A Smoothed Least Squares Estimator For Threshold Regression Models*

M. Seo† and O. Linton‡

London School of Economics and Political Science

October 5, 2006

Abstract

We propose a smoothed least squares estimator of the parameters of a threshold regression model. Our model generalizes that considered in Hansen (2000) to allow the thresholding to depend on a linear index of observed regressors, thus allowing discrete variables to enter. We also do not assume that the threshold effect is vanishingly small. Our estimator is shown to be consistent and asymptotically normal thus facilitating standard inference techniques based on estimated standard errors or standard bootstrap for the slope and threshold parameters. In the working paper version of this paper, Seo and Linton (2006), we compared the finite sample performance of our confidence intervals with those of Hansen (2000) and show that our methods outperformed his for large values of the threshold.

Some key words: Index model; Sample Splitting; Segmented Regression; Smoothing; Threshold Estimation.

JEL Classification Number: C12, C13, C14.

---

*We would like to thank the co-editor and associate-editor, two referees for helpful comments. We are very grateful to Søren Johansen for pointing out a mistake in an earlier draft.

†Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: `m.seo@lse.ac.uk`. This research was supported through a grant from the Economic and Social Science Research Council.

‡Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: `o.linton@lse.ac.uk`. This research was supported through a grant from the Economic and Social Science Research Council.

1 Introduction

The threshold model (often called sample splitting or segmented regression) has wide application in economics. Hansen (2000) brought many of those applications to the attention of econometricians. The literature divides according to autoregression and regression, according to smooth, continuous, or discontinuous threshold, and according to nonparametric or parametric functional form. It is different from the regime-switching literature, see e.g., Kim and Nelson (1999) for a review, in that the switching variable is observable. The smooth transition autoregressive models have been widely used in macro and financial applications, see the recent review paper of van Dijk, Teräsvirta, and Franses (2000). The discontinuous threshold effect has found applications in macro and in cross-sectional growth regressions, see Hansen (2000) for discussion. There is also a nonparametric literature in applied economics associated with the concept of 'regression discontinuity design', see for example Hahn, Todd and van der Klauw (2001). In fact, a whole methodology has been built around this, and there are many applications. In that case the threshold point is usually assumed known. The paper of Delgado and Hidalgo (2000) work with the more general case of multiple unknown threshold points in a nonparametric regression and obtain a full set of results for estimation and inference.

This paper is about the parametric threshold regression model. Unfortunately, this model does not have a satisfactory basis for inference even in the case of least-squares estimation. It has been established that the threshold parameter estimate converges faster than the slope parameter estimates and that its asymptotic distribution is not Normal. On the other hand, the slope parameter estimates converges to a Normal distribution independently of the threshold parameter estimate. In the context of threshold autoregression, Chan (1993) establishes that the threshold parameter estimate converges to a functional of a compound Poisson process; the distribution is too complicated to be used in practice due to the dependence on the marginal distribution of the covariates. Hansen (2000) develops an asymptotic distribution for the threshold parameter estimate based on the diminishing threshold effect assumption, in which the threshold model becomes the linear model asymptotically. The limiting distribution is symmetric about zero and has moderate tails but is unbounded at zero. Although the distribution is readily available through a simulation, the validity of the asymptotic distribution may be limited to the "small effect" case, as he calls it. It should be noted, however, that it provides a conservative confidence interval for the threshold estimate for the case where the threshold effect is held fixed, under the auxiliary assumption of the normality of and the independence of the error from the regressors. These, however, are strong assumptions.

Recently, Gonzalo and Wolf (2005) have proposed using subsampling to conduct inference in threshold autoregressive models. They consider the set-up of Tong (1990) and Chan (1993) but also allow for the continuous threshold case of Chan and Tsay (1998). They allow for regime specific heteroskedasticity as in Chan (1993) (this was excluded in Hansen (2000)) but otherwise the innovation process is i.i.d. They establish consistency of tests about and confidence intervals for the threshold parameters based on the least squares estimator under constant threshold assumption.

We consider a threshold model that is more general than the one in Hansen (2000), which permits only a pre-assigned continuous variable. In contrast, we allow the threshold variable to be a linear combination of the regressors and/or other variables, validating the use of discontinuous variables for sample splitting in addition to continuous variables. This may be of interest because it allows different threshold values for subsamples divided by a discrete variable like gender. Furthermore, we can make decisions about the inclusion of a (some) variable(s) based on a test such as the t- or Wald test.[1]

This paper proposes the least squares estimation of the threshold model after smoothing the objective function in the spirit of the smoothed maximum score estimator of Horowitz (1992). It is based on the replacement of the indicator function in the objective function with an integrated kernel. While the maximum score estimator by Manski (1975) is asymptotically distributed as the random variable that maximizes a certain Gaussian process, the smoothed maximum score estimator exhibits asymptotic normality. The smoothing also brings about a change in the convergence rate. Under smoothness conditions the smoothed maximum score estimator converges faster than the maximum score estimator.

We develop an asymptotic theory for the smoothed least-squares estimation of the threshold model in the regression context. Unlike the previous literature, the threshold estimate is distributed as asymptotically normal. Its convergence rate to ensure the normality is slower than that obtained in Chan (1993) and depends on the choice of bandwidth. Unlike in the maximum score case, smoothing reduces the rate of convergence. It is worth noting that Hansen (2000) also attains a manageable distribution at the expense of the convergence rate. The slope estimates are square root n consistent and asymptotically normally distributed, and independent of the threshold estimate. Our development allows for time series data, a special case being the threshold autoregression of Tong (1983,1990). For inference we provide two new results. First, the consistency of the HAC estimation in Andrews (1991) is extended to allow for the discontinuity in the threshold model. Second, we establish the consistency of the standard bootstrap resampling algorithm in the i.i.d. case.

Our set-up is more general than Gonzalo and Wolf (2005) in that we allow both regime specific heteroskedasticity and covariate dependent heteroskedasticity as would be common in cross-sectional regression applications. Also, our method has the usual advantage over subsampling that if we work with pivotal test statistics we can expect to obtain asymptotic refinements, Horowitz (1998, 2001, 2002).

[1] But we should include at least one continuous variable and the coefficient is normalized to one.

We also investigate two slightly different implementations of the 'smoothing over' approach. Although the two different methods result in the same asymptotics for the slope estimates, the limiting distribution of the threshold estimates are different, and not in general rankable.

We provide some simulation evidence on the rate of convergence and the finite sample distribution of our procedures. Confidence intervals based on our procedure perform better than those of Hansen (2000) in his design in the larger threshold case.

The paper is organized as follows. Section 2 introduces the smoothed LS estimators and their consistency and asymptotic Normality is established in Section 3. Section 4 provides methods to construct the asymptotic and bootstrap confidence intervals. Section 5 discusses the continuous case. Numerical results are presented in Section 6. Section 7 concludes. The proofs of theorems are collected in an Appendix.

The following notations are used. The integral $\int$ is taken over $(-\infty, \infty)$ unless specified otherwise. Let $||g||_2^2 = \int g(s)^2 ds$ for any function g. For any matrix A, let $||A|| = \mathrm{tr}(A^\top A)^{1/2}$. All the convergences are as $n \to \infty$.

2 The Smoothed LS estimator

2.1 The Model

Write the model

$$y_t = x_t^\top \beta + \delta^\top \widetilde{x}_t 1\left\{q_t^\top \boldsymbol{\psi} > 0\right\} + \varepsilon_t, \tag{1}$$

where x_t, $\widetilde{x}_t$, and q_t may have common variables. A leading case is where $\widetilde{x}_t = x_t$ but $\widetilde{x}_t$ can also be a strict proper subset of x_t. Let q_{1t} be the first element of q_t, and q_{2t} the other elements of q_t. Let X_t whose first element is q_{1t} denote all the regressors. Furthermore, assume the first element of q_{2t} is the constant 1. Similarly, X_{1t} denotes q_{1t} and X_{2t} the other elements in X_t. The first element of $\boldsymbol{\psi}$ is normalized to 1, and the others are denoted as ψ, so that $q_t^\top \boldsymbol{\psi} = q_{1t} + q_{2t}^\top \psi$. This model includes many considered in the literature as special cases, for example, the threshold autoregression of Tong (1983) as used by Potter (1995). Hansen (2000) considered the special case where q_{2t} is only a constant. It may be the case in practice where only a few variables are employed to construct the threshold index.

2.2 Estimators

The least squares (LS) estimator minimizes the objective function

$$
\begin{aligned}
S_n^*(\theta) &= \frac{1}{n}\sum_{t=1}^{n}\left(y_t - x_t^\top\beta - \tilde{x}_t^\top\delta 1\left\{q_{1t}+q_{2t}^\top\psi>0\right\}\right)^2 \qquad (2)\\
&= \frac{1}{n}\sum_{t=1}^{n}\left(y_t - x_t^\top\beta\right)^2 + \frac{1}{n}\sum_{t=1}^{n}\left\{\left(\tilde{x}_t^\top\delta\right)^2 - 2\tilde{x}_t^\top\delta\left(y_t - x_t^\top\beta\right)\right\}1_t(\psi),
\end{aligned}
$$

where $\theta=\left(\beta^\top,\delta^\top,\psi^\top\right)^\top\in\Theta\subset\mathbb{R}^k$ and $1_t(\psi)=1\left\{q_{1t}+q_{2t}^\top\psi>0\right\}$. We assume that the parameter space $\Theta=\Theta_\beta\times\Theta_\delta\times\Theta_\psi$ is compact and that the true parameter $\theta_0=\left(\beta_0^\top,\delta_0^\top,\psi_0^\top\right)^\top$ is an interior point of Θ. The solution is obtained by profiled least squares, see Hansen (2000). Let θ_n^* denote the least squares estimator.

Define a bounded function $\mathcal{K}(\cdot)$ satisfying that

$$\lim_{s\to-\infty}\mathcal{K}(s)=0,\ \lim_{s\to+\infty}\mathcal{K}(s)=1.$$

It is worth noting that this function is analogous to a cumulative distribution function rather than a density function. Then, define a smoothed objective function

$$S_n(\theta;\sigma_n)=\frac{1}{n}\sum_{t=1}^{n}\left(y_t-x_t^\top\beta\right)^2+\frac{1}{n}\sum_{t=1}^{n}\left\{\left(\tilde{x}_t^\top\delta\right)^2-2\tilde{x}_t^\top\delta\left(y_t-x_t^\top\beta\right)\right\}\mathcal{K}\left(\frac{q_{1t}+q_{2t}^\top\psi}{\sigma_n}\right), \qquad (3)$$

and a smoothed least squares (SLS) estimator

$$\theta_n=\left(\beta_n^\top,\delta_n^\top,\psi_n^\top\right)^\top=\arg\min_{\theta\in\Theta}S_n(\theta;\sigma_n). \qquad (4)$$

To distinguish the slope parameters, let $\theta^s=(\beta^\top,\delta^\top)^\top$ and $\theta_0^s=(\beta_0^\top,\delta_0^\top)^\top$. In practice, one solves the optimization problem by computing $\beta_n(\psi),\delta_n(\psi)$ by an explicit least squares formula for a given ψ, this is

$$\begin{bmatrix}\beta_n(\psi)\\ \delta_n(\psi)\end{bmatrix}=\begin{bmatrix}\sum_{t=1}^n x_tx_t^\top & \sum_{t=1}^n x_t\tilde{x}_t^\top\mathcal{K}_t(\psi)\\ \sum_{t=1}^n \tilde{x}_tx_t^\top\mathcal{K}_t(\psi) & \sum_{t=1}^n \tilde{x}_t\tilde{x}_t^\top\mathcal{K}_t(\psi)\end{bmatrix}^{-1}\begin{bmatrix}\sum_{t=1}^n x_ty_t\\ \sum_{t=1}^n \tilde{x}_ty_t\mathcal{K}_t(\psi)\end{bmatrix},$$

where $\mathcal{K}_t(\psi)=\mathcal{K}\left((q_{1t}+q_{2t}^\top\psi)/\sigma_n\right)$, and then optimizing the profiled criterion over ψ. Practical difficulty arises only in the case of large dimensional ψ.

There is an alternative approach, which is based on just replacing $1_t(\psi)$ in (2) by $\mathcal{K}_t(\psi)$, thus instead of (3) one has

$$S_n^+(\theta;\sigma_n)=\frac{1}{n}\sum_{t=1}^{n}\left(y_t-x_t^\top\beta-\tilde{x}_t^\top\delta\mathcal{K}\left(\frac{q_{1t}+q_{2t}^\top\psi}{\sigma_n}\right)\right)^2 \qquad (5)$$

and another smoothed least squares (SLS) estimator

$$\theta_n^+ = \left(\beta_n^{+\top}, \delta_n^{+\top}, \psi_n^{+\top}\right)^\top = \arg\min_{\theta\in\Theta} S_n^+(\theta;\sigma_n).$$

As before this optimization is done in two stages with the profiled least squares estimators

$$\begin{bmatrix} \beta_n^+(\psi) \\ \delta_n^+(\psi) \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^n x_t x_t^\top & \sum_{t=1}^n x_t \tilde{x}_t^\top \mathcal{K}_t(\psi) \\ \sum_{t=1}^n \tilde{x}_t x_t^\top \mathcal{K}_t(\psi) & \sum_{t=1}^n \tilde{x}_t \tilde{x}_t^\top \mathcal{K}_t^2(\psi) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^n x_t y_t \\ \sum_{t=1}^n \tilde{x}_t y_t \mathcal{K}_t(\psi) \end{bmatrix},$$

which are then plugged back into (5) for optimization over ψ. Note that although $1_t^2(\psi) = 1_t(\psi)$, $\mathcal{K}_t^2(\psi) \neq \mathcal{K}_t(\psi)$ and the estimators defined by (3) and (5) are different. In the case of the slope coefficients this difference vanishes asymptotically, but in the case of the threshold parameters it does not. In the exposition we concentrate mainly on the estimator θ_n, although similar comments apply to θ_n^+.

3 Asymptotic Properties

3.1 Consistency

We assume the following conditions to show the consistency of the SLS estimators.

Assumption 1 *(a) $\{X_t, \varepsilon_t\}$ is a sequence of strictly stationary strong mixing random variables with mixing numbers α_m, $m = 1, 2, \ldots$, that satisfy $\alpha_m = o\left(m^{-\alpha_0/(\alpha_0-1)}\right)$ as $m \to \infty$ for some $\alpha_0 \geq 1$.*
(b) For some $\xi > \alpha_0$, $0 < E\left\|X_t X_t^\top\right\|^\xi < \infty$, and $E\left\|X_t\varepsilon_t\right\|^\xi < \infty$. Furthermore, for $\mathbf{x}_t = \left(x_t^\top, \tilde{x}_t^\top\right)^\top$, $E\left[\mathbf{x}_t\mathbf{x}_t^\top|q_t\right] > 0$ a.s.
(c) $E(\varepsilon_t|X_t) = 0$ a.s. and $\delta_0 \neq 0$.
(d) For almost every X_{2t}, the probability distribution of X_{1t} (q_{1t}) conditional on X_{2t} has everywhere positive density with respect to Lebesgue measure.

Condition (a) corresponds to Assumption B1 of Andrews (1987). Given a compact parameter space, the generic uniform law of large numbers of Andrews (1987) is applied for the development of the consistency proof, supported by the strong law of large numbers of de Jong (1995, Theorem 4). Condition (b) is common in threshold models for identification. For example, Chan (1993) assumes that the error in a threshold autoregression has a Lebesgue density that is positive everywhere, which implies that $E\left[\mathbf{x}_t\mathbf{x}_t^\top|q_t\right] > 0$ *a.s.* as well as condition (d). If $\delta_0 = 0$, then the threshold parameter ψ is not identified.

The following theorem establishes the strong consistency of the SLS estimators.

Theorem 1 *Let Assumption 1 hold. Then, $\theta_n \to \theta_0$ and $\theta_n^+ \to \theta_0$ almost surely.*

3.2 Asymptotic Normality

The asymptotic distribution is developed based on the standard Taylor series expansion. Suppose that $S_n(\theta;\sigma_n)$ is twice differentiable with respect to θ, and define

$$\begin{aligned} T_n(\theta;\sigma_n) &= \partial S_n(\theta;\sigma_n)/\partial\theta \\ Q_n(\theta;\sigma_n) &= \partial^2 S_n(\theta;\sigma_n)/\partial\theta\partial\theta^\top. \end{aligned}$$

The superscript s and ψ on T_n and Q_n, when applied, indicate the obvious partitions of T_n and Q_n according to the slope parameter θ^s and the threshold parameter ψ.

