ESTIMATION OF SEMIPARAMETRIC MODELS WHEN THE CRITERION FUNCTION IS NOT SMOOTH

X. CHEN, O. LINTON and I. VAN KEILEGOM

Estimation of Semiparametric Models when the Criterion Function is not Smooth*

Xiaohong Chen†
London School of Economics

Oliver Linton‡
London School of Economics

Ingrid Van Keilegom§
Université catholique de Louvain

January 24, 2002

Abstract

We provide easy to verify sufficient conditions for the consistency and asymptotic normality of a class of semiparametric optimization estimators where the criterion function does not obey standard smoothness conditions and simultaneously depends on some preliminary nonparametric estimators. Our results extend existing theories like those of Pakes and Pollard (1989), Andrews (1994a), and Newey (1994). We apply our results to an example.

*We would like to thank Richard Blundell, Han Hong, Bo Honore, Hide Ichimura, Costas Meghir, Bob Sherman, and Elie Tamer for helpful comments. The research is completed while Chen is visiting Department of Economics, Princeton University from September 2001 to July 2002. Chen would like to thank the great hospitality provided by Princeton University, especially the Gregory Chow Econometrics Research Center. Chen and Linton would also like to acknowledge financial support from ESRC grant R00023952. The research of Van Keilegom was supported by 'Projet d'Actions de Recherche Concertées', No. 98/03–217, and by an IAP research network from the Belgian government'.

†Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: x.chen1@lse.ac.uk

‡Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: O.Linton@lse.ac.uk

§Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, B1348 Louvain-la-Neuve, Belgium. E-mail address: vankeilegom@stat.ucl.ac.be

1 Introduction

In this note, we investigate a class of semiparametric estimation and testing problems that involve non-smooth criterion functions that contain both finite and infinite dimensional unknown parameters. These sorts of problems arise in many micro-econometrics applications where there is a need for a flexible functional form combined with either censoring, switching, simulation, or counting. The necessity of allowing for flexible functional form is well understood in econometrics, see Powell (1994) and Horowitz (1998b) for reviews of semiparametric models.

We give some examples. The first case is a partially linear censored regression $y = \max\{x_1\beta + h(x_2) + \epsilon, 0\}$, where $\mathrm{med}(\epsilon|x_1, x_2) = 0$. This generalization of the usual censored regression might be needed when some covariates have nonlinear effects in the latent model. The usual CLAD estimation method of Powell (1984) can be generalized to this case either by the profiling method or through simultaneous estimation using sieves. In either case, the first order condition depends on a function $\widehat{h}(.)$ that lies inside an indicator function.[1] Han and Tamer (2001) study a linear regression model $y = x\beta + u$ in which the covariates can be endogenous and where the dependent variable is subject to a sort of interval censoring, i.e., you only observe a lower and upper bound y_0, y_1 on y. This sampling scheme arises in many data sets, see Manski and Tamer (2000) for some examples. The estimator they propose is based on the condition that $ET(\beta, z) = 0$ if and only if $\beta = \beta_0$, where z are some instruments and $T(\beta, z) = (0.5 - F_{y_0 - x\beta|z}(0))^2 1(F_{y_0 - x\beta|z}(0) > 0.5) + (0.5 - F_{y_1 - x\beta|z}(0))^2 1(F_{y_1 - x\beta|z}(0) < 0.5)$. Their estimator is based on a sample version of $T(\beta, z)$ using kernel estimators of the conditional c.d.f.'s $F_{y_j - x\beta|z}$, $j = 1, 2$.[2] Manski (1983) proposed an estimator for semiparametric models with separable error terms based on minimizing the mean squared distance of the empirical c.d.f.'s from independence [see also Manski (1994)]. This has been further investigated in Brown and Wegkamp (2000) who apply this method to estimating nonlinear simultaneous equations [see also Gozalo and Linton (1996) for a similar idea using conditional independence]. This is an example of the mini-

[1] There may exist other methods for estimating this model. Lee (2001) considers this model in the absence of censoring and proposes an estimator based on integrating a multivariate conditional quantile estimator. The Lewbel and Linton (2001) procedure for estimating censored nonparametric regression can be used to obtain estimates of the unknown quantities here. However, both these approaches rely on preliminary high-dimensional nonparametric regression and therefore potentially perform poorly in small samples.

[2] They actually replace the indicator functions in T by smooth approximations because of the presence of $\widehat{F}_{y_j - x\theta|z}$ inside them. By smoothing over the criterion function, they are able to apply Taylor series expansion arguments and derive the consistency and asymptotic normality of their estimation procedure.

mum distance class of estimators reviewed in Koul (2001); the criterion function here is generically discontinuous in the parameters. An extension of interest here is to allow semiparametric covariate effects. We also see the potential for nonparametric threshold models, generalizing those of Caner and Hansen (2001), in which the switching function contains a nonparametric regression. Matzkin (1994) also contains some discussion of nonparametric threshold models.

There have been many econometrics papers devoted to general theories of estimation, following Huber (1967). The existing theories allow for non-smooth objective functions of finite dimensional parameters (without infinite dimensional parameters) [e.g., Pakes and Pollard (1989), Pötscher and Prucha (1991ab), Newey and McFadden (1994, Section 7)], or smooth objective functions of both finite and infinite dimensional parameters [e.g., Bickel, Klaassen, Ritov, and Wellner (1993), Andrews (1994a), Newey (1994), Newey and McFadden (1994, Section 8), Pakes and Olley (1995), Chen and Shen (1998) and Ai and Chen (1999)]. To the best of our knowledge, there does not exist a general theory on non-smooth objective functions with both finite and infinite dimensional parameters, or rather the existing high level conditions for consistency and asymptotic normality have not been verified in this less regular setting.

A viable strategy here is to change the estimator so that it satisfies the usual regularity conditions. In many cases, this can be achieved by smoothing over the criterion function. Horowitz (1992) applied this approach successfully in the binary choice model of Manski (1975). He proposed a smoothed maximum score estimator, in which the Manski criterion function was convoluted with a smooth kernel. Horowitz showed that his estimator converges faster than the original unsmoothed maximum score [which was shown to be $n^{1/3}$ consistent by Kim and Pollard (1990)] and moreover the limiting distribution is normal so that standard inference techniques can be applied.[3] See Honoré and Kyriazidou (2000) for a similar problem. Of course, these are cases where the semiparametric efficiency bound is zero and the problem is correctly viewed as being nonparametric so that the smoothing idea fits in quite naturally. In other contexts, the extra smoothing is not strictly necessary and is rather invoked because it makes the asymptotic properties easier to work out or perhaps because the computations can be carried out by derivative based algorithms. Horowitz has applied this smoothing idea to a number of other 'standard' problems including LAD regression estimation and gives some additional justifications in terms of higher order properties.[4] The comparison in that

[3]Although, as Pollard (1993) forthrightly points out, Horowitz's conditions are stronger than Manski's. Under only Manski's conditions, Horowitz's estimator is: "badly biased and no longer rate optimal".

[4]Horowitz (1998a) establishes a nice result about asymptotic refinements of bootstrap based test statistics in median

case is analogous to whether one should use the smoothed empirical distribution function instead of the usual unsmoothed empirical distribution function. Although there are some statistical reasons for so doing,[5] most investigators would be content with using the unsmoothed empirical distribution.

In this note, we provide sufficient conditions to ensure $\sqrt{n}-$asymptotic normality of the finite dimensional parameters obtained from a non-smooth criterion that depends on a preliminary infinite dimensional parameter estimate. Our approach and results extend those of Pakes and Pollard (1989), Andrews (1994a), Newey (1994) and Pakes and Olley (1995). The theory we present here rely on certain key empirical process results regarding the stochastic equicontinuity properties of the non-smooth objective function, (especially with respect to preliminary nonparametric estimators). We provide these results below by extending the work of Andrews (1994b), Van Keilegom and Akritas (1999), and by applying the work of van der Vaart and Wellner (1996). We then show how the stochastic equicontinuity condition is satisfied in our 'hit rate' example.

In section 2 we introduce a general class of estimators involving non-smooth criterion functions and preliminary nonparametric estimation. In section 3 we present the consistency and asymptotic normality results. The key condition here is a stochastic equicontinuity property; in section 4 we provide a new result that gives sufficient conditions for this property to hold in our 'unsmooth' case. In section 5 we verify the conditions in an example.

2 A General Class of Estimators

Throughout the paper, we assume that the data $\{Z_i\}_{i=1}^n$ is an i.i.d. sample from a unknown distribution P whose support is $\mathbf{Z} \subset \mathbb{R}^d$. Write $Z_i' = (Y_i', X_i')$, where $Y_i \in \mathbb{R}^{d_y}$ and $X_i \in \mathbb{R}^{d_x}$. We denote Θ for a finite dimensional parameter set (a compact subset of $\mathbb{R}^k$) and $\mathcal{H}$ for an infinite dimensional parameter set. In this paper, we assume that $\mathcal{H}$ is a linear vector space of functions

regression. By using the smoothed LAD criterion function and the standard bootstrap resample, the corresponding t and χ^2 statistics are closer to their limiting distributions than when an unsmoothed LAD criterion function is used with the same bootstrap algorithm. Sakov and Bickel (1998) establish a similar result for the test statistic where the unsmoothed criterion function is used for estimation but the bootstrap algorithm is the sub-sample bootstrap.

[5]For example, Azzalini (1981) shows that the smoothed empirical has smaller mean squared error [in the second order] than the unsmoothed one provided an optimal bandwidth is chosen and provided there is sufficient smoothness present. Of course, you have to sacrifice unbiasedness, and for any sub-optimal bandwidth there may exist a c.d.f. for which the mean squared error of the unsmoothed estimator is better.

endowed with a pseudo-metric $||\cdot||_{\mathcal{H}}$, (i.e., $\mathcal{H}$ is a metric space except that $||h||_{\mathcal{H}} = 0$ does not necessarily imply that $h = 0$ *almost everywhere*). For example when $\mathcal{H}$ is a class of continuous functions mapping from $\mathbf{Z}$ to $\mathbb{R}$ and having finite sup-norms, we can take $||h||_{\mathcal{H}} = ||h||_{\infty} = \sup_{\bullet} |h(\bullet)|$ or $||h||_{\mathcal{H}} = ||h||_{L_2(P)} = \sqrt{\int |h|^2 dP}$. We also denote $\theta_o \in \Theta$ and $h_o \in \mathcal{H}$ as the true unknown finite and infinite dimensional parameters. Finally we denote $||A|| = (tr(A'A))^{1/2}$ for any matrix A.

Suppose there exists a non-random measurable vector-valued function $M : \mathbb{R}^k \times \mathcal{H} \to \mathbb{R}^l$, with $l \geq k$, such that

$$M(\theta, h_o) = 0 \quad \text{at } \theta = \theta_o \in \Theta \subset \mathbb{R}^k.$$

Suppose there also exists a random vector-valued function $M_n : \mathbb{R}^k \times \mathcal{H} \to \mathbb{R}^l$ depending on data $\{Z_i : i = 1, \ldots, n\}$, such that $||M_n(\theta, h_o)||$ is close to $||M(\theta, h_o)||$, and $\min_{\theta\in\Theta} ||M_n(\theta, h)||$ is easy to compute. We allow that $M_n(\theta, h)$ could be non-smooth with respect to (θ, h) but will assume that $M(\theta, h)$ is smooth at (θ_o, h_o) in a sense to be defined later. As in Newey (1994), Pakes and Olley (1995), and Ai and Chen (1999), we allow that Z, θ_o could enter $h_o(.)$ as the arguments.

