.1-3.6 are satisfied then $f(\tilde{z}|\theta)^{1/2}$ is H-continuously H-differentiable in a neighborhood of $\theta_0 = (\beta_0', 0)'$, in $L_2(\tilde{\mu})$. Also, if $s(\tilde{z})$ is any vector such that $E[\|s(\tilde{z}_t)\|^2]$ is finite then $s(\tilde{z})f(\tilde{z}|\beta_0,\eta)^{1/2}$ is H-continuous at η_0.

Proof: To prove the first conclusion it suffices to prove H-continuous H-differentiability of $R(-\tilde{x}'\alpha, \tilde{x}, \eta)$ in $L_2(p_x)$ and $f(\varepsilon + x'\delta|\tilde{x},\eta)^{1/2}$ in $L_2(\ell \times p_x)$ respectively, where $R(m,\tilde{x},\eta) = [1-F(m|\tilde{x},\eta)]^{1/2} = [\int_m^\infty f(\varepsilon|\tilde{x},\eta)d\varepsilon]^{1/2}$. For notational convenience, suppress the x argument in v. Let N_η be an open ball around zero such that $\inf_{\eta\in N_\eta}[1+(v'\eta)q(\varepsilon,\tilde{x})] \geq b > 0$ a.s. $\ell \times p_x$, and restrict attention to N_η and drop the a.s. qualifier henceforth. Then, $f(\varepsilon|\tilde{x},\eta) \geq bf(\varepsilon|\tilde{x})$ and by boundedness $f(\varepsilon|\tilde{x},\eta) \leq Bf(\varepsilon|\tilde{x})$ for some $B > 0$.

Consider first $R(-\tilde{x}'\alpha, \tilde{x}, \eta)$. Continuous differentiability of

$R(m,\tilde{x},\eta)$ in m and η when $F(m|\tilde{x}) < 1$ follows immediately from the chain rule and $\int_{-\infty}^{m} f(\varepsilon|\tilde{x})q(\varepsilon,\tilde{x})d\varepsilon$ finite for all m. The derivative is given by $-[2R(m,\tilde{x},\eta)]^{-1}(f(m|\tilde{x},\eta), v'\int_{-\infty}^{m} f(\varepsilon|\tilde{x})q(\varepsilon,\tilde{x})d\varepsilon)'$. Consider a point $(\bar{m},\tilde{x},\bar{\eta})$ such that $F(\bar{m}|\tilde{x}) = 1$, and sequences $\{v_i\}_{i=1}^{\infty}$, $\{\eta_i\}_{i=1}^{\infty}$, with $m_i \to \bar{m}$ and $\eta_i \to \bar{\eta}$. Note that by $F(m|\tilde{x}) = 1$ for $m > \bar{m}$ and Assumption 2.4, $R(m,\tilde{x})$ is differentiable at $\bar{m}$ with derivative zero. Then

$$(5.2) \qquad 0 \le R(m_i,\tilde{x},\eta_i)/(|m_i-\bar{m}|+\|\eta_i-\bar{\eta}\|) \le B^{1/2}R(m_i,\tilde{x})/|m_i-\bar{m}| \to 0.$$

Thus, $R(m,\tilde{x},\eta)$ is differentiable in m and η, with derivative zero, at $(\bar{m},\tilde{x},\bar{\eta})$. To check continuity of the derivative at a point with $F(\bar{m}|\tilde{x}) = 1$, consider sequences as before, and without loss of generality (w.l.g.) suppose that $F(m_i|\tilde{x}) < 1$ for all i. Then

$$(5.3) \qquad |\partial R(m_i,\tilde{x},\eta_i)/\partial m| = |[2R(v_i,\tilde{x},\eta_i)]^{-1}f(m_i|\tilde{x},\eta_i)|$$
$$\le [B/(b^{1/2})]|[2R(m_i,\tilde{x})]^{-1}f(m_i|\tilde{x})| = [B/(b^{1/2})]|R_m(m_i,\tilde{x})| \to 0$$

by continuity of $R_m(m,\tilde{x})$ and $R_m(\bar{m},\tilde{x}) = 0$. Furthermore, for some positive constant C,

$$(5.4) \qquad \|\partial R(m_i,\tilde{x},\eta_i)/\partial\eta\| = [2R(m_i,\tilde{x},\eta_i)]^{-1}\|v'\int_{-\infty}^{m_i} f(\varepsilon|\tilde{x})q(\varepsilon,\tilde{x})d\varepsilon\|$$
$$= [2R(m_i,\tilde{x},\eta_i)]^{-1}\|-v'\int_{m_i}^{\infty} f(\varepsilon|\tilde{x})q(\varepsilon,\tilde{x})d\varepsilon\|$$
$$\le [2bR(m_i,\tilde{x})]^{-1}\sup(\|v\||q(\varepsilon,\tilde{x})|)\int_{m_i}^{\infty} f(\varepsilon|\tilde{x})d\varepsilon \le C[1-F(m_i|\tilde{x})]^{1/2} \to 0.$$

Thus $R(-\tilde{x}'\alpha,\tilde{x},\eta)^{1/2}$ is continuously differentiable in β and η. To show H-continuous H-differentiability it suffices (by a mean-value expansion argument and the dominated convergence theorem (DCT)) to show that the derivative is dominated by an element of $L_2(\rho_x)$. This dominance follows from Assumption 2.4 and the same calculation used in the above pair of equations. For example, $\|\partial F(-\tilde{x}'\alpha,\tilde{x},\eta)^{1/2}/\partial\eta\| \le CR(-\tilde{x}'\alpha,\tilde{x})^{1/2}$ (as in equation (5.4)) yields dominance of $\|\partial F(-\tilde{x}'\alpha,\tilde{x},\eta)^{1/2}/\partial\eta\|$ by $C\gamma(\tilde{x})$.

Next, we consider $f(\varepsilon+x'\delta|x,\eta)$. For the moment let $\theta = (\delta',\eta')$. Also, let

$$(5.5)\quad r(\delta) = r(\varepsilon+x'\delta,\tilde{x}) = f(\varepsilon+x'\delta|\tilde{x})^{1/2},$$

$$d(\delta,\eta) = [1+(v'\eta)q(\varepsilon+x'\delta,\tilde{x})]^{1/2}, \quad r_\varepsilon(\delta) = r_\varepsilon(\varepsilon+x'\delta,\tilde{x}),$$

$$d_\varepsilon(\delta,\eta) = \partial d(\delta,\eta)/\partial\varepsilon = [4\{1+(v'\eta)q(\varepsilon+x'\delta,\tilde{x})\}]^{-1/2}(v'\eta)q_\varepsilon(\varepsilon+x'\delta,\tilde{x}),$$

$$d_\eta(\delta,\eta) = \partial d(\delta,\eta)/\partial\eta = [4\{1+(v'\eta)q(\varepsilon+x'\delta,\tilde{x})\}]^{-1/2}q(\varepsilon+x'\delta,\tilde{x})v,$$

$$r(\theta) = r(\delta)d(\delta,\eta), \quad r_\delta(\theta) = [r_\varepsilon(\delta)d(\delta,\eta) + r(\delta)d_\varepsilon(\delta,\eta)]x,$$

$$r_\eta(\theta) = r(\delta)d_\eta(\delta,\eta),$$

$$r_\theta(\theta) = (r_\delta(\theta)',r_\eta(\theta)')',$$

where dependence on ε and x has been suppressed for notational convenience. Note that by Assumptions 2.2 and 2.3 and by construction $r(\delta)$, $r_\varepsilon(\delta)$, $d(\delta,\eta)$, $d_\varepsilon(\delta,\eta)$, and $d_\eta(\delta,\eta)$ are each continuous at (δ,η) a.s. $\ell x \rho_x$, so that $r_\delta(\delta,\eta)$ and $r_\eta(\delta,\eta)$ are also continuous a.e. $\ell x \rho_x$. Furthermore, it follows from the dominance conditions of Assumption 2.3 that $r_\delta(\delta,\eta)$ and $r_\eta(\delta,\eta)$ are each dominated by square integrable functions. For example, on $N \times N_\eta$, $\|r_\eta(\delta,\eta)\| \le r(\delta)(4b)^{-1/2}\sup\|vq(\varepsilon,\tilde{x})\| \le C\gamma(\varepsilon,\tilde{x})$ for some constant C. It follows from DCT that $r_\delta(\delta,\eta)$ and $r_\eta(\delta,\eta)$ are mean-square continuous. To show H-differentiability, for the moment suppress the subscript on $\|\circ\|_{\ell x p_x}$, and let $\theta_i = (\delta_i',\eta_i')' \to (\bar{\delta}',\bar{\eta}')' = \bar{\theta}$ where w.l.g. $\delta_i \ne \bar{\delta}$ and $\eta_i \ne \bar{\eta}$. Then by $d(\delta,\eta)$ continuously differentiable in η and H-continuity of $r_\eta(\theta)$ a mean value expansion of $r(\delta_i,\eta_i)$ around $\bar{\eta}$ yields, for the mean value $\mathring{\eta}_i = \mathring{\eta}_i(\varepsilon,\tilde{x})$,