We make a reparameterization to express the limiting distributions conveniently. Let $z_t = q_{1t} + q_{2t}^\top\psi_0$. This involves decomposing $\widetilde{x}_t$ into the part measurable with respect to z_t and the part that is not so. There is a one-to-one relation between $\left(z_t, X_{2t}^\top\right)^\top$ and X_t for any ψ_0. Let $\mathcal{T}$ be the mapping such that $\left(z_t, X_{2t}^\top\right)^\top = \mathcal{T}X_t$ and let $\mathcal{S}$ be the selection matrix such that $\tilde{x}_t = \mathcal{S}X_t$. Let $\dot{\delta} = \mathcal{T}^{-1^\top}\mathcal{S}^\top\delta$ so that

$$\tilde{x}_t^\top\delta = \left(z_t, X_{2t}^\top\right)\dot{\delta} = z_t\dot{\delta}_1 + X_{2t}^\top\dot{\delta}_2. \tag{6}$$

For example, if $x_t = \tilde{x}_t = q_t$, whose dimension is k, then $\mathcal{S} = I_k$,

$$\mathcal{T} = \begin{pmatrix} 1 & \psi_0^\top \\ 0 & I_{k-1} \end{pmatrix} \text{ and } \dot{\delta} = \begin{pmatrix} \delta_1 \\ -\delta_1\psi_0 + \delta_2 \end{pmatrix}.$$

We then have $\delta^\top\widetilde{x}_t 1\{z_t > 0\} = (z_t\dot{\delta}_1 + X_{2t}^\top\dot{\delta}_2)1(z_t > 0) = z_t\dot{\delta}_1 1(z_t > 0) + X_{2t}^\top\dot{\delta}_2 1(z_t > 0)$, where the first term on the right hand side is continuous in z_t at $z_t = 0$ with probability one, while the second term is not unless $\dot{\delta}_2 = 0$.

By Assumption 1, the distribution of z_t conditional on X_{2t} has everywhere positive density with respect to Lebesgue measure for almost every X_{2t}. Let $f(\cdot|X_2)$ denote this density given $X_{2t} = X_2$ and $f(\cdot)$ the density of z_t. For each positive integer i, define

$$f^{(i)}(z|X_2) = \partial^i f(z|X_2)/\partial z^i$$

whenever the derivative exists, and let $\tilde{\mathcal{K}}(s) = (1\{s > 0\} - \mathcal{K}(s))\mathcal{K}'(s)$.

Define

$$\overline{E}\left(\varepsilon_t^2|X_{2t}\right) = \left(\lim_{z\to 0^+} E\left(\varepsilon_t^2|z_t = z, X_{2t}\right) + \lim_{z\to 0^-} E\left(\varepsilon_t^2|z_t = z, X_{2t}\right)\right)/2, \tag{7}$$

and

$$\begin{aligned}
V^s &= \begin{pmatrix} \sum_{s=-\infty}^{\infty} E x_1 x_s^\top \varepsilon_1 \varepsilon_s & \sum_{s=-\infty}^{\infty} E x_1 \tilde{x}_s^\top \varepsilon_1 \varepsilon_s 1\{z_s > 0\} \\ \sum_{s=-\infty}^{\infty} E \tilde{x}_1 x_s^\top \varepsilon_1 \varepsilon_s 1\{z_1 > 0\} & \sum_{s=-\infty}^{\infty} E \tilde{x}_1 \tilde{x}_s^\top \varepsilon_1 \varepsilon_s 1\{z_1 > 0, z_s > 0\} \end{pmatrix} \\
V^\psi &= \|\mathcal{K}'\|_2^2 \, E\left[\left((X_{2t}^\top \dot{\delta}_2)^4/4 + (X_{2t}^\top \dot{\delta}_2)^2 \overline{E}\left(\varepsilon_t^2 | X_{2t}\right)\right) q_{2t} q_{2t}^\top | z_t = 0\right] f(0) \\
V^{\psi+} &= E\left[\left(\left\|\tilde{\mathcal{K}}\right\|_2^2 (X_{2t}^\top \dot{\delta}_2)^4 + \|\mathcal{K}'\|_2^2 (X_{2t}^\top \dot{\delta}_2)^2 \overline{E}\left(\varepsilon_t^2 | X_{2t}\right)\right) q_{2t} q_{2t}^\top | z_t = 0\right] f(0) \\
Q^s &= \begin{pmatrix} E x_t x_t^\top & E x_t \tilde{x}_t^\top 1\{z_t > 0\} \\ E \tilde{x}_t x_t^\top 1\{z_t > 0\} & E \tilde{x}_t \tilde{x}_t^\top 1\{z_t > 0\} \end{pmatrix} \\
Q^\psi &= \mathcal{K}'(0) \, E\left[(X_{2t}^\top \dot{\delta}_2)^2 q_{2t} q_{2t}^\top | z_t = 0\right] f(0) .
\end{aligned}$$

If we impose a stronger assumption that $\{\varepsilon_t\}$ is a martingale difference sequence, then all the autocovariances drop out of V^s. By contrast, V^ψ and $V^{\psi+}$ do not involve the long-run variance as is the case in the dynamic binary choice model of de Jong and Woutersen (2004) and in the threshold LAD model of Caner (2002) .

The assumptions we need are collected in the following.

Assumption 2 (a) *For all vectors* ζ *such that* $|\zeta| = 1$ *and* $r > 4$, $E|X_{2t}^\top \dot{\delta}_2 \varepsilon_t q_{2t}^\top \zeta|^r < \infty$ *and* $E|(X_{2t}^\top \dot{\delta}_2)^2 q_{2t}^\top \zeta|^r < \infty$; (b) $\{X_t, \varepsilon_t\}$ *is a sequence of strictly stationary strong mixing random variables with mixing numbers* α_m, $m = 1, 2, \ldots$, *that satisfy* $\alpha_m \leq C m^{-(2r-2)/(r-2)-\eta}$ *for positive* C *and* η, *as* $m \to \infty$; (c) *For an integer* h *that will be specified later and each integer* i *such that* $1 \leq i \leq h$, *all* z *in a neighborhood of 0, almost every* X_2, *and some* $M < \infty$, $f^{(i)}(z|X_2)$ *exists and is a continuous function of* z *satisfying* $\left|f^{(i)}(z|X_2)\right| < M$. *In addition,* $f(z|X_2) < M$ *for all* z *and almost every* X_2. *Furthermore,* $E\left(\varepsilon_t^4 | X_t\right) < M$ *for almost every* X_t; (d) *The conditional joint density* $f(z_t, z_{t-m} | X_{2t}, X_{2t-m}) < M$, *for all* (z_t, z_{t-m}) *and almost all* (X_{2t}, X_{2t-m}); (e) *The matrices* V^s, V^ψ, Q^s, *and* Q^ψ *are finite and positive definite.*

In Hansen's model where $z_t = q_{1t} + \psi$, V^ψ and Q^ψ are defined without $q_{2t} q_{2t}^\top$. The moment conditions are to ensure the consistency of the variance covariance matrix estimators that are introduced later.

The mixing condition (b) is more general than $\rho-$ mixing in Hansen (2000), which includes many nonlinear time series such as TAR processes as discussed there. The conditions (c) - (e) are common in the smoothed estimation as in Horowitz (1992), only (d) being an analogue of an i.i.d. sample to a dependent sample. The smoothness condition here is stronger than that of Chan (1993) since the boundedness of the first derivative of the density implies the uniform continuity. While (e) is standard, the positivity of Q^ψ excludes a continuous threshold model, that is, $\dot{\delta}_2 \neq 0$, so does

Assumption 1.7 of Hansen (2000). The finiteness of V^s can be implied by the α-mixing condition with a slightly stronger assumption on the mixing coefficient α_m plus a moment condition. See Andrews (1991, Lemma 1).

Unlike Hansen (2000), we do not impose the continuity of $E\left(\varepsilon_t^2|z_t\right)$ at $z_t = 0$, thus allowing for a regime specific heteroskedasticity. This type of heteroskedasticity is quite plausible in applications and we would certainly want to allow for it. In such a case, one may want to employ a weighted least squares although this requires further estimation. It is expected that the asymptotics in Hansen (2000) can be modified to allow such discontinuity, but then the studentizing of the threshold estimate seems to become more cumbersome.[2]

We make the following assumptions regarding the smoothing function $\mathcal{K}$ and the bandwidth parameter σ_n.

Assumption 3 *(a) $\mathcal{K}$ is twice differentiable everywhere, $\mathcal{K}'$ is symmetric around zero, $|\mathcal{K}'(\cdot)|$ and $|\mathcal{K}''(\cdot)|$ are uniformly bounded, and: $\int |\mathcal{K}'(v)|^4 dv < \infty$, $\int |\mathcal{K}''(v)|^2 dv < \infty$, $\int |v^2\mathcal{K}''(v)\, dv| < \infty$. (b) For an integer h that will be given later and each integer i $(1 \le i \le h)$, $\int |v^i \mathcal{K}'(v)\, dv| < \infty$, and*

$$\int s^{i-1}\mathrm{sgn}(s)\,\mathcal{K}'(s)\,ds = 0, \text{ and} \int s^h\mathrm{sgn}(s)\,\mathcal{K}'(s)\,ds \neq 0,$$

and $\mathcal{K}(x) - \mathcal{K}(0) \gtreqless 0$ if $x \gtreqless 0$.
(c) For each integer $i\,(0 \le i \le h)$, and $\eta > 0$, and any sequence $\{\sigma_n\}$ converging to 0,

$$\lim_{n\to\infty} \sigma_n^{i-h} \int_{|\sigma_n s|>\eta} \left|s^i\mathcal{K}'(s)\right| ds = 0, \text{ and } \lim_{n\to\infty} \sigma_n^{-1} \int_{|\sigma_n s|>\eta} |\mathcal{K}''(s)|\, ds = 0.$$

(d) For some $\mu \in (0,1]$, a positive constant C, and all $x, y \in \mathbb{R}$,

$$|\mathcal{K}''(x) - \mathcal{K}''(y)| \le C\,|x-y|^{\mu}.$$

(e) For some sequence $m_n \ge 1$ and $\varepsilon > 0$, $n\sigma_n^3 \to 0$ and

$$\begin{aligned} \log(nm_n)\left(n^{1-6/r}\sigma_n^2 m_n^{-2}\right)^{-1} &\to 0 \\ \sigma_n^{-9k/2-3} n^{(9k/2+2)/r+\varepsilon}\alpha_{m_n} &\to 0. \end{aligned}$$

These conditions are similar to those in Horowitz (1992). Condition (b) is an analogous condition to that defining the so-called h^{th} order kernel and requires a kernel $\mathcal{K}'$ that permits negative values

[2] The limit distribution in Theorem 1 of Hansen (2000) is expected to change to

$$\arg\max_r \left\{\omega_1\left(-|r|/2 + W(r)\right)1\{r>0\} + \omega_2\left(-|r|/2+W(r)\right)1\{r<0\}\right\},$$

where ω_1 and ω_2 are the right and left limit in (7). Thus, ω_1 and ω_2 does not average out as it does in our case.

if $h \geq 2$. A kernel that satisfies these conditions for $h = 2$ is $\mathcal{K}(x) = \Phi(x) + x\phi(x)$, where Φ and ϕ are the distribution function and the density of the standard normal, respectively. For this kernel $\mathcal{K}'(0) = \sqrt{2/\pi} \approx 0.798$, $||\tilde{\mathcal{K}}||_2^2 \approx 0.064$, and $||\mathcal{K}'||_2^2 \approx 0.776$.

Condition (e) serves to determine the rate for σ_n. When the data are i.i.d. and the regressors possess a moment generating function, the conditions can be weakened to

$$\frac{\log(n)}{n\sigma_n^2} \to 0, \tag{8}$$

since $\alpha_{m_n} = 0$ and we can set $m_n = 1$ in this case. Contrary to the smoothed maximum score estimation, we choose the bandwidth that converges to zero as fast as permissible.

Although condition (e) in Assumption 3 provides permissible rates for the bandwidth selection, it may not be sharp. In fact, Delgado and Hidalgo (2000) study the nonparametric estimation of the locations and sizes of the discontinuities in conditional expectation. They establish asymptotic normality at rate $\sqrt{n\sigma_n^{p-1}}$, where p is the number of covariates in the nonparametric regression, under the restrictions that $n\sigma_n^{p+1} \to \infty$ and $\limsup_n n\sigma_n^{p+5} < \infty$. If one could take $p = 0$ (one cannot in their theory), which would correspond to parametric regression, in their results, this would suggest asymptotic normality holds at rates arbitrarily close to n.

Theorem 2 *Let Assumptions 1 - 3 hold for $h \geq 2$. Then*

$$\begin{aligned}
\sqrt{n}\left(\theta_n^s - \theta_0^s\right) &\implies N\left(0, Q^{s^{-1}} V^s Q^{s^{-1}}\right), \\
\sqrt{n\sigma_n^{-1}}\left(\psi_n - \psi_0\right) &\implies N\left(0, Q^{\psi^{-1}} V^\psi Q^{\psi^{-1}}\right),
\end{aligned}$$

and they are asymptotically independent.

Similarly, we have

Corollary 3 *Let Assumptions 1 - 3 hold for $h \geq 1$. Then*

$$\begin{aligned}
\sqrt{n}\left(\theta_n^{s+} - \theta_0^s\right) &\implies N\left(0, Q^{s^{-1}} V^s Q^{s^{-1}}\right), \\
\sqrt{n\sigma_n^{-1}}\left(\psi_n^+ - \psi_0\right) &\implies N\left(0, Q^{\psi^{-1}} V^{\psi+} Q^{\psi^{-1}}\right),
\end{aligned}$$

and they are asymptotically independent.

Remarks.

1. The convergence rate of ψ_n is $\sqrt{n\sigma_n^{-1}}$, which means that faster convergence of σ_n to zero accelerates the convergence of ψ_n. This is in contrast to the smoothed maximum score estimator for which the faster convergence of the bandwidth reduces the convergence rate of the estimator.

In the i.i.d. case, the bandwidth $\sigma_n = \log n/\sqrt{n}$ satisfies the condition (8) and $n\sigma_n^3 \to 0$. In this case we obtain that ψ_n is (apart from a logarithmic factor) $n^{-3/4}$ consistent. However, the bandwidth restrictions are sufficient and not necessary and it is quite plausible that one obtains $\sqrt{n\sigma_n^{-1}}$ convergence but perhaps not asymptotic normality for smaller bandwidths.

2. As in the least squares estimation of the threshold model, the slope estimate θ_n^s is not affected asymptotically by the estimation of the threshold parameter ψ in either case.

3. The assumption that $n\sigma_n^3 \to 0$ is imposed to ensure the asymptotic independence of ψ_n from θ_n^s. With a bandwidth that converges slower, we may obtain the covariances between these estimators, which may prove beneficial for finite sample inference on the slope parameters. It is also likely, however, that it may introduce an asymptotic bias for ψ_n as is the case for the smoothed maximum score estimator. The convergence rate of θ_n^s is not affected by this change in the rate of convergence of the bandwidth.

4. Our conditions are stronger than those of Hansen (2000) and Chan (1993) with regard to smoothness. Specifically, they do not require the distribution of $z_t|X_{2t}$ to be smooth. When the smoothness conditions do not hold, our estimator converges at a slower rate due to the presence of a bias term of large order. This is as found in Pollard (1993) regarding the smoothed maximum score estimator of Horowitz (1992).

5. Although we do not explicitly treat it, the small threshold case of Hansen (2000) can be analyzed within the same framework. Specifically, when δ_2 is replaced by $\delta_2/n^{\pi} \to 0$ one still obtains asymptotic normality, provided $\delta_2 > 0$ and π is not too large, but at a slower rate of convergence reflecting the presence of $n^{-\pi}$ in the score and Hessian functions. Notice that the asymptotic variance of the score function (of ψ_n) is somewhat simpler in this case because the term $E[(X_{2t}^{\top}\dot{\delta}_2)^4 q_{2t}q_{2t}^{\top}|z_t = 0]$ is of smaller order relative to $E[4(X_{2t}^{\top}\dot{\delta}_2)^2 \overline{E}\,(\varepsilon_t^2|X_{2t})\, q_{2t}q_{2t}^{\top}|z_t = 0]$ and that V^{ψ} becomes $V^{\psi+}$. Compare with Hansen (2000).

6. If q_{2t} consists of the constant only, then ψ_n is the threshold estimate in the usual sense. If a dummy such as gender or region is included in addition to the constant, then the coefficient estimate for the dummy means the difference in the threshold values between two subsamples. Therefore, the t -test on the coefficient examines whether the threshold points are the same across two subsamples or not.

7. The case where the thresholding variable is time can also be handled in this framework. The results obtained above apply to the estimation of the break fraction, say $\kappa \in (0, 1)$, with some modifications. The terms constituting the asymptotic variances are defined with $f(0) = 1$, $q_{2t} = 1$, and the conditional expectations replaced with the unconditional ones.