Suppose for now that there is an initial nonparametric estimator $\widehat{h}(.)$ for $h_o(.)$. We estimate θ_o by $\widehat{\theta}$, which solves the sample minimization problem:

$$\min_{\theta\in\Theta} ||M_n(\theta, \widehat{h})|| + o_p(1).$$

There are now many algorithms available for computing the optimum of general non-smooth functions, e.g., the Nelder-Mead, and the more recent genetic and evolutionary algorithms. Koenker (1997) gives a review of some methods targeted at quantile regression problems. Following Pakes and Pollard (1989), and Bickel, Klaassen, Ritov, and Wellner (1993, Section A.10), it is not necessary that $M(\theta, h) = E[M_n(\theta, h)]$ in the general limiting theorems we present below. Nevertheless many econometric applications correspond to this case:

$$M(\theta, h) = E[m(Z_i, \theta, h)] \quad \text{and } M_n(\theta, h) = \frac{1}{n}\sum_{i=1}^{n} m(Z_i, \theta, h)$$

where $m : \mathbb{R}^d \times \mathbb{R}^k \times \mathcal{H} \to \mathbb{R}^l$, with $l \geq k$, is a measurable vector-valued function such that

$$E[m(Z_i, \theta, h_o)] = 0 \quad \text{if } \theta = \theta_o \in \Theta \subset \mathbb{R}^k.$$

Hence the notations $M(\theta, h)$ and $M_n(\theta, h)$ implicitly correspond to population and sample moment conditions. Usually the function h enters m only through $h(X_i)$, but there are some case where

$h(Z_1), \ldots, h(Z_n)$ enter m, and our results permit this case. Note that the estimator of h can be profiled, that is, $\widehat{h}$ can also depend on θ, but we have chosen not to emphasize this to keep the notation simpler.[6] There are also some cases of interest where unique identification does not apply [Manski and Tamer (2000)], and our consistency results can be extended to that case too, although the distribution theory has not been solved for that case.

3 The Distribution Theory

There are many approaches to asymptotic theory for what are now called M and Z estimators in statistics, and van der Vaart and Wellner's (1996) book contains a modern treatment. We present distribution theory that follows the line of the well known paper of Pakes and Pollard (1989). See Newey and McFadden (1994) for discussion.

3.1 Consistency

Theorem 1. *Suppose that $\theta_o \in \Theta$ satisfies $M(\theta_o, h_o) = 0$, and that:*

1. $\|M_n(\widehat{\theta}, \widehat{h})\| \leq \inf_{\theta\in\Theta} \|M_n(\theta, \widehat{h})\| + o_p(1)$.

2. *For all $\delta > 0$, there exists $\epsilon(\delta) > 0$ such that*

$$\inf_{\|\theta-\theta_o\|>\delta} \|M(\theta, h_o)\| \geq \epsilon(\delta) > 0.$$

3. $||\widehat{h} - h_o\|_{\mathcal{H}} = o_p(1)$ *uniformly over $\theta \in \Theta$.*

4. *Uniformly for all $\theta \in \Theta$, $M(\theta, h)$ is continuous [with respect to the pseudo-metric $|| \cdot ||_{\mathcal{H}}$] in h at $h = h_o$.*

5. *For all sequences of positive numbers $\{\delta_n\}$ with $\delta_n = o(1)$,*

$$\sup_{\theta\in\Theta, \|h-h_o\|_{\mathcal{H}}\leq\delta_n} \frac{\|M_n(\theta, h) - M(\theta, h)\|}{1 + \|M_n(\theta, h)\| + \|M(\theta, h)\|} = o_p(1).$$

[6] There is an interesting extension to the case where θ, h are estimated simultaneously as in Chen and Shen (1998). This can be considered an alternative estimation strategy to profiling as was pointed out in Chen and Shen (1998). However, this approach requires stronger results than we are currently able to provide, and so we have not described this case.

Then,

$$\widehat{\theta} - \theta_o = o_p(1).$$

Remark 1: (i) Obviously condition 5 is implied by the **Condition 5'**: *for any* $\delta_n = o(1)$,

$$\sup_{\theta \in \Theta, \|h - h_o\|_{\mathcal{H}} \leq \delta_n} \|M_n(\theta, h) - M(\theta, h)\| = o_p(1).$$

(ii) If $M_n(\theta, h) = M_n(\theta, h_o)$ and $M(\theta, h) = M(\theta, h_o)$ for all h, then conditions 3 and 4 are automatically satisfied, and our Theorem 1 becomes Pakes and Pollard's (1989) Corollary 3.2.

(iii) The proof of Theorem 1 does not depend on the i.i.d. assumption at all, and Theorem 1 is still valid for non-identically distributed, dependent data. This theorem is comparable to Andrews' (1994a) Theorem A-1.

(iv) Comparing our Theorem 1 to Newey's (1994) Lemma 5.2 in the case $M(\theta, h) = E[M_n(\theta, h)] = E[m(Z, \theta, h)]$, the main difference is that while we impose continuity assumption on $E[m(Z, \theta, h)]$ (with respect to θ, h), Newey imposes a continuity assumption directly on $m(Z, \theta, h)$ (with respect to θ, h). More precisely, instead of our conditions 2 and 4, Newey's (1994) assumption 5.4 assumes: (i) $m(Z, \theta, h_o)$ is continuous in θ; and (ii) $m(Z, \theta, h)$ is Hölder continuous in h uniformly for all θ.

3.2 Asymptotic Normality

We need some smoothness property in both θ and $h(.)$ to hold for the population function $M(\theta, h)$. For any $\theta \in \Theta$, we say that $M(\theta, h)$ is pathwise differentiable at h in the direction $[h' - h]$ if there is a continuous-path $\{g_t(\cdot) : t \in [0, 1]\}$ in $\mathcal{H}$ with $g_0(\cdot) = h(\cdot)$ and $g_1(\cdot) = h'(\cdot)$ such that the limit

$$\lim_{t \longrightarrow 0} \frac{M(\theta, g_t) - M(\theta, h)}{t}$$

is well-defined, and is denoted as $\Gamma_2(\theta, h)[h' - h]$.[7]

The following theorem extends the Theorem 3.3 of Pakes and Pollard (1989) to the case where $M_n(\theta, h)$ depends on h and could be non-smooth in (θ, h).

Theorem 2. *Suppose that* $\theta_o \in int(\Theta)$ *satisfies* $M(\theta_o, h_o) = 0$, *that* $\widehat{\theta} - \theta_o = o_p(1)$, *and* $||\widehat{h} - h_o||_{\mathcal{H}} = o_p(1)$, *and that:*

[7] This is alternatively called tangential Hadamard differentiable, see Bickel, Klaassen, Ritov, and Wellner (1993, p456).

1. $\|M_n(\widehat{\theta},\widehat{h})\| = \inf_{||\theta-\theta_o||\leq\delta_n} \|M_n(\theta,\widehat{h})\| + o_p(1/\sqrt{n})$ *for some positive sequence* $\delta_n = o(1)$.

2. $M(\theta, h_o)$ *is differentiable in* θ *at* $\theta = \theta_o$ *with derivative matrix* Γ_1 *of full (column) rank.*

3. *There is a functional* $\Gamma_2(\theta, h_o)[h - h_o]$ *that is linear in* $h - h_o$ *such that:*

$$(i) \quad ||M(\theta, h) - M(\theta, h_o) - \Gamma_2(\theta, h_o)[h - h_o]|| \leq c||h - h_o||_{\mathcal{H}}^{1+\epsilon}$$

   for all θ *with* $||\theta - \theta_o|| = o(1)$, *all* h *with* $||h - h_o||_{\mathcal{H}} = o(1)$, *some constants* $c \geq 0$, $\epsilon \in (0, 1]$;

$$(ii) \quad ||\Gamma_2(\theta, h_o)[h - h_o] - \Gamma_2(\theta_o, h_o)[h - h_o]|| \leq c_1||\theta - \theta_o|| \times ||h - h_o||_{\mathcal{H}}^{\epsilon_1}$$

   for all θ *with* $||\theta - \theta_o|| = o(1)$, *all* h *with* $||h - h_o||_{\mathcal{H}} = o(1)$, *some constants* $c_1 \geq 0$, $\epsilon_1 \in (0, 1]$.

4. $c \times ||\widehat{h} - h_o||_{\mathcal{H}}^{1+\epsilon} = o_p(n^{-1/2})$ *uniformly over* θ *with* $||\theta - \theta_o|| = o(1)$.

5. *For all sequences of positive numbers* $\{\delta_n\}$ *with* $\delta_n = o(1)$,

$$\sup_{\|\theta-\theta_o\|<\delta_n, \|h-h_o\|_{\mathcal{H}}<\delta_n} \frac{\|M_n(\theta, h) - M(\theta, h) - M_n(\theta_o, h_o)\|}{n^{-1/2} + ||M_n(\theta, h)|| + ||M(\theta, h)||} = o_p(1).$$

6. *For some positive finite matrix* V_1, $\sqrt{n}\{M_n(\theta_o, h_o) + \Gamma_2(\theta_o,h_o)[\widehat{h} - h_o]\} \Longrightarrow \mathcal{N}[0, V_1]$.

Then

$$\sqrt{n}(\widehat{\theta} - \theta_o) \Longrightarrow \mathcal{N}[0, (\Gamma_1'\Gamma_1)^{-1}\Gamma_1'V_1\Gamma_1(\Gamma_1'\Gamma_1)^{-1}].$$

Remark 2: (i) Given condition 2, we have that condition 5 is implied by the **Condition 5'**: *for all positive values* $\delta_n = o(1)$,

$$\sup_{\|\theta-\theta_o\|<\delta_n, \|h-h_o\|_{\mathcal{H}}<\delta_n} \|\sqrt{n}\{M_n(\theta, h) - M(\theta, h) - M_n(\theta_o, h_o)\}\| = o_p(1).$$

(ii) If $M_n(\theta, h) = M_n(\theta, h_o)$ and $M(\theta, h) = M(\theta, h_o)$ for all h, then conditions 3 and 4 are automatically satisfied, and our Theorem 2 becomes Pakes and Pollard's (1989) Theorem 3.3 and is also very similar to Theorem 7.2 of Newey and McFadden (1994).

(iii) Comparing our Theorem 2 to Newey's (1994) Lemma 5.3 in the case $M(\theta, h) = E[M_n(\theta, h)] = E[m(Z, \theta, h)]$, the main difference is that while we impose smoothness assumption on $E[m(Z, \theta, h)]$ (with respect to θ, h), Newey imposes smoothness assumption directly on $m(Z, \theta, h)$ (with respect to θ, h). More precisely, our conditions 2 and 3(i) on $E[m(Z, \theta, h)]$ is similar to Newey's (1994) assumptions 5.6(ii)(iii) and 5.1(i) on $m(Z, \theta, h)$.

(iv) While most commonly seen assumptions such as Newey's (1994) assumption 5.1(i) require $\epsilon = 1$, our condition 3(i) allows for $\epsilon \in (0, 1]$, that is $M(\theta, h)$ could be not very smooth with respect to h. However the smaller ϵ has to be compensated by the faster rate of convergence of $\widehat{h}$ to h_o whenever $c > 0$ in condition 3(i). Since when $c > 0$, our condition 4 becomes $||\widehat{h} - h_o||_{\mathcal{H}} = o_p(n^{-\frac{1}{2(1+\epsilon)}})$.[8] Notice that when $\epsilon = 1$ and $c > 0$, our condition 4 becomes $||\widehat{h} - h_o||_{\mathcal{H}} = o_p(n^{-1/4})$ which is the well-known assumption 5.1(ii) in Newey (1994).

(v) Besides that we directly impose smoothness conditions on $E[m(Z, \theta, h)]$, our approach and assumptions are very similar to those in Newey (1994), Andrews (1994a) and Pakes and Olley (1995). Although we do not provide any examples where h explicitly depends on θ, our results do cover such cases; see the above mentioned papers for such examples.

