A Smoothed Least Squares Estimator For Threshold Regression Models

M. Seoy and O. Lintonz

London School of Economics

Contents:

A Smoothed Least Squares Estimator For Threshold Regression Models[*]

M. Seo[†]and O. Linton[‡]

London School of Economics

July 20, 2005

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 threshold parameters themselves. We compare our confidence intervals with those of Hansen (2000) in a simulation study and show that our methods outperform his for large values of the threshold. We also include an application to cross-country growth regressions.

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

JEL Classification Number: C12, C13, C14.

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, Terasvirta, and Franses (2000). The discontinuous threshold effect has found applications in macro and in cross-section 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 het-

eroskedasticity 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. It may be of interest because it allows different threshold values for subsamples divided by a discrete variable like gender. Furthermore, we can make decision on 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). The consistency of the HAC estimation in Andrews (1991) is extended to allow for the discontinuity in the threshold model.

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 we can work with pivotal test statistics and hence expect to obtain asymptotic refinements.

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.

---

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

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 some extensions. Numerical results are presented in Section 6 and an application to a cross-section growth model in Section 7. Section 8 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 $||A|| = \text{tr}(A^\top A)^{1/2}$.

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 and $E(\varepsilon_t|X_t) = 0$. 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

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

3

where $\theta = \left(\beta^\top, \delta^\top, \psi^\top\right)^\top \in \Theta \subset \mathbb{R}^k$ and $1_t(\psi) = 1\left\{q_{1t} + q_{2t}^\top \psi > 0\right\}$. The solution is obtained by profiled least squares, see Hansen (2000). Let θ_n^{LS} 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 worthwhile noting that this function is analogous to a cumulative distribution function rather than a density function. Then, define a smoothed objective function

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

and a smoothed least squares (SLS) estimator

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

We assume that the parameter space Θ 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. 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 given ψ, this is

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

where $\mathcal{K}_t(\psi) = \mathcal{K}(q_{1t} + q_{2t}^\top \psi)/\sigma_n)$, 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

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

and the 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 estimator.

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* $\xi > \alpha_0$, $E\left\|X_t X_t^\top\right\|^\xi < \infty$ *and* $E\left\|X_t \varepsilon_t\right\|^\xi < \infty$.
(c) $E\left(\varepsilon_t | X_t\right) = 0$ *a.s.*
(d) *For almost every* X_{2t}, *the probability distribution of* X_{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 by Andrews (1987) is applied for the following development of the consistency proof, supported by the strong law of large numbers of de Jong (1995, Theorem 4). For the asymptotic normality, we need to strengthen the mixing condition. The following theorem establishes the strong consistency of the SLS estimator.

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 $S_n(\theta; \sigma_n)$ is twice differentiable with respect to θ, we 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 to 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 $\left(z_t, X_{2t}^\top\right) \dot{\delta} = \tilde{x}_t^\top \delta$. As above, we denote the first element of $\dot{\delta}$ as $\dot{\delta}_1$ and the others as $\dot{\delta}_2$. For example, if $x_t = \tilde{x}_t = q_t$, whose dimension is k, then $\mathcal{S} = I_k$,

$$
\mathcal{T} = \begin{pmatrix} 1 & \psi' \\ 0 & I_{k-1} \end{pmatrix} \text{ and } \dot{\delta} = \begin{pmatrix} \delta_1 \\ -\delta_1 \psi + \delta_2 \end{pmatrix}.
$$

We then have $\delta^\top \widetilde{x}_t 1 \left\{ q_t^\top \psi > 0 \right\} = (z_t \dot{\delta}_1 + X_{2t}^\top \dot{\delta}_2) 1(z_t > 0) = z_t \dot{\delta}_1 1(z_t > 0) + X_{2t}^\top \dot{\delta}_2 1(z_t > 0)$, 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.

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.

Define

$$\overline{E}\left(\varepsilon_t^2|X_{2t}\right) = \lim_{z \to 0^+} E\left(\varepsilon_t^2|z_t = z, X_{2t}\right) \int_{s>0} \mathcal{K}'(s)^2 \, ds + \lim_{z \to 0^-} E\left(\varepsilon_t^2|z_t = z, X_{2t}\right) \int_{s<0} \mathcal{K}'(s)^2 \, ds, \quad (6)$$

and

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

If we impose a stronger assumption that $\{\varepsilon_t\}$ is a martingale difference sequence, then all the autocovariances drop out of V^s. In contrast, the threshold estimates 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). If $\mathcal{K}'$ is symmetric around zero,

$$\overline{E}\left(\varepsilon_t^2|X_{2t}\right) = \|\mathcal{K}'\|_2^2 \left(\lim_{z \to 0^+} E\left(\varepsilon_t^2|z_t = z, X_{2t}\right) + \lim_{z \to 0^-} E\left(\varepsilon_t^2|z_t = z, X_{2t}\right) \right)/2.$$