8. The asymptotic distributions of ψ_n and ψ_n^+ do not depend on the error autocorrelation function, whereas the asymptotic distributions of the slope parameter estimates do.

9. The two estimators ψ_n and ψ_n^+ have different asymptotic variances. The difference when $\mathcal{K}(x) = \Phi(x) + x\phi(x)$ is, for $C > 0$, $C(4||\tilde{\mathcal{K}}||_2^2 - ||\mathcal{K}'||_2^2) = C(0.256 - 0.776) < 0$. Therefore, the estimator ψ_n^+ has the smaller asymptotic variance. Also, in the case where $|\mathcal{K}(s)| < 1/2$ for $s \in (-\infty, 0)$, then $4||\tilde{\mathcal{K}}||_2^2 - \|\mathcal{K}'\|_2^2 < 0$ as $(1\{s > 0\} - \mathcal{K}(s))^2 < 1/4$ for all $s \neq 0$ and $\mathcal{K}(0) = 1/2$. Therefore, ψ_n^+ dominates ψ_n in this case also.

4 Inference Methods

The construction of an asymptotic confidence set is straightforward by inverting the t or Wald statistic given the asymptotic normality. Ways to estimate the asymptotic variances are described below. We also discuss the likelihood ratio statistics. We also discuss bootstrap confidence intervals.

4.1 Asymptotic Variance Estimation, t and Wald Statistics

We now discuss various estimators of the asymptotic variance of our estimators. As usual there are many alternative estimators of the asymptotic variance depending on which information is imposed. In the simulation experiments below we investigate some of the proposals made here.

Let

$$e_t(\theta) = y_t - x_t^\top \beta - \tilde{x}_t^\top \delta \mathcal{K}\left(\frac{q_{1t} + q_{2t}^\top \psi}{\sigma_n}\right), \tag{9}$$

and $e_t^+ = e_t\left(\theta_n^+\right)$. Also let

$$\begin{aligned} \tau_{n,t}^{\psi}(\theta) &= \left\{\left(\tilde{x}_t^\top \delta\right)^2 - 2\tilde{x}_t^\top \delta\left(y_t - x_t^\top \beta\right)\right\} \frac{q_{2t}}{\sigma_n} \mathcal{K}'\left(\frac{q_{1t} + q_{2t}^\top \psi}{\sigma_n}\right) \\ \tau_{n,t}^{\psi+}(\theta) &= 2e_t^+ \tilde{x}_t^\top \delta \frac{q_{2t}}{\sigma_n} \mathcal{K}'\left(\frac{q_{1t} + q_{2t}^\top \psi}{\sigma_n}\right). \end{aligned} \tag{10}$$

Then, the variance estimators for the threshold parameter ψ are defined, respectively:

$$\hat{V}^{\psi} = \frac{1}{n}\sum_{t=1}^{n} \tau_{n,t}^{\psi}(\theta_n)\, \tau_{n,t}^{\psi}(\theta_n)^\top \text{ and } \hat{V}^{\psi+} = \frac{1}{n}\sum_{t=1}^{n} \tau_{n,t}^{\psi+}\left(\theta_n^+\right) \tau_{n,t}^{\psi+}\left(\theta_n^+\right)^\top. \tag{11}$$

These impose the absence of any theoretical autocorrelation but allow for heteroskedasticity. We may also make some degrees of freedom adjustment replacing n (in the denominator) by $n-k$, where k is the total number of estimated regression parameters. One may wish to impose homoskedasticity, which can be achieved by separating out the residuals, for example replace

$$\hat{V}_H^{\psi+} = \frac{1}{n}\sum_{t=1}^{n}(e_t^+)^2 \times \frac{4}{n\sigma_n^2}\sum_{t=1}^{n}(\tilde{x}_t^\top \delta_n)^2 \mathcal{K}'\left(\frac{q_{1t} + q_{2t}^\top \psi_n}{\sigma_n}\right)^2 q_{2t} q_{2t}^\top.$$

For the estimation of Q^{ψ} and $Q^{\psi+}$, just take $Q_n^{\psi}(\theta_n;\sigma_n)$ and $Q_n^{\psi+}\left(\theta_n^{+};\sigma_n\right)$. Unlike Hansen (2000) we do not need to explicitly do nonparametric estimation of density and conditional expectation.

We now turn to V^s, which requires HAC estimation because the effect of error autocorrelation does not die out. Let

$$\begin{aligned} \tau_{n,t}^{s}(\theta) &= \left(x_t^{\top}, \tilde{x}_t^{\top}\mathcal{K}\left(\frac{q_{1t}+q_{2t}^{\top}\psi}{\sigma_n}\right)\right)^{\top} \qquad (12) \\ \hat{\Gamma}_j &= \begin{cases} \frac{1}{n}\sum_{t=j+1}^{n}\tau_{n,t}^{s}(\theta_n)\,\tau_{n,t-j}^{s}(\theta_n)^{\top} e_t(\theta_n)\,e_{t-j}(\theta_n) & for\ j\geq 0, \\ \frac{1}{n}\sum_{t=-j+1}^{n}\tau_{n,t+j}^{s}(\theta_n)\,\tau_{n,t}^{s}(\theta_n)^{\top} e_{t+j}(\theta_n)\,e_t(\theta_n) & for\ j<0. \end{cases} \end{aligned}$$

Let $w(\cdot):\mathbb{R}\rightarrow[-1,1]$ be a continuous function such that $w(0)=1$, $w(x)=w(-x)$, $\sup_x|w(x)|<\infty$ and $\int\bar{w}(x)\,dx<\infty$ where $\bar{w}(x)=\sup_{y\geq x\geq 0}|w(y)|$. Then, define

$$\hat{V}^s=\sum_{j=-n+1}^{n-1} w\left(\frac{j}{l_n}\right)\hat{\Gamma}_j,$$

where l_n is a lag truncation parameter that is $o\left(\sigma_n^{-1}\right)$. Similarly we can define $\hat{V}^{s+}$. For more discussion regarding the choice of the kernel and lag truncation parameter, see Andrews (1991) and Jasson (2002). It should be noted, however, that his consistency results regarding the HAC estimator do not hold for the threshold models due to the lack of smoothness. Finally Q^s and Q^{s+} can be estimated by

$$\begin{aligned} \hat{Q}^s &= \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n}x_t x_t^{\top} & \frac{1}{n}\sum_{t=1}^{n}x_t\tilde{x}_t^{\top}\mathcal{K}\left(\frac{q_{1t}+q_{2t}^{\top}\psi_n}{\sigma_n}\right) \\ \frac{1}{n}\sum_{t=1}^{n}\tilde{x}_t x_t^{\top}\mathcal{K}\left(\frac{q_{1t}+q_{2t}^{\top}\psi_n}{\sigma_n}\right) & \frac{1}{n}\sum_{t=1}^{n}\tilde{x}_t\tilde{x}_t^{\top}\mathcal{K}\left(\frac{q_{1t}+q_{2t}^{\top}\psi_n}{\sigma_n}\right) \end{pmatrix} \\ \hat{Q}^{s+} &= \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n}x_t x_t^{\top} & \frac{1}{n}\sum_{t=1}^{n}x_t\tilde{x}_t^{\top}\mathcal{K}\left(\frac{q_{1t}+q_{2t}^{\top}\psi_n^{+}}{\sigma_n}\right) \\ \frac{1}{n}\sum_{t=1}^{n}\tilde{x}_t x_t^{\top}\mathcal{K}\left(\frac{q_{1t}+q_{2t}^{\top}\psi_n^{+}}{\sigma_n}\right) & \frac{1}{n}\sum_{t=1}^{n}\tilde{x}_t\tilde{x}_t^{\top}\mathcal{K}\left(\frac{q_{1t}+q_{2t}^{\top}\psi_n^{+}}{\sigma_n}\right)^2 \end{pmatrix}. \end{aligned}$$

We can replace $\mathcal{K}$ with the indicator function.

The above standard errors have imposed the block diagonal structure between the estimates of ψ,θ^s found in the asymptotics. In small samples it may be preferable to not impose this restriction; indeed, Hansen (2000) proposed to use Bonferoni-type bands to take account of the small sample effect of estimation error in ψ on the the estimation of θ^s. We have a much more natural and simple way of doing this. Consider the case where the error is the martingale difference sequence as in Hansen (2000) and we do not need the HAC estimation. Let $\tau_{n,t}(\theta)=\left(\tau_{n,t}^{s}(\theta)^{\top}e_t(\theta),\tau_{n,t}^{\psi}(\theta)^{\top}\right)^{\top}$ and $\hat{V}_{NB}=n^{-1}\sum_{t=1}^{n}\tau_{n,t}(\theta_n)\,\tau_{n,t}(\theta_n)^{\top}$, and use the diagonal elements of the matrix

$$D_n^{-1}Q(\theta_n)^{-1}\hat{V}_{NB}Q(\theta_n)^{-1}D_n^{-1},$$

where D_n is a diagonal matrix whose elements are 1 and $\sqrt{\sigma_n}$, and the latter is associated with the threshold parameter.

The following theorem establishes the consistency of the proposed standard errors.

Theorem 4 *Let $l_n \sigma_n \to 0$ and Assumption 1-3 hold. Then, $\hat{V}^s, \sigma_n \hat{V}^\psi, \hat{Q}^s$ and $\sigma_n Q_n^\psi(\theta_n)$ converge in probability to V^s, V^ψ, Q^s and Q^ψ, respectively.*

It follows that t and Wald statistics based on any of the above estimates are asymptotically correctly sized.

4.2 Likelihood Ratio

Dufour (1997) argues that t and Wald statistics behave poorly when the parameter space contains a region where identification fails. Therefore, Hansen (2000), in which the threshold parameter is not identified asymptotically, proposes the confidence interval for the threshold parameter ψ inverted from the LR statistic that is constructed under the auxiliary assumption that the error is i.i.d. normal. We may define

$$LR(\psi) = n \frac{S_n(\psi) - S_n(\psi_n)}{S_n(\psi_n)}, \tag{13}$$

and similarly $LR(\psi)^+$ using S_n^+. If ψ is one-dimensional, the statistics are distributed as $s \times \mathcal{X}_1^2$ asymptotically where the scaling factors are $s = V^\psi / 2Q^\psi \sigma^2$ or $s = V^{\psi+}/2Q^{\psi+}\sigma^2$, where $\sigma^2 = \mathrm{var}(\varepsilon_t)$. We need to adjust the critical values or repivot the test statistics by dividing through by an estimate of s obtained in the previous section. The resulting confidence region is the set $C_\alpha = \{\psi : LR(\psi)/\widehat{s} \leq \mathcal{X}_1^2(\alpha)\}$, where $\mathcal{X}_1^2(\alpha)$ is the upper α-critical value of the $\mathcal{X}_1^2$ distribution and $\widehat{s}$ is a consistent estimate of s. Note that in finite samples C_α is not necessarily an interval and may be a union of disjoint intervals, as happens quite often in practice, see Hansen (2000, Figure 2). In this case, one may prefer the interval $C_\alpha^{int} = [\psi_{\min}, \psi_{\max}]$, where $\psi_{\min} = \inf_{\psi \in C_\alpha} \psi$ and $\psi_{\max} = \sup_{\psi \in C_\alpha} \psi$. Asymptotically, C_α^{int} and C_α are the same, but in finite samples $C_\alpha^{int} \supseteq C_\alpha$. When ψ is multidimensional, the adjustment for conditional heteroskedasticity is more complicated and this reduces the attractiveness of the likelihood ratio.

4.3 Bootstrap

An alternative approach to inference is based on the bootstrap. This can be expected to produce consistent confidence intervals here because we have an asymptotically normal estimator, see Horowitz (2001). We provide a theorem on this below. However, the bootstrap does not work for the unsmoothed estimator according to our simulations.

There are two possible approaches to bootstrapping in regression models for i.i.d. data: residual resampling and i.i.d. resampling. We propose to use i.i.d. resampling. Let $\{W_t\}_{t=1}^n$ be the dataset, where $W_t = (y_t, X_t)$. Then let $\{W_t^*\}_{t=1}^n$ be a random sample drawn with replacement from $\{W_t\}_{t=1}^n$. Compute $\theta_n^* = (\theta_n^{*s\top}, \psi_n^{*\top})^\top$ from $\{W_t^*\}_{t=1}^n$ in the same way as θ_n was computed from $\{W_t\}_{t=1}^n$. Suppose that one wants a two-sided symmetric level α confidence interval for the scalar quantity $\tau(\theta)$ for some function τ. The percentile method is to obtain the empirical quantiles $x_{n,\alpha}$ of the distribution of $\tau(\theta_n^*)$ conditional on $\{W_t\}_{t=1}^n$, and then let the interval be $[\tau(\theta_n) + x_{n,\alpha/2}, \tau(\theta_n) + x_{n,1-\alpha/2}]$. A perhaps more desirable approach, usually called the bootstrap-t method, is based on the pivotal statistic $T = (\tau(\theta_n) - \tau(\theta))/s_n^\tau$, where s_n^τ is an estimate of the asymptotic standard deviation of $\tau(\theta_n)$. In the event that τ is differentiable we would take $s_n^\tau = \nabla\tau(\theta_n)^\top \widehat{Q}^{-1}\widehat{V}\widehat{Q}^{-1}\nabla\tau(\theta_n)$, where $\widehat{V}$ and $\widehat{Q}$ are the matrices with sub-blocks $\hat{V}^s$ and $\hat{V}^\psi$ and $\widehat{Q}^s$ and $\widehat{Q}^\psi$ described above (in the i.i.d. case one does not compute the covariances). From the bootstrap simulation one obtains the critical values $z_{n,\alpha/2}$ of $T^* = (\tau(\theta_n^*) - \tau(\theta_n))/s_n^{\tau*}$ and then the required interval is $[\tau(\theta_n) - z_{n,1-\alpha/2}s_n^\tau, \tau(\theta_n) - z_{n,\alpha/2}s_n^\tau]$.

The following theorem establishes the asymptotically correctness of the bootstrap. The proof of this theorem follows from the arguments of Horowitz (2001, 2002), the latter of which establishes a higher-order refinement of the bootstrap of the smoothed maximum score estimator.

Theorem 5 *Let Assumptions 1 - 3 hold. Furthermore, assume $\{W_t\}$ are i.i.d. sequence of bounded random variables. Then, conditional on $\{W_t\}$ the following holds almost surely:*

$$\begin{aligned}
\sqrt{n}\,(\theta_n^{s*} - \theta_n^s) &\implies N\left(0, Q^{s^{-1}} V^s Q^{s^{-1}}\right), \\
\sqrt{n\sigma_n^{-1}}\,(\psi_n^* - \psi_n) &\implies N\left(0, Q^{\psi^{-1}} V^\psi Q^{\psi^{-1}}\right),
\end{aligned}$$

and they are asymptotically independent. The same is true for θ_n^{+*}. Furthermore, conditional on $\{W_t\}$, the following also holds almost surely: $\hat{V}^{s*}, \sigma_n\hat{V}^{\psi*}, \hat{Q}^s$ and $\sigma_n\hat{Q}^{\psi*}$ converge in probability to V^s, V^ψ, Q^s and Q^ψ, respectively.

Following Horowitz (2002) we consider bounded random variables. However, this restriction is not essential but simplifies the proof as he comments. Additional conditions on the smoothness of $\mathcal{K}$ and the distribution imposed there are irrelevant in our case, as we do not consider higher-order expansions. The bounded support for $\mathcal{K}$ is also irrelevant as he remarks. Since the asymptotic distribution of T does not depend on nuisance parameters, we may expect the bootstrap to achieve asymptotic refinements under stronger conditions as in Horowitz (1998, 2002) , who establishes such refinements for the cases of smoothed median estimation and smoothed maximum score estimation, respectively. Although we do not pursue higher-order expansions here, we note that the expansions in Horowitz (2002) can be used to develop the expansion of a key component of our statistics. Similar comments apply to the likelihood ratio statistics or the repivoted likelihood ratio statistics.

In the time series case, one generally has to use a more complicated resampling method like the block bootstrap (Carlstein (1986)) and Künsch (1989)) to capture the effect of the dependence structure on the limiting distribution. However, in the special case of the threshold parameter or functions thereof, one can obtain consistent confidence intervals from the i.i.d. resampling because the limiting distribution of the estimator is not affected by the dependence structure. On the other hand, one does not obtain asymptotic refinements by this method.