3.3 A formula for the asymptotic variance

Denote the linear approximation error for $M(\theta, h)$ as

$$R(\theta, h_o, h - h_o) \equiv M(\theta, h) - M(\theta, h_o) - \Gamma_2(\theta, h_o)[h - h_o],$$

and let $\tau_n(\theta, h) \equiv M_n(\theta, h) - M(\theta, h)$. Then conditions 3(i) and 4 imply $||R(\theta, h_o, \widehat{h} - h_o)|| = o_p(n^{-1/2})$, and conditions 1-6 imply $\tau_n(\theta, \widehat{h}) - \tau_n(\theta, h_o) = o_p(n^{-1/2})$, both uniformly over θ within a $o(1)$ neighborhood of θ_o. Therefore

$$\begin{aligned}
M_n(\theta, \widehat{h}) &= M(\theta, h_o) + M(\theta, \widehat{h}) - M(\theta, h_o) + M_n(\theta, \widehat{h}) - M(\theta, \widehat{h}) \\
&= M(\theta, h_o) + \Gamma_2(\theta, h_o)[\widehat{h} - h_o] + R(\theta, h_o, \widehat{h} - h_o) + M_n(\theta, \widehat{h}) - M(\theta, \widehat{h}) \\
&= \mathcal{M}_n(\theta) + R(\theta, h_o, \widehat{h} - h_o) + \tau_n(\theta, \widehat{h}) - \tau_n(\theta, h_o) \\
&= \mathcal{M}_n(\theta) + o_p(n^{-1/2}),
\end{aligned}$$

where $\mathcal{M}_n(\theta)$ denotes the linear approximation:

$$\mathcal{M}_n(\theta) \equiv M_n(\theta, h_o) + \Gamma_2(\theta, h_o)[\widehat{h} - h_o]. \tag{1}$$

[8]Notice that $c = 0$ in condition 3(i) if and only if $M(\theta, h) - M(\theta, h_o) - \Gamma_2(\theta, h_o)[h - h_o] \equiv 0$ for any direction $[h - h_o]$, which is not likely to be true in many applications.

Condition 6 in Theorem 2 becomes $\sqrt{n}\mathcal{M}_n(\theta_o)\Longrightarrow \mathcal{N}(0,V_1)$ for some positive finite V_1, which is implied by the following **Condition 6'**:

(i) $M_n(\theta,h)=\frac{1}{n}\sum_{i=1}^n m(Z_{i,}\theta,h)$, $E[m(Z,\theta_o,h_o)]=0$, $E[||m(Z,\theta_o,h_o)||^2]<\infty$;

(ii) There is $\gamma_2(Z)$ *such that* $E[\gamma_2(Z)]=0$, $E[||\gamma_2(Z)||^2]<\infty$ *and*

$$\frac{1}{n}\sum_{i=1}^n \gamma_2(Z_i)-\Gamma_2(\theta_o,h_o)[\widehat{h}-h_o]=o_p(n^{-1/2});$$

(iii) The matrix $V_1\equiv E[\{m(Z,\theta_o,h_o)+\gamma_2(Z)\}\{m(Z,\theta_o,h_o)+\gamma_2(Z)\}']$ *is finite, positive definite.*

Note that our condition 6'(ii) is essentially the same as Newey's (1994) assumption 5.3.

4 Empirical Processes

In this section, we introduce some empirical process terms; we use these concepts to define sufficient conditions for condition 5' of Theorem 1 and condition 5' of Theorem 2. To assist in this we state some high level theorems of van der Vaart and Wellner (1996). In the sequel we shall suppose that $M(\theta,h)=E[M_n(\theta,h)]=E[m(Z,\theta,h)]$, and define the class $\mathcal{F}=\{m(Z,\theta,h):\theta\in\Theta,h\in\mathcal{H}\}$ of measurable functions indexed by (θ,h).

4.1 An alternative condition for ULLN

Condition 5' of Theorem 1 is satisfied when $\mathcal{F}$ is a $P-$*Glivenko-Cantelli class* (i.e., $\sup_{m\in\mathcal{F}}||\int m\cdot d(P_n-P)||\to 0$ *almost surely,* where P,P_n are the population and empirical measures respectively). Whether or not $\mathcal{F}$ is a $P-$Glivenko-Cantelli class is closely linked to its $L_1(P)-$covering numbers with bracketing, see van der Vaart and Wellner (1996, p. 81-83). Let $L_r(Q), 1\le r<\infty$, be the space of measurable functions with finite r-th moments against probability measure Q; this is a separable Banach space under the norm $||f||_{L_r(Q)}=\{\int|f|^r dQ\}^{1/r}<\infty$. Finally for a class of real-valued measurable functions $\mathcal{F}=\{f:\mathbf{Z}\to\mathbb{R}\}$, we call F its envelope function if $\sup_{f\in\mathcal{F}}|f(z)|\le F(z)$ for all $z\in\mathbf{Z}$.

Definition 1: Let $(\mathcal{G},||\cdot||_{\mathcal{G}})$ be a subset of a normed space of real-valued functions $g:\mathcal{X}\to\mathbb{R}$ on some set.

(i) The *covering number* $N(\varepsilon,\mathcal{G},||\cdot||_{\mathcal{G}})$ is the minimal number of N for which there exist $\varepsilon-$balls $\{\{f:||f-g_j||_{\mathcal{G}}\le\varepsilon\}, ||g_j||_{\mathcal{G}}<\infty, j=1,\ldots,N\}$ to cover $\mathcal{G}$ (i.e., for each $g\in\mathcal{G}$, there is a $j=j(g)\in\{1,\ldots,N\}$ such that $||f-g_j||_{\mathcal{G}}\le\varepsilon$).

(ii) The *covering number with bracketing* $N_{[]}(\varepsilon, \mathcal{G}, ||\cdot||_{\mathcal{G}})$ is the minimal number of N for which there exist $\varepsilon-$brackets $\{[l_j, u_j] : ||l_j - u_j||_{\mathcal{G}} \leq \varepsilon, ||l_j||_{\mathcal{G}}, ||u_j||_{\mathcal{G}} < \infty, j = 1, \ldots, N\}$ to cover $\mathcal{G}$ (i.e., for each $g \in \mathcal{G}$, there is a $j = j(g) \in \{1, \ldots, N\}$ such that $l_j \leq g \leq u_j$).

There are many sufficient conditions to ensure that $\mathcal{F}$ is a $P-$Glivenko-Cantelli class. Here we state Theorem 2.4.1 of van der Vaart and Wellner (1996, p.84):

Lemma VW1: *Let $\mathcal{F} = \{m(Z, \theta, h) : \theta \in \Theta, h \in \mathcal{H}\}$ be $P-$measurable and $\{Z_i : i = 1, \ldots, n\}$ be i.i.d. If either (a) or (b) holds:*

(a) $N_{[]}(\varepsilon, \mathcal{F}, ||\cdot||_{L_1(P)}) < \infty$ for every $\varepsilon > 0$.

(b) $\mathcal{F}$ has envelope F such that $\int F dP < \infty$ and

$$\sup_{Q:||F||_{L_1(Q)}<\infty} N(\varepsilon ||F||_{L_1(Q)}, \mathcal{F}, ||\cdot||_{L_1(Q)}) < \infty \quad \textit{for every } \varepsilon > 0,$$

where the $\sup_{Q:||F||_{L_1(Q)}<\infty}$ is taken over all the probability measures on $\mathbf{Z}$ that concentrate on a finite set and have $||F||_{L_1(Q)} < \infty$.
Then $\mathcal{F}$ is a $P-$Glivenko-Cantelli class.

Lemma VW1(b) implies Pakes and Pollard's Lemma 2.8.[9] We will later verify the conditions of this theorem under primitive conditions.

4.2 An alternative condition for stochastic equicontinuity

Condition 5' of Theorem 2 is implied by the assumption that $\mathcal{F}$ is a $P-$*Donsker class* or satisfies *stochastic equicontinuity*, see van der Vaart and Wellner (1996, p. 81-90, corollary 2.3.12).

Definition 2: Let $\{Z_i\}$ be an i.i.d. sequence drawn from P. A measurable class $\mathcal{G}$ is $P-$Donsker if and only if: $\mathcal{G}$ is totally bounded for the pseudo norm $||g||_{\mathcal{G}} = [\mathrm{var}(g)]^{1/2}$ [i.e., $N(\varepsilon, \mathcal{G}, [\mathrm{var}(\cdot)]^{1/2}) < \infty$ for every $\varepsilon > 0$]; and $\mathcal{G}$ is asymptotically equicontinuous: for every $\varepsilon > 0$,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} P^* \left(\sup_{g_1, g_2 \in \mathcal{G}: \mathrm{var}(g_1 - g_2) < \delta} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{g_1(Z_i) - g_2(Z_i) - E[g_1(Z_i) - g_2(Z_i)]\} \right| > \varepsilon \right) = 0,$$

where P^* denotes the outer-measure.

[9] For the special case $h \equiv h_o$, Pakes and Pollard (1989)'s Lemma 2.8 states that the class $\mathcal{F} = \{m(Z, \theta, h_o) : \theta \in \Theta\}$ is a $P-$Glivenko-Cantelli class if it is a *Euclidean class* for envelope F with $\int F dP < \infty$. Recall that $\mathcal{F} = \{m(Z, \theta, h_o) : \theta \in \Theta\}$ is *Euclidean* if and only if there exist some positive finite constants A, a such that $\sup_Q N(\varepsilon ||F||_{L_1(Q)}, \mathcal{F}, ||\cdot||_{L_1(Q)}) \leq A\varepsilon^{-a}$ for every $\varepsilon > 0$.

Again there are many sufficient conditions to ensure that $\mathcal{F}$ is a $P-$Donsker class. Here we state Theorem 2.5.6 of van der Vaart and Wellner (1996):

Lemma VW2: *Let* $\mathcal{F} = \{m(Z, \theta, h) : \theta \in \Theta, h \in \mathcal{H}\}$ *be* $P-$*measurable and* $\{Z_i : i = 1, \ldots, n\}$ *be i.i.d. If either (a) or (b) holds:*

(a) $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, ||\cdot||_{L_2(P)})} d\varepsilon < \infty$.

(b) $\mathcal{F}$ *has envelope* F *with* $\int F^2 dP < \infty$ *and*

$$\int_0^\infty \sup_{Q:||F||_{L_2(Q)}<\infty} \sqrt{\log N(\varepsilon ||F||_{L_2(Q)}, \mathcal{F}, ||\cdot||_{L_2(Q)})} d\varepsilon < \infty,$$

where the $\sup_{Q:||F||_{L_2(Q)}<\infty}$ *is taken over all the probability measures on* $\mathbf{Z}$ *that concentrate on a finite set and have* $||F||_{L_2(Q)} < \infty$.

Then $\mathcal{F}$ *is a* $P-$*Donsker class.*

For the special case $h \equiv h_o$, Pakes and Pollard (1989)'s Lemma 2.16 states that the class $\mathcal{F} = \{m(Z, \theta, h_o) : \theta \in \Theta\}$ is a $P-$Donsker class if it is a Euclidean class for envelope F with $\int F^2 dP < \infty$. Again the above Lemma VW2(b) implies Pakes and Pollard's Lemma 2.16.

4.3 Primitive Conditions for Stochastic Equicontinuity

We now verify condition 5' of Theorem 1 and condition 5' of Theorem 2. Since Lemma VW2 implies Lemma VW1, we will focus on verifying (a) or (b) of Lemma VW2.