If additionally, $E\left(\varepsilon_t^2|z_t, X_{2t}\right)$ is continuous at $z_t = 0$, this expression simplifies further to $\|\mathcal{K}'\|_2^2 \cdot E\left(\varepsilon_t^2|z_t = 0, X_{2t}\right)$.

The assumptions we need are collected in the following.

6

Assumption 2 (*a*) *For all vectors* ξ *such that* $|\xi| = 1$ *and* $r > 4$, $E\left|X_{2t}^\top \dot{\delta}_2 \varepsilon_t q_{2t}^\top \xi\right|^r < \infty$ *and* $E\left|(X_{2t}^\top \dot{\delta}_2)^2 q_{2t}^\top \xi\right|^r < \infty$, *or* ($a^+$) *the condition* $E\left|(X_{2t}^\top \dot{\delta}_2)^2 q_{2t}^\top \xi\right|^r < \infty$ *in* (*a*) *is replaced by* $E\left|X_{2t}^\top \dot{\delta}_2 q_{2t}^\top \xi\right|^r < \infty$. (*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 some integer* $h \geq 2$ *and each integer* i *such that* $1 \leq i \leq h-1$, *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*) *and 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}), *and the conditional expectation* $E(\varepsilon_{t-m} | X_t, X_{t-m}) < M$ *for almost all* (X_t, X_{t-m}). (*e*) θ_0 *is an interior point of a compact parameter space* Θ. (*f*) *And* V^s, V^ψ, Q^s, *and* Q^ψ *are finite and positive definite.*

In case of Hansen's model where $z_t = q_{1t} + \psi$, V^ψ and Q^ψ are defined without $q_{2t} q_{2t}^\top$ and the condition (*a*) is simplified to $E|X_{2t}^\top \dot{\delta}_2 \varepsilon_t|^r < \infty$ and $E|(X_{2t}^\top \dot{\delta}_2)^2|^r < \infty$, or to $E|X_{2t}^\top \dot{\delta}_2|^r < \infty$ in (a^+). The moment conditions are to ensure the consistency of the variance covariance matrix estimators that are introduced later. The condition (a^+) is analogous to de Jong and Woutersen (2004) and is slightly stronger than that of Chan (1993) or Hansen (2000), which requires a finite fourth moment.

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*) - (*f*) are common in the smoothed estimation as in Horowitz (1992), only (*d*) being an analogue of an *iid* 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 (*f*) is standard, the positivity of Q^ψ excludes a continuous threshold model, 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 $E(\varepsilon_t^2 | z_t)$ 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 followings assumptions regarding the smoothing function $\mathcal{K}$ and the bandwidth parameter σ_n.

Assumption 3 (a) $\mathcal{K}$ *is twice differentiable everywhere,* $|\mathcal{K}'(\cdot)|$ *and* $|\mathcal{K}''(\cdot)|$ *are uniformly bounded, and each of the following integrals is finite:* $\int |\mathcal{K}'|^4 , \int |\mathcal{K}''|^2 , \int |v^2 \mathcal{K}''(v) \, dv|$. (b) *For some integer* $h \geq 2$ *and each integer* i $(1 \leq i \leq h)$, $\int |v^i \mathcal{K}'(v) \, dv| < \infty$, *and*

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

and $\mathcal{K}(x) - \mathcal{K}(0) \gtrless 0$ *if* $x \gtrless 0$.

(c) *For each integer* i $(0 \leq i \leq h)$, *and* $\eta > 0$, *and any sequence* $\{\sigma_n\}$ *converging to 0,*

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

(d) $\limsup_{n \to \infty} n \sigma_n^{2h} < \infty$ *and*

$$\lim_{n \to \infty} \sigma_n^{-2h} \int_{|\sigma_n s| > \eta} |\mathcal{K}'(s)| \, ds = 0.$$

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

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

(f) *For some sequence* $m_n \geq 1$, *and* $\varepsilon > 0$,

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

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. A kernel that satisfies these conditions 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 $\mathcal{K}'(0) = \sqrt{2/\pi} = 0.798$ and $||\mathcal{K}'||_2^2 = 0.776$.