$$(5.6)\quad \|r(\delta_i,\eta_i) - r(\delta_i,\bar{\eta}) - r_\eta(\bar{\theta})'(\eta_i-\bar{\eta})\|/\|\theta_i-\bar{\theta}\|$$

$$= \|[r_\eta(\delta_i,\mathring{\eta}_i) - r_\eta(\bar{\theta})]'(\eta_i-\bar{\eta})\|/\|\theta_i-\bar{\theta}\|$$

$$\le \|r_\eta(\delta_i,\mathring{\eta}_i) - r_\eta(\bar{\theta})\| \to 0.$$

Similarly, a mean value expansion of $d(\delta_i,\bar{\eta})$ around $\bar{\delta}$ yields

$$(5.7)\quad \|r(\delta_i)d(\delta_i,\bar{\eta}) - r(\delta_i)d(\bar{\delta},\bar{\eta}) - r(\bar{\delta})d_\varepsilon(\bar{\delta},\bar{\eta})x'(\delta_i-\bar{\delta})\|/\|\theta_i-\bar{\theta}\|$$

$$= \|[r(\delta_i)d_\varepsilon(\delta_i,\mathring{\eta}) - r(\bar{\delta})d_\varepsilon(\bar{\delta},\bar{\eta})]x'(\delta_i-\bar{\delta})\|/\|\theta_i-\bar{\theta}\|$$

$$\le \|[r(\delta_i)d_\varepsilon(\delta_i,\mathring{\eta}) - r(\bar{\delta})d_\varepsilon(\bar{\delta},\bar{\eta})]x\| \to 0,$$

by continuity and dominance of $r(\delta)d_\varepsilon(\delta,\eta)x$. Finally, it follows from the H-differentiability of $r(\delta)$ and $d(\delta,\eta)$ bounded that

$$(5.8)\quad \|r(\delta_i)d(\bar{\delta},\bar{\eta}) - r(\bar{\delta})d(\bar{\delta},\bar{\eta}) - r_\varepsilon(\bar{\delta})d(\bar{\delta},\bar{\eta})x'(\delta_i-\bar{\delta})\|/\|\theta_i-\bar{\theta}\|$$

$$\le B\|r(\delta_i) - r(\bar{\delta}) - r_\varepsilon(\bar{\delta})x'(\delta_i-\bar{\delta})\|/\|\delta_i-\bar{\delta}\| \to 0.$$

The first conclusion of the lemma now follows from equations (5.6) - (5.8) and the triangle inequality.

The second conclusion follows by $R(m,\tilde{x},\eta)$ and $f(\varepsilon|\tilde{x},\eta)^{1/2}$ continous at $\eta_0 = 0$, $[1-F(-\tilde{x}'\alpha_0|\tilde{x},\eta)]^{1/2} \le B^{1/2}[1-F(-\tilde{x}'\alpha_0|\tilde{x})]^{1/2}$, $f(\varepsilon|\tilde{x},\eta)^{1/2} \le B^{1/2}f(\varepsilon|\tilde{x})^{1/2}$, and the DCT.

The following two Lemmas will be used to show $V_\beta = V^*$. Define $\mathcal{P}_R$ to be the parametric submodel corresponding to $f(\tilde{z}|\theta)$ such that θ lies in an open set contained in $([N+\{\beta_0\}]\cap N_0)\times N_\eta$ that contains θ_0. Let $r_\theta = (r_\beta', r_\eta')'$ denote the $L_2(\tilde{\mu})$ H-derivative of $f(\tilde{z}|\theta)^{1/2}$ at θ_0.

Lemma 5.2: Suppose that there is $\rho \in L_2(\tilde{\mu})$ such that for all $\mathcal{P}_R \in Q$, $\int \rho r_\eta' d\tilde{\mu} = 0$. Also suppose there exists a sequence $\mathcal{P}_R^j \in Q$ and a sequence A_j of conformable constant matrices such that $\int (r_\beta - A_j r_\eta^j - \rho)'(r_\beta - A_j r_\eta^j - \rho)d\tilde{\mu} \to 0$. Then $\int \rho\rho' d\tilde{\mu} = \inf_{\mathcal{P}_R \in Q, A} \int (r_\beta - Ar_\eta)(r_\beta - Ar_\eta)' d\tilde{\mu}$.

Proof: The hypotheses imply that $\int (r_\beta - A_j r_\eta^j)(r_\beta - A_j r_\eta^j)' d\tilde{\mu}$ converges to $\int \rho\rho' d\tilde{\mu}$. Therefore it suffices to show that for any $\mathcal{P}_R \in Q$ and conformable λ, $\lambda'[\int \rho\rho' d\tilde{\mu}]\lambda \le \lambda'[\int (r_\beta - Ar_\eta)(r_\beta - Ar_\eta)' d\tilde{\mu}]\lambda$. Let $\bar{\rho} = \lambda'\rho$, $\bar{r}_\beta = \lambda' r_\beta$, $a = \lambda' A$, and $a_j = \lambda' A_j$. Note that

$$(5.9)\qquad \lambda'[\int(r_\beta - Ar_\eta)(r_\beta - Ar_\eta)'d\tilde{\mu}]\lambda = \|\bar{r}_\beta - a_j r_\eta^j + a_j r_\eta^j - ar_\eta\|^2$$

$$= \|\bar{r}_\beta - a_j r_\eta^j\|^2 + \|a_j r_\eta^j - ar_\eta\|^2 + 2\int(\bar{r}_\beta - a_j r_\eta^j)(a_j r_\eta^j - ar_\eta)d\tilde{\mu},$$

where the $\tilde{\mu}$ subscript has been suppressed for notational convenience. Since $\lim(\|\bar{r}_\beta - a_j r_\eta^j\|^2) = \|\bar{\rho}\|^2 = \lambda'[\int\rho\rho' d\tilde{\mu}]\lambda$ by hypothesis and $\liminf(\|a_j r_\eta^j - ar_\eta\|^2)$ is nonnegative, it suffices to show that $\int(\bar{r}_\beta - a_j r_\eta^j)(a_j r_\eta^j - ar_\eta)d\tilde{\mu}$ converges to zero. By orthogonality of ρ and r_η^j and the Cauchy-Schwartz inequality it follows that

$$(5.10)\qquad |\int(\bar{r}_\beta - a_j r_\eta^j)(a_j r_\eta^j)d\tilde{\mu}| = |\int(\bar{r}_\beta - a_j r_\eta^j - \bar{\rho})(a_j r_\eta^j)d\tilde{\mu}|$$

$$\le \|\bar{r}_\beta - a_j r_\eta^j - \bar{\rho}\| \circ \|a_j r_\eta^j\|.$$

Since boundedness of $\|a_j r_\eta^j\|$ follows from convergence of $a_j r_\eta^j$ to $\bar{r}_\beta - \bar{\rho}$, it follows from this inequality that $\int(\bar{r}_\beta - a_j r_\eta^j)(a_j r_\eta^j)d\tilde{\mu}$ converges to zero. A similar argument can then be used to show that $\int(\bar{r}_\beta - a_j r_\eta^j)(a_j r_\eta^j)d\tilde{\mu}$ converges to zero.

Lemma 5.3: Suppose $s(\tilde{z})$ is a vector such that the components of $s(\tilde{z})f(\tilde{z}|\beta_0,\eta)^{1/2}$ are elements of $L_2(\tilde{\mu})$ for η in a neighborhood of η_0 and are H-continuous at η_0. If $\int s(\tilde{z})f(\tilde{z}|\beta_0,\eta)d\tilde{\mu} = 0$ for all η in a neighborhood of η_0 then $\int s(\tilde{z})f(\tilde{z}|\theta_0)^{1/2}r_\eta' d\tilde{\mu} = 0$.