The key is to compute the covering numbers with bracketing for the moment class $\{m(Z, \theta, h)\}$ based on the covering numbers with/without bracketing for the parameter class $\{\theta \in \Theta, h \in \mathcal{H} : \|h - h_o\|_{\mathcal{H}} \leq \delta_n\}$. Since Θ is a compact subset of $\mathbb{R}^k$, the covering number of Θ is known. Since in most applications, we estimate h_o by some nonparametric smoothing methods such as kernel and sieve procedures, h_o is often assumed to be in $\mathcal{H}$, a space of smooth functions such as a Sobolev, Hölder or Besov class, or at least lie there with probability tending to one. Therefore the covering number of the function space $\mathcal{H}$ can be found in many books and papers on approximation theory. When the moment function $m(Z, \theta, h)$ is Lipschitz continuous with respect to (θ, h), we can directly bound $N_{[]}(\varepsilon, \mathcal{F}, ||\cdot||_{L_2(P)})$ from above by the covering number of the parameter class $\{\theta \in \Theta, h \in \mathcal{H} : \|h - h_o\|_{\mathcal{H}} \leq \delta_n\}$, see e.g., Theorem 2.7.11 of van der Vaart and Wellner (1996). This is the approach taken in Chen and Shen (1998) and many others. When the moment function $m(Z, \theta, h)$ is Lipschitz continuous with respect to h but not in θ, we can sometimes still apply the results in Andrews

(1994b). However we are unaware of general results to handle the case where the moment function $m(Z,\theta,h)$ is not Lipschitz continuous with respect to h. In the following, Theorem 3 extends the work of Andrews (1994b) to the case where the moment function $m(Z,\theta,h)$ could be not (pointwise) continuous with respect to h and θ.

Theorem 3. *Suppose that each component m_j of $m=(m_1,\ldots,m_l)'$ takes the form $m_j(Z,\theta,h(\cdot)) = m_{cj}(Z,\theta,h(\cdot))+m_{lcj}(Z,\theta,h(\cdot))$, and satisfies:*

1. *$m_{cj}(Z,\theta,h(\cdot))$ is Hölder continuous with respect to θ,h in the sense:*

$$|m_{cj}(Z,\theta_1,h_1)-m_{cj}(Z,\theta_2,h_2)| \le b_j(Z)\{\|\theta_1-\theta_2\|^{s_{1j}}+\|h_1-h_2\|_{\mathcal{H}}^{s_j}\}$$

   for some constants s_{1j}, $s_j\in(0,1]$, a measurable function $b_j(Z)$ with $E[b_j(Z)]^r<\infty$ for some $r\ge 2$.

2. *$m_{lcj}(Z,\theta,h(\cdot))$ is locally uniformly $L_r(P),(r\ge 2)-$continuous with respect to θ,h in the sense:*

$$\left(E\left[\sup_{(\theta',h'):||\theta'-\theta||<\delta,||h'-h||_{\mathcal{H}}<\delta}|m_{lcj}(Z,\theta',h')-m_{lcj}(Z,\theta,h)|^r\right]\right)^{1/r} \le K_j\delta^{s_j}$$

   for all $(\theta,h)\in\Theta\times\mathcal{H}$, all small positive value $\delta=o(1)$, and for some constants $s_j\in(0,1]$, $K_j>0$.

3. *Θ is a compact subset of $\mathbb{R}^k$, and $\int_0^\infty \sqrt{\log N(\varepsilon^{1/s_j},\mathcal{H},||\cdot||_{\mathcal{H}})}d\varepsilon<\infty$ for $j=1,...,l$.*

Then: (i) the class $\mathcal{F}=\{m(Z,\theta,h):\theta\in\Theta,h\in\mathcal{H}\}$ is $P-$Donsker, and

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon,\mathcal{F},||\cdot||_{L_r(P)})}d\varepsilon<\infty.$$

(ii) for all positive δ_n with $\delta_n=o(1)$,

$$\sup_{\|\theta-\theta_o\|<\delta_n,\|h-h_o\|_{\mathcal{H}}<\delta_n} \|\frac{1}{\sqrt{n}}\sum_{i=1}^n\{m(Z_i,\theta,h(\cdot))-Em(Z_i,\theta,h(\cdot))-m(Z_i,\theta_o,h_o(\cdot))\}\|=o_p(1).$$

Remark 3: (i) Condition 1 of Theorem 3 is an extension of the "type II class" of Andrews (1994b) from $\theta\in\Theta$ to $(\theta,h)\in\Theta\times\mathcal{H}$; Condition 2 of Theorem 3 is an extension of the "type IV class" of Andrews (1994b) from $\theta \in \Theta$ to $(\theta, h) \in \Theta \times \mathcal{H}$. Condition 2 allows for discontinuous moment functions in (θ, h) such as sign and indicator functions of (θ, h).

(ii) Condition 3 of Theorem 3 allows for many nonparametric estimators of h_o. As an example, we present a popular nonparametric class $\mathcal{H}$. For any vector $a = (a_1, \ldots, a_{d_x})$ of d_x integers, define the differential operator

$$D^a = \frac{\partial^{|a|}}{\partial x_1^{a_1} \ldots \partial x_{d_x}^{a_{d_x}}},$$

where $|a| = \sum_{i=1}^{d_x} a_i$. Let R_X be a bounded, convex subset of $\mathbb{R}^{d_x}$ with nonempty interior. For any smooth function $h : R_X \to \mathbb{R}$ and some $\alpha > 0$, let $\underline{\alpha}$ be the largest integer smaller than α, and

$$\|h\|_{\infty,\alpha} = \max_{|a| \leq \underline{\alpha}} \sup_x |D^a h(x)| + \max_{|a| = \underline{\alpha}} \sup_{x \neq x'} \frac{|D^a h(x) - D^a h(x')|}{\|x - x'\|^{\alpha - \underline{\alpha}}}$$

Further, let $C_c^\alpha(R_X)$ be the set of all continuous functions $h : R_X \to \mathbb{R}$ with $\|h\|_{\infty,\alpha} \leq c$. If $\mathcal{H} = C_c^\alpha(R_X)$ with $|| \cdot ||_{\mathcal{H}} = || \cdot ||_\infty$, then $\log N(\delta, C_c^\alpha(R_X), || \cdot ||_\infty) \leq K(c/\delta)^{d_x/\alpha}$, where K is a constant depending only on d_x, α and Lebesgue measure of the domain R_X, hence

$$\int_0^\infty \sqrt{\log N(\varepsilon^{1/s}, C_c^\alpha(R_X), || \cdot ||_\infty)} d\varepsilon < \infty \quad \text{if} \quad \alpha > \frac{d_x}{2s}.$$

That is, when the sample moment function $m(Z, \theta, h)$ is less smooth in h (i.e., smaller $s < 1$), we need $h \in \mathcal{H}$ to be a "smaller function space" (i.e., $\alpha > d_x/2s$ or higher smoothness of h) to satisfy the stochastic equicontinuity condition 5'. Thus there is a trade-off between the smoothness of the criterion function and the smoothness of the nuisance function.

We conclude this section with another theorem which is based on the idea in Van Keilegom and Akritas (1999). Notice that Theorem 4 could be regarded as a consequence of Theorem 3, but is directly applicable for many examples.

Theorem 4. *Suppose that each component m_j of $m = (m_1, \ldots, m_l)'$ satisfies:*

1. *For fixed h, $m_j(Z, \theta, h(\cdot))$ is monotone with respect to θ; And for fixed θ, $m_j(Z, \theta, h(\cdot))$ is monotone with respect to h.*

2. *$m_j(Z, \theta, h(\cdot))$ is $L_r(P), (r \geq 2)-$Hölder continuous with respect to θ, h in the sense:*

$$\|m_j(Z, \theta, h) - m_j(Z, \theta', h')\|_{L_r(P)} \leq K_j\{\|\theta - \theta'\|^{s_{1j}} + \|h - h'\|_{\mathcal{H}}^{s_j}\}$$

for some constants s_{1j}, $s_j \in (0, 1]$, $K_j > 0$.

3. Θ *is a compact subset of* $\mathbb{R}^k$, $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon^{1/s_j}, \mathcal{H}, ||\cdot||_{\mathcal{H}})} d\varepsilon < \infty$ *for* $j = 1, ..., l$.

Then all the conclusions of Theorem 3 hold.

Remark 4: (i) For the same $(\mathcal{H}, ||\cdot||_{\mathcal{H}})$, $N(\varepsilon^{1/s_j}, \mathcal{H}, ||\cdot||_{\mathcal{H}}) \leq N_{[]}(\varepsilon^{1/s_j}, \mathcal{H}, ||\cdot||_{\mathcal{H}})$. Hence condition 3 of Theorem 4 is a stronger requirement than condition 3 of Theorem 3. For example, for the space $\mathcal{H} = C_c^\alpha(R_X)$, if we still take $||\cdot||_{\mathcal{H}} = ||\cdot||_\infty$, the upper bound for $\log N_{[]}(\delta, C_c^\alpha(R_X), ||\cdot||_\infty)$ is unknown. Hence we do not know whether $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon^{1/s}, C_c^\alpha(R_X), ||\cdot||_\infty)} d\varepsilon < \infty$ or $= \infty$. However, since

$$\log N_{[]}(\delta, C_c^\alpha(R_X), ||\cdot||_{L_r(P)}) \leq const. \log N(\delta, C_c^\alpha(R_X), ||\cdot||_\infty) \leq K(\frac{c}{\delta})^{d_x/\alpha} \quad \text{for any } \delta > 0,$$

for any $r \in [1, \infty)$, see e.g., Corollary 2.7.2 in van der Vaart and Wellner (1996), if we take $||\cdot||_{\mathcal{H}} = ||\cdot||_{L_2(P)}$, then $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon^{1/s}, C_c^\alpha(R_X), ||\cdot||_{L_2(P)})} d\varepsilon < \infty$ if $\alpha > d_x/2s$.

(ii) Conditions 1 and 2 of Theorem 4 imply that condition 2 of Theorem 3 is satisfied, condition 3 of Theorem 4 implies that condition 3 of Theorem 3 is satisfied. Hence Theorem 4 is a direct consequence of Theorem 3.

The conditions of these theorems are relatively easy to verify; in the next section we verify them in some specific examples.

5 An Example

Here we present an example that is of interest in itself but is simple enough to make clear the argument without requiring too much detail. Suppose that one wants to estimate the parameter

$$\theta_o = \Pr[h_o(X) \in A(Z)],$$

where $A(Z)$ is a random set that depends on the data Z, and h_o is an unknown function. The natural estimator of θ_o is the sample analogue

$$\widehat{\theta} = \frac{1}{n} \sum_{i=1}^n 1(\widehat{h}(X_i) \in A(Z_i)), \tag{2}$$

where $\widehat{h}(X_i)$ is some nonparametric estimate of $h(X_i)$. The estimator $\widehat{\theta}$ can be interpreted as a GMM estimator by taking the sample moment condition to be $M_n(\theta, h) = \frac{1}{n} \sum_{i=1}^n 1(h(X_i) \in A(Z_i)) - \theta$.

One potential application of this is in testing for Revealed Preference as in Blundell, Browning, and Crawford (1997, section 4.3). Their test involves comparing a weighted combination of budget shares [estimated nonparametrically] with some relative prices. They do a test that uses the pointwise confidence interval and then counts violations. Instead one could do a global test based on a count like (2). Another application is in evaluating the fit of a nonparametric procedure.