5 The Continuous Case

Suppose that

$$\dot{\delta}_2 = 0, \tag{14}$$

where $\dot{\delta}_2$ was defined in section 3.2. Then, the model (1) becomes continuous, since $\tilde{x}_t^\top \delta = \left(z_t, X_{2t}^\top\right) \dot{\delta}$. In this case, the formula $Q^{\psi+^{-1}} V^{\psi+} Q^{\psi+^{-1}}$ we gave for the asymptotic variance of the threshold parameter estimate is not well-defined, since V^ψ and Q^ψ are zero; however, lower order terms can be found that are non-zero in both quantities. Let

$$\begin{aligned}
\mathcal{V}^\psi &= 4\dot{\delta}_1^2 \cdot \int s^2 \mathcal{K}'(s)^2 \, ds \cdot E\left[\bar{E}\left(\varepsilon_t^2 | X_{2t}\right) q_{2t} q_{2t}^\top | z_t = 0\right] f(0), \\
\mathcal{Q}^\psi &= \dot{\delta}_1^2 \int -s^2 \operatorname{sgn}(s) \mathcal{K}''(s) \, ds \cdot E\left[q_2 q_2^\top | z_t = 0\right] f(0).
\end{aligned}$$

Then, Corollary 3 can be modified as follows.

Corollary 6 *Let Assumptions 1 - 3 hold with $h = 1$ and with V^ψ and Q^ψ replaced by $\mathcal{V}^\psi$ and $\mathcal{Q}^\psi$ respectively. Furthermore, assume* (14) *. Then,*

$$\begin{aligned}
\sqrt{n}\left(\theta_n^{s+} - \theta_0^s\right) &\Longrightarrow N\left(0, Q^{s^{-1}} V^s Q^{s^{-1}}\right), \\
\sqrt{n\sigma_n}\left(\psi_n^+ - \psi_0\right) &\Longrightarrow N\left(0, \mathcal{Q}^{\psi^{-1}} \mathcal{V}^\psi \mathcal{Q}^{\psi^{-1}}\right),
\end{aligned}$$

and they are asymptotically independent.

Note that the convergence rate of the threshold estimate ψ_n is changed from $\sqrt{n\sigma_n^{-1}}$ to $\sqrt{n\sigma_n}$. This rate is slower than that of the unsmoothed LSE of a TAR model in Gonzalo and Wolf (2005), where both the slope and threshold estimates are jointly asymptotically normally distributed with the $\sqrt{n}$ rate and they are correlated.

Since Assumption 3 (b) with $h \geq 2$ implies that $\int -s^2 \operatorname{sgn}(s) \mathcal{K}''(s) \, ds = 0$, the leading term of Q_n^ψ is asymptotically stochastic in this case, destroying the asymptotic normality of ψ_n^+.

An asymptotically correct confidence interval for threshold parameter can still be constructed by the same method described above

When the confidence interval is constructed as in Section 4 with the bandwidth σ_n satisfying $n\sigma_n^3 \to 0$ and with $\mathcal{K}$ satisfying Assumption 3 for $h = 1$, it will be an asymptotically correct even when the true model is continuous. We can also construct a test for the continuity of the model. Since $\dot{\delta} = \mathcal{T}^{-1^\top} \mathcal{S}^\top \delta$, we can test the hypothesis (14) by the $\mathcal{X}^2$ test, utilizing the delta-method and Corollary 6.

6 Discussion and Conclusions

We have shown that the smoothed threshold estimator is asymptotically normal albeit at a slower rate than the corresponding unsmoothed estimator. In Seo and Linton (2006) we carried out simulations to investigate the finite sample performance of the estimator and the accuracy of the distributional approximation in the context of the design of Hansen (2000). We report briefly on those results here. Regarding the estimators of ψ, the performance of both estimators improves with sample size and as the size of the threshold increases. The small sample variability of all estimates is much higher than predicted by the asymptotic theory, but this overprediction reduces considerably with sample size and as the size of the threshold increases. This overprediction is also implicitly true for the unsmoothed least squares estimator. The estimator ψ_n^+ is nearly always better than ψ_n. We also found that normality is a good approximation for the largest sample size considered. Regarding the confidence intervals for ψ, the estimators we consider converge at a slower rate than the unsmoothed least squares estimator and so should lead to asymptotically longer confidence intervals than those of Hansen (2000) and Gozalo and Wolf (2005) [although neither of these authors investigate the length of their confidence intervals in finite samples]. On the other hand, the coverage rate of the respective confidence intervals is asymptotically the same and may be better or worse in finite samples - the first order asymptotic theory has nothing to say on this issue. In fact, we found that our confidence intervals especially those based on the standard bootstrap can be more accurate than the confidence intervals of Hansen (2000) especially for larger thresholds.[3] It might be possible to provide a theoretical justification for this finding. Specifically, it may be possible to show that the rate at which the estimator (or corresponding test statistics) approaches its limit is quite fast, see Hall (1992) for corresponding results for density and regression estimators and Horowitz (1998) for

[3]On the other hand the bootstrap for the unsmoothed estimator does not appear to work well in simulations. The coverage rates were very low (and are not reported here) even in the largest sample sizes. We take this as evidence of inconsistency, which would agree with results of Abrevaya and Huang (2005) for the unsmoothed maximum score procedure.

results for smoothed LAD (SLAD) estimators, and perhaps faster than is the case for the unsmoothed estimator. Furthermore, we expect that the smoothed estimation will enable higher-order correction by the pivotal bootstrap, as is the case in the SLAD estimation in Horowitz (1998). He shows that the SLAD estimator has much simpler higher-order asymptotics than the LAD estimator and thus the bootstrap can correct the second-order term. Since smoothing also makes the objective function of the threshold estimation differentiable, which is necessary for the Taylor-series expansion, we can expect a simpler expansion and the higher-order correctibility of the bootstrap. This would provide a theoretical rationale for the simulation results and give one motivation for preferring our estimator/test statistic over the unsmoothed one.

In practice, it is important to have some strategy for choosing the smoothing parameter σ_n. The answer is likely to depend on the criterion employed. For estimation, the smaller the bandwidth the faster the convergence rate, so there is no optimal bandwidth. We have used the smallest bandwidth that is consistent with our theory i.e., of the order $(\log n)n^{-1/2}$ and it seems to perform well in simulations. For testing problems bandwidth is likely to affect size distortion (accuracy of the distributional approximation) in a complex way and it may be possible to find an optimal bandwidth based on higher order expansions. However, bandwidth can also affect the power in a different way and so the size distortion optimal bandwidth may yield poor power and vice versa. See Fan and Linton (2003).

A Proofs of Theorems

A word on notation. Every convergence is as $n \to \infty$. The integral $\int$ is taken over whole sample space unless specified otherwise.

LEMMA 1. *Suppose that $\{w_t, q_t\}$ satisfies Assumption 1(a) and $E\,|w_t|^{\xi} < \infty$ for ξ in Assumption 1(b). Then, for a given η,*

$$\frac{1}{n}\sum_{t=1}^{n} w_t 1\left\{q_{1t} + q_{2t}^{\top}\psi > \eta\right\} \to E w_t 1\left\{q_{1t} + q_{2t}^{\top}\psi > \eta\right\}$$

almost surely uniformly in $\psi \in \Theta_{\psi}$.

PROOF OF LEMMA 1. We apply the generic uniform law of large numbers by Andrews (1987, Corollary 1). We need to check Assumption A1, B1, B2, and A3 in the Corollary. Assumption A1 and B1 of that paper are assumed in Assumption 1. Assumption B2 is satisfied since

$$E\left(\sup_{\psi\in\Theta_{\psi}} \left|w_t 1\left\{q_{1t} + q_{2t}^{\top}\psi > \eta\right\}\right|\right)^{\xi} \le E\,|w_t|^{\xi} < \infty.$$

For Assumption A3 note that

$$
\begin{aligned}
& \left|\frac{1}{n}\sum_{t=1}^{n} E\left[w_t 1\left\{\sup_{\grave{\psi}:\left|\psi-\grave{\psi}\right|<\rho} q_{1t}+q_{2t}^{\top}\psi>\eta\right\}\right]-E\left[w_t 1\left\{\inf_{\grave{\psi}:\left|\psi-\grave{\psi}\right|<\rho} q_{1t}+q_{2t}^{\top}\psi>\eta\right\}\right]\right| \\
\leq\ & \left|E\left[w_t\left(1\left\{\sup_{\grave{\psi}:\left|\psi-\grave{\psi}\right|<\rho} q_{1t}+q_{2t}^{\top}\psi>\eta\right\}-1\left\{\inf_{\grave{\psi}:\left|\psi-\grave{\psi}\right|<\rho} q_{1t}+q_{2t}^{\top}\psi>\eta\right\}\right)\right]\right| \\
\leq\ & \left(Ew_t^{\xi}\right)^{1/\xi}\left(E\left(1\left\{\sup_{\grave{\psi}:\left|\psi-\grave{\psi}\right|<\rho} q_{1t}+q_{2t}^{\top}\psi>\eta\right\}-1\left\{\inf_{\grave{\psi}:\left|\psi-\grave{\psi}\right|<\rho} q_{1t}+q_{2t}^{\top}\psi>\eta\right\}\right)^{\zeta}\right)^{1/\zeta},
\end{aligned}
$$

for $1/\zeta+1/\xi=1$. But, the last term converges to zero as $\rho\to 0$ by de Jong and Woutersen (2004, Lemma 4).[4] ■

Proof of Theorem 1. First, we show that $\left|S_n^*(\theta)-S_n(\theta;\sigma_n)\right|\to 0$ almost surely uniformly in $\theta\in\Theta$. To do that, let q satisfy $1/q+1/\grave{\xi}=1$ for $\xi>\grave{\xi}>1$ and ξ in Assumption 1(b) and note that

$$
\begin{aligned}
& \sup_{\theta\in\Theta}\left|S_n^*(\theta)-S_n(\theta;\sigma_n)\right| \\
=\ & \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\left\{\left(\tilde{x}_t^{\top}\delta\right)^2-2\tilde{x}_t^{\top}\delta\left(y_t-x_t^{\top}\beta\right)\right\}\left[1\left\{q_{1t}+q_{2t}^{\top}\psi>0\right\}-\mathcal{K}\left(\frac{q_{1t}+q_{2t}^{\top}\psi}{\sigma_n}\right)\right]\right| \\
\leq\ & \left(\sup_{\theta\in\Theta}\frac{1}{n}\sum_{t=1}^{n}\left\{\left(\tilde{x}_t^{\top}\delta\right)^2-2\tilde{x}_t^{\top}\delta\left(y_t-x_t^{\top}\beta\right)\right\}^{\grave{\xi}}\right)^{1/\grave{\xi}} \\
& \times\left(\sup_{\theta\in\Theta}\frac{1}{n}\sum_{t=1}^{n}\left|1\left\{q_{1t}+q_{2t}^{\top}\psi>0\right\}-\mathcal{K}\left(\frac{q_{1t}+q_{2t}^{\top}\psi}{\sigma_n}\right)\right|^{q}\right)^{1/q},
\end{aligned}
$$

where the first term converges to its expectation, which is finite, almost surely as in Lemma 1 and the second term to zero almost surely. The latter follows from Lemma 4 of Horowitz (1992), provided that, for any $\eta>0$,

$$
\frac{1}{n}\sum_{t=1}^{n}1\left\{\left|q_{1t}+q_{2t}^{\top}\psi\right|<\eta\right\}
$$

converges to $\Pr\left\{\left|q_{1t}+q_{2t}^{\top}\psi\right|<\eta\right\}$, almost surely uniformly over $\psi\in\Theta_{\psi}$, which follows from Lemma 1 since

$$
1\left\{\left|q_{1t}+q_{2t}^{\top}\psi\right|<\eta\right\}=1\left\{q_{1t}+q_{2t}^{\top}\psi<\eta\right\}-1\left\{q_{1t}+q_{2t}^{\top}\psi\leq-\eta\right\}.
$$

Next, we show that $\theta_n^*=\arg\min_{\theta\in\Theta}S_n^*(\theta)$ is consistent, which also yields the consistency of θ_n. Let

$$
r_t(\theta)=x_t^{\top}(\beta-\beta_0)+\tilde{x}_t^{\top}\left(\delta 1\left\{q_{1t}+q_{2t}^{\top}\psi>0\right\}-\delta_0 1\left\{z_t>0\right\}\right).
$$

[4] Lemma 4 in de Jong and Woutersen is shown for the case where w_t is bounded.

First, by Lemma 1,

$$S_n^*(\theta) \xrightarrow{p} S^*(\theta) = E(\varepsilon_t - r_t(\theta))^2,$$

uniformly in $\theta \in \Theta$. As $S^*(\theta)$ is continuous, it remains to show that it is uniquely minimized at $\theta = \theta_0$. But, as $E(\varepsilon_t | X_t) = 0$, we have

$$S^*(\theta) = S^*(\theta_0) + Er_t(\theta)^2.$$

Then, we show that there is a set A that has positive probability and that for any $\theta \neq \theta_0$, $E\left[r_t(\theta)^2 | A\right] > 0$, which in turn implies that $Er_t(\theta)^2 > 0$ for any $\theta \neq \theta_0$, as $r_t(\theta)^2 \geq 0$. Let

$$\begin{aligned} A_1 &= \{z_t > 0\}, \\ A_2 &= \left\{z_t > 0, q_{1t} + q_{2t}^\top \psi \leq 0\right\}, \\ A_3 &= \left\{z_t \leq 0, q_{1t} + q_{2t}^\top \psi \leq 0\right\}. \end{aligned}$$

Note that the probabilities of these sets are all positive by Assumption 1 (d). As $\theta = (\beta, \delta, \psi)$, we need to consider the following three cases: (i) $\psi = \psi_0, (\beta, \delta) \neq (\beta_0, \delta_0)$. Then,

$$E\left[r_t(\theta)^2 | A_1\right] = E\left[\left(x_t^\top(\beta - \beta_0) + \tilde{x}_t^\top(\delta - \delta_0)\right)^2 | A_1\right].$$

(ii) $\psi \neq \psi_0, \beta = \beta_0$. Then,

$$E\left[r_t(\theta)^2 | A_2\right] = E\left[\left(\tilde{x}_t^\top \delta_0\right)^2 | A_2\right].$$

(iii) $\psi \neq \psi_0, \beta \neq \beta_0$. Then,

$$E\left[r_t(\theta)^2 | A_3\right] = E\left[\{x_t^\top(\beta - \beta_0)\}^2 | A_3\right].$$

By Assumption 1 (b) and (c), all these three quantities are positive. ■

Proof of Theorem 2. The asymptotic distribution developed here is based on the Taylor series expansion of $T_n(\theta; \sigma_n)$: for large n

$$T_n(\theta_n; \sigma_n) = T_n(\theta_0; \sigma_n) + Q_n(\tilde{\theta}; \sigma_n)(\theta_n - \theta_0) = 0,$$

where $\tilde{\theta} = (\tilde{\beta}^\top, \tilde{\delta}^\top, \tilde{\psi}^\top)^\top$ lies between θ_n and θ_0. Let the dimension of θ^s be k_s and define a k-dimensional diagonal matrix D_n whose first k_s elements are 1 and the others are $\sqrt{\sigma_n}$. Then, we may write

$$\sqrt{n} D_n^{-1}(\theta_n - \theta_0) = \begin{pmatrix} Q_n^s\left(\tilde{\theta}, \sigma_n\right) & \sqrt{\sigma_n} Q_n^{s\psi}\left(\tilde{\theta}, \sigma_n\right) \\ \sqrt{\sigma_n} Q_n^{s\psi\top}\left(\tilde{\theta}, \sigma_n\right) & \sigma_n Q_n^{\psi}\left(\tilde{\theta}, \sigma_n\right) \end{pmatrix}^{-1} \begin{pmatrix} \sqrt{n} T_n^s(\theta_0, \sigma_n) \\ \sqrt{n\sigma_n} T_n^{\psi}(\theta_0, \sigma_n) \end{pmatrix}.$$