Proof: Consider an element s of $s(z)$, and let $f(\eta) = f(\tilde{z}|\beta_0,\eta)$ and $f_0 = f(\tilde{z}|\theta_0)$. Consider a vector λ with $\|\lambda\| = 1$. For $\nu_j > 0$ such that $\nu_j \to 0$ let $\eta_j = \eta_0 + \nu_j\lambda$. Note that by H-differentiability and the Cauchy-Schwartz and triangle inequalities,

$$(5.11)\qquad |\int[sf(\eta_j)^{1/2} + sf_0^{1/2}][f(\eta_j)^{1/2} - f_0^{1/2} - r_\eta'(\eta_j - \eta_0)]d\tilde{\mu}|/\|\eta_j - \eta_0\|$$

$$\le (\|sf(\eta_j)^{1/2}\| + \|sf_0^{1/2}\|)\|f(\eta_j)^{1/2} - f_0^{1/2} - r_\eta'(\eta_j - \eta_0)\|/\|\eta_j - \eta_0\|$$

$$\longrightarrow 0.$$

where the last line follows by $\|s\ell(\eta)^{1/2}\|$ continuous (and therefore bounded). It follows from this equation and H-continuity of $sf(\eta)^{1/2}$, that

(5.12) $$0 = [\int s\ell(\eta_j)d\tilde{\mu} - \int s\ell_0 d\tilde{\mu}]/\nu_j$$
$$= \{\int s[\ell(\eta_j)^{1/2}+\ell_0^{1/2}][\ell(\eta_j)^{1/2}-\ell_0^{1/2}-r_\eta'(\eta_j-\eta_0)]d\tilde{\mu}/\|\eta_j-\eta_0\|\}$$
$$+ \{\int s[\ell(\eta_j)^{1/2}+\ell_0{}^{1/2}]r_\eta' d\tilde{\mu}\}\lambda \rightarrow 2\{\int s\ell_0^{1/2} r_\eta' d\tilde{\mu}\}\lambda.$$

The conclusion follows from the arbitrary choice of λ.

The following Lemma is useful for verifying that the $q(\varepsilon,\tilde{x})$ given in equation (2.3) can be approximated by a function satisfying Assumption 2.1. For the moment let μ be the probability measure of $(\varepsilon_t,\tilde{x}_t')$.

Lemma 5.4: If Assumption 3.1 is satisfied, then the functions $q(\varepsilon,\tilde{x})$ that are bounded and continuously differentiable in ε with bounded derivative are dense in the subset M of $L_2(\mu)$ satisfying $E[q(\varepsilon,\tilde{x})|\tilde{x}] = 0$ and $E[sgn(\varepsilon)q(\varepsilon,\tilde{x})|\tilde{x}] = 0$.

Proof: Let $p_0(\varepsilon) = \varepsilon/[1+|\varepsilon|]$, and note that $p_0(\varepsilon)$ is a bounded, continuously differentiable function with bounded derivative. For a positive integer J, let $p_J(\varepsilon) = (1, p_0(\varepsilon), p_0(\varepsilon)^2, \ldots, p_0(\varepsilon)^{J-1})'$. Note that if $d_J(\tilde{x})$ is a $J \times 1$ vector of bounded functions of x then $q_J(\varepsilon,\tilde{x}) = d_J(\tilde{x})'p_J(\varepsilon)$ will be bounded and continuously differentiable in ε with bounded derivative. Therefore, the result will be proved if for any $q(\varepsilon,\tilde{x}) \in M$ we can construct bounded vectors $d_J(\tilde{x})$ such that $q_J(\varepsilon,\tilde{x}) \in M$ and $\lim_{J\rightarrow\infty}E[\{q(\varepsilon,\tilde{x}) - q_J(\varepsilon,\tilde{x})\}^2] = 0$.

Let $\tilde{H}(\tilde{x})$ denote the Hilbert space of measurable functions $m(\varepsilon)$ such that $E[m(\varepsilon)^2|\tilde{x}]$ exists and is finite at $\tilde{x}$, with inner product $\langle m_1|m_2\rangle_x = E[m_1(\varepsilon)m_2(\varepsilon)|\tilde{x}]$. Similarly let $N_1(\tilde{x})$ denote the linear span of the components of $p_J(\varepsilon)$ and let $N_2(\tilde{x}) = \{m(\varepsilon)|m(\varepsilon) \in H(\tilde{x}), E[m(\varepsilon)|\tilde{x}] = E[sgn(\varepsilon)m(\varepsilon)|\tilde{x}] = 0\}$. Consider $q(\varepsilon,\tilde{x}) \in M$. Note that $E[q(\varepsilon,\tilde{x})^2|\tilde{x}]$, $E[\|p_J(\varepsilon)\|^2|\tilde{x}]$, $E[sgn(\varepsilon)^2|\tilde{x}]$ and are finite a.s., so that

$q(\varepsilon,\tilde{x})$, the elements of $p_J(\varepsilon)$, and $sgn(\varepsilon)$ are each in $H(\tilde{x})$ a.s.. Let $\bar{q}_J(\varepsilon,\tilde{x}) = \bar{d}_J(\tilde{x})'p_J(\varepsilon)$ and $sgn_J(\varepsilon,\tilde{x}) = d^s_J(\tilde{x})'p_J(\varepsilon)$ denote the projection on $N_1(\tilde{x})$ of $q(\varepsilon,\tilde{x})$ and $sgn(\varepsilon)$ respectively, and let $\hat{q}_J(\varepsilon,\tilde{x}) = c_J(\tilde{x})'(1,sgn_J(\varepsilon,\tilde{x}))$, denote the projection of $\bar{q}_J(\varepsilon,\tilde{x})$ on 1 and $sgn_J(\varepsilon,\tilde{x})$. Also, let $q^*_J(\varepsilon,\tilde{x}) = d^*_J(\tilde{x})'p_J(\varepsilon)$ denote the projection of $q(\varepsilon,\tilde{x})$ on $N_1(\tilde{x})\cap N_2(\tilde{x})$.

The fact that each vector of functions of $\tilde{x}$ can be chosen to be measurable follows in a straightforward way. For example, $\bar{d}_J(\tilde{x})$ can be computed as $\{E[p^x_J(\varepsilon)p^x_J(\varepsilon)'|\tilde{x}]\}^{-1}E[p^x_J(\varepsilon)q(\varepsilon,\tilde{x})|\tilde{x}]$, where $p^x_J(\varepsilon)$ is the largest subvector of $p_J(\varepsilon)$ with nonsingular conditional second moment matrix at $\tilde{x}$. Measurability of $\bar{d}_J(\tilde{x})$ then follows from measurability of the conditional second moment matrices, continuity of the matrix inverse function where it is nonsingular, and the fact that the composition of $p^x_J(\varepsilon)$ is determined by which determinants (which are continuous functions) of submatrices of the conditional second moment matrix of $p_J(\varepsilon)$ is zero. Also, it is straightforward to show that each conditional projection is square integrable. For example, by construction $\int(q(\varepsilon,\tilde{x})-\bar{q}_J(\varepsilon,\tilde{x}))^2f(\varepsilon|\tilde{x})d\varepsilon \le \int(q(\varepsilon,\tilde{x}))^2f(\varepsilon|\tilde{x})d\varepsilon$ a.e. so that by $\bar{q}_J(\varepsilon,\tilde{x})^2 \le 2\{[q(\varepsilon,\tilde{x})-\bar{q}_J(\varepsilon,\tilde{x})]^2 + q(\varepsilon,\tilde{x})^2\}$ it follows that $\int\bar{q}_J(\varepsilon,\tilde{x})^2f(\varepsilon|\tilde{x})d\varepsilon \le 4\int q(\varepsilon,\tilde{x})^2f(\varepsilon|\tilde{x})d\varepsilon$ a.e., so that $E[\bar{q}_J(\varepsilon,\tilde{x})^2] = \int[\int\bar{q}_J(\varepsilon,\tilde{x})^2f(\varepsilon|\tilde{x})d\varepsilon]d\mu_x$ exists.