Bliss (1997) uses a criterion like (2) to evaluate yield curve fits, and we shall pursue his example here. In this case, one observes a bid and an ask quote on a bond, $\{p_{Li}, p_{Ui}\}$, along with maturity and payment information. For convenience, the mid-point of the bid and ask price is taken as a proxy for the actual price p_i. From this one can estimate nonparametrically the discount function and the yield curve [see Linton, Mammen, Nielsen, and Tanggaard (2000)], and hence obtain a predicted price $\widehat{p}_i$ for each bond. One way of evaluating the performance of the procedure is to calculate the so-called hit rate, which is $n^{-1}\sum_{i=1}^n 1(p_{Li} \leq \widehat{p}_i \leq p_{Ui})$. A high hit rate corresponds to a good procedure. Smoothing over the criterion function is not a particularly attractive option here, although it would lead to straightforward but messy distribution theory.[10]

We will suppose that the bonds are all normalized to have unit payoff at maturity. Suppose that $p_i = h_o(\tau_i) + \varepsilon_i$, where ε_i is a mean zero error term. The function $h_o(.)$ is the discount function. We shall suppose that (τ_i, ε_i) are i.i.d. and mutually independent. Suppose that we observe (p_{Li}, p_{Ui}, τ_i) for each bond, where $p_{Li} = p_i - \eta_{1i}$ and $p_{Ui} = p_i + \eta_{2i}$, where η_{ji}, $j = 1, 2$ are from the same distribution F_η, which has support $[0, \infty)$. Define the mid-point $p_{Mi} = (p_{Li} + p_{Ui})/2$; then $p_{Mi} = h_o(\tau_i) + u_i$, where $u_i = \varepsilon_i + (\eta_{2i} - \eta_{1i})/2$ is a mean zero i.i.d. error term. We estimate $\theta_o = \Pr[p_{Li} \leq h_o(\tau_i) \leq p_{Ui}]$ based on the moment function:

$$M_n(\theta, \widehat{h}) = \frac{1}{n}\sum_{i=1}^n 1\left(p_{Li} \leq \widehat{p}_i \leq p_{Ui}\right) - \theta,$$

where $\widehat{p}_i = \widehat{h}(\tau_i)$, and $\widehat{h}(\tau_i)$ is some nonparametric estimate of $h_o(\tau_i)$, specifically the local linear

[10] Suppose we take instead

$$\widehat{\theta}_b = \frac{1}{n}\sum_{i=1}^n \mathcal{K}\left(\frac{p_{Li} - \widehat{p}_i}{b}\right)\mathcal{K}\left(\frac{\widehat{p}_i - p_{Ui}}{b}\right),$$

where $\mathcal{K}$ is a c.d.f. and b is a bandwidth sequence. As $b \to 0$ we get our estimator, but as $b \to \infty$, $\widehat{\theta}_b \to 1$. For intermediate values of b, we could have $\widehat{\theta}_b < \widehat{\theta}$. Under suitable conditions on the rate at which $b(n) \to 0$ as $n \to \infty$, $\widehat{\theta}_b$ is consistent and asymptotically normal.

kernel estimator $\widehat{h}(\tau_i) = \widehat{\beta}_0$, where $(\widehat{\beta}_0, \widehat{\beta}_1)$ minimize

$$\sum_{\substack{j=1 \\ j\neq i}}^{n} K\left(\frac{\tau_i - \tau_j}{\lambda_n}\right) \{p_{Mj} - \beta_0 - \beta_1(\tau_j - \tau_i)\}^2 ,$$

where $K(.)$ is some kernel function and $\{\lambda_n\}$ is a bandwidth sequence. Under mild conditions we have $||\widehat{h} - h_o||_{\mathcal{H}} = \sup_{\tau\in[0,T]} |\widehat{h}(\tau) - h_o(\tau)| = o_p(1)$. Let $\widehat{\theta}$ be the resulting estimator of θ_o,

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} 1\left(p_{Li} \leq \widehat{h}(\tau_i) \leq p_{Ui}\right).$$

In the proposition below we give primitive conditions on the data distributions that imply the high level conditions of Theorems 1 and 2 are satisfied. In turn, we verify the stochastic equicontinuity conditions by taking $s = 1/2$ in Theorem 3. We take $\mathcal{H} = C_c^{\alpha}([0,T])$ with $\alpha > 1$. Denote by $F_\varepsilon, f_\varepsilon$ the c.d.f. and density of ε, while let F_η, f_η be the c.d.f. and density function of η.

Proposition 1. *Suppose that* $\theta_o = \Pr[p_{Li} \leq h_o(\tau_i) \leq p_{Ui}] \in (0,1)$, *and that:*

1. h_o *is twice continuously differentiable*
2. $(\tau_i, \varepsilon_i, \eta_{1i}, \eta_{2i})$ *are i.i.d., and* $(\varepsilon_i, \eta_{1i}, \eta_{2i})$ *are mutually independent and independent of* τ_i.
3. *The payment times* τ_i *have a Lebesgue density* f_τ *that is bounded away from zero on its compact support* $[0,T]$.
4. f_ε *is continuously differentiable and* $\sup_\bullet |f'_\varepsilon(\bullet)| < \infty$.
5. $u_i = \varepsilon_i + (\eta_{2i} - \eta_{1i})/2$ *satisfies* $E(u_i) = 0$ *and* $\mathrm{var}(u_i) < \infty$.
6. *The kernel function is a symmetric probability density function with bounded support* $[-1,1]$ *such that* $K \in C_c^{\alpha}([-1,1])$.
7. *The bandwidth sequence* λ_n *satisfies* $n\lambda_n^{1+2\alpha}/\log n \to \infty$ *and* $n\lambda_n^4 \to 0$.

Then

$$\sqrt{n}(\widehat{\theta} - \theta_o) \Longrightarrow \mathcal{N}[0, \mathrm{var}(\zeta_i)],$$

where $\zeta_i \equiv 1(-\eta_{1i} \leq -\varepsilon_i \leq \eta_{2i}) - \theta_o + \psi \cdot u_i$ *with* $\psi \equiv \int_0^\infty \{f_\varepsilon(\eta) + f_\varepsilon(-\eta)\} f_\eta(\eta) d\eta$.

To conclude this section we report the results of a small simulation experiment. We took $h_0(\tau) = \exp(-0.1\tau)$, with $\tau \sim U[0,1]$, $\varepsilon \sim N(0, 0.01)$, and $\eta_j \sim U[0, 0.2]$. In this case $\theta = 0.60$. We used a local linear estimator with bandwidth determined by a Silverman's rule of thumb on the covariate density; this bandwidth rate does not satisfy our regularity conditions so we also recomputed the estimator with scaling factor $n^{-2/7}$, which does. Table 1 gives the results based on 1000 replications. The standard deviations decline at the predicted rate; the bias, although small numerically, seems to decline very slowly. Figure 1 gives a density plot of the relocated and scaled hit rates against a standard normal density and shows that the estimator is very close to being normally distributed.

Table 1

	$\lambda(n) = 1.06 s_\tau n^{-1/5}$		$\lambda(n) = 1.06 s_\tau n^{-2/7}$	
Sample Size n	Mean	Std	Mean	Std
100	0.5840	0.0516	0.5803	0.0520
200	0.5842	0.0368	0.5835	0.0365
400	0.5833	0.0264	0.5840	0.0260
800	0.5837	0.0187	0.5843	0.0184
1600	0.5844	0.0128	0.5847	0.0135

6 Conclusion

We presented our results for the semiparametric estimators close to those in Andrews (1994a) and Newey (1994), except that our criterion functions may be non-smooth with respect to both finite and infinite dimensional parameters. Our new results on stochastic equicontinuity for the non-smooth case are applicable beyond the estimation problems we have addressed in this paper. For example, Breiman (1996) and Granger (2000) have discussed ways to improve forecasting by combining different predictors (say $\widehat{h}_j(\cdot), j = 1, \ldots, J$), and some of the procedures involve terms like $1(Y_i \widehat{h}_j(X_i) \geq 0)$. A large class of goodness-of-fit testing problems also involve functions that depend on nonparametric estimators non-smoothly. For example, a natural extension of Andrews' (1997) Kolmogorov test of parametric conditional distribution specification $F_{Y|X}(\bullet; \theta_o)$ is to the case where the null hypothesis itself is semiparametric or nonparametric; the resulting test statistic is constructed using residuals from this more general null hypothesis and would not satisfy the conditions of Andrews (1997).

7 Appendix

We give the proofs of Theorems 1-4 here.

Proof of Theorem 1. The proof is similar to that of Corollary 3.2 in Pakes and Pollard (1989). By condition 2, for all $\delta > 0$,

$$\Pr(||\widehat{\theta} - \theta_o|| > \delta) \leq \Pr(||M(\widehat{\theta}, h_o)|| \geq \epsilon(\delta)),$$

hence it suffices to show that $||M(\widehat{\theta}, h_o)|| = o_p(1)$.

Now by the triangle inequality,

$$||M(\widehat{\theta}, h_o)|| \leq ||M(\widehat{\theta}, h_o) - M(\widehat{\theta}, \widehat{h})|| + ||M(\widehat{\theta}, \widehat{h}) - M_n(\widehat{\theta}, \widehat{h})|| + ||M_n(\widehat{\theta}, \widehat{h})||.$$

Hence by conditions 3 and 4,

$$||M(\widehat{\theta}, h_o) - M(\widehat{\theta}, \widehat{h})|| = o_p(1), \tag{3}$$

and by conditions 3 and 5,

$$\begin{aligned} ||M(\widehat{\theta}, \widehat{h}) - M_n(\widehat{\theta}, \widehat{h})|| &\leq o_p(1) \times \{1 + ||M_n(\widehat{\theta}, \widehat{h})|| + ||M(\widehat{\theta}, \widehat{h})||\} \\ \text{(by (3))} &\leq o_p(1) \times \{1 + ||M_n(\widehat{\theta}, \widehat{h})|| + ||M(\widehat{\theta}, h_o)|| + o_p(1)\}. \end{aligned}$$

Hence

$$||M(\widehat{\theta}, h_o)|| \times \{1 - o_p(1)\} \leq o_p(1) + ||M_n(\widehat{\theta}, \widehat{h})|| \times \{1 + o_p(1)\}.$$

By condition 1, we have

$$||M_n(\widehat{\theta}, \widehat{h})|| \leq o_p(1) + \inf_{\theta \in \Theta} ||M_n(\theta, \widehat{h})||.$$

Again under conditions 3, 4, 5 and $M(\theta_o, h_o) = 0$, we have:

$$\begin{aligned} ||M_n(\theta, \widehat{h})|| &\leq ||M_n(\theta, \widehat{h}) - M(\theta, \widehat{h})|| + ||M(\theta, \widehat{h}) - M(\theta, h_o)|| + ||M(\theta, h_o) - M(\theta_o, h_o)|| \\ &\leq o_p(1) \times \{1 + ||M_n(\theta, \widehat{h})|| + ||M(\theta, h_o)||\} + o_p(1) + ||M(\theta, h_o) - M(\theta_o, h_o)||, \end{aligned}$$

and hence

$$||M_n(\theta, \widehat{h})|| \times \{1 - o_p(1)\} \leq o_p(1) + ||M(\theta, h_o) - M(\theta_o, h_o)|| \times \{1 + o_p(1)\},$$

where all the $o_p(1)$'s hold uniformly with respect to $\theta \in \Theta$. Now by $M(\theta_o, h_o) = 0$,

$$
\begin{aligned}
\inf_{\theta\in\Theta} ||M_n(\theta, \widehat{h})|| &\leq \sup_{\theta\in\Theta} o_p(1) + \inf_{\theta\in\Theta} ||M(\theta, h_o) - M(\theta_o, h_o)|| \times \{1 + \sup_{\theta\in\Theta} o_p(1)\} \\
&= o_p(1)
\end{aligned}
$$

and the result follows.

Proof of Theorem 2. We first establish $\sqrt{n}-$consistency of $\widehat{\theta}$ to θ_o. We choose a positive sequence $\delta_n = o(1)$ such that $\Pr(||\widehat{\theta} - \theta_o|| \geq \delta_n, ||\widehat{h} - h_o||_{\mathcal{H}} \geq \delta_n) \to 0$. Hence we just need to look at θ, h such that $\|\theta - \theta_o\| < \delta_n, ||h - h_o||_{\mathcal{H}} < \delta_n$.