---

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

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

where ω_1 and ω_2 are the right and left limit in (6). Thus, ω_1 and ω_2 does not average out as it does in our case.

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

$$\frac{\log(n)}{n\sigma_n^2} \to 0, \tag{7}$$

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^{-1}.

Theorem 2 *Let Assumptions 1 - 3 hold with Assumption 2(a). Then*

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

and they are asymptotically independent.

Similarly, we have

Corollary 3 *Let Assumptions 1 - 3 hold with Assumption 2(a^+). Then*

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

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 (7) and $\limsup_{n\to\infty} n\sigma_n^{2h} < \infty$ for any $h \geq 2$. In this case we obtain that ψ_n is (apart from a logarithmic factor) $n^{-3/4}$ consistent. However,

9

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 $n\sigma_n^{2h}$ is bounded is imposed to ensure the asymptotic independence of ψ_n from θ_n^s. With a bandwidth converging slower, we may obtain the covariances between them, which may prove beneficial for finite sample inference on the slope parameters since θ_n^s depend on ψ_n regardless of the choice of σ_n in finite samples. It is also likely, however, that it may introduce an asymptotic bias for ψ_n as it is the case in 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 $\delta_2 \mapsto \delta_2/n^\pi \to 0$ 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 $E[4(X_{2t}^\top \dot\delta_2)^2 \overline{E}(\varepsilon_t^2|X_{2t}) q_{2t} q_{2t}^\top | z_t = 0]$. 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 estimate of the break fraction $\tau \in (0,1)$ 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 does.

9. The two estimators ψ_n and ψ_n^+ have different asymptotic variances. The ranking could go either way, as the following example illustrates, and so there is nothing a priori to favour one approach over

10

the other. Consider the design of Hansen (2000)

$$y_t = \beta^\top x_t + \delta^\top x_t 1(q_t \leq \psi) + \varepsilon_t,$$

where $x_t = (1, x_{2t})$, $q_t \sim N(2,1)$, $\varepsilon_t \sim N(0,1)$, $\delta = (\delta_1, \delta_2)^\top$, $\beta = 0$, $\delta_1 = 0$, and $\psi = 2$. In case I, $x_{2t} = q_t$ and in case II, $x_{2t} \sim N(0,1)$. The theoretical asymptotic variance of the two smoothed estimators ψ_n and ψ_n^+ in these designs is given below. This shows that as $\delta_2 \to 0$ the asymptotic variance increases for both estimators. For small δ_2, ψ_n has slightly lower asymptotic variance but for δ_2 bigger than about 0.25, ψ_n^+ has smaller variance in cases I and II.

	$\mathrm{avar}(\psi_n)$	$\mathrm{avar}(\psi_n^+)$	$\mathrm{avar}(\psi_n)/\mathrm{avar}(\psi_n^+)$
Case I	$\frac{11\pi\sqrt{2}}{64}\left(1 + \frac{1}{\delta_2^2}\right)$	$\frac{2\sqrt{2}\pi}{11\delta_2^2}$	$\frac{121}{128}\delta_2^2 + \frac{121}{128}$
Case II	$\frac{11\pi\sqrt{2}}{64}\left(3 + \frac{4}{\delta_2^2}\right)$	$\frac{\pi 8\sqrt{2}}{11\delta_2^2}$	$\frac{363}{512}\delta_2^2 + \frac{121}{128}$

4   Inference Methods

The construction of the 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 the 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

$$\tau_{n,t}^\psi(\theta_n) = \left\{ \left(\tilde{x}_t^\top \delta_n\right)^2 - 2\tilde{x}_t^\top \delta_n \left(y_t - x_t^\top \beta_n\right) \right\} \frac{q_{2t}}{\sigma_n} \mathcal{K}'\left(\frac{q_{1t} + q_{2t}^\top \psi_n}{\sigma_n}\right)$$

$$\tau_{n,t}^{\psi+}(\theta_n^+) = 2e_t^+ \tilde{x}_t^\top \delta_n \frac{q_{2t}}{\sigma_n} \mathcal{K}'\left(\frac{q_{1t} + q_{2t}^\top \psi_n}{\sigma_n}\right),$$

where $e_t^+ = y_t - x_t^\top \beta_n^+ - \tilde{x}_t^\top \delta_n^+ \mathcal{K}\left(\frac{q_{1t} + q_{2t}^\top \psi_n^+}{\sigma_n}\right)$. Then, the variance estimators for the threshold parameter ψ are defined, respectively:

$$\hat{V}^\psi = \frac{1}{n}\sum_{t=1}^n \tau_{n,t}^\psi(\theta_n)\tau_{n,t}^\psi(\theta_n)^\top \text{ and } \hat{V}^{\psi+} = \frac{1}{n}\sum_{t=1}^n \tau_{n,t}^{\psi+}(\theta_n^+)\tau_{n,t}^{\psi+}(\theta_n^+)^\top. \tag{8}$$