Now, note that from the fact that $N_1(\tilde{x})\cap N_2(\tilde{x})$ grows with J (and the positive integers are countable so that one can consider a.s. statements that apply for all J) $E[\{q(\varepsilon,\tilde{x}) - q^*_J(\varepsilon,\tilde{x})\}^2|\tilde{x}]$ is monotonically decreasing in J a.s.. Also, it is a simple exercise in projection theory to show that

$$(5.13)\qquad E[\{q(\varepsilon,\tilde{x}) - q^*_J(\varepsilon,\tilde{x})\}^2|\tilde{x}]$$

$$= E[\{q(\varepsilon,\tilde{x}) - \bar{q}_J(\varepsilon,\tilde{x})\}^2|\tilde{x}] + E[\hat{q}_J(\varepsilon,\tilde{x})^2|\tilde{x}].$$

(A detailed proof of this equality is available from the authors upon request.) By $p_0(\varepsilon)$ monotonic, continuously differentiable with everywhere nonzero derivative, and bounded, the sequence $\{p_0(\varepsilon)^j\}_{j=1}^{\infty}$ is a basis for $\tilde{H}(\tilde{x})$ a.e. (e.g. see Lemma 2 of Newey (1987)). Therefore,

$$(5.14) \qquad 0 = \lim_{J\to\infty}E[\{q(\varepsilon,\tilde{x}) - \bar{q}_J(\varepsilon,\tilde{x})\}^2|\tilde{x}]$$

$$= \lim_{J\to\infty}E[\{sgn(\varepsilon) - sgn_J(\varepsilon,\tilde{x})\}^2|\tilde{x}],$$

a.e.. Also, by Assumption 2.5 the conditional second moment matrix of $(1,sgn(\varepsilon))$ is nonsingular a.s.. It is then a simple exercise in projection theory to show that it follows from equation (5.14) and the fact that 1 is the projection of 1 on $N_1(\tilde{x})$, that the projection of $\bar{q}_J(\varepsilon,\tilde{x})$ on $(1,sgn_J(\varepsilon,\tilde{x}))$ converges in $\tilde{H}(\tilde{x})$ to the projection of $q(\varepsilon,\tilde{x})$ $(1,sgn(\varepsilon))$ a.s., which is zero by $q(\varepsilon,\tilde{x}) \in M$. Therefore, $\lim_{J\to\infty}E[\hat{q}_J(\varepsilon,\tilde{x})^2|\tilde{x}] = 0$ a.s.. Thus, by equations (5.13) and (5.14), $\lim_{J\to\infty}E[\{q(\varepsilon,\tilde{x}) - q_J^*(\varepsilon,\tilde{x})\}^2|\tilde{x}] = 0$ a.s., so that by the monotone convergence theorem $\lim_{J\to\infty}E[\{q(\varepsilon,\tilde{x}) - q_J^*(\varepsilon,\tilde{x})\}^2] = 0$.

For each $D > 0$ and J let $q_J^D(\varepsilon,\tilde{x}) = q_J^*(\varepsilon,\tilde{x})$ for $\|d_J(\tilde{x})\| \le D$ and $q_J^D(\varepsilon,\tilde{x}) = 0$ otherwise. Note that by construction $q_J^D(\varepsilon,\tilde{x}) \in M$ and by the choice of $p_J(\varepsilon)$, $q_J^D(\varepsilon,\tilde{x})$ is bounded and continuously differentiable in ε with bounded derivative. Also, $|q_J^D(\varepsilon,\tilde{x})| \le |q_J^*(\varepsilon,\tilde{x})|$ and $\lim_{B\to\infty}q_J^D(\varepsilon,\tilde{x}) = q_J^*(\varepsilon,\tilde{x})$ so that by the dominated convergence theorem for each J there is $D(J)$ such that $E[\{q_J^{D(J)}(\varepsilon,\tilde{x}) - q_J^*(\varepsilon,\tilde{x})\}^2] \le 1/J$. The conclusion then follows by taking $q_J(\varepsilon,\tilde{x}) = q_J^{D(J)}(\varepsilon,\tilde{x})$.

Proof of Theorem 2.1: For each parametric submodel $\mathcal{P}_R \in Q$ the information matrix $I(\mathcal{P}_R,\theta_0)$ is given by

$$(5.15) \qquad I(\mathcal{P}_R,\theta_0) = 4\int r_\theta r_\theta' d\tilde{\mu}.$$

which is well defined by Lemma 5.1. By hypothesis the information matrix is nonsingular, so that for each $\mathcal{P}_R \in Q$, $I^{\beta\beta}(\mathcal{P}_R,\theta_0)$ is well defined, and can be calculated as

(5.16) $$I^{\beta\beta}(\mathcal{P}_R,\theta_0) = [4\int(r_\beta-Ar_\eta)(r_\beta-Ar_\eta)'d\tilde{\mu}]^{-1},$$

$$A = [\int r_\eta r_\eta' d\tilde{\mu}]^{-1}\int r_\eta r_\beta' d\tilde{\mu}.$$

Let $s(\tilde{z}) = 2f(0|\tilde{x})1(\tilde{x}'\alpha_0<0)\mathrm{sgn}(\tilde{y}-\tilde{x}'\alpha_0)x$ and $\rho(\tilde{z}) = s(\tilde{z})f(\tilde{z}|\theta_0)^{1/2}$. To complete the proof, by Lemmas 5.1 - 5.3 it suffices to show that for each $\mathcal{P}_R \in Q$, $\int s(\tilde{z})f(\tilde{z}|\beta_0,\eta)d\tilde{\mu} = 0$ in a neighborhood of η_0, and that there exists $\mathcal{P}_R$ and A such that $\|r_\beta-Ar_\eta-\rho\|_p$ is arbitrarily close to zero, where $\|\circ\|_p = [\int(\circ)'(\circ)d\tilde{\mu}]^{1/2}$.

We will first show $\int s(\tilde{z})f(\tilde{z}|\beta_0,\eta)d\tilde{\mu} = 0$ in a neighborhood of η_0. Note that $\mathrm{sgn}(\tilde{y}-\tilde{x}'\alpha_0) = 1$ for $\tilde{x}'\alpha_0 < 0$ and $\tilde{y} = 0$, so that

(5.17) $$\int s(\tilde{z})f(\tilde{z}|\beta_0,\eta)d\tilde{\mu} = \int x\circ f(0|\tilde{x})\circ 1(\tilde{x}'\alpha_0<0) \circ$$

$$\{ 1-F(-\tilde{x}'\alpha_0|\tilde{x},\eta) + \int_{-\infty}^{0}\mathrm{sgn}(\tilde{y}-\tilde{x}'\alpha_0)f(\tilde{y}-\tilde{x}'\alpha_0|\tilde{x},\eta)d\tilde{y} \}\circ d\mathcal{P}_x$$

$$= \int x\circ f(0|\tilde{x})\circ 1(\tilde{x}'\alpha_0<0)\circ\{\int_{-\infty}^{\infty}\mathrm{sign}(\varepsilon)f(\varepsilon|\tilde{x},\eta)d\varepsilon\}\circ d\mathcal{P}_x = 0,$$

where the last equality follows by Assumptions 2.1 and 2.2.

We now show that there exists $\mathcal{P}_R \in Q$ and A such that $\|\rho-(r_\beta-Ar_\eta)\|_p$ is arbitrarily small. Let $v = x$ for $\|x\| \le D$ and $v = 0$ otherwise, for some positive constant D. Let ρ^v and r_β^v be the vectors obtained by replacing the vector component x of ρ and r_β, respectively, by v. Note that $\|\rho-\rho^v-(r_\beta-r_\beta^v)\|_p^2 \le 2(\|\rho\|^2+\|r_\beta\|^2)$, so that by the DCT and $\lim_{D\to\infty}1(v\ne x) = 0$ everywhere,

(5.18) $$\lim_{D\to\infty}\|\rho-\rho^v-(r_\beta-r_\beta^v)\|_p^2 \le 2\lim_{D\to\infty}\int 1(v\ne x)(\|\rho\|^2+\|r_\beta\|^2)d\tilde{\mu} = 0,$$

Also, for fixed D it follows by $\mathrm{Prob}(\tilde{x}_t'\alpha_0 = 0) = 0$ that for $\nu > 0$ and $1_\nu = 1(0 \le \tilde{x}'\alpha_0 < \nu)$, $\lim_{\nu\to 0}1_\nu = 0$ a.e. $\tilde{\mu}$. Then by $\|r_\beta^v\| \le \|r_\beta\|$ and $\|\rho^v\| \le \|\rho\|$, the triangle inequality, and DCT