From (10) , we write

$$T_n^{\psi}(\theta_0,\sigma_n)=\frac{1}{n}\sum_{t=1}^{n}\tau_{n,t}^{\psi}(\theta_0)=-\frac{1}{n}\sum_{t=1}^{n}\left(\left(\tilde{x}_t^{\top}\delta_0\right)^2\operatorname{sgn}(z_t)+2\tilde{x}_t^{\top}\delta_0\varepsilon_t\right)\frac{q_{2t}}{\sigma_n}\mathcal{K}'\left(\frac{z_t}{\sigma_n}\right),$$

since

$$\begin{aligned}\left\{\left(\tilde{x}_t^{\top}\delta\right)^2-2\tilde{x}_t^{\top}\delta\left(y_t-x_t^{\top}\beta\right)\right\} &= \left(\tilde{x}_t^{\top}\delta_0\right)^2-2\tilde{x}_t^{\top}\delta_0\left(y_t-x_t^{\top}\beta_0\right)+R_{nt}(\theta)\\ &= \left(\tilde{x}_t^{\top}\delta_0\right)^2-2\tilde{x}_t^{\top}\delta_0\left(\tilde{x}_t^{\top}\delta_0 1\{z_t>0\}+\varepsilon_t\right)+R_{nt}(\theta)\\ &= \left(\tilde{x}_t^{\top}\delta_0\right)^2\left(1-2\cdot 1\{z_t>0\}\right)-2\tilde{x}_t^{\top}\delta_0\varepsilon_t+R_{nt}(\theta)\\ &= -\left(\left(\tilde{x}_t^{\top}\delta_0\right)^2\operatorname{sgn}(z_t)+2\tilde{x}_t^{\top}\delta_0\varepsilon_t\right)+R_{nt}(\theta),\end{aligned}\tag{15}$$

where

$$R_{nt}(\theta)=2\delta^{\top}\tilde{x}_t x_t^{\top}(\beta-\beta_0)+\left((\delta+\delta_0)^{\top}\tilde{x}_t\tilde{x}_t^{\top}-2\tilde{x}_t^{\top}1\{z_t>0\}-2\tilde{x}_t^{\top}\varepsilon_t\right)(\delta-\delta_0),\tag{16}$$

and $\operatorname{sgn}(s)=1$ if s is positive, and -1 otherwise. And,

$$\begin{aligned}T_n^{s}(\theta_0,\sigma_n) &= \begin{pmatrix}\frac{1}{n}\sum_{t=1}^{n}2\left(y_t-x_t^{\top}\beta_0\right)(-x_t)+\frac{1}{n}\sum_{t=1}^{n}2\tilde{x}_t^{\top}\delta_0 x_t\mathcal{K}\left(\frac{z_t}{\sigma_n}\right)\\ \frac{1}{n}\sum_{t=1}^{n}\left\{2\left(\tilde{x}_t^{\top}\delta_0\right)\tilde{x}_t-2\tilde{x}_t\left(y_t-x_t^{\top}\beta_0\right)\right\}\mathcal{K}\left(\frac{z_t}{\sigma_n}\right)\end{pmatrix}\\ &= \begin{pmatrix}-\frac{2}{n}\sum_{t=1}^{n}x_t\varepsilon_t-\frac{2}{n}\sum_{t=1}^{n}x_t\tilde{x}_t^{\top}\delta_0\left(1\{z_t>0\}-\mathcal{K}\left(\frac{z_t}{\sigma_n}\right)\right)\\ -\frac{2}{n}\sum_{t=1}^{n}\tilde{x}_t\varepsilon_t\mathcal{K}\left(\frac{z_t}{\sigma_n}\right)+\frac{2}{n}\sum_{t=1}^{n}\tilde{x}_t\tilde{x}_t^{\top}\delta_0 1\{z_t\le 0\}\mathcal{K}\left(\frac{z_t}{\sigma_n}\right)\end{pmatrix}.\end{aligned}$$

Furthermore,

$$\begin{aligned}Q_n^{s}\left(\tilde{\theta},\sigma_n\right) &= 2\begin{pmatrix}\frac{1}{n}\sum_{t=1}^{n}x_t x_t^{\top} & \frac{1}{n}\sum_{t=1}^{n}x_t\tilde{x}_t^{\top}\mathcal{K}\left(\frac{q_{1t}+q_{2t}^{\top}\tilde{\psi}}{\sigma_n}\right)\\ \frac{1}{n}\sum_{t=1}^{n}\tilde{x}_t x_t^{\top}\mathcal{K}\left(\frac{q_{1t}+q_{2t}^{\top}\tilde{\psi}}{\sigma_n}\right) & \frac{1}{n}\sum_{t=1}^{n}\tilde{x}_t\tilde{x}_t^{\top}\mathcal{K}\left(\frac{q_{1t}+q_{2t}^{\top}\tilde{\psi}}{\sigma_n}\right)\end{pmatrix}\\ Q_n^{s\tilde{\psi}}\left(\tilde{\theta},\sigma_n\right) &= 2\begin{pmatrix}\frac{1}{n}\sum_{t=1}^{n}x_t\tilde{x}_t^{\top}\tilde{\delta}\frac{q_{2t}^{\top}}{\sigma_n}\mathcal{K}'\left(\frac{q_{1t}+q_{2t}^{\top}\tilde{\psi}}{\sigma_n}\right)\\ \frac{1}{n}\sum_{t=1}^{n}\left(\tilde{x}_t\tilde{x}_t^{\top}\tilde{\delta}+\tilde{x}_t x_t^{\top}\tilde{\beta}-\tilde{x}_t y_t\right)\frac{q_{2t}^{\top}}{\sigma_n}\mathcal{K}'\left(\frac{q_{1t}+q_{2t}^{\top}\tilde{\psi}}{\sigma_n}\right)\end{pmatrix}\\ Q_n^{\tilde{\psi}}\left(\tilde{\theta},\sigma_n\right) &= \frac{1}{n}\sum_{t=1}^{n}\left\{\left(\tilde{x}_t^{\top}\delta\right)^2-2\tilde{x}_t^{\top}\delta(y_t-x_t^{\top}\tilde{\beta})\right\}\frac{q_{2t}q_{2t}^{\top}}{\sigma_n^2}\mathcal{K}''\left(\frac{q_{1t}+q_{2t}^{\top}\tilde{\psi}}{\sigma_n}\right)\\ &= \frac{1}{n}\sum_{t=1}^{n}\left\{-\left(\left(\tilde{x}_t^{\top}\delta_0\right)^2\operatorname{sgn}(z_t)+2\tilde{x}_t^{\top}\delta_0\varepsilon_t\right)+R_{nt}(\tilde{\theta})\right\}\frac{q_{2t}q_{2t}^{\top}}{\sigma_n^2}\mathcal{K}''\left(\frac{z_t+q_{2t}^{\top}(\tilde{\psi}-\psi_0)}{\sigma_n}\right),\end{aligned}$$

where the last equality follows from (15) . We show the convergence of each term in the following sequence of Lemmas.

Lemma 2. *The covariances between $\sqrt{n}T_n^{s}(\theta_0,\sigma_n)$ and $\sqrt{n\sigma_n}T_n^{\psi}(\theta_0,\sigma_n)$ are asymptotically negligible and*

$$\begin{aligned}\lim_{n\to\infty}E\left[\sqrt{n\sigma_n}T_n^{\psi}(\theta_0,\sigma_n)\right] &= 0\\ \lim_{n\to\infty}\operatorname{var}\left[\sqrt{n\sigma_n}T_n^{\psi}(\theta_0,\sigma_n)\right] &= 4V^{\psi}.\end{aligned}$$

Proof of Lemma 2. Recall that $E\left[\varepsilon_t|X_t\right]=0$. By (6) we may write

$$\begin{aligned}
-E\left[\sigma_n^{-h}\tau_{n,t}^{\psi}\left(\theta_0\right)\right] &= \sigma_n^{-h}\int\left\{\left(\tilde{x}^{\top}\delta_0\right)^2\operatorname{sgn}\left(z\right)\right\}\frac{q_2}{\sigma_n}\mathcal{K}'\left(\frac{z}{\sigma_n}\right)f_{z|X_2}\left(z|X_2\right)dzdF_{X_2}(X_2)\\
&= \sigma_n^{-h}\int\left\{\left(\sigma_n s\dot{\delta}_{10}+X_2^{\top}\dot{\delta}_{20}\right)^2\operatorname{sgn}\left(s\right)\right\}q_2\mathcal{K}'\left(s\right)f_{z|X_2}\left(\sigma_n s|X_2\right)dsdF_{X_2}(X_2),
\end{aligned}$$

where $s=z/\sigma_n$ and $F_{X_2}(X_2)$ is the distribution of X_{2t}. We show that

$$I_{n1}=\sigma_n^{-h}\int\left\{\left(X_2^{\top}\dot{\delta}_{20}\right)^2\operatorname{sgn}\left(s\right)\right\}q_2\mathcal{K}'\left(s\right)f_{z|X_2}\left(\sigma_n s|X_2\right)dsdF_{X_2}(X_2)=O\left(1\right).$$

The other terms can be analyzed similarly. A Taylor series expansion of $f_{z|X_2}$ about $\sigma_n s=0$ yields, for ζ between 0 and $\sigma_n s$,

$$f_{z|X_2}\left(\sigma_n s|X_2\right)=\sum_{j=0}^{h-1}\left(1/j!\right)f_{z|X_2}^{(j)}\left(0|X_2\right)\left(\sigma_n s\right)^j+\left(1/h!\right)f_{z|X_2}^{(h)}\left(\zeta|X_2\right)\left(\sigma_n s\right)^h.$$

Then, using Assumption $3(b)$, we have for the constant M in Assumption 2

$$\begin{aligned}
I_{n1} &= \int s^h\operatorname{sgn}\left(s\right)\mathcal{K}'\left(s\right)f_{z|X_2}^{(h)}\left(\zeta|X_2\right)ds\left(X_2^{\top}\dot{\delta}_{20}\right)^2q_2dF_{X_2}(X_2)/h!\\
&\leq M\int s^h\operatorname{sgn}\left(s\right)\mathcal{K}'\left(s\right)ds\int\left(X_2^{\top}\dot{\delta}_{20}\right)^2q_2dF_{X_2}(X_2)<\infty.
\end{aligned}$$

As $\sqrt{n\sigma_n}\sigma_n^h\to 0$, the first part of the lemma is proved.

To study var $\left[\sqrt{n\sigma_n}T_n^{\psi}\left(\theta_0,\sigma_n\right)\right]$, note that

$$\sigma_n\operatorname{var}\left[\tau_{n,t}^{\psi}\right]=\sigma_n E\left[\left(\left(\tilde{x}_t^{\top}\delta_0\right)^4+4\left(\tilde{x}_t^{\top}\delta_0\varepsilon_t\right)^2\right)\frac{q_{2t}q_{2t}^{\top}}{\sigma_n^2}\mathcal{K}'\left(\frac{z_t}{\sigma_n}\right)^2\right],$$

where $\tau_{n,t}^{\psi}=\tau_{n,t}^{\psi}\left(\theta_0\right)$. Since $\left|f_{z|X_2}\right|$ and $E\left(X_{2t}^{\top}\dot{\delta}_{20}\right)^4q_{2t}q_{2t}^{\top}$ are bounded by M by Assumption 2 and $\int s^4\mathcal{K}'\left(s\right)^2ds<\infty$ by Assumption 3, it follows from the dominated convergence theorem that

$$\begin{aligned}
\sigma_n E\left[\left(\tilde{x}_t^{\top}\delta_0\right)^4\frac{q_{2t}q_{2t}^{\top}}{\sigma_n^2}\mathcal{K}'\left(\frac{z_t}{\sigma_n}\right)^2\right] &= \int\left(\tilde{x}^{\top}\delta_0\right)^4\frac{q_2q_2^{\top}}{\sigma_n}\mathcal{K}'\left(\frac{z}{\sigma_n}\right)^2f_{z|X_2}\left(z|X_2\right)dzdF_{X_2}(X_2)\\
&= \int\left(\sigma_n s\dot{\delta}_{10}+X_2^{\top}\dot{\delta}_{20}\right)^4q_2q_2^{\top}\mathcal{K}'\left(s\right)^2f_{z|X_2}\left(\sigma_n s|X_2\right)dsdF_{X_2}(X_2)\\
&\to ||\mathcal{K}'||_2^2\cdot E\left[\left(X_{2t}^{\top}\dot{\delta}_{20}\right)^4q_{2t}q_{2t}^{\top}f_{z|X_2}\left(0|X_2\right)\right],
\end{aligned}$$

where $s=z/\sigma_n$.

Similarly, as $E\left(\varepsilon_t^2|X_t\right) < M$ a.s. by Assumption 2,

$$
\begin{aligned}
& \sigma_n E\left[4\left(\tilde{x}_t^\top \delta\varepsilon_t\right)^2 \frac{q_{2t}q_{2t}^\top}{\sigma_n^2}\mathcal{K}'\left(\frac{z_t}{\sigma_n}\right)^2\right] \\
= & \int 4\left(\sigma_n s\dot{\delta}_{10} + X_2^\top \dot{\delta}_{20}\right)^2 E\left(\varepsilon^2|z=\sigma_n s, X_2\right) q_2 q_2^\top \mathcal{K}'(s)^2 f_{z|X_2}\left(\sigma_n s|X_2\right) ds dF_{X_2}(X_2) \\
= & \int_{s>0} 4\left(\sigma_n s\dot{\delta}_{10} + X_2^\top \dot{\delta}_{20}\right)^2 E\left(\varepsilon^2|z=\sigma_n s, X_2\right) q_2 q_2^\top \mathcal{K}'(s)^2 f_{z|X_2}\left(\sigma_n s|X_2\right) ds dF_{X_2}(X_2) \\
& + \int_{s<0} 4\left(\sigma_n s\dot{\delta}_{10} + X_2^\top \dot{\delta}_{20}\right)^2 E\left(\varepsilon^2|z=\sigma_n s, X_2\right) q_2 q_2^\top \mathcal{K}'(s)^2 f_{z|X_2}\left(\sigma_n s|X_2\right) ds dF_{X_2}(X_2) \\
\to & \ ||\mathcal{K}'||_2^2 \cdot E\left[4\left(X_{2t}^\top \dot{\delta}_{20}\right)^2 \overline{E}\left(\varepsilon_t^2|X_{2t}\right) q_{2t}q_{2t}^\top f_{z|X_2}\left(0|X_2\right)\right],
\end{aligned}
$$

where $\overline{E}\left(\varepsilon_t^2|X_{2t}\right)$ is defined in (7). Note that, for any measurable g, we have $E\left[g\left(X_{2t}\right) f_{z|X_2}\left(0|X_{2t}\right)\right] = E\left[g\left(X_{2t}\right)|z_t=0\right] f_z(0)$ by the law of iterated expectation and reversing the order of expectations. Thus, we get the expression in V^ψ.

Next, we proceed to show that the covariance terms die out as in de Jong and Woutersen (2004). Let a vector ζ satisfy $\|\zeta\| = 1$. First, by the mixing inequality (Davidson 1994, corollary 14.3), for $p \geq 2$,

$$\sigma_n \mathrm{cov}\left(\zeta^\top \tau_{n,t}^\psi, \zeta^\top \tau_{n,t-m}^\psi\right) \leq \sigma_n 2\left(2^{1-1/p}+1\right)\alpha_m^{1-2/p}\left\|\zeta^\top \tau_{n,t}^\psi\right\|_p^2,$$

and

$$\left\|\zeta^\top \tau_{n,t}^\psi\right\|_p^2 = \sigma_n^{-2+2/p} E\left[\left(\left(\tilde{x}_t^\top \delta_0\right)^2 \mathrm{sgn}\left(z_t\right) + 2\tilde{x}_t^\top \delta_0 \varepsilon_t\right)^p \frac{1}{\sigma_n}\left(q_{2t}^\top \zeta \mathcal{K}'\left(\frac{z_t}{\sigma_n}\right)\right)^p\right] = O\left(\sigma_n^{-2+2/p}\right).$$

Second, note that

$$
\begin{aligned}
\sigma_n \mathrm{cov}\left(\zeta^\top \tau_{n,t}^\psi, \zeta^\top \tau_{n,t-m}^\psi\right) &= \sigma_n E\zeta^\top \tau_{n,t-m}^\psi \zeta^\top \tau_{n,t}^\psi + \sigma_n\left(E\zeta^\top \tau_{n,t}^\psi\right)^2 \\
&= \sigma_n E\zeta^\top \tau_{n,t-m}^\psi \zeta^\top \tau_{n,t}^\psi + O\left(\sigma_n^{2h+1}\right),
\end{aligned}
$$

as shown in the first part of the proof of this lemma. Then, Lemma 7 of de Jong and Wouterson $(2004, p.24)$ shows that these two results yield

$$\sigma_n \sum_{m=1}^\infty \left|\mathrm{cov}\left(\zeta^\top \tau_{n,t}^\psi, \zeta^\top \tau_{n,t-m}^\psi\right)\right| \to 0.$$

Therefore, we conclude

$$\mathrm{var}\left[\sqrt{n\sigma_n} T_n^\psi\left(\theta_0, \sigma_n\right)\right] \to V^\psi.$$

By the same reasoning as for this, we can show that the covariances between $\sqrt{n} T_n^s\left(\theta_0, \sigma_n\right)$ and $\sqrt{n\sigma_n} T_n^\psi\left(\theta_0, \sigma_n\right)$ are asymptotically negligible. ■

LEMMA 3. *$\sqrt{n\sigma_n} T_n^{\psi}(\theta_0, \sigma_n)$ converges in distribution to $N(0, 4V^{\psi})$.*

PROOF OF LEMMA 3. See Lemma 6 of Horowitz (1992) and Theorem 2 of de Jong (1997). ∎