By condition 2 and the triangle inequality, there is a constant $C > 0$ such that

$$
\begin{aligned}
& ||\widehat{\theta} - \theta_o||C \\
\leq\ & \|M(\widehat{\theta}, h_o)\| \\
\leq\ & ||M(\widehat{\theta}, h_o) - M(\widehat{\theta}, \widehat{h})|| + ||M(\widehat{\theta}, \widehat{h}) - M_n(\widehat{\theta}, \widehat{h}) + M_n(\theta_o, h_o)|| + ||M_n(\widehat{\theta}, \widehat{h})|| + ||M_n(\theta_o, h_o)|| \\
=\ & ||M(\widehat{\theta}, h_o) - M(\widehat{\theta}, \widehat{h})|| + ||M(\widehat{\theta}, \widehat{h}) - M_n(\widehat{\theta}, \widehat{h}) + M_n(\theta_o, h_o)|| + ||M_n(\widehat{\theta}, \widehat{h})|| + O_p(n^{-1/2}), \quad (4)
\end{aligned}
$$

where the last equation is due to condition 6.

By conditions 3, 4 and 6, and the fact that $||\widehat{\theta} - \theta_o|| < \delta_n, ||\widehat{h} - h_o||_{\mathcal{H}} < \delta_n$ with probability approaching one, we have

$$
\begin{aligned}
& ||M(\widehat{\theta}, h_o) - M(\widehat{\theta}, \widehat{h})|| \\
\leq\ & ||M(\widehat{\theta}, \widehat{h}) - M(\widehat{\theta}, h_o) - \Gamma_2(\widehat{\theta}, h_o)[\widehat{h} - h_o]|| \\
& + ||\{\Gamma_2(\widehat{\theta}, h_o) - \Gamma_2(\theta_o, h_o)\}[\widehat{h} - h_o]|| + ||\Gamma_2(\theta_o, h_o)[\widehat{h} - h_o]|| \\
\leq\ & c\{||\widehat{h} - h_o||_{\mathcal{H}}\}^{1+\epsilon} + c_1||\widehat{\theta} - \theta_o|| \times \{||\widehat{h} - h_o||_{\mathcal{H}}\}^{\epsilon_1} + ||\Gamma_2(\theta_o, h_o)[\widehat{h} - h_o]|| \\
=\ & o_p(n^{-1/2}) + ||\widehat{\theta} - \theta_o|| \times o_p(1) + O_p(n^{-1/2}) \\
\leq\ & \|M(\widehat{\theta}, h_o)\| \times o_p(1) + O_p(n^{-1/2}), \quad \text{by condition 2.} \qquad (5)
\end{aligned}
$$

And by condition 5,

$$
\begin{aligned}
& ||M(\widehat{\theta}, \widehat{h}) - M_n(\widehat{\theta}, \widehat{h}) + M_n(\theta_o, h_o)|| \\
\leq\ & o_p(1) \times \{n^{-1/2} + ||M_n(\widehat{\theta}, \widehat{h})|| + ||M(\widehat{\theta}, \widehat{h})||\} \\
\text{(by (5))} \leq\ & o_p(1) \times \{n^{-1/2} + ||M_n(\widehat{\theta}, \widehat{h})|| + ||M(\widehat{\theta}, h_o)|| + \|M(\widehat{\theta}, h_o)\| \times o_p(1) + O_p(n^{-1/2})\} \\
=\ & o_p(n^{-1/2}) + o_p(1) \times \{||M_n(\widehat{\theta}, \widehat{h})|| + ||M(\widehat{\theta}, h_o)|| \times (1 + o_p(1))\}.
\end{aligned}
$$

Hence

$$||M(\widehat{\theta}, h_o)|| \times \{1 - o_p(1)\} \leq o_p(n^{-1/2}) + ||M_n(\widehat{\theta}, \widehat{h})|| \times \{1 + o_p(1)\} + O_p(n^{-1/2}).$$

By condition 1, we have

$$||M_n(\widehat{\theta}, \widehat{h})|| \leq o_p(n^{-1/2}) + \inf_{\theta \in \Theta, ||\theta - \theta_o|| \leq \delta_n} ||M_n(\theta, \widehat{h})||.$$

Again under conditions 3, 4, 5 and 6, we have:

$$\begin{aligned} ||M_n(\theta, \widehat{h})|| &\leq ||M_n(\theta, \widehat{h}) - M(\theta, \widehat{h}) - M_n(\theta_o, h_o)|| + ||M(\theta, \widehat{h}) - M(\theta, h_o)|| + ||M(\theta, h_o)|| + ||M_n(\theta_o, h_o)|| \\ &\leq o_p(n^{-1/2}) + o_p(1) \times \{||M_n(\theta, \widehat{h})|| + ||M(\theta, h_o)|| \times (1 + o_p(1))\} + ||M(\theta, h_o)|| + O_p(n^{-1/2}) \end{aligned}$$

hence with $M(\theta_o, h_o) = 0$, we have:

$$||M_n(\theta, \widehat{h})|| \times \{1 - o_p(1)\} \leq o_p(1) \times ||M(\theta, h_o) - M(\theta_o, h_o)|| + O_p(n^{-1/2}),$$

where all the $o_p(1), O_p(n^{-1/2})$'s hold uniformly with respect to $\theta \in \Theta$ with $\|\theta - \theta_o\| < \delta_n$. Hence by condition 2

$$\inf_{\theta \in \Theta} ||M_n(\theta, \widehat{h})|| = O_p(n^{-1/2}),$$

and

$$||\widehat{\theta} - \theta_o||C \leq ||M(\widehat{\theta}, h_o)|| \leq O_p(n^{-1/2}).$$

The rest of the proof is very similar to that of Theorem 3.3 in Pakes and Pollard (1989) for $\sqrt{n}-$normality, hence we just sketch the main steps here. Let $\mathcal{L}_n(\theta) = \Gamma_1(\theta - \theta_o) + \mathcal{M}_n(\theta_o)$. By similar arguments as above, we have, for any sequence θ_n with $\sqrt{n}(\theta_n - \theta_o) = O_p(1)$, $\sqrt{n}||\mathcal{L}_n(\theta_n) - \mathcal{M}_n(\theta_n)|| = o_p(1)$. Let $\theta^* = \arg\min_{\theta \in \Theta} ||\mathcal{L}_n(\theta)||$, then $\sqrt{n}(\theta^* - \theta_o) = -(\Gamma_1'\Gamma_1)^{-1}\Gamma_1'\sqrt{n}\mathcal{M}_n(\theta_o)$; this follows with probability tending to one because θ_o is in the interior of Θ and so the constrained optimization is asymptotically equivalent to the unconstrained problem whose solution is θ^*. Under conditions 2 and 6, a standard central limit theorem now implies $\sqrt{n}(\theta^* - \theta_o) \Longrightarrow N[0, (\Gamma_1'\Gamma_1)^{-1}\Gamma_1'V_1\Gamma_1(\Gamma_1'\Gamma_1)^{-1}]$.

Finally, by following the arguments of Pakes and Pollard (1989), with their Γ, $G_n(\theta_o)$ being ours Γ_1, $\mathcal{M}_n(\theta_o)$, we obtain $\sqrt{n}(\widehat{\theta} - \theta^*) = o_p(1)$. Hence $\sqrt{n}(\widehat{\theta} - \theta_o) \Longrightarrow N[0, (\Gamma_1'\Gamma_1)^{-1}\Gamma_1'V_1\Gamma_1(\Gamma_1'\Gamma_1)^{-1}]$.

The following Lemma 1 is needed in the proofs of Theorems 3 and 4. This result is a direct extension of Pakes and Pollard's Lemma 2.17 from the case $m(Z, \theta)$ to the case $m(Z, \theta, h)$, and can be proved the same way as theirs.

Lemma 1: *Suppose that* $\{Z_i : i = 1, \ldots, n\}$ *is i.i.d., and that:*

(1) either (a) or (b) of Lemma VW2 holds.

(2) $m(\cdot, \theta, h)$ *is* $L_2(P)-$*continuous at* (θ_o, h_o)*, i.e.,* $E[||m(Z, \theta, h) - m(Z, \theta_o, h_o)||^2] \to 0$ *as* $\|\theta - \theta_o\| \to 0, \|h - h_o\|_{\mathcal{H}} \to 0$.

Then for all positive sequences $\{\delta_n\}$ *with* $\delta_n = o(1)$,

$$\sup_{\|\theta-\theta_o\|<\delta_n, \|h-h_o\|_{\mathcal{H}}<\delta_n} \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{m(Z_i, \theta, h(\cdot)) - Em(Z_i, \theta, h(\cdot)) - m(Z_i, \theta_o, h_o(\cdot))\}\| = o_p(1).$$

Proof of Theorem 3: We obtain the result by applying Lemma 1. Since either condition 1 or condition 2 of Theorem 3 implies that condition (2) of Lemma 1 is satisfied. We just need to check that (a) of Lemma VW2 holds. By example 2.10.7 in van der Vaart and Wellner (1996) or Theorem 6 in Andrews (1994b), it suffices to show that for each $j = 1, \ldots, l$, the followings are true:

(1) $\mathcal{F}_{1j} = \{m_{cj}(Z, \theta, h) : \theta \in \Theta, h \in \mathcal{H}\}$ is $P-$Donsker, or $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}_{1j}, || \cdot ||_{L_r(P)})} d\varepsilon < \infty$;

(2) $\mathcal{F}_{2j} = \{m_{lcj}(Z, \theta, h) : \theta \in \Theta, h \in \mathcal{H}\}$ is $P-$Donsker, or $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}_{2j}, || \cdot ||_{L_r(P)})} d\varepsilon < \infty$.

For **(1)**. Let $\{\theta_k : k = 1, \ldots, N_1\}$ be an $\eta^{1/s_{1j}}-$cover for $(\Theta, || \cdot ||)$, and $\{h_k : k = 1, \ldots, N_2\}$ be an $\eta^{1/s_j}-$cover for $(\mathcal{H}, || \cdot ||_{\mathcal{H}})$. Then under condition 1, for any $m_{cj}(Z, \theta, h)$, there exist $k_1 \in \{1, \ldots, N_1\}$ and $k_2 \in \{1, \ldots, N_2\}$ such that

$$m_{cj}(Z, \theta_{k_1}, h_{k_2}) - 2\eta b_j(Z) \leq m_{cj}(Z, \theta, h) \leq m_{cj}(Z, \theta_{k_1}, h_{k_2}) + 2\eta b_j(Z).$$

Since $E[b_j(Z)]^r < \infty$ for some $r \geq 2$, we have that $\{[m_{cj}(Z, \theta_{k_1}, h_{k_2}) - 2\eta b_j(Z), m_{cj}(Z, \theta_{k_1}, h_{k_2}) + 2\eta b_j(Z)] : k_1 \in \{1, \ldots, N_1\}, k_2 \in \{1, \ldots, N_2\}\}$ forms an $\varepsilon = 4\eta ||b_j(Z)||_{L_r(P)}-$bracket for $(\mathcal{F}_{1j}, || \cdot ||_{L_r(P)})$. Hence

$$N_{[]}(\varepsilon, \mathcal{F}_{1j}, || \cdot ||_{L_r(P)}) \leq N([\frac{\varepsilon}{4||b_j(Z)||_{L_r(P)}}]^{1/s_{1j}}, \Theta, || \cdot ||) \times N([\frac{\varepsilon}{4||b_j(Z)||_{L_r(P)}}]^{1/s_j}, \mathcal{H}, || \cdot ||_{\mathcal{H}}).$$

This and condition 3 imply **(1)**.