(5.19) $$\lim_{\nu\to 0}\|1_\nu(\rho^v-r_\beta^v)\|_p \le \lim_{\nu\to 0}(\|1_\nu\rho^v\|_p + \|1_\nu r_\beta^v\|_p) = 0.$$

For $\nu > 0$ consider the function

$$(5.20)\qquad q_\nu(\varepsilon,\tilde{x}) = 1(\tilde{x}'\alpha_0 \geq \nu)\{-1(\varepsilon < -\tilde{x}'\alpha_0)[f_\varepsilon(\varepsilon|\tilde{x})/f(\varepsilon|\tilde{x})]$$

$$+ 1(-\tilde{x}'\alpha_0 \leq \varepsilon \leq 0)f(-\tilde{x}'\alpha_0|\tilde{x})/[F(0|\tilde{x})-F(-\tilde{x}'\alpha_0|\tilde{x})]\}$$

$$+ 1(\tilde{x}'\alpha_0 < 0)\{-[f_\varepsilon(\varepsilon|\tilde{x})/f(\varepsilon|\tilde{x})] - 2f(0|\tilde{x})\text{sign}(\varepsilon)\},$$

By Assumption 3.1 it follows as in Hajek and Sidak (1967) that for any measurable function $v(\tilde{x})$, $E[f_\varepsilon(\varepsilon|\tilde{x})/f(\varepsilon|\tilde{x})|\tilde{x}] = 0$ and

$$(5.21)\qquad E[1(\varepsilon < v(\tilde{x}))\{f_\varepsilon(\varepsilon|\tilde{x})/f(\varepsilon|\tilde{x})\}|\tilde{x}] = f(v(\tilde{x})|\tilde{x})$$

$$= -E[1(\varepsilon > v(\tilde{x}))\{f_\varepsilon(\varepsilon|\tilde{x})/f(\varepsilon|\tilde{x})\}|\tilde{x}].$$

Finiteness of $E[f(v(\tilde{x})|\tilde{x})^2]$ then follows from Assumption 3.1 and (the conditional) Holders inequality (which implies $f(v(\tilde{x})|\tilde{x})^2 \leq E[\{f_\varepsilon(\varepsilon|\tilde{x})/f(\varepsilon|\tilde{x})\}^2|\tilde{x}]$). Also, by Assumption 2.5, $1(\tilde{x}'\alpha_0{\geq}\nu)/[F(0|\tilde{x})-F(-\tilde{x}'\alpha_0|\tilde{x})]$ is bounded. Therefore $E[q_\nu(\varepsilon,\tilde{x})^2]$ is finite. Also, note that

$$(5.22)\qquad E[q_\nu(\varepsilon,\tilde{x})|\tilde{x}] = 1(\tilde{x}'\alpha_0 \geq \nu)\{-E[1(\varepsilon < -\tilde{x}'\alpha_0)[f_\varepsilon(\varepsilon|\tilde{x})/f(\varepsilon|\tilde{x})]|\tilde{x}]$$

$$+ E[1(-\tilde{x}'\alpha_0 \leq \varepsilon \leq 0)|\tilde{x}]\circ f(-\tilde{x}'\alpha_0|x)/[F(0|\tilde{x})-F(-\tilde{x}'\alpha_0|\tilde{x})]\}$$

$$+ 1(\tilde{x}'\alpha_0 < 0)\{-E[f_\varepsilon(\varepsilon|\tilde{x})/f(\varepsilon|\tilde{x})|\tilde{x}] - 2f(0|\tilde{x})E[\text{sgn}(\varepsilon)|\tilde{x}]\}$$

$$= 0,$$

$$(5.23)\qquad E[\text{sgn}(\varepsilon)q_\nu(\varepsilon,x)|x] = 1(\tilde{x}'\alpha_0 \geq \nu)\circ 0$$

$$+ 1(\tilde{x}'\alpha_0 < 0)\{-E[\text{sgn}(\varepsilon)f_\varepsilon(\varepsilon|x)/f(\varepsilon|x)|x] - 2f(0|x)\}$$

$$= 1(\tilde{x}'\alpha_0 < 0)\circ\{E[1(\varepsilon{<}0)f_\varepsilon(\varepsilon|x)/f(\varepsilon|x)|x] -$$

$$E[1(\varepsilon{>}0)f_\varepsilon(\varepsilon|x)/f(\varepsilon|x)|x] \quad - 2f(0|x)\} = 0.$$

Then by Lemma 5.4 there is a bounded function $\tilde{q}_\nu(\varepsilon,x)$ that is continuously differentiable in ε with bounded derivative and satisfies $\int\tilde{q}_\nu(\varepsilon,x)f(\varepsilon|x)d\varepsilon = 0$ and $\int\text{sgn}(\varepsilon)\tilde{q}_\nu(\varepsilon,x)f(\varepsilon|x)d\varepsilon = 0$, such that

$E[\{q_\nu(\varepsilon,x)-\tilde{q}_\nu(\varepsilon,x)\}^2]$ is arbitrarily small. Straightforward calculation shows that the corresponding derivative $\tilde{r}^\nu_\eta$ of the root likelihood with respect to η at θ_0 is given by

(5.24) $$\tilde{r}^\nu_\eta = (v/2)\{[1(\tilde{y} = 0)[1-F(-\tilde{x}'\alpha_0|\tilde{x})]^{-1/2}\int_{-\tilde{x}'\alpha_0}^{\infty} \tilde{q}_\nu(\varepsilon,\tilde{x})f(\varepsilon|\tilde{x})d\varepsilon] + 1(\tilde{y} < 0)\tilde{q}_\nu(\tilde{y}-\tilde{x}'\alpha_0,\tilde{x})f(\tilde{y}-\tilde{x}'\alpha_0|\tilde{x})^{1/2}\}.$$

For given v define

(5.25) $$r^\nu_\eta = (v/2)\{[1(\tilde{y} = 0)[1-F(-\tilde{x}'\alpha_0|\tilde{x})]^{-1/2}\int_{-\tilde{x}'\alpha_0}^{\infty} q_\nu(\varepsilon,\tilde{x})f(\varepsilon|\tilde{x})d\varepsilon] + 1(\tilde{y} < 0)q_\nu(\tilde{y}-\tilde{x}'\alpha_0,\tilde{x})f(\tilde{y}-\tilde{x}'\alpha_0|\tilde{x})^{1/2}\}.$$

By (the conditional) Holders inequality,

(5.26) $$\|\tilde{r}^\nu_\eta-r^\nu_\eta\|_p$$
$$\le (D/2)E[\{1-F(-\tilde{x}'\alpha_0|\tilde{x})\}^{-1}\{\int_{-\tilde{x}'\alpha_0}^{\infty} (\tilde{q}_\nu(\varepsilon,\tilde{x})-q_\nu(\varepsilon,\tilde{x}))f(\varepsilon|\tilde{x})d\varepsilon\}^2] + (D/2)E[1(\varepsilon < -\tilde{x}'\alpha_0)(\tilde{q}_\nu(\varepsilon,x)-q_\nu(\varepsilon,x))^2].$$
$$= (D/2)E[\{1-F(-\tilde{x}'\alpha_0|\tilde{x})\}\{E[\tilde{q}_\nu(\varepsilon,x)-q_\nu(\varepsilon,x)|\varepsilon > -\tilde{x}'\alpha_0,\ \tilde{x}]\}^2] + (D/2)E[1(\varepsilon < -\tilde{x}'\alpha_0)(\tilde{q}_\nu(\varepsilon,x)-q_\nu(\varepsilon,x))^2]$$
$$\le (D/2)E[\{1-F(-\tilde{x}'\alpha_0|\tilde{x})\}E[\{\tilde{q}_\nu(\varepsilon,x)-q_\nu(\varepsilon,x)\}^2|\varepsilon > -\tilde{x}'\alpha_0,\ \tilde{x}]] + (D/2)E[1(\varepsilon < -\tilde{x}'\alpha_0)(\tilde{q}_\nu(\varepsilon,x)-q_\nu(\varepsilon,x))^2]$$
$$= (D/2)E[\{\tilde{q}_\nu(\varepsilon,x)-q_\nu(\varepsilon,x)\}^2]\}.$$