11

These impose the absence of any theoretical autocorrelation but allow for heteroskesdasticity. We may also make some degrees of freedom adjustment replacing n 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

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

Regarding the estimation of Q^ψ and $Q^{\psi+}$ there are several possibilities. First, just take $Q_n^\psi (\theta_n; \sigma_n)$ and $Q_n^{\psi+} (\theta_n^+; \sigma_n)$. Second, as with nonlinear least squares one can drop some terms that are asymptotically zero. For example, the Hessian is

$$Q_n^{\psi+} \left(\theta_n^+; \sigma_n \right) = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial e_t^+}{\partial \psi} \frac{\partial e_t^+}{\partial \psi^\top} + e_t^+ \frac{\partial^2 e_t^+}{\partial \psi \partial \psi^\top}, \tag{9}$$

and the second term is asymptotically zero. Instead therefore, compute the OPG (outer product of the gradient) estimate

$$\hat{Q}^{\psi+} = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial e_t^+}{\partial \psi} \frac{\partial e_t^+}{\partial \psi^\top}. \tag{10}$$

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

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

Let $w (\cdot) : \mathbb{R} \to [-1, 1]$ be a continuous function such that $w (0) = 1$, $w (x) = w (-x)$, and $\|w\|_2^2 < \infty$. 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 $o(n)$. Similarly we can define $\hat{V}^{s+}$. For more discussion regarding the choice of the kernel and lag truncation parameter, see Andrews (1991). 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

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

$$
\hat{Q}^{s+} = \frac{1}{n}\sum_{t=1}^n \tau_{n,t}^{s+}\left(\theta_n^+\right)\tau_{n,t}^{s+}\left(\theta_n^+\right)^\top .
$$

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. Instead, compute the diagonal elements of the matrix

$$
\hat{Q}_{NB}^{-1}\widehat{V}_{NB}\hat{Q}_{NB}^{-1},
$$

where:

$$
\widehat{V}_{NB} = \begin{pmatrix} \widehat{V}^s & \hat{V}^{s\psi} \\ \hat{V}^{s\psi\top} & \hat{V}^\psi \end{pmatrix} \quad ; \quad \hat{Q}_{NB} = \begin{pmatrix} \hat{Q}^s & \sqrt{\sigma_n}\hat{Q}^{s\psi} \\ \hat{Q}^{s\psi\top} & \sigma_n\hat{Q}^\psi \end{pmatrix}, \tag{11}
$$

$$
\hat{V}^{s\psi} = \hat{Q}^{s\psi} = 2 \begin{pmatrix} \frac{1}{n}\sum_{t=1}^n x_t \tilde{x}_t^\top \delta_n \frac{q_{2t}^\top}{\sigma_n}\mathcal{K}'\left(\frac{q_{1t}+q_{2t}^\top \psi_n}{\sigma_n}\right) \\ \frac{1}{n}\sum_{t=1}^n \left(\tilde{x}_t \tilde{x}_t^\top \delta_n + \tilde{x}_t x_t^\top \beta_n - \tilde{x}_t y_t\right)\frac{q_{2t}^\top}{\sigma_n}\mathcal{K}'\left(\frac{q_{1t}+q_{2t}^\top \psi_n}{\sigma_n}\right) \end{pmatrix}.
$$

Similarly, we may define $\hat{Q}^+$ with

$$
\hat{Q}^{s\psi+} = 2 \begin{pmatrix} \frac{1}{n}\sum_{t=1}^n x_t \tilde{x}_t^\top \delta_n^+ \frac{q_{2t}^\top}{\sigma_n}\mathcal{K}'\left(\frac{q_{1t}+q_{2t}^\top \psi_n^+}{\sigma_n}\right) \\ \frac{1}{n}\sum_{t=1}^n \left(2\tilde{x}_t \tilde{x}_t^\top \delta_n \mathcal{K}\left(\frac{q_{1t}+q_{2t}^\top \psi_n^+}{\sigma_n}\right) + \tilde{x}_t x_t^\top \beta_n^+ - \tilde{x}_t y_t\right)\frac{q_{2t}^\top}{\sigma_n}\mathcal{K}'\left(\frac{q_{1t}+q_{2t}^\top \psi_n^+}{\sigma_n}\right) \end{pmatrix}.
$$