For **(2)**. Let $\{\theta_k : k = 1, \ldots, N_1\}$ be an $\delta-$cover for $(\Theta, || \cdot ||)$, and $\{h_k : k = 1, \ldots, N_2\}$ be an $\delta-$cover for $(\mathcal{H}, || \cdot ||_{\mathcal{H}})$. Then under condition 2, for any $m_{lcj}(Z, \theta, h)$, there exist $k_1 \in \{1, \ldots, N_1\}$ and $k_2 \in \{1, \ldots, N_2\}$ such that

$$\begin{aligned} & |m_{lcj}(Z, \theta, h) - m_{lcj}(Z, \theta_{k_1}, h_{k_2})| \\ \leq & \sup_{(\theta', h'): ||\theta' - \theta_{k_1}|| < \delta, ||h' - h_{k_2}||_{\mathcal{H}} < \delta} |m_{lcj}(Z, \theta', h') - m_{lcj}(Z, \theta_{k_1}, h_{k_2})| \\ \equiv & b_j(Z, \theta_{k_1}, h_{k_2}, \delta), \end{aligned}$$

hence

$$m_{lcj}(Z,\theta_{k_1},h_{k_2}) - b_j(Z,\theta_{k_1},h_{k_2},\delta) \le m_{lcj}(Z,\theta,h) \le m_{lcj}(Z,\theta_{k_1},h_{k_2}) + b_j(Z,\theta_{k_1},h_{k_2},\delta).$$

Again by condition 2, $\{E[b_j(Z,\theta_{k_1},h_{k_2},\delta)]^r\}^{1/r} \le K_j\delta^{s_j}$ for all (θ_{k_1},h_{k_2}) and all positive value $\delta = o(1)$. Therefore, $\{[m_{lcj}(Z,\theta_{k_1},h_{k_2}) - b_j(Z,\theta_{k_1},h_{k_2},\delta), m_{lcj}(Z,\theta_{k_1},h_{k_2}) + b_j(Z,\theta_{k_1},h_{k_2},\delta)] :$ $k_1 \in \{1,\ldots,N_1\}, k_2 \in \{1,\ldots,N_2\}\}$ forms an $\varepsilon = 2K_j\delta^{s_j}-$bracket for $(\mathcal{F}_{2j}, ||\cdot||_{L_r(P)})$. Hence

$$N_{[]}(\varepsilon,\mathcal{F}_{2j},||\cdot||_{L_r(P)}) \le N([\frac{\varepsilon}{2K_j}]^{1/s_j},\Theta,||\cdot||) \times N([\frac{\varepsilon}{2K_j}]^{1/s_j},\mathcal{H},||\cdot||_{\mathcal{H}}).$$

This and condition 3 imply **(2)**.

Proof of Theorem 4: Again it suffices to show that for each $j = 1,\ldots,l$, the following is true: $\mathcal{F}_j = \{m_j(Z,\theta,h) : \theta \in \Theta, h \in \mathcal{H}\}$ is $P-$Donsker, or $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon,\mathcal{F}_j,||\cdot||_{L_r(P)})}d\varepsilon < \infty$.

We assume without loss of generality that $m_j(Z,\theta,h)$ is non-decreasing in θ and h in condition 1.

Let $\{[\theta_k^L,\theta_k^U] : k = 1,\ldots,N_1\}$ be an $\eta^{1/s_{1j}}-$ bracket for $(\Theta,||\cdot||)$, and $\{[h_k^L,h_k^U] : k = 1,\ldots,N_2\}$ be an $\eta^{1/s_j}-$ bracket for $(\mathcal{H},||\cdot||_{\mathcal{H}})$. Since $m_j(Z,\theta,h)$ is non-decreasing in θ,h, we have for any $m_j(Z,\theta,h)$, there exist $k_1 \in \{1,\ldots,N_1\}$ and $k_2 \in \{1,\ldots,N_2\}$ such that $\theta_{k_1}^L \le \theta \le \theta_{k_1}^U, h_{k_2}^L \le h \le h_{k_2}^U$, and

$$m_j(Z,\theta_{k_1}^L,h_{k_2}^L) \le m_j(Z,\theta,h_{k_2}^L) \le m_j(Z,\theta,h) \le m_j(Z,\theta,h_{k_2}^U) \le m_j(Z,\theta_{k_1}^U,h_{k_2}^U).$$

By condition 2 we have

$$\|m_j(Z,\theta_{k_1}^U,h_{k_2}^U) - m_j(Z,\theta_{k_1}^L,h_{k_2}^L)\|_{L_r(P)} \le 2\eta K_j.$$

Therefore $\{[m_j(Z,\theta_{k_1}^L,h_{k_2}^L), m_j(Z,\theta_{k_1}^U,h_{k_2}^U)] : k_1 \in \{1,\ldots,N_1\}, k_2 \in \{1,\ldots,N_2\}\}$ forms an $\varepsilon = 2\eta K_j-$bracket for $(\mathcal{F}_j, ||\cdot||_{L_r(P)})$. Hence

$$N_{[]}(\varepsilon,\mathcal{F}_j,||\cdot||_{L_r(P)}) \le N_{[]}([\frac{\varepsilon}{2K_j}]^{1/s_{1j}},\Theta,||\cdot||) \times N_{[]}([\frac{\varepsilon}{2K_j}]^{1/s_j},\mathcal{H},||\cdot||_{\mathcal{H}}).$$

Now this and condition 3 imply result (i):

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon,\mathcal{F}_j,L_r(P))}d\varepsilon < \infty.$$

This shows that the class $\mathcal{F}_j$ (and hence $\mathcal{F}$) is Donsker.

Moreover condition (2) of Lemma 1 is satisfied given condition 2, hence result (ii) follows.

Proof of Proposition 1. We establish the $\sqrt{n}-$consistency and asymptotic normality by applying Theorem 2. Condition 1 of Theorem 2 is automatically satisfied. Notice that

$$\begin{aligned}\theta_o &= \Pr\left[p_{Li} \leq h_o(\tau_i) \leq p_{Ui}\right] = \Pr\left[p_i - \eta_{1i} \leq p_i - \varepsilon_i \leq p_i + \eta_{2i}\right] \\ &= \Pr\left[-\eta_{1i} \leq -\varepsilon_i \leq \eta_{2i}\right] = \int_0^\infty F_\varepsilon(\eta) f_\eta(\eta) d\eta + \int_0^\infty F_\varepsilon(-\eta) f_\eta(\eta) d\eta - 1.\end{aligned}$$

$$M(\theta, h) = E\left[1\left(p_{Li} \leq h(\tau_i) \leq p_{Ui}\right)\right] - \theta,$$

hence condition 2 of Theorem 2 is satisfied with $\Gamma_1 = -1$. By definition,

$$\begin{aligned}&\Gamma_2(\theta, h_o)[h - h_o] \\ =\ &\lim_{t\to 0^+} \frac{E\left[1\left(p_{Li} \leq \{h_o(\tau_i) + t[h(\tau_i) - h_o(\tau_i)]\} \leq p_{Ui}\right) - 1\left(p_{Li} \leq h_o(\tau_i) \leq p_{Ui}\right)\right]}{t}.\end{aligned}$$

Since

$$\begin{aligned}&\Pr\left[p_{Li} \leq \{h_o(\tau_i) + t[h(\tau_i) - h_o(\tau_i)]\} \leq p_{Ui} \mid \tau_i\right] \\ =\ &\Pr\left[p_i - \eta_{1i} \leq p_i - \varepsilon_i + t[h(\tau_i) - h_o(\tau_i)] \leq p_i + \eta_{2i} \mid \tau_i\right] \\ =\ &\Pr\left[-\eta_{1i} \leq t[h(\tau_i) - h_o(\tau_i)] - \varepsilon_i \leq \eta_{2i} \mid \tau_i\right] \\ =\ &\int_0^\infty F_\varepsilon(\eta + t[h(\tau_i) - h_o(\tau_i)]) f_\eta(\eta) d\eta + \int_0^\infty F_\varepsilon(t[h(\tau_i) - h_o(\tau_i)] - \eta) f_\eta(\eta) d\eta - 1\end{aligned}$$

we obtain

$$\begin{aligned}&\Gamma_2(\theta, h_o)[h - h_o] \\ =\ &E\left[\left\{\frac{d}{dt}\Pr\left[p_{Li} \leq b_i\{h_o(\tau_i) + t[h(\tau_i) - h_o(\tau_i)]\} \leq p_{Ui} \mid \tau_i\right]\right\} |_{t=0}\right] \\ =\ &\left\{\int_0^\infty f_\varepsilon(\eta) f_\eta(\eta) d\eta + \int_0^\infty f_\varepsilon(-\eta) f_\eta(\eta) d\eta\right\} E\left([h(\tau_i) - h_o(\tau_i)]\right) \\ \equiv\ &\psi \cdot E\left([h(\tau_i) - h_o(\tau_i)]\right).\end{aligned}$$

Hence condition 3(ii) of Theorem 2 is satisfied with $c_1 = 0$. Because

$$\begin{aligned}&E\left[1\left(p_{Li} \leq h(\tau_i) \leq p_{Ui}\right)\right] \\ =\ &E\left(E\left[p_{Li} \leq \{h_o(\tau_i) + t[h(\tau_i) - h_o(\tau_i)]\} \leq p_{Ui} \mid \tau_i\right]\right) \\ =\ &E\left[\int_0^\infty F_\varepsilon(\eta + [h(\tau_i) - h_o(\tau_i)]) f_\eta(\eta) d\eta + \int_0^\infty F_\varepsilon([h(\tau_i) - h_o(\tau_i)] - \eta) f_\eta(\eta) d\eta - 1\right],\end{aligned}$$

we have by Taylor expansion up to second term, uniformly over h with $||h-h_o||_{\mathcal{H}} = o(1)$,

$$\begin{aligned}
& M(\theta,h) - M(\theta,h_o) - \Gamma_2(\theta,h_o)[h-h_o] \\
= \;& E\left[1\left(p_{Li} \leq h(\tau_i) \leq p_{Ui}\right) - 1\left(p_{Li} \leq h_o(\tau_i) \leq p_{Ui}\right)\right] - \psi \cdot E\left([h(\tau_i) - h_o(\tau_i)]\right) \\
= \;& \frac{1}{2} E\left[\{[h(\tau_i) - h_o(\tau_i)]\}^2 \Psi(\tau_i)\right],
\end{aligned}$$

where

$$\Psi(\tau_i) \equiv \int_0^\infty \left\{f'_\varepsilon(\eta + \overline{g}(\tau_i)) + f'_\varepsilon(-\eta + \widetilde{g}(\tau_i))\right\} f_\eta(\eta) d\eta,$$

with $\overline{g}(\tau_i)$, $\widetilde{g}(\tau_i)$ some intermediate values. Since $\sup_\bullet |f'_\varepsilon(\bullet)| < \infty$, we have

$$\begin{aligned}
& |M(\theta,h) - M(\theta,h_o) - \Gamma_2(\theta,h_o)[h-h_o]| \\
\leq \;& \frac{1}{2} \sup_\tau |\Psi(\tau)| \times \{\sup_\tau |h(\tau) - h_o(\tau)|\}^2 \\
\leq \;& \sup_\bullet |f'_\varepsilon(\bullet)| \times \int_0^\infty f_\eta(\eta) d\eta \times \{||h-h_o||_{\mathcal{H}}\}^2,
\end{aligned}$$

hence condition 3(i) of Theorem 2 is satisfied with $c = \sup_\bullet |f'_\varepsilon(\bullet)| > 0$ and $\epsilon = 1$.

Condition 4 of theorem 2 is satisfied since under our conditions, $||\widehat{h} - h_o||_{\mathcal{H}} = O_p(\sqrt{\log n / n\lambda_n} + \lambda_n^2)$. Note also that $\widehat{h} \in \mathcal{H}$ with probability tending to one provided also $n\lambda_n^{1+2\alpha}/\log n \to \infty$.