LEMMA 4. *$\sigma_n^{-1}(\psi_n - \psi_0) = o_p(1)$.*

PROOF OF LEMMA 4. Since θ_n is consistent there exists a sequence $r_n \to 0$ such that $P\{\|\theta_n - \theta_0\| > r_n\} \to 0$. We first show the following two claims:

$$\sup_{\theta \in \Theta_n} \left\| T_n^{\psi}(\theta) - ET_n^{\psi}(\theta) \right\| \xrightarrow{P} 0, \tag{17}$$

$$\sup_{\theta \in \Theta_n} \left\| ET_n^{\psi}(\theta) - \ddot{Q}_n(\theta) \right\| \to 0, \tag{18}$$

where $\Theta_n = \{\theta : \|\theta - \theta_0\| \le r_n\}$ and

$$\ddot{Q}_n(\theta) = 2\int \left(\mathcal{K}\left(q_2^{\top}(\psi - \psi_0)/\sigma_n\right) - \mathcal{K}(0)\right)\left(X_2^{\top}\dot{\delta}_2\right)^2 q_2 f_{z|X_2}(0|X_2)\, dF_{X_2}(X_2).$$

PROOF OF (17). Note that,

$$T_n^{\psi}(\theta) = -\left(\left(\tilde{x}_t^{\top}\delta\right)^2 \operatorname{sgn}(z_t) + 2\tilde{x}_t^{\top}\delta\varepsilon_t - R_{nt}(\theta)\right)\frac{q_{2t}}{\sigma_n}\mathcal{K}'\left(\frac{z_t + (\psi - \psi_0)^{\top} q_{2t}}{\sigma_n}\right), \tag{19}$$

where R_{nt} is defined in (15). Let $\tilde{q}_{2t} = q_{2t}/\sigma_n$ and define

$$g_{nt}(\theta) = \left(\tilde{x}_t^{\top}\delta\right)^2 \operatorname{sgn}(z_t)\, \tilde{q}_{2t}\mathcal{K}'\left(\frac{z_t}{\sigma_n} + (\psi - \psi_0)^{\top}\tilde{q}_{2t}\right),$$

and $g_{nt}^{C_n}(\theta) = g_{nt}(\theta)\, 1\left\{\left\|\tilde{x}_t\tilde{x}_t^{\top}\right\| \le C_n^{1/2}, \|\tilde{q}_{2t}\| \le C_n^{1/2}\right\}$. Note that $g_{nt}^{C_n}(\theta)$ is uniformly continuous in θ as $\mathcal{K}''$ is bounded. Then, decompose $\frac{1}{n\sigma_n}\sum_{t=1}^n (g_{nt}(\theta) - Eg_{nt}(\theta))$ into

$$\frac{1}{n\sigma_n}\sum_{t=1}^n \left\{\left(g_{nt}(\theta) - g_{nt}^{C_n}(\theta)\right) + \left(g_{nt}^{C_n}(\theta) - Eg_{nt}^{C_n}(\theta)\right) + \left(Eg_{nt}^{C_n}(\theta) - Eg_{nt}(\theta)\right)\right\}. \tag{20}$$

Let $n\sigma_n^{-r}/C_n^{r/2} \to 0$ and consider the first term of (20). By the Markov inequality,

$$\begin{aligned}
&\Pr\left\{\sup_{\theta \in \Theta_n}\left\|\frac{1}{n\sigma_n}\sum_{t=1}^n \left(g_{nt}(\theta) - g_{nt}^{C_n}(\theta)\right)\right\| > 0\right\} \\
\le\ & n\left(\Pr\left\{\left\|\tilde{x}_t\tilde{x}_t^{\top}\right\| > C_n^{1/2}\right\} + \Pr\left\{\|\tilde{q}_{2t}\| > C_n^{1/2}\right\}\right) \\
\le\ & n\left(E\left\|\tilde{x}_t\tilde{x}_t^{\top}\right\|^r + E\|\tilde{q}_{2t}\|^r\right)/C_n^{r/2} \to 0.
\end{aligned}$$

Next, for any $\eta > 0$, define a partition of $\Theta_n, \Theta_{ni}, i = 1, \ldots, \Gamma_n$ and $\theta_{ni} \in \Theta_{ni}$ as in Lemma 7 of Horowitz (1992). In particular, let the distance between any two points in Θ_{ni} does not exceed $\eta C_n^{-3/2}$

and Γ_n does not exceed $C_0 C_n^{9k/4}$ for some $C_0 > 0$. Then, from $(A17)$ of Horowitz (1992),

$$\begin{aligned}
& \Pr\left\{\sup_{\theta\in\Theta_n}\left\|\frac{1}{n\sigma_n}\sum_{t=1}^{n}\left(g_{nt}^{C_n}(\theta)-Eg_{nt}^{C_n}(\theta)\right)\right\|>\varepsilon\right\} \\
\leq\ & \sum_{i=1}^{\Gamma_n}\Pr\left\{\left\|\frac{1}{n\sigma_n}\sum_{t=1}^{n}\left(g_{nt}^{C_n}(\theta_{ni})-Eg_{nt}^{C_n}(\theta_{ni})\right)\right\|>\varepsilon/2\right\} \qquad (21)\\
& +\sum_{i=1}^{\Gamma_n}\Pr\left\{\frac{1}{n\sigma_n}\sum_{t=1}^{n}\sup_{\theta\in\Theta_n}\left\|g_{nt}^{C_n}(\theta)-Eg_{nt}^{C_n}(\theta)-\left(g_{nt}^{C_n}(\theta_{ni})-Eg_{nt}^{C_n}(\theta_{ni})\right)\right\|>\varepsilon/2\right\}.
\end{aligned}$$

Consider the first term in (21). Let a sequence m_n and σ_n be as in Assumption 3(e). As $g_{nt}^{C_n}$ is strong mixing,it follows from Lemma 10 of De Jong and Woutersen (2004) that

$$\begin{aligned}
& \sum_{i=1}^{\Gamma_n}\Pr\left\{\left\|\frac{1}{n\sigma_n}\sum_{t=1}^{n}\left(g_{nt}^{C_n}(\theta_{ni})-Eg_{nt}^{C_n}(\theta_{ni})\right)\right\|>\varepsilon\right\} \\
\leq\ & \sum_{i=1}^{\Gamma_n}\Pr\left\{\left\|\frac{1}{n\sigma_n}\sum_{t=1}^{n}\left(g_{nt}^{C_n}(\theta_{ni})-E\left(g_{nt}^{C_n}(\theta_{ni})\,|\mathcal{F}_{t-m_n}\right)\right)\right\|>\varepsilon/2\right\} \\
& +\sum_{i=1}^{\Gamma_n}\Pr\left\{\left\|\frac{1}{n\sigma_n}\sum_{t=1}^{n}\left(E\left(g_{nt}^{C_n}(\theta_{ni})\,|\mathcal{F}_{t-m_n}\right)-Eg_{nt}^{C_n}(\theta_{ni})\right)\right\|>\varepsilon/2\right\} \\
\leq\ & \Gamma_n 2m_n\exp\left(-\varepsilon^2 n\sigma_n^2 C_n^{-2}m_n^{-2}\right)+\Gamma_n O\left(\sigma_n^{-1}C_n\alpha_{m_n}\right) \\
=\ & o(1).
\end{aligned}$$

For the second term of (21), note that $g_{nt}^{C_n}(\theta)$ is uniformly continuous. In particular, for a constant $c > 0$

$$\left\|g_{nt}^{C_n}(\theta)-g_{nt}^{C_n}(\theta_{ni})\right\|\leq c\cdot C_n^{3/2}\left\|\theta_{ni}-\theta\right\|\leq c\eta,$$

because $\mathcal{K}''$are bounded. Let $\eta < \varepsilon/8c$, then the second term of (21) is zero.

Next, it follows from the same reasoning for the proof of (18) below that

$$\frac{1}{n\sigma_n}\sum_{t=1}^{n}\left(Eg_{nt}^{C_n}(\theta)-Eg_{nt}(\theta)\right)\to 0,$$

uniformly in θ, provided that $C_n\to\infty$. In the same manner, we can proceed for the parts associated with $2\tilde{x}_t^{\top}\delta\varepsilon_t$ and $R_{nt}(\theta)$.

Proof of (18). Let $\theta\in\Theta_n, \ddot{\psi}_n = (\psi-\psi_0)/\sigma_n$, and $s = \frac{z}{\sigma_n}+\ddot{\psi}_n^{\top}q_2$. We start with the first term in (19).

$$
\begin{aligned}
&-E\left\{\left(\tilde{x}_t^\top \delta\right)^2 \operatorname{sgn}\left(z_t\right) \frac{q_{2t}}{\sigma_n} \mathcal{K}'\left(\frac{z_t+\left(\psi-\psi_0\right)^\top q_{2t}}{\sigma_n}\right)\right\} \qquad (22)\\
= & -\int\left(\sigma_n\left(s-\ddot{\psi}_n^\top q_2\right)+X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}\left(s-\ddot{\psi}_n^\top q_2\right) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s-\ddot{\psi}_n^\top q_2\right)|X_2\right) ds dF_{X_2}(X_2)\\
= & -\int\left(X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}\left(s-\ddot{\psi}_n^\top q_2\right) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s-\ddot{\psi}_n^\top q_2\right)|X_2\right) ds dF_{X_2}(X_2)\\
& -\sigma_n^2 \int\left(s-\ddot{\psi}_n^\top q_2\right)^2 \operatorname{sgn}\left(s-\ddot{\psi}_n^\top q_2\right) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s-\ddot{\psi}_n^\top q_2\right)|X_2\right) ds dF_{X_2}(X_2)\\
& -2\sigma_n \int\left(s-\ddot{\psi}_n^\top q_2\right) X_2^\top \dot{\delta}_2 \operatorname{sgn}\left(s-\ddot{\psi}_n^\top q_2\right) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s-\ddot{\psi}_n^\top q_2\right)|X_2\right) ds dF_{X_2}(X_2)\\
= & \ I_1+I_2+I_3.
\end{aligned}
$$

Due to Assumption 2 and 3, the integrals are bounded uniformly in $\theta \in \Theta_n$. Furthermore, for a constant $C>0$,

$$
\sup_{\theta \in \Theta_n} I_3 \leq \sigma_n\left(C+\left|r_n/\sigma_n\right| C\right) \rightarrow 0,
$$

and

$$
\sup_{\theta \in \Theta_n} I_2 \leq \sigma_n^2\left(C+\left|r_n/\sigma_n\right| C+\left|r_n/\sigma_n\right|^2 C\right) \rightarrow 0.
$$

Let $-I_1=J_1+J_2$, where $J_i's$ are defined below. Let $A_n=\left\{\left|\sigma_n s\right|<\eta,\left|\sigma_n \ddot{\psi}_n^\top q_2\right|<\eta\right\}$ for some $\eta>0$. Then,

$$
\begin{aligned}
J_1 = & \int\left(X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}(s) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s-\ddot{\psi}_n^\top q_2\right)|X_2\right) ds dF_{X_2}(X_2)\\
\leq & \int_{A_n}\left(X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}(s) q_2 \mathcal{K}'(s) f_{z|X_2}\left(-\sigma_n \ddot{\psi}_n^\top q_2|X_2\right) ds dF_{X_2}(X_2)\\
& +\sigma_n \int_{A_n}\left(X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}(s) s q_2 \mathcal{K}'(s) f_{z|X_2}\left(-\sigma_n \ddot{\psi}_n^\top q_2+\zeta|X_2\right) ds dF_{X_2}(X_2)\\
& +\int_{\left|\sigma_n s\right| \geq \eta}\left|\left(X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}(s) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s-\ddot{\psi}_n^\top q_2\right)|X_2\right)\right| ds dF_{X_2}(X_2)\\
& +\int_{\left|\sigma_n \ddot{\psi}_n^\top q_2\right| \geq \eta}\left|\left(X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}(s) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s-\ddot{\psi}_n^\top q_2\right)|X_2\right)\right| ds dF_{X_2}\left(X_2\right),
\end{aligned}
$$

where ζ lies between zero and $\sigma_n s$. Due to Assumption 2 and 3, the integrals are bounded uniformly in $\theta \in \Theta_n$. Therefore, the first term is $o(1)$ uniformly in $\theta \in \Theta_n$ by the bounded convergence theorem since $\int \operatorname{sgn}(s) \mathcal{K}'(s) ds=0$. Similarly, all the other terms are $o(1)$ uniformly in $\theta \in \Theta_n$, since

$\sigma_n \to 0$, $\int_{|\sigma_n s|\geq \eta} \mathrm{sgn}\,(s)\,\mathcal{K}'(s)\,ds \to 0$, and $\Pr\left\{\sup_{\theta\in\Theta_n}\left|\sigma_n \ddot{\psi}_n^{\top} q_2\right| \geq \eta\right\} \to 0$. By the same reasoning,

$$
\begin{aligned}
J_2 &= \int \left(X_2^{\top}\dot{\delta}_2\right)^2 \left(\mathrm{sgn}\left(s-\ddot{\psi}_n^{\top} q_2\right)-\mathrm{sgn}\,(s)\right) q_2 \mathcal{K}'(s)\, f_{z|X_2}\left(\sigma_n\left(s-\ddot{\psi}_n^{\top}q_2\right)|X_2\right) ds dF_{X_2}(X_2) \\
&= 2\int (X_2^{\top}\dot{\delta}_2)^2 (1\{0<s<\ddot{\psi}_n^{\top}q_2\} - 1\{\ddot{\psi}_n^{\top}q_2<s<0\}) q_2\mathcal{K}'(s)\, f_{z|X_2}(\sigma_n(s-\ddot{\psi}_n^{\top}q_2)|X_2) ds dF_{X_2}(X_2) \\
&= 2\int \left(\mathcal{K}\left(\ddot{\psi}_n^{\top}q_2\right)-\mathcal{K}(0)\right)\left(X_2^{\top}\dot{\delta}_2\right)^2 q_2 f_{z|X_2}(0|X_2)\, dF_{X_2}(X_2) + o(1) \\
&= \ddot{Q}_n(\theta) + o(1),
\end{aligned}
$$

uniformly in $\theta \in \Theta_n$.

In the same manner, we can show that the last term in (19) are uniformly negligible. And note that $E(\varepsilon_t|X_t) = 0$ for the second term in (19). This completes the proof of (18).

Now that $T_n^{\psi}(\theta_n) = 0$ by the first order condition of the minimization and that $\theta_n \in \Theta_n$ with probability approaching one, we conclude that $\mathbf{E}T_n^{\psi}(\theta_n) = o_p(1)$ due to (17) and thus $\ddot{Q}_n(\theta_n) = o_p(1)$ due to (18). We next use a standard argument found in Pakes and Pollard (1989) for example. Write

$$
\ddot{Q}_n(\theta_n) = \ddot{Q}_n(\psi_0;\delta_{2n}) + \frac{\partial}{\partial\psi}\ddot{Q}_n\left(\tilde{\psi};\delta_{2n}\right)(\psi_n-\psi_0),
$$

and note that $\ddot{Q}_n(\psi_0;\delta_{2n}) = 0$. Furthermore,

$$
\begin{aligned}
\sigma_n\frac{\partial}{\partial\psi}\ddot{Q}_n\left(\tilde{\psi};\delta_{2n}\right) &= 2\int \mathcal{K}'\left(q_2^{\top}(\psi_n-\psi_0)\right) q_2 q_2^{\top}\left(X_2^{\top}\dot{\delta}_{2n}\right)^2 f_{z|X_2}(0|X_2)\, dF_{X_2}(X_2) \\
&\xrightarrow{p} \mathcal{K}'(0)\,\mathbf{E}\left(q_{2t}q_{2t}^{\top}\left(X_{2t}^{\top}\dot{\delta}_2\right)^2 |z_t=0\right) f(0) \\
&> 0,
\end{aligned}
$$

due to the consistency of θ_n and Assumption 2 (e). Therefore $\sigma_n^{-1}(\psi_n-\psi_0) = o_p(1)$. ■

Lemma 5 *Let* $\theta_n \xrightarrow{P} \theta_0$. *Then,*

$$
\sqrt{n}T_n^s(\theta_0,\sigma_n) \Longrightarrow N(0,4V^s)
$$

$$
Q_n^s(\theta_n,\sigma_n) \xrightarrow{P} 2Q^s
$$

$$
\sup_{\theta\in\Theta}\sqrt{\sigma_n}Q_n^{s\psi}(\theta,\sigma_n) \xrightarrow{P} 0.
$$

Proof of Lemma 5. Assumption 3 (a) and (b) imply that

$$
\int_{-\infty}^{\infty}\left(1\{s>0\}-\mathcal{K}(s)\right)ds = 0, \tag{23}
$$

and that

$$
\int_{-\infty}^{\infty} 1\{s\leq 0\}\,\mathcal{K}(s)\,ds = -\int_{-\infty}^{0} s\mathcal{K}'(s)\,ds = \int_{-\infty}^{\infty} s\cdot\mathrm{sgn}\,(s)\,\mathcal{K}'(s)\,ds/2 = 0,
$$

where the first equality follows from the integration by parts. Then, it follows from the same line of argument in Lemma 2 that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}x_t\tilde{x}_t^\top\left(1\left\{z_t>0\right\}-\mathcal{K}\left(\frac{z_t}{\sigma_n}\right)\right)=o_p\left(1\right). \tag{24}$$