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

Theorem 4 *Under Assumption 1-3, $\hat{V}^s, \sigma_n\hat{V}^\psi, \hat{Q}^s$ and $\sigma_n Q_n^\psi\left(\theta_n\right)$ 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 the t or Wald statistic behaves 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

$$LR(\psi) = n\frac{S_n(\psi) - S_n(\psi_n)}{S_n(\psi_n)}, \tag{12}$$

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 $\sigma^2 = \text{var}(\varepsilon_t)$. Under homoskedasticity, the scaling factor of $LR(\psi)^+$ is equal to one. Apart form this special case, one must 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 heteroskedasticity is more complicated and this reduces the attractiveness of the likelihood ratio.

4.3 Bootstrap

An alternative approach to inference here is based on the bootstrap. In the i.i.d. case this is particularly simple. 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 θ_n^* 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 $\tau(\theta)$. The first method is to just 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 $[\tau(\theta_n) - x_{n,\alpha/2}, \tau(\theta_n) + x_{n,1-\alpha/2}]$. This would be called the percentile method. A perhaps more desirable approach is based on the 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 have

$$s_n^\tau = \nabla\tau(\theta_n)^\top \widehat{Q}^{-1}\widehat{V}\widehat{Q}^{-1}\nabla\tau(\theta_n),$$

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). By 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 interval $[\tau(\theta_n) - z_{n,\alpha/2}s_n^\tau, \tau(\theta_n) + z_{n,1-\alpha/2}s_n^\tau]$. This is usually called the bootstrap-t method. This confidence interval is asymptotically correct, refer to Theorem 2.2 of Horowitz (2001). Since the asymptotic distribution of T does not depend on nuisance parameters, we can expect the bootstrap to achieve asymptotic refinements, see Shao

and Tu (1995) and Horowitz (2001). 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 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.

In order to obtain asymptotic refinements for the threshold parameters or to compute consistent intervals for the slopes we may use the non-overlapping (viz., Carlstein (1986)) and overlapping (viz., Künsch (1989)) block bootstrap procedures. The observations to be bootstrapped are the vectors $\{W_t : t = 1, \ldots, n\}$ as before. Let L denote the length of the blocks satisfying $L \propto n^\gamma$ for some $0 < \gamma < 1$. With non-overlapping blocks, block 1 is observations $\{W_j : j = 1, \ldots, L\}$, block 2 is observations $\{W_{L+j} : j = 1, \ldots, L\}$, and so forth. There are B different blocks, where $BL = n$. With overlapping blocks, block 1 is observations $\{W_j : j = 1, \ldots, L\}$, block 2 is observations $\{W_{1+j} : j = 1, \ldots, L\}$, and so forth. There are $T - L + 1$ different blocks. The bootstrap sample $\{W_t^* : t = 1, \ldots, n\}$ are obtained by sampling B blocks randomly with replacement from either the B non-overlapping blocks or the $n - L + 1$ overlapping blocks and laying them end-to-end in the order sampled.

5 Some Extensions

5.1 The Continuous Case

Suppose that

$$\dot{\delta}_2 = 0, \tag{13}$$

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

$$
\begin{aligned}
\mathcal{V}^\psi &= 4\dot{\delta}_1^2 \cdot E\left[\bar{\mathcal{E}}\left(\varepsilon_t^2 | X_{2t}\right) q_{2t} q_{2t}^\top | z_t = 0\right] f(0) \\
\mathcal{Q}^\psi &= \dot{\delta}_1^2 \int -s^2 \mathrm{sgn}(s) \mathcal{K}''(s) \, ds \cdot E\left[q_2 q_2^\top | z_t = 0\right] f(0) \\
A &= \alpha_A \dot{\delta}_1^2 \left(1/(h-2)!\right) \int f_{z|X_2}^{(h-2)}(0|X_2) q_2 dF_{X_2}(X_2),
\end{aligned}
$$

15

where $\bar{\mathcal{E}}\left(\varepsilon_t^2|X_{2t}\right) = \int_{s>0} s^2 \mathcal{K}'(s)^2 ds \lim_{z\to 0^+} E\left(\varepsilon^2|z, X_2\right) + \int_{s<0} s^2 \mathcal{K}'(s)^2 ds \lim_{z\to 0^-} E\left(\varepsilon^2|z, X_2\right)$ and $\alpha_A = \int s^h \text{sgn}\,(s)\,\mathcal{K}'(s)\,ds$. Then, Theorem 2 can be modified as follows.