Next, note that

(5.27) $$\int_{-\tilde{x}'\alpha_0}^{\infty} q_\nu(\varepsilon,\tilde{x})f(\varepsilon|\tilde{x})d\varepsilon$$
$$= 1(\tilde{x}'\alpha_0 \ge \nu)\circ f(-\tilde{x}'\alpha_0|\tilde{x})$$

$$+ 1(\tilde{x}'\alpha_0 < 0)\circ\{f(-\tilde{x}'\alpha_0|\tilde{x}) - 2f(0|\tilde{x})[1-F(-\tilde{x}'\alpha_0|\tilde{x})]\}.$$

Therefore, $r^{v}_{\beta} - r^{\nu}_{\eta} - \rho^{v} = 1_{\nu}\circ(r^{v}_{\beta} - \rho^{v})$, so that there is a constant C such that

$$\text{(5.28)} \quad \|r_{\beta}-\tilde{r}^{\nu}_{\eta}-\rho\|^2_p \le C\{\|\rho-\rho^{v}-(r_{\beta}-r^{v}_{\beta})\|^2_p + \|r^{v}_{\beta}-r^{\nu}_{\eta}-\rho^{v}\|^2_p + \|\tilde{r}^{\nu}_{\eta}-r^{\nu}_{\eta}\|^2_p\}$$

$$= C\{\|\rho-\rho^{v}-(r_{\beta}-r^{v}_{\beta})\|^2_p + \|1_{\nu}(r^{v}_{\beta}-\rho^{v})\|^2_p + \|\tilde{r}^{\nu}_{\eta}-r^{\nu}_{\eta}\|^2_p\}.$$

and the conclusion follow by eqs (5.18), (5.19), and (5.26).

Before proving Theorem 3.1 we will first prove some preliminary Lemmas. Let $s(z,\beta) = 2f(0|\tilde{x})\circ 1(\tilde{x}'\alpha < 0)\circ \mathrm{sgn}(y - x'\beta)\circ x$ and $\hat{s}_j(z,\beta) = \hat{g}_j(x,\beta)1(\tilde{x}'\alpha<0)\mathrm{sgn}(y-x'\beta)x/\|x\|$. Also, for a parametric density of the form given in (2.1) term "under local drift" will refer to the environment where $\tilde{z}_t$, $t \le n$, are i.i.d. with density $f(\tilde{z}|\theta_n)$ and $\theta_n = (\beta_0',0)'$.

Lemma 5.5: If Assumptions 2.1-2.6 and 3.1-3.5 are satisfied then $\tilde{\beta}_j$ is locally regular for Q. Furthermore, under local drift, for any $\tilde{\beta}_n = \beta_0 + O_p(1/\sqrt{n})$,

$$\text{(5.29)} \quad \Sigma_{t\in I_j}[s(z_t,\tilde{\beta}) - s(z_t,\beta_n) + (V^*)^{-1}(\tilde{\beta}-\beta_n)]/\sqrt{n} = o_p(1), \quad (j=1,2).$$

Proof: In the absence of local drift (i.e. $\theta_n = (\beta_0',0)$ for all n), $\mathrm{plim}(\tilde{\beta}_j) = \beta_0$ follows exactly as in Powell's (1984) treatment of the CLAD estimator. By Lemma 5.1 and nonsingularity of the information matrix the model is locally regular in the sense of Hajek (1970), so that $\mathrm{plim}(\tilde{\beta}_j) = \beta_0$ under local drift follows by contiguity. Also, it follows exactly as in Powell (1984) that the asymptotic first-order condition of equation (1.3) is satisfied.

By Assumption 2.4 and $q(\varepsilon,\tilde{x})$ bounded, $f(\varepsilon|\tilde{x},\eta)$ is bounded and bounded away from zero in a neighborhood of $\varepsilon = 0$, uniformly in $\tilde{x}$ and η close enough to $\eta_0 = 0$. Also, by Assumption 3.3, $q(\varepsilon,\tilde{x})$

bounded, and $q(\varepsilon,\tilde{x})$ Lipshitz in ε uniformly in $\tilde{x}$ (which follows from continuous differentiability in ε with bounded derivative), it follows that $f(\varepsilon|\tilde{x},\eta)$ is Lipshitz in ε, uniformly in $\tilde{x}$ and η close enough to $\eta_0 = 0$. It then follows as in Powell (1984) that conditions N-1 - N-4 of Lemma 3 of Huber (1967) are satisfied uniformly in Θ_n, for

(5.30) $\quad \psi(\tilde{z},\beta) = 1(x'\beta < u)\mathrm{sgn}(y - x'\beta)x.$

Note also, that by $q(\varepsilon,\tilde{x})$ bounded,

(5.31) $$V_{1n} = \{E[2f(0|\tilde{x}_t,\eta_n)1(x_t'\beta_0<u_t)x_tx_t']\}^{-1} \circ$$

$$E[1(x_t'\beta_0<u_t)x_tx_t']\circ\{E[2f(0|\tilde{x}_t,\eta_n)1(x_t'\beta_0<u_t)x_tx_t']\}^{-1} \to V_1.$$

It then follows from $n/n_j \to 2$, as in Powell (1984), that

(5.32) $$\sqrt{n}(\tilde{\beta}_j-\beta_n) = (\sqrt{n/n_j})\sqrt{n_j}(\tilde{\beta}_j-\beta_n) \xrightarrow{d} N(0,\ 2V_1).$$

Next, note that by $2f(0|\tilde{x})$ bounded and bounded away from zero, it follows (as before for $\psi(\tilde{z},\beta)$ in equation (5.30)) that conditions N-1 - N-4 of Lemma 3 of Huber (1967) are satisfied uniformly in Θ_n for $\psi(\tilde{z},\beta) = 2\circ f(0|\tilde{x})\circ 1(x'\beta < u)\mathrm{sgn}(y - x'\beta)x$. Equation (5.29) then follows from the conclusion of Huber's (1967) Lemma 3 and an argument similar to that for equation (5.31).

Define $\tilde{X}_n = (\tilde{x}_1,\ldots,\tilde{x}_n)$.

Lemma 5.6: Suppose that Assumptions 2.1-2.6 and 3.1-3.5 are satisfied and suppose the absence of local drift. Then for any positive constant D, $\int E[\sup_{\|\beta-\beta_0\|\le D/\sqrt{n}}\{\hat{g}_j(\tilde{x},\beta) - g(\tilde{x})\}^2]d\rho_x \to 0$, $(j=1,2)$.

Proof: Write $\hat{g}_j(\tilde{x},\beta) - g(\tilde{x}) = A_1(\tilde{x},\beta) + A_2(\tilde{x}) + A_3(\tilde{x})$, where

(5.33) $$A_1(\tilde{x},\beta) = \hat{g}_j(\tilde{x},\beta) - \hat{g}_j(\tilde{x},\beta_0),$$

$$A_2(\tilde{x}) = \hat{g}_j(\tilde{x},\beta_0) - \Sigma_{t=1}^n W_{ntj}(\tilde{x})1(\tilde{x}_t'\alpha_0<0)f(0|\tilde{x}_t)\|x_t\|,$$

$$A_3(\tilde{x}) = \Sigma_{t=1}^{n} W_{ntj}(\tilde{x})\{1(\tilde{x}_t'\alpha_0<0)f(0|\tilde{x}_t)\|x_t\| - 1(\tilde{x}'\alpha_0<0)f(0|\tilde{x})\|x\|\},$$

It follows immediately from $E[g(\tilde{x})^2]$ finite and Proposition 1 of Stone (1977) that $\int E[A_3(\tilde{x})^2]d\mu_x \to 0$.