In particular, for $s=z/\sigma_n$ and for ζ between 0 and $\sigma_n s$, we have

$$\begin{aligned}
&\sqrt{n}Ex_t\tilde{x}_t^\top\left(1\left\{z_t>0\right\}-\mathcal{K}\left(\frac{z_t}{\sigma_n}\right)\right)\\
=\;&\sqrt{n}\int_{-\infty}^{\infty}x\tilde{x}^\top\left(1\left\{z>0\right\}-\mathcal{K}\left(\frac{z}{\sigma_n}\right)\right)f_{z|X_2}\left(z|X_2\right)dzdF_{X_2}(X_2)\\
=\;&\sigma_n\sqrt{n}\int_{-\infty}^{\infty}x\tilde{x}^\top\left(1\left\{s>0\right\}-\mathcal{K}\left(s\right)\right)f_{z|X_2}\left(\sigma_n s|X_2\right)dsdF_{X_2}(X_2)\\
=\;&\sigma_n^2\sqrt{n}\int_{-\infty}^{\infty}x\tilde{x}^\top s\left(1\left\{s>0\right\}-\mathcal{K}\left(s\right)\right)f^{(1)}_{z|X_2}\left(\zeta|X_2\right)dsdF_{X_2}(X_2)\\
\to\;&0,
\end{aligned}$$

where the second equality is due to the change of variables and the third equality is due to the mean value theorem and (23) and the convergence follows because the integral is bounded and $\sigma_n^4 n\to 0$. Furthermore, the variance of the term in (24) is $O\left(\sqrt{\sigma_n}\right)$ by the same reasoning as in Lemma 2, which yields (24). Similarly, $n^{-1/2}\sum_{t=1}^n\tilde{x}_t\tilde{x}_t^\top 1\{z_t\le 0\}\mathcal{K}\left(\frac{z_t}{\sigma_n}\right)$ and $n^{-1/2}\sum_{t=1}^n\tilde{x}_t\varepsilon_t(1\{z_t>0\}-\mathcal{K}\left(\frac{z_t}{\sigma_n}\right))$ are both $o_p\left(1\right)$. Therefore,

$$\sqrt{n}T_n^s\left(\theta_0,\sigma_n\right)=\begin{pmatrix}-\frac{2}{\sqrt{n}}\sum_{t=1}^n x_t\varepsilon_t\\ -\frac{2}{\sqrt{n}}\sum_{t=1}^n\tilde{x}_t\varepsilon_t 1\left\{z_t>0\right\}\end{pmatrix}+o_p\left(1\right)\Longrightarrow N\left(0,4V^s\right).$$

Next, note that

$$\frac{1}{n}\sum_{t=1}^{n}\tilde{x}_t\tilde{x}_t^\top\left(\mathcal{K}\left(\frac{q_{1t}+q_{2t}^\top\psi}{\sigma_n}\right)-1\left\{q_{1t}+q_{2t}^\top\psi>0\right\}\right)=o_p\left(1\right),$$

uniformly in $\psi\in\Theta_\psi$, by the Cauchy-Schwarz inequality since, $E\left|\left|\tilde{x}_t\right|\right|^4<\infty$ and, for $s=\left(z+q_2^\top\left(\psi-\psi_0\right)\right)/\sigma_n$,

$$\begin{aligned}
&E\left|1\left\{q_{1t}+q_{2t}^\top\psi>0\right\}-\mathcal{K}\left(\frac{q_{1t}+q_{2t}^\top\psi}{\sigma_n}\right)\right|^2\\
=\;&\int\left|1\left\{z+q_2^\top\left(\psi-\psi_0\right)>0\right\}-\mathcal{K}\left(\frac{z+q_2^\top\left(\psi-\psi_0\right)}{\sigma_n}\right)\right|^2 f_{z|X_2}\left(z|X_2\right)dzdF_{X_2}\left(X_2\right)\\
=\;&\sigma_n\int\left|1\left\{s>0\right\}-\mathcal{K}\left(s\right)\right|^2 f\left(\sigma_n s+q_2^\top\left(\psi-\psi_0\right)|q_2\right)dzdF\left(q_2\right)\\
=\;&O\left(\sigma_n\right).
\end{aligned}$$

Then, by Lemma 1,

$$\frac{1}{n}\sum_{t=1}^{n}\tilde{x}_t\tilde{x}_t^\top 1\left\{q_{1t}+q_{2t}^\top\psi>0\right\}\xrightarrow{P}E\tilde{x}_t\tilde{x}_t^\top 1\left\{q_{1t}+q_{2t}^\top\psi>0\right\},$$

uniformly in $\psi\in\Theta_\psi$. Therefore, the convergence of $Q_n^s\left(\theta_n,\sigma_n\right)$ follows.

For $\sqrt{\sigma_n}Q_n^{s\psi}(\tilde{\theta},\sigma_n)=o_p\left(1\right),$ we show $\sup_{\theta\in\Theta}Q_n^{s\psi}(\theta,\sigma_n)=O_p\left(1\right),$ for which it is sufficient that, by Assumption 2 and 3, for any $\theta\in\Theta$ and a constant C,

$$\begin{aligned}
&\int\left|x\tilde{x}^\top\delta q_2^\top\mathcal{K}'\left(\frac{z}{\sigma_n}+q_2^\top\ddot{\psi}_n\right)\right|f_{z|X_2}\left(z|X_2\right)\frac{dz}{\sigma_n}dF_{X_2}(X_2)\\
=&\int\left|x\tilde{x}^\top\delta q_2^\top\mathcal{K}'\left(s\right)\right|f_{z|X_2}\left(\left(s-q_2^\top\ddot{\psi}_n\right)\sigma_n|X_2\right)dsdF_{X_2}(X_2)\\
\leq\ & C,
\end{aligned}$$

where $\ddot{\psi}_n=\left(\psi-\psi_0\right)/\sigma_n$ and $s=z/\sigma_n+q_2^\top\ddot{\psi}_n$. ∎

LEMMA 6. *Let $\{e_n\}$ be a sequence such that $e_n\to0$ and that $\Pr\left\{|\theta_n-\theta_0|>\sigma_ne_n\right\}\to0$. Then,*

$$\sigma_nQ_n^\psi\left(\theta,\sigma_n\right)\xrightarrow{P}2Q^\psi,$$

uniformly in $\theta\in\Theta_n^\sigma=\left\{\theta:|\theta-\theta_0|\leq\sigma_ne_n\right\}.$

PROOF OF LEMMA 6: Recall that

$$Q_n^\psi\left(\tilde{\theta},\sigma_n\right)=\frac{1}{n}\sum_{t=1}^{n}\left\{-\left(\left(\tilde{x}_t^\top\delta_0\right)^2\operatorname{sgn}\left(z_t\right)+2\tilde{x}_t^\top\delta_0\varepsilon_t\right)+R_{nt}(\tilde{\theta})\right\}\frac{q_{2t}q_{2t}^\top}{\sigma_n^2}\mathcal{K}''\left(\frac{z_t+q_{2t}^\top\left(\tilde{\psi}-\psi_0\right)}{\sigma_n}\right),$$

where R_{nt} is defined in (16). Since $\mathcal{K}''$ is bounded and $\left(\theta-\theta_0\right)/\sigma_n\to0$ for $\theta\in\Theta_n^\sigma$, it is straightforward that

$$\sup_{\theta\in\Theta_n^\sigma}E\left|\sigma_n\frac{1}{n}\sum_{t=1}^{n}R_{nt}\left(\theta\right)\frac{q_{2t}q_{2t}^\top}{\sigma_n^2}\mathcal{K}''\left(\frac{z_t+q_{2t}^\top\left(\psi-\psi_0\right)}{\sigma_n}\right)\right|=o\left(1\right).$$

As $E\left[\varepsilon_t|X_t\right]=0$ and δ and $\mathcal{K}''$ are bounded, it follows from the same reasoning as Lemma 2 that

$$\sup_{\theta\in\Theta_n^\sigma}\sigma_n\frac{1}{n}\sum_{t=1}^{n}\tilde{x}_t^\top\delta_0\varepsilon_t\frac{q_{2t}q_{2t}^\top}{\sigma_n^2}\mathcal{K}''\left(\frac{z_t+q_{2t}^\top\left(\psi-\psi_0\right)}{\sigma_n}\right)=o_p\left(1\right).$$

Let $\theta\in\Theta_n$ and $\ddot{\psi}_n=\left(\psi-\psi_0\right)/\sigma_n$ and $s=z/\sigma_n+q_2^\top\ddot{\psi}_n$. Define

$$\chi_{nt}^\psi\left(\theta\right)=-\sigma_n\left(\tilde{x}_t^\top\delta_0\right)^2\operatorname{sgn}\left(z_t\right)\frac{q_{2t}q_{2t}^\top}{\sigma_n^2}\mathcal{K}''\left(\frac{z_t+q_{2t}^\top\left(\psi-\psi_0\right)}{\sigma_n}\right),$$

and note that $E\chi_{nt}^\psi$ is the same as the one in (22) except that it contains $q_{2t}q_{2t}^\top$ and $\mathcal{K}''$ instead of q_{2t} and $\mathcal{K}'$. But, as $\mathcal{K}''$ and $E\left\|\left(\tilde{x}_t^\top\delta_0\right)^2q_{2t}q_{2t}^\top\right\|$ are bounded, it follows from the same reasoning for the convergence of (22) that

$$\sup_{\theta\in\Theta_n^\sigma}\left|E\chi_{nt}^\psi\left(\theta\right)-2\mathcal{K}'\left(0\right)\int\left(X_2^\top\dot{\delta}_{20}\right)^2q_2q_2^\top f_{z|X_2}\left(0|X_2\right)dF_{X_2}(X_2)\right|=o\left(1\right).$$

Note that $-\int_{-\infty}^{\infty} \operatorname{sgn}(s)\,\mathcal{K}''(s)\,ds = 2\mathcal{K}'(0)$. Furthermore,

$$\sup_{\theta\in\Theta_n^{\sigma}} \left| \frac{1}{n}\sum_{t=1}^{n} \chi_{nt}^{\psi}(\theta) - E\chi_{nt}^{\psi}(\theta) \right| \xrightarrow{P} 0,$$

by the same reasoning as for (17), which finishes the proof. ■

Proof of Corollary 3. The proof is almost identical to that of Theorem 2. We point out some differences which result in different formula in the asymptotic variance of θ_n^{+}. Recall (9) and note that

$$\frac{\partial e_t(\theta)}{\partial\theta^{\top}} = -\left(x_t^{\top}, -\tilde{x}_t^{\top}\mathcal{K}\left(\frac{z_t + q_{2t}^{\top}(\psi-\psi_0)}{\sigma_n}\right), -\tilde{x}_t^{\top}\delta\frac{q_{2t}^{\top}}{\sigma_n}\mathcal{K}'\left(\frac{z_t + q_{2t}^{\top}(\psi-\psi_0)}{\sigma_n}\right)\right)$$

and

$$T_n^{+}(\theta_0) = \frac{\partial S_n^{+}(\theta_0)}{\partial\theta} = \frac{2}{n}\sum_{t=1}^{n}\frac{\partial e_t(\theta)}{\partial\theta}\left(\varepsilon_t + \tilde{x}_t^{\top}\delta\left(1\{z_t>0\} - \mathcal{K}(z_t/\sigma_n)\right)\right).$$

We compare T_n^{+} to T_n. First, we observe that $\partial S_n^{+}(\theta_0)/\partial\beta$ is identical to $\partial S_n(\theta_0)/\partial\beta$. Second,

$$\begin{aligned} &\partial S_n^{+}(\theta_0)/\partial\delta \\ =\ & \frac{2}{n}\sum_{t=1}^{n}\mathcal{K}\left(\frac{z_t}{\sigma_n}\right)\tilde{x}_t\tilde{x}_t^{\top}\varepsilon_t + \frac{2}{n}\sum_{t=1}^{n}\mathcal{K}\left(\frac{z_t}{\sigma_n}\right)\tilde{x}_t\tilde{x}_t^{\top}\delta\left(1\{z_t>0\} - \mathcal{K}(z_t/\sigma_n)\right), \end{aligned}$$

which yields the same limit as $\partial S_n(\theta_0)/\partial\delta$ by the same argument as Lemma 5.[5]

However, we see that we need to replace $\operatorname{sgn}(z_t)$ in

$$T_n^{\psi} = -\frac{1}{n}\sum_{t=1}^{n}\left(\left(\tilde{x}_t^{\top}\delta_0\right)^2\operatorname{sgn}(z_t) + 2\tilde{x}_t^{\top}\delta_0\varepsilon_t\right)\frac{q_{2t}}{\sigma_n}\mathcal{K}'\left(\frac{z_t}{\sigma_n}\right),$$

with $2\left(1\{z_t>0\} - \mathcal{K}(z_t/\sigma_n)\right)$. This difference does not disappear due to σ_n in the denominator but the same line of proof as in Lemma 2 yields the limit variance $V^{\psi+}$. In particular, note that

$$\begin{aligned} \int_{-\infty}^{\infty}\left(1\{s>0\} - \mathcal{K}(s)\right)\mathcal{K}'(s)\,ds &= -\int_{-\infty}^{0}\mathcal{K}(s)\,\mathcal{K}'(s)\,ds + \int_{0}^{\infty}\left(1-\mathcal{K}(s)\right)\mathcal{K}'(s)\,ds \\ &= -\int_{-\infty}^{0}\mathcal{K}(s)\,\mathcal{K}'(s)\,ds + \int_{0}^{\infty}\mathcal{K}(-s)\,\mathcal{K}'(-s)\,ds \\ &= 0, \end{aligned}$$

due to the symmetry of $\mathcal{K}'$.

Next, write

$$Q_n^{+}\left(\theta_n^{+};\sigma_n\right) = \frac{2}{n}\sum_{t=1}^{n}\frac{\partial e_t^{+}}{\partial\theta}\frac{\partial e_t^{+}}{\partial\theta^{\top}} + \frac{2}{n}\sum_{t=1}^{n}e_t^{+}\frac{\partial^2 e_t^{+}}{\partial\theta\partial\theta^{\top}},$$

[5] Note that $\partial S_n(\theta_0)/\partial\delta$ involves $1\{z_t\le 0\}\,\mathcal{K}\{z_t/\sigma_n\}$, which requires $h\ge 2$ to be negligible.

where $e_t^+ = e_t\left(\theta_n^+\right)$. Following Lemma 5 and 6, it is straightforward to see that the part associated with (β, δ) is the same as before. For the part associated with ψ, we observe that, following Lemma 6,

$$
\begin{aligned}
-\frac{2\sigma_n}{n}\sum_{t=1}^{n} e_t^+ \frac{\partial^2 e_t^+}{\partial\psi\partial\psi^\top} &= \frac{2\sigma_n}{n}\sum_{t=1}^{n} e_t^+ \tilde{x}_t^\top \delta_n^+ \frac{q_{2t}q_{2t}^\top}{\sigma_n^2}\mathcal{K}''\left(\frac{z_t + q_{2t}^\top\left(\psi_n^+ - \psi_0\right)}{\sigma_n}\right) \\
&\xrightarrow{P} \int_{-\infty}^{\infty} \mathcal{K}''(s)\left(1\{s>0\} - \mathcal{K}(s)\right) ds \cdot E\left[(X_{2t}^\top\dot{\delta}_2)^2 q_{2t}q_{2t}^\top | z_t = 0\right] f(0).
\end{aligned}
$$

Similarly,

$$
\frac{2\sigma_n}{n}\sum_{t=1}^{n} \frac{\partial e_t^+}{\partial\psi}\frac{\partial e_t^+}{\partial\psi^\top} \xrightarrow{P} \int_{-\infty}^{\infty} \mathcal{K}'(s)^2\, ds \cdot E\left[(X_{2t}^\top\dot{\delta}_2)^2 q_{2t}q_{2t}^\top | z_t = 0\right] f(0).
$$

But, the integration by parts yields

$$
\int_{-\infty}^{\infty} \mathcal{K}'(s)^2\, ds - \int_{-\infty}^{\infty} \mathcal{K}''(s)\left(1\{s>0\} - \mathcal{K}(s)\right) ds = -\int_{-\infty}^{\infty} \mathcal{K}''(s)\, 1\{s>0\}\, ds = \mathcal{K}'(0),
$$

which completes the proof. ■

Now, we prove the consistency of the variance estimators.