Condition 5' of Theorem 2 is satisfied by applying Theorem 3. This is because $m(Z,\theta,h) = 1(p_L \leq h(\tau) \leq p_U) - \theta$, and

$$\begin{aligned}
& |m(Z,\theta',h') - m(Z,\theta,h)|^2 \\
\leq \;& 2\,|1(p_L \leq h'(\tau) \leq p_U) - 1(p_L \leq h(\tau) \leq p_U)| + 2\,|\theta' - \theta|^2,
\end{aligned}$$

hence for any small $\delta \in (0,1]$,

$$\begin{aligned}
& E\left[\sup_{(\theta',h'):|\theta'-\theta| \leq \delta, ||h'-h||_{\mathcal{H}} \leq \delta} |m(Z,\theta',h') - m(Z,\theta,h)|^2\right] \\
\leq \;& 2E\left[\sup_{h':||h'-h||_{\mathcal{H}} \leq \delta} |1(p_L \leq h'(\tau) \leq p_U) - 1(p_L \leq h(\tau) \leq p_U)|\right] + 2\delta^2.
\end{aligned}$$

Since for all $h' \in \mathcal{H}$ with $||h'-h||_{\mathcal{H}} \leq \delta \leq 1$, we have conditional on (p_L, τ, p_U),

$$1(p_L \leq h'(\tau) \leq p_U) \leq 1(p_L \leq h(\tau) - \delta) \times 1(h(\tau) + \delta \leq p_U).$$

hence either

$$\begin{aligned}
0 &\leq 1(p_L \leq h'(\tau) \leq p_U) - 1(p_L \leq h(\tau) \leq p_U) \\
&\leq 1(p_L \leq h(\tau) - \delta \leq p_U) - 1(p_L \leq h(\tau) \leq p_U) \leq 1
\end{aligned}$$

or

$$\begin{aligned}
0 &\leq 1(p_L \leq h(\tau) \leq p_U) - 1(p_L \leq h'(\tau) \leq p_U) \\
&\leq 1(p_L \leq h(\tau) \leq p_U) - 1(p_L \leq h(\tau) + \delta \leq p_U) \leq 1
\end{aligned}$$

The term above is either one or zero. $1(p_L \leq h(\tau) \leq p_U) - 1(p_L \leq h(\tau) + \delta \leq p_U) = 1$ when for example

$$h(\tau) + \delta > p_U \text{ and } h(\tau) \leq p_U.$$

The probability of such an event occurring is

$$\begin{aligned}
& \Pr\left[h(\tau) + \delta > p_U \geq h(\tau)\right] \\
= & \Pr\left[h(\tau) + \delta > h_o(\tau) + \varepsilon + \eta_2 \geq h(\tau)\right] \\
= & E\left(\Pr\left[h(\tau) - h_o(\tau) + \delta > \varepsilon + \eta_2 \geq h(\tau) - h_o(\tau) \mid \tau\right]\right) \\
\leq & E\left(\Pr\left[2\delta > \varepsilon + \eta_2 \geq -\delta \mid \tau\right]\right) \\
= & E\left(\int_0^\infty \left[\int_{-\delta-\eta_2}^{2\delta-\eta_2} dF_\varepsilon(\varepsilon)\right] dF_{\eta_2}(\eta_2)\right) \\
\leq & E\left(3\delta \int_0^\infty f_\varepsilon(-\delta - \eta_2) dF_{\eta_2}(\eta_2) + \frac{1}{2}[3\delta]^2 \times \sup_\bullet |f'_\varepsilon(\bullet)|\right) \\
\leq & \; const. \times \delta
\end{aligned}$$

where the last inequality is due to the assumptions that $\sup_\bullet |f'_\varepsilon(\bullet)| < \infty$. Hence

$$\begin{aligned}
& E\left[\sup_{h': ||h'-h||_{\mathcal{H}} \leq \delta} |1(p_L \leq h'(\tau) \leq p_U) - 1(p_L \leq h(\tau) \leq p_U)|\right] \\
\leq & E\left[E\left(\max\left\{\begin{array}{c} |1(p_L \leq h(\tau) + \delta \leq p_U) - 1(p_L \leq h(\tau) \leq p_U)|, \\ |1(p_L \leq h(\tau) - \delta \leq p_U) - 1(p_L \leq h(\tau) \leq p_U)| \end{array}\right\} \mid p_L, \tau, p_U\right)\right] \\
\leq & \; const. \times \delta
\end{aligned}$$

Therefore, condition 2 of Theorem 3 is satisfied with $s = 1/2$, and condition 3 of Theorem 3 is satisfied by Remark 3(ii) and the assumption that $h_o \in \mathcal{H} = C_c^\alpha([0,T])$ with $||\cdot||_{\mathcal{H}} = ||\cdot||_\infty$ and $\alpha > 1$.

Notice that

$$\begin{aligned}
& \psi \cdot \frac{1}{n}\sum_{i=1}^{n}[h(\tau_i) - h_o(\tau_i)] - \Gamma_2(\theta, h_o)[h - h_o] \\
= \; & \psi \cdot \frac{1}{n}\sum_{i=1}^{n}\left\{[h(\tau_i) - h_o(\tau_i)] - E\left([h(\tau_i) - h_o(\tau_i)]\right)\right\} \\
= \; & o_p(n^{-1/2}) \text{ uniformly over } h.
\end{aligned}$$

Then, we have

$$T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\widehat{h}(\tau_i) - h_o(\tau_i)] = \sqrt{n}\int[\widehat{h}(\tau) - h_o(\tau)]f_\tau(\tau)d\tau + o_p(1)$$

by a result of Andrews (1989, Theorem 7) and the uniform rate of convergence of $\widehat{h}$. Furthermore, because the bias term is $O_p(\lambda_n^2\sqrt{n}) = o_p(1)$:

$$T_n = \sqrt{n}\int \frac{1}{n\lambda_n}\sum_{j=1}^{n}K\left(\frac{\tau - \tau_j}{\lambda_n}\right)u_j d\tau + o_p(1) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}u_j + o_p(1)$$

by a change of variables and dominated convergence argument.

We have that condition 6'(ii) of Theorem 2 is satisfied with $\gamma_2(Z_i) = \psi \cdot u_i$. Finally condition 6'(i)(iii) of Theorem 2 is satisfied with

$$\begin{aligned}
& \sqrt{n}\{M_n(\theta_o, h_o) + \frac{1}{n}\sum_{i=1}^{n}\gamma_2(Z_i)\} \\
= \; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{1\left(p_{Li} \le h_o(\tau_i) \le p_{Ui}\right) - \theta_o\} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi \cdot u_i) \\
= \; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{1\left(-\eta_{1i} \le -\varepsilon_i \le \eta_{2i}\right) - \theta_o + \psi \cdot u_i\} \\
= \; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\zeta_i \Longrightarrow \mathcal{N}[0, \text{var}(\zeta_i)],
\end{aligned}$$

where $\zeta_i = 1\left(-\eta_{1i} \le -\varepsilon_i \le \eta_{2i}\right) - \theta_o + \psi \cdot u_i$ is mean zero and has finite variance.

References

Ai, C., and X. Chen (1999): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions". Manuscript, London School of Economics.

Andrews, D.W.K. (1989). Stochastic Equicontinuity for Semiparametric Models. Unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D.W.K., (1994a): "Asymptotics for semiparametric econometric models via stochastic equicontinuity", *Econometrica* 62, 43-72.

Andrews, D.W.K., (1994b): "Empirical process method in econometrics," in *The Handbook of Econometrics*, vol. IV, eds. D.F. McFadden and R.F. Engle III. North Holland.

Andrews, D.W.K., (1997): "A Conditional Kolmogorov Test," Econometrica 65, 1097-1128.

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. Biometrika 68, 326-8.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and J. A. Wellner (1993). *Efficient and adaptive estimation for semiparametric models.* The John Hopkins University Press, Baltimore and London.

Bliss, R. (1997). Testing Term Structure Estimation Methods. *Advances in Futures and Options Research* 9, 197-231.

Blundell, R., M. Browning, and I. Crawford (1997): Non-parametric Engel Curves and Revealed Preference. IFS working paper 97/14.

Breiman, L. (1996). Bagging Predictors. *Machine Learning* 26, 123-140.

Brown, D.J., and M.H. Wegkamp (2000): "Weighted Minimum Mean-Square Distance from Independence Estimation." Unpublished manuscript, available at http://www.stat.yale.edu.

Caner, M., and B.E. Hansen (2001): "Instrumental Variable Estimation of a Threshold Model. Manuscript available at http://www.ssc.wisc.edu/~bhansen/

Chen, X. and X. Shen (1998): "Sieve Extremum Estimates for Weakly Dependent Data," *Econometrica* 66, 289-314.

Gozalo, P., and O. Linton (1996) Conditional independence Restrictions: Testing and Estimation Cowles Foundation Discussion Paper no. 1140.

Granger, C.W.J. (2000): "Thick modeling", UCSD Dept. of Economics Working Paper.

Han, H., and E.T. Tamer (2000): "Inference in Censored Models with Endogenous Regressors". Manuscript, Princeton University.

Honore, B. and E. Kyriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica* 68, 839-874.

Horowitz, J. (1992): "A smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica* 60, 505-531.

Horowitz, J. (1998a): "Bootstrap methods for median regression models," Econometrica 66, 1327-1352.

Horowitz, J. (1998b): "Semiparametric Methods in Econometrics," Springer Verlag: Berlin.

Huber, P.J., (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1, 221-233. University of California, Berkeley.

Kim, J., and D. Pollard (1990): "Cube Root Asymptotics," *The Annals of Statistics* 18, 191-219.

Koenker, R. (1997): "ℓ_1 computation: An interior monologue. in IMS Lecture Notes, Volume 31, pp 15-32.

Koul, H.L. (2001). *Weighted Empirical Processes in Regression and Autoregression Models.* Springer.

Lee, S. (2001). Efficient Semiparametric Estimation of a Partially Linear Quantile Regression Model. Forthcoming in Econometric Theory.

Lewbel, A., and O. Linton (2001). Nonparametric Censored and Truncated Regression. Forthcoming in Econometrica.

Linton, O., Mammen, Nielsen, and Tanggaard (2000): "Estimating the Yield Curve by Kernel Smoothing Methods" *Journal of Econometrics* 105/1 185-223.

Manski, C.F. (1975): "The Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics* 3, 205-228.

Manski, C.F., (1983): "Closest Empirical Distribution Estimation," *Econometrica* 51, 305-319.

Manski, C.F. (1994): "Analog Estimation of Econometric Models," in *The Handbook of Econometrics*, vol. IV, eds. D.F. McFadden and R.F. Engle III. North Holland.

Manski, C.F., and E. Tamer (2000): "Inference on Regressions with Interval Data on a Regressor or Outcome," Forthcoming in *Econometrica*.

Matzkin, R.L. (1994): "Restrictions of Economic Theory in Nonparametric Methods," in *The Handbook of Econometrics*, vol. IV, eds. D.F. McFadden and R.F. Engle III. North Holland.

Newey, W.K., (1994): "The asymptotic variance of semiparametric estimators," *Econometrica* 62, 1349-1382.

Newey, W.K., and D. F. McFadden (1994): "Large sample estimation and hypothesis testing," in *The Handbook of Econometrics*, vol. IV, eds. D.F. McFadden and R.F. Engle III. North Holland.

Pakes, A., and S. Olley (1995): "A limit theorem for a smooth class of semiparametric estimators", *Journal of Econometrics* 65, 295-332.

Pakes, A., and D. Pollard (1989): "Simulation and the asymptotics of optimization estimators," *Econometrica* 57, 1027-1057.

Pollard, D. (1993): "The asymptotics of a binary choice model," Unpublished manuscript, available at http://www.stat.yale.edu.

Pötscher, B.M. and I. R. Prucha (1991a): "Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part i: Consistency and Approximation Concepts," Econometric Reiew 10, 125-216.

Pötscher, B.M. and I. R. Prucha (1991b): "Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part ii: Asymptotic Normality," Econometric Reiew 10, 253-325.

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

Powell, J.L. (1994): "Estimation in semiparametric models." in The *Handbook of Econometrics*, vol. IV. eds R.F. Engle III and D.F. McFadden. North Holland.

Sakov, A., and P.J. Bickel (1998). An Edgeworth expansion for the m out of n bootstrapped median.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer-Verlag, New York.

Van Keilegom, I., and M. G. Akritas (1999): "Transfer of tail information in censored regression models", *Annals of Statistics* 27, 1745-1784.