Note that $W_{ntj}(\tilde{x}) \le W_0/k_n$, and for all $\|\beta-\beta_0\| \le D/\sqrt{n}$,

(5.34) $$|1(\tilde{x}_t'\alpha<0)1(x_t'\beta-c_n<y_t<x_t'\beta) - 1(\tilde{x}_t'\alpha_0<0)1(x_t'\beta_0-c_n<y_t<x_t'\beta_0)|$$

$$\le 1(|\tilde{x}_t'\alpha_0| \le \|x_t\|\cdot\|\beta-\beta_0\|)$$

$$+ 1(|\varepsilon_t| \le \|x_t\|\cdot\|\beta-\beta_0\|) + 1(|\varepsilon_t+c_n| \le \|x_t\|\cdot\|\beta-\beta_0\|) \equiv \ell_t,$$

$$\le 1(|\tilde{x}_t'\alpha_0| \le \|x_t\|\cdot D/\sqrt{n})$$

$$+ 1(|\varepsilon_t| \le \|x_t\|\cdot D/\sqrt{n}) + 1(|\varepsilon_t+c_n| \le \|x_t\|\cdot D/\sqrt{n}) \equiv \ell_{tn},$$

so that for n large enough (by $f(0|\tilde{x})$ bounded and Assumption 3.4),

(5.35) $$E[\sup_{\|\beta-\beta_0\|\le D/\sqrt{n}} A_1(\tilde{x},\beta)^2] \le n^2(W_0/k_n)^2 c_n^{-2} E[\|x_t\|\ell_{tn}]$$

$$\le 2W_0[n/(c_n k_n)]^2\{2f_0 E[\|x_t\|^2] + K_0\}D/\sqrt{n} = O(n^{3/2}/(c_n{}^2 k_n{}^2)) = o(1).$$

Since this equation holds uniformly in $\tilde{x}$, $E[\sup_{\|\beta-\beta_0\|\le D/\sqrt{n}} A_1(\tilde{x},\beta)^2] \to 0$.

Next, note that

(5.36) $$|E[A_2(\tilde{x})|\tilde{X}_n]|$$

$$\le \Sigma_{t=1}^{n} W_{ntj}(\tilde{x})1(\tilde{x}_t'\alpha_0<0)\cdot\|x_t\|\cdot|[\mathrm{Prob}(-c_n<\varepsilon_t<0|\tilde{x}_t)/c_n] - f(0|\tilde{x}_t)|$$

$$\le M_0 c_n \Sigma_{t=1}^{n} W_{ntj}(\tilde{x})\cdot\|x_t\|,$$

and

(5.37) $$\mathrm{Var}(A_2(\tilde{x})|\tilde{X}_n)$$

$$= \Sigma_{t=1}^{n} W_{ntj}(\tilde{x})^2\cdot\|x_t\|^2\cdot\mathrm{Var}[\{1(-c_n<\varepsilon_t<0|\tilde{x}_t)/c_n\}-f(0|\tilde{x}_t)|\tilde{X}_n]$$

$$\le (W_0/k_n)^2\{\Sigma_{t=1}^{n}\|x_t\|^2 E[1(-c_n<\varepsilon_t<0|\tilde{x}_t)|\tilde{X}_n]/(c_n{}^2)\}$$

$$\leq (W_0/k_n)^2(\Sigma_{t=1}^n \|x_t\|^2) f_0/c_n.$$

Therefore, by $E[\|x_t\|^2]$ finite and Proposition 1 of Stone (1977), (which imply $\int E[(\Sigma_{t=1}^n W_{ntj}(\tilde{x})\circ\|x_t\|)^2]d\mu_x = O(1)$),

$$\text{(5.38)} \quad \int E[A_2(\tilde{x})^2]d\mu_x = \int E[E[A_2(\tilde{x})^2|X_n]]d\mu_x$$

$$= \int E[\mathrm{Var}(A_2(\tilde{x})|\tilde{X}_n) + E(A_2(\tilde{x})^2|\tilde{X}_n)]d\mu_x$$

$$\leq (W_0/k_n)^2 nE[\|x_t\|^2]f_0/c_n + M_0{}^2 c_n{}^2 \int E[(\Sigma_{t=1}^n W_{ntj}(\tilde{x})\circ\|x_t\|)^2]d\mu_x$$

$$= O(n/(k_n{}^2 c_n)) + O(c_n{}^2) = o(1).$$

The conclusion then follows from $(a_1+a_2+a_3)^2 \leq 3(a_1{}^2+a_2{}^2+a_3{}^2)$.

Lemma 5.7: Suppose that Assumptions 2.1-2.6 and 3.1-3.5 are satisfied and suppose the absence of local drift. Then

$$\text{(5.39)} \quad \Sigma_{t=1}^n 1(t\notin I_j)[\hat{s}_j(z_t,\tilde{\beta}_j) - s(z_t,\tilde{\beta}_j)]/\sqrt{n} = o_p(1), \quad (j = 1, 2).$$

Proof: Without loss of generality consider $j = 2$. Note first that for $x_t'\beta < u_t$ and $u_t = y_t$ it is the case that $\varepsilon_t + x_t'\beta_0 > u_t$, so that $1 = \mathrm{sgn}(y_t - x_t'\beta) = \mathrm{sgn}(u_t - x_t'\beta) = \mathrm{sgn}(\varepsilon_t + x_t'\beta_0 - x_t'\beta)$. Thus,

$$\text{(5.40)} \quad 1(x_t'\beta<u_t)\circ\mathrm{sgn}(y_t-x_t'\beta) = 1(x_t'\beta<u_t)\circ\mathrm{sgn}(\varepsilon_t+x_t'\beta_0-x_t'\beta).$$

Then, by $|\mathrm{sgn}(\varepsilon)-\mathrm{sgn}(\varepsilon+\gamma)| \leq 2\circ 1(|\varepsilon| \leq |\gamma|)$ and $E[\mathrm{sgn}(\varepsilon_t)|\tilde{x}_t] = 0$, for any $\lambda \in \mathbb{R}^p$ with $\|\lambda\| = 1$,

$$\text{(5.41)} \quad |E[\lambda'\Sigma_{t=1}^n 1(t\in I_1)[\hat{s}_2(z_t,\tilde{\beta}_2)-s(z_t,\tilde{\beta}_2)]/\sqrt{n}|I_2]|$$

$$= |\sqrt{n}E[(g_2(x_1,\tilde{\beta}_2)-g(x_1))\circ(\lambda'x_1/\|x_1\|)\circ 1(\tilde{x}_1'\tilde{\alpha}_2<0) \circ$$

$$E[\mathrm{sgn}(\varepsilon_1+x_1'\beta_0-x_1'\tilde{\beta}_2) - \mathrm{sgn}(\varepsilon_1)|x_1,I_2] \mid I_2]|$$

$$\leq 2\sqrt{n}E[|g(x_1,\tilde{\beta}_2)-g(x_1)|\circ E[1(|\varepsilon_1| \leq \|x_1\|\circ\|\tilde{\beta}_2-\beta_0\|)|x_1,I_2] \mid I_2]$$

$$\leq 4f_0E[|g(x_1,\tilde{\beta}_2)-g(x_1)|\|x_1\||I_2]\sqrt{n}\|\tilde{\beta}_2-\beta_0\|$$

$$\leq \{\int[\hat{g}_2(x,\tilde{\beta}_2)-g(x)]^2 d\rho_x\}\cdot O_p(1),$$

where the last inequality follows by the Cauchy-Schwartz inequality. and $\sqrt{n}$-consistency of $\tilde{\beta}_2$. Similarly,

(5.42) $$\mathrm{Var}[\lambda'\Sigma_{t=1}^n 1(t\in I_1)[\hat{s}_2(z_t,\tilde{\beta}_2)-s(z_t,\tilde{\beta}_2)]/\sqrt{n}\,|I_2]$$
$$\leq \int[g_2(x,\tilde{\beta}_2)-g(x)]^2 d\rho_x.$$

Also, for any $\epsilon > 0$ there exists D such that $\mathrm{Prob}(\|\tilde{\beta}_2-\beta_0\| > D/\sqrt{n}) < \epsilon$ for all n large enough, while by the conclusion of Lemma 5.6,

(5.43) $$1(\|\tilde{\beta}_2-\beta_0\|\leq D/\sqrt{n})\cdot\int[g_2(x,\tilde{\beta}_2)-g(x)]^2 d\rho_x$$
$$\leq \int\sup_{\|\beta-\beta_0\|\leq D/\sqrt{n}}[g_2(x,\tilde{\beta}_2)-g(x)]^2 d\rho_x = o_p(1).$$

so that by the arbitrary choice of ϵ, $\int[g_2(x,\tilde{\beta}_2)-g(x)]^2 d\rho_x = o_p(1)$. The conclusion then follows from equations (5.41) and (5.42) and the arbitrary choice of λ.