Proof of Theorem 4. We first examine the convergence of $\hat{V}^s$. Let

$$
\tau_t^s = \begin{pmatrix} -x_t \\ -\tilde{x}_t 1\{z_t > 0\} \end{pmatrix}, \Gamma_j = E\tau_t^s \tau_{t-j}^{s\top}\varepsilon_t\varepsilon_{t-j}, \text{ and } V_n^s = \sum_{j=-n+1}^{n-1}\Gamma_j.
$$

And also define $\tilde{V}_n^s = \sum_{j=-n+1}^{n-1} w\left(\frac{j}{l_n}\right)\tilde{\Gamma}_j$ and $\vec{V}_n^s = \sum_{j=-n+1}^{n-1} w\left(\frac{j}{l_n}\right)\vec{\Gamma}_j$, where

$$
\begin{aligned}
\tilde{\Gamma}_j &= \begin{cases} \frac{1}{n}\sum_{t=j+1}^{n} \tau_t^s \tau_{t-j}^{s\top}\varepsilon_t\varepsilon_{t-j} & for\ j \geq 0, \\ \frac{1}{n}\sum_{t=-j+1}^{n} \tau_{t+j}^s \tau_t^{s\top}\varepsilon_{t+j}\varepsilon_t & for\ j < 0, \end{cases} \\
\vec{\Gamma}_j &= \begin{cases} \frac{1}{n}\sum_{t=j+1}^{n} \tau_{n,t}^s(\theta_0)\,\tau_{n,t-j}^s(\theta_0)^\top e_t(\theta_0)\, e_{t-j}(\theta_0) & for\ j \geq 0, \\ \frac{1}{n}\sum_{t=-j+1}^{n} \tau_{n,t+j}^s(\theta_0)\,\tau_{n,t}^s(\theta_0)^\top e_{t+j}(\theta_0)\, e_t(\theta_0) & for\ j < 0, \end{cases}
\end{aligned}
$$

for $\tau_{n,t}^s$ and e_t defined in (12) and (9). It follows from Andrews (1991 : Proposition 1 (c) and Theorem 1 (c)) and a correction by Jansson (2002) that $\tilde{V}_n^s - V_n^s \xrightarrow{P} 0$ and $\hat{V}_n^s - \vec{V}_n^s \xrightarrow{P} 0$. Then, it remains to show that $\tilde{V}_n^s - \vec{V}_n^s \xrightarrow{P} 0$. Since

$$
\tilde{V}_n^s - \vec{V}_n^s = \sum_{j=-n+1}^{n-1} w\left(\frac{j}{l_n}\right)\left(\tilde{\Gamma}_j - \vec{\Gamma}_j\right),
$$

and $\sum_{j=-n+1}^{n-1} |w(j/l_n)| = O(l_n)$, it suffices to show that $\sup_j \mathbf{E}\left|\tilde{\Gamma}_j - \vec{\Gamma}_j\right| = O(\sigma_n)$ as $l_n\sigma_n \to 0$.

Assume $j \geq 0$ and let $1_t = 1\{z_t > 0\}$ and $\mathcal{K}_t = \mathcal{K}(z_t/\sigma_n)$. As $e_t(\theta_0) = \varepsilon_t - \tilde{x}_t^\top \delta_0 (\mathcal{K}_t - 1_t)$, we write

$$
\begin{aligned}
\tilde{\Gamma}_j - \vec{\Gamma}_j &= \frac{1}{n}\sum_{t=j+1}^{n} \varepsilon_t \varepsilon_{t-j} \begin{pmatrix} 0 & x_t \tilde{x}_{t-j}^\top (1_{t-j} - \mathcal{K}_{t-j}) \\ \tilde{x}_t x_{t-j}^\top (1_t - \mathcal{K}_t) & \tilde{x}_t \tilde{x}_{t-j}^\top (1_t 1_{t-j} - \mathcal{K}_t \mathcal{K}_{t-j}) \end{pmatrix} \\
&\quad + \frac{1}{n}\sum_{t=j+1}^{n} \tau_{n,t}^s(\theta_0)\, \tau_{n,t-j}^s(\theta_0)^\top \left(\tilde{x}_t^\top \delta_0 \tilde{x}_{t-j}^\top \delta_0 (\mathcal{K}_t - 1_t)(\mathcal{K}_{t-j} - 1_{t-j})\right) \\
&\quad + \frac{1}{n}\sum_{t=j+1}^{n} \tau_{n,t}^s(\theta_0)\, \tau_{n,t-j}^s(\theta_0)^\top \left(\tilde{x}_t^\top \delta_0 (\mathcal{K}_t - 1_t)\varepsilon_{t-j} + \varepsilon_t \tilde{x}_{t-j}^\top \delta_0 (\mathcal{K}_{t-j} - 1_{t-j})\right).
\end{aligned}
$$

We show that $\sup_j \mathbf{E}\left|\tilde{x}_t \tilde{x}_{t-j}^\top \varepsilon_t \varepsilon_{t-j} (1_t 1_{t-j} - \mathcal{K}_t \mathcal{K}_{t-j})\right| = O(\sigma_n)$ and the other terms can be studied in the same way. As $\mathcal{K}$ is bounded and $\mathcal{K}(x) = 1 - \mathcal{K}(-x)$,

$$
\begin{aligned}
\left|1 - \mathcal{K}\left(\frac{x}{\sigma_n}\right)\mathcal{K}\left(\frac{y}{\sigma_n}\right)\right| &= \left|1 - \mathcal{K}\left(\frac{x}{\sigma_n}\right)\left(1 - \mathcal{K}\left(\frac{-y}{\sigma_n}\right)\right)\right| \\
&\leq \left|1 - \mathcal{K}\left(\frac{x}{\sigma_n}\right)\right| + \left|\mathcal{K}\left(\frac{x}{\sigma_n}\right)\mathcal{K}\left(\frac{-y}{\sigma_n}\right)\right| \\
&\leq \left|1 - \mathcal{K}\left(\frac{x}{\sigma_n}\right)\right| + C\left|\mathcal{K}\left(\frac{-y}{\sigma_n}\right)\right|.
\end{aligned}
$$

for some constant $C > 1$. Thus,

$$
\begin{aligned}
&\left|1\{x > 0, y > 0\} - \mathcal{K}\left(\frac{x}{\sigma_n}\right)\mathcal{K}\left(\frac{y}{\sigma_n}\right)\right| \\
= &\left|1 - \mathcal{K}\left(\frac{x}{\sigma_n}\right)\mathcal{K}\left(\frac{y}{\sigma_n}\right)\right| 1\{x > 0, y > 0\} + \left|\mathcal{K}\left(\frac{x}{\sigma_n}\right)\mathcal{K}\left(\frac{y}{\sigma_n}\right)\right| 1\{x < 0 \text{ or } y < 0\} \\
\leq &\left|1 - \mathcal{K}\left(\frac{x}{\sigma_n}\right)\right| 1\{x > 0\} + C\left|\mathcal{K}\left(\frac{-y}{\sigma_n}\right)\right| 1\{y > 0\} \\
&+ C\left|\mathcal{K}\left(\frac{x}{\sigma_n}\right)\right| 1\{x < 0\} + C\left|\mathcal{K}\left(\frac{y}{\sigma_n}\right)\right| 1\{y < 0\} \\
\leq &\; C\left|1\{x > 0\} - \mathcal{K}\left(\frac{x}{\sigma_n}\right)\right| + C\left|1\{y > 0\} - \mathcal{K}\left(\frac{y}{\sigma_n}\right)\right|.
\end{aligned}
$$

Then, for any w_t such that $E\left|w_t\right|^4<\infty$,

$$
\begin{aligned}
& \sup_j \left| n^{-1} \sum_{t=j+1}^{n} w_t w_{t-j} \left(1_t 1_{t-j} - \mathcal{K}_t \mathcal{K}_{t-j}\right) \right| \\
\leq \; & \sup_j \frac{C}{n} \sum_{t=j+1}^{n} \left|w_t w_{t-j}\right| \left(\left|1_t - \mathcal{K}_t\right| + \left|1_{t-j} - \mathcal{K}_{t-j}\right|\right) \\
\leq \; & \sup_j \sqrt{\frac{C}{n} \sum_{t=j+1}^{n} \left|w_t w_{t-j}\right|^2} \left(\sqrt{\frac{C}{n} \sum_{t=j+1}^{n} \left|1_t - \mathcal{K}_t\right|^2} + \sqrt{\frac{C}{n} \sum_{t=j+1}^{n} \left|1_{t-j} - \mathcal{K}_{t-j}\right|^2} \right) \\
\leq \; & 2C \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left|w_t\right|^4} \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left|1_t - \mathcal{K}_t\right|^2} = O_p\left(\sigma_n\right),
\end{aligned}
$$

since

$$
\mathbf{E}\left|1_t - \mathcal{K}_t\right|^2 = \int \left| 1\{z>0\} - \mathcal{K}\left(\frac{z}{\sigma_n}\right) \right|^2 f(z)\, dz = \sigma_n \int \left|1\{s>0\} - \mathcal{K}(s)\right|^2 f(s\sigma_n)\, ds \to 0,
$$

by the dominated convergence theorem. The case with $j<0$ can be done similarly.

Next, from (10) and (15),

$$
\tau_{n,t}^{\psi}(\theta_n) = -\left(\left(\tilde{x}_t^{\top}\delta_0\right)^2 \operatorname{sgn}(z_t) + 2\tilde{x}_t^{\top}\delta_0\varepsilon_t + R_{nt}(\theta_n) \right) \frac{q_{2t}}{\sigma_n} \mathcal{K}' \left(\frac{z_t + q_{2t}^{\top}(\psi_n - \psi_0)}{\sigma_n} \right).
$$

The convergence of

$$
\hat{V}^{\psi} = \frac{\sigma_n}{n} \sum_{t=1}^{n} \tau_{n,t}^{\psi}(\theta_n)\, \tau_{n,t}^{\psi}(\theta_n)^{\top}
$$

follows from the same reasoning for Lemma 6 given appropriate moment conditions, which are assumed in Assumption 2. ■

Proof of Theorem 5. In this proof, every quantity with superscript $*$ stands for the bootstrap quantity conditional on a realized data. For example, P^* stands for the empirical distribution of a give sample.

The proof follows the arguments of Horowitz $(2001, 2002)$. First, Lemma 6 of Horowitz (2002) corresponds to the bootstrap version of Lemma 1. Therefore, the same argument as Theorem 1 yields:

$$
\text{for any } \delta>0,\; P^*\left(\left\|\theta_n^* - \theta_n\right\| > \delta\right) \to 0
$$

almost surely. It follows from Theorem 2.2 of Horowitz (2001) and Lemma 2 and 3 that

$$
T_n^*(\theta_0) \Rightarrow N(0, V),
$$

almost surely. The uniform convergence of $D_n Q_n^*(\theta) D_n$ and that of the variance estimates $\hat{V}^{\psi *}$, $\hat{V}^{\psi + *}$, and $\hat{V}^{s*}$ follow from uniform convergences $(a), (b)$ and (f) in Lemma 8 of Horowitz (2002). While we have the slope parameters that lies outside $\mathcal{K}$, which are not directly covered by this lemma, the model is linear in these parameters and therefore the uniformity in these parameters is trivially satisfied given the compactness of the parameter space and boundedness of the variables.

Note that we do not need all the results in Lemma 8 of Horowitz (2002) since we do not consider higher-order expansions.[6] Thus, we do not need to impose the existence of higher than the second order derivatives of $\mathcal{K}$. Furthermore, we need to multiply σ_n^{-1} (h_n^{-1} for the notation there) to the right hand side of (b) of Lemma 8 to reflect the differences in the leading terms of second and higher order derivatives of the objective functions. That is, in case of the smoothed maximum score estimation, the expectation of the derivatives of the objective function involves a term $1 - 2F(h_n s)$, which goes to zero as $n \to \infty$ due to the median restriction, which results in the change of convergence rates in the order of h_n. Therefore, one term expansion of $F(\cdot)$ around zero yields

$$1 - 2F(h_n s) = \frac{\partial F(\xi)}{\partial \xi} h_n s.$$

On the other hand, we need to multiply h_n in our case to get the convergence as in Horowitz. This is the reason why we get different convergence rates for our estimator from that for the smoothed maximum score estimator. ■

Seo and Linton (2006)

References

Abrevaya, J., Huang, J., 2005. On the bootstrap of the maximum score estimator. Econometrica 73 (4), 1175–1204.

Andrews, D. W. K., 1987. Consistency in nonlinear econometric models: a generic uniform law of large numbers. Econometrica 55 (6), 1465–1471.

Andrews, D. W. K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Caner, M., 2002. A note on least absolute deviation estimation of a threshold model. Econometric Theory 18 (3), 800–814.

Carlstein, E., 1986. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. The Annals of Statistics 14 (3), 1171–1179.

[6]Not all the conditions on the kernel imposed in Assumption 6 of Horowitz (2002) are required for the proof of the consistency of the bootstrap. For example, he notes that the condition (c) in the assumption is only required for higher-order expansion and is irrelevant for the consistency of the bootstrap.

Chan, K. S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. The Annals of Statistics 21, 520–533.

Chan, K. S., Tsay, R. S., 1998. Limiting properties of the least squares estimator of a continuous threshold autoregressive model. Biometrika 85 (2), 413–426.

Davidson, j., 1994. Stochastic limit theory. Oxford University Press.

De Jong, R., Woutersen, T., 2004. Dynamic time series binary choice, manuscript.

de Jong, R. M., 1995. Laws of large numbers for dependent heterogeneous processes. Econometric Theory 11 (2), 347–358.

de Jong, R. M., 1997. Central limit theorems for dependent heterogeneous random variables. Econometric Theory 13 (3), 353–367.

Delgado, M. A., Hidalgo, J., 2000. Nonparametric inference on structural breaks. Journal of Econometrics 96 (1), 113–144.

Dufour, J.-M., 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. Econometrica 65 (6), 1365–1387.

Fan, Y., Linton, O., 2003. Some higher-order theory for a consistent non-parametric model specification test. Journal of Statistical Planning and Inference 109 (1-2), 125–154, c. R. Rao 80th birthday felicitation volume, Part III.

Gonzalo, J., Wolf, M., 2005. Subsampling inference in threshold autoregressive models. Journal of Econometrics.

Hahn, J., Todd, P., Van der Klaauw, W., 2001. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69 (1), 201–09.

Hall, P., 1992. The bootstrap and Edgeworth expansion. Springer Series in Statistics. Springer-Verlag, New York.

Hansen, B., 2000. Sample splitting and threshold estimation. Econometrica 68, 575–603.

Horowitz, J. L., 1992. A smoothed maximum score estimator for the binary response model. Econometrica 60 (3), 505–531.

Horowitz, J. L., 1998. Bootstrap methods for median regression models. Econometrica 66 (6), 1327–1351.

Horowitz, J. L., 2001. The Bootstrap. Vol. 5 of Handbook of Econometrics. Elsevier, Ch. 52, pp. 3159–3228.

Horowitz, J. L., 2002. Bootstrap critical values for tests based on the smooth maximum score estimator. Journal of Econometrics 111 (2), 141–167.

Jansson, M., 2002. Consistent covariance matrix estimation for linear processes. Econometric Theory 18 (6), 1449–1459.

Kim, C.-J., Nelson, C. R., 1999. State-Space Models with Regime-Switching: Classical and Gibbs-Sampling Approaches with Applications. MIT Press.

Koenker, R., 1997. l_1 computation: an interior monologue. In: L_1-statistical procedures and related topics (Neuchâtel, 1997). Vol. 31 of IMS Lecture Notes Monogr. Ser. Inst. Math. Statist., Hayward, CA, pp. 15–32.

Künsch, H. R., 1989. The jackknife and the bootstrap for general stationary observations. The Annals of Statistics 17, 1217–1241.

Manski, C. F., 1975. Maximum score estimation of the stochastic utility model of choice. Journal of Econometrics 3 (3), 205–228.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57 (5), 1027–1057.

Pollard, D., 1993. The asymptotics of a binary choice model. Tech. rep.

Potter, S. M., 1995. A nonlinear approach to US GNP. Journal of Applied Econometrics 10 (2), 109–25.

Seo, M. H., Linton, O., 2006. A smoothed least squares estimator for threshold regression models, london School of Economics.

Shao, J., Tu, D. S., 1995. The jackknife and bootstrap. Springer Series in Statistics. Springer-Verlag, New York.

Tong, H., 1983. Threshold models in nonlinear time series ananlysis. Lecture Notes in Statistics. Springer, Berlin.

Tong, H., 1990. Nonlinear time series. Vol. 6 of Oxford Statistical Science Series. The Clarendon Press Oxford University Press, New York.

van Dijk, D., Terasvirta, T., Franses, P., 2000. Smooth transition autoregressive models - a survey of recent developments. Tech. rep.