Lemma 5.8: Suppose that Assumptions 2.1-2.6 and 3.1-3.5 are satisfied and suppose the absence of local drift. Then

(5.44) $$\hat{V}_j^* = V^* + o_p(1).$$

Proof: Without loss of generality consider $j = 2$. Let $\hat{V}_2 = \Sigma_{t=1}^n 1(t\in I_2)s(z_t,\tilde{\beta}_2)s(z_t,\tilde{\beta}_2)'/n_1$. It follows immediately from the proof of Lemma 5.7 and that for any λ with $\|\lambda\| = 1$,

(5.45) $$E[\,|\lambda'(\hat{V}_2^* - \hat{V}_2)\lambda|\,|I_2] \leq E[\,|[\lambda'\hat{s}_2(z_1,\tilde{\beta}_2)]^2-[\lambda' s(z_1,\tilde{\beta}_2)]^2|\,|I_2]$$
$$\leq E[\{|\lambda'\hat{s}_2(z_1,\tilde{\beta}_2)|+|\lambda' s(z_1,\tilde{\beta}_2)|\}|\lambda'\hat{s}_2(z_1,\tilde{\beta}_2)-\lambda' s(z_1,\tilde{\beta}_2)|\,|I_2]$$
$$\leq 2E[\,|\lambda'\hat{s}_2(z_1,\tilde{\beta}_2)|^2+|\lambda' s(z_1,\tilde{\beta}_2)|^2|I_2]\cdot E[\{\lambda'\hat{s}_2(z_1,\tilde{\beta}_2)-\lambda' s(z_1,\tilde{\beta}_2)\}^2|I_2]$$
$$= O_p(1)o_p(1) = o_p(1),$$

so that $\hat{V}_2^* - \hat{V}_2 = o_p(1)$. It also follows from arguments previously used that $\hat{V}_2 - V^* = o_p(1)$, giving the conclusion.

Proof of Theorem 2: Lemma 5.5 gives $(\hat{V}_1^*+\hat{V}_2^*)/2 = o_p(1)$. Also, it follows similarly to equation (5.41) that in the absence of local drift,

$$(5.46) \qquad \Sigma_{t=1}^{n} 1(t\notin I_j)s(z_t,\tilde{\beta}_j)/\sqrt{n} = O_p(1), \quad (j=1,2),$$

Therefore, by Lemmas 5.7 and 5.8 it follows that in the absence of drift

$$(5.47) \qquad [\hat{V}_2^*\Sigma_{t\in I_1}\hat{s}_2(z_t) + \hat{V}_1^*\Sigma_{t\in I_2}\hat{s}_1(z_t)]/\sqrt{n}$$

$$- V^*[\Sigma_{t\in I_1}s_2(z_t,\tilde{\beta}_2) + \Sigma_{t\in I_2}\hat{s}_1(z_t,\tilde{\beta}_1)]/\sqrt{n} = o_p(1),$$

and by contiguity this equation continues to hold when local drift is present. It then follows from Lemma 5.5 that under local drift,

$$(5.48) \qquad \sqrt{n}(\hat{\beta}-\beta_n) = \sqrt{n}\Sigma_{j=1}^{2}[(\tilde{\beta}_j-\beta_n)/2 + \hat{V}_j^*\Sigma_{t=1}^{n}1(t\notin I_j)\hat{s}_j(z_t)/n]$$

$$= \sqrt{n}\Sigma_{j=1}^{2}[(\tilde{\beta}_j-\beta_n)/2 + V^*\Sigma_{t=1}^{n}1(t\notin I_j)s(z_t,\tilde{\beta}_j)/n] + o_p(1)$$

$$= V^*\Sigma_{t=1}^{n}s(z_t,\beta_n)/\sqrt{n}$$

$$+ \sqrt{n}(\tilde{\beta}_1-\beta_n)[1-(n_2/n)] + \sqrt{n}(\tilde{\beta}_2-\beta_n)[1-(n_1/n)] + o_p(1)$$

$$= V^*\Sigma_{t=1}^{n}s(z_t,\beta_n)/\sqrt{n} + o_p(1).$$

Let $E_n[\circ]$ denote the expectation taken at θ_n. Note that by Assumption 3.3 and $q(\varepsilon,\tilde{x})$ and $v(x)$ bounded, $f(0|\tilde{x},\eta_n)$ is bounded uniformly in $\tilde{x}$ and n. Then by Assumption 2.3 and $f(\varepsilon|\tilde{x},\eta_n)$ having conditional median zero, $E_n[s(z_1,\beta_n)] = 0$. Furthermore,

$$(5.49) \qquad \lim_{D\to\infty}\sup_n E_n[1(\|s(z_t,\beta_n)\| > D)\circ\|s(z_t,\beta_n)\|^2]$$

$$\le \lim_{D\to\infty}E[1(\|x_t\| > D)\circ\|x_t\|^2] = 0.$$

Let $V_n^* = \{E_n[s(z_t,\beta_n)s(z_t,\beta_n)']\}^{-1}$. It follows as in the proof of Lemma 5.8 that $E_n[s(z_t,\beta_n)s(z_t,\beta_n)'] - E_n[s(z_t)s(z_t)'] = o(1)$, while

$E_n[s(z_t)s(z_t)'] - E[s(z_t)s(z_t)']$ follows by boundedness of $v(x)$ and $q(\varepsilon,\tilde{x})$ and the dominated convergence theorem, implying $V_n^* - V^* = o(1)$. Then by (5.49) it is possible to apply the central limit theorem to conclude that in the presence of local drift,

$$(5.50)\quad \sqrt{n}(\hat{\beta}-\beta_n) = (V^*)^{1/2}[V^*(V_n^*)^{-1}]^{1/2}\{(V_n^*)^{1/2}\Sigma_{t=1}^n s(z_t,\beta_n)/\sqrt{n}\} + o_p(1)$$

$$\xrightarrow{d} (V^*)^{1/2}N(0,I_p) \overset{d}{=} N(0,V^*).$$

Acknowledgement: This research was supported by NSF Grants SES-8606545 to Princeton University and SES-8641096 to the University of Wisconsin. The authors are grateful to Joel Horowitz for comments.

References

Begun, J., W. Hall, W. Huang, and J. Wellner, 1983, "Information and Asymptotic Efficiency in Parametric-Nonparametric Models," *Annals of Statistics*, 11, 432-452.

Bickle, P., 1982, "On Adaptive Estimation," *Annals of Statistics*, 10, 647-671.

Bickel P., Klaasen, Y. Ritov, and J. Wellner, 1987, *Efficiency in Semiparametric Models*, monograph, forthcoming.

Chamberlain, G., 1987, "Asymptotic Efficiency in Semiparametric Models with Censoring," *Journal of Econometrics*, forthcoming.

Cosslett, S. R., 1984, "Efficiency Bounds for Distribution-Free Estimators of the Binary Choice and Censored Regression Models," manuscript, Center for Econometrics and Decision Sciences, University of Florida.

Hajek, J., 1970, "A Characterization of Limiting Distributions of Regular Estimates, Z. *Wahrscheinlichkeitstheorie verw. Geb.*, 14, 323-330.

Hajek, J., and Z. Sidak, 1967, *Theory of Rank Tests*, Academic Press, New York.

Huber, P., 1967, "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press.

Koshevnik and Levitt, 1976, "On a Non-parametric Analouge of the Information Matrix," *Theory of Probability and Applications*, 21, 738-753.

Newey, W. K., 1987, "Adaptive Estimation of Regression Models Via Moment Restrictions," Econometric Research Program Memo. No. 330, Princeton University.

Pfanzagl, J, 1982, *Contributions to a General Asymptotic Statistical Theory*, New York: Springer-Verlag.

Powell, L. L., 1984, "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303-325.

Powell, J. L., 1986a, "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143-155.

Powell, J. L., 1986b, "Symmetrically Trimmed Least Squares Estimation for Tobit Models," *Econometrica*, 54, 1435-1460.

Ritov, Y., 1984, "Efficient and Unbiased Estimation in Nonparametric Linear Regression with Censored Data," manuscript, University of California, Berkeley.

Schick, A., 1986, "On Asymptotically Efficient Estimation in Semiparametric Models," *Annals of Statistics*, 14, 1139-1151.

Stein, C., 1956, "Efficient Nonparametric Testing and Estimation," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Berkeley: University of California Press.

Stone, C., 1977, "Consistent Nonparametric Regression," *Annals of Statistics*, 5, 549-645.