Corollary 5 *Let Assumptions 1 - 3 hold with V^ψ and Q^ψ replaced by $\mathcal{V}^\psi$ and $\mathcal{Q}^\psi$ respectively. Furthermore, assume (13) and $\sqrt{n\sigma_n^{2h-1}}$ has a finite limit λ. Then,*

$$\sqrt{n}\left(\theta_n^s - \theta_0^s\right) \implies N\left(0, Q^{s^{-1}} V^s Q^{s^{-1}}\right),$$
$$\sqrt{n\sigma_n}\left(\psi_n - \psi_0\right) \implies N\left(-\lambda \mathcal{Q}^{\psi^{-1}} A, \mathcal{Q}^{\psi^{-1}} \mathcal{V}^\psi \mathcal{Q}^{\psi^{-1}}\right),$$

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.

The bias correction is straightforward since $\sqrt{n\sigma_n^{-1}}T_n\left(\theta_n\right)$ is a consistent estimator of A and the studentizing can be done as described in Section 4. When the confidence interval is constructed as in Section 4 with the bandwidth σ_n satisfying (7), it will be an asymptotically correct one even when the true model is continuous since $\lambda = 0$ in that case. 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 (13) by the $\mathcal{X}^2$ test, utilizing the delta-method and Theorem 5.

5.2 Multiple Threshold Case

Suppose that there are multiple thresholds determined by variables $q_{tj}^\top \psi_j$ that enter in an additively separable fashion

$$y_t = x_t^\top \beta + \sum_{j=1}^{p} \tilde{x}_t^\top \delta_j 1\left\{q_{tj}^\top \psi_j > 0\right\} + \varepsilon_t.$$

Then, the estimation strategy and theoretical results are essentially as before. Specifically, let

$$\mathcal{K}_{tj} = \mathcal{K}_j\left(\frac{q_{j1t} + q_{j2t}^\top \psi_j}{\sigma_{nj}}\right), \quad j = 1, \ldots, p$$

and define for given $\psi = (\psi_1, \ldots, \psi_p)$, $\widehat{\theta}^s(\psi) = (W(\psi)^\top W(\psi))^{-1} W(\psi)^\top y$, where $W(\psi)$ is the $n \times k(1+p)$ matrix with rows $w_t = (x_t^\top, \tilde{x}_t^\top \mathcal{K}_{t1}, \ldots, \tilde{x}_t^\top \mathcal{K}_{tp})^\top$. Then define ψ_n to minimize

$$S_n^+(\psi) = \sum_{t=1}^{n}\left(y_t - w_t^\top(\psi)\widehat{\theta}^s(\psi)\right)^2.$$

In this case, we expect the rate of convergence of ψ_n to be the same as before, although the asymptotic variance will be different. Bai (1997) has shown, in the structural change context, that a sequential strategy can work: estimate a single threshold model and then a second threshold conditioning on the first one and so on. This is very convenient computationally. Simulations show that this approach also works in this case: the dominant threshold is identified in the first round etc.

On the other hand if one has thresholds of the type

$$y_t = x_t^\top \beta + \tilde{x}_t^\top \delta 1 \left\{ q_{t1}^\top \boldsymbol{\psi}_1 > 0, \ldots, q_{tp}^\top \boldsymbol{\psi}_p > 0 \right\} + \varepsilon_t,$$

then the smoothing based method will suffer severely from curse of dimensionality because the smoothing operation is of dimension p.

5.3 Alternative Estimation Criteria

The least squares method can sometimes be strongly influenced by outliers and one may wish to use a more robust method for estimating parameters like the LAD. Our second method can easily be adapted to this case. Thus for example consider the criterion

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

and let $\theta_n^{LAD+} = (\beta_n^{LAD+\top}, \delta_n^{LAD+\top}, \psi_n^{LAD+\top})^\top = \arg\min_{\theta \in \Theta} S_n^{LAD+}(\theta; \sigma_n)$. Although there is not an explicit formula for the profiled slope estimators in this case, the profiled slopes are regular LAD regression estimators and can be computed efficiently by linear programming, Koenker (1997). The threshold estimate can easily be computed in the scalar case by grid search but otherwise it requires some care. It can be shown that θ_n^{LAD+} is consistent and asymptoptically normal with the same rates as the least squares estimators, under some conditions.

6 Numerical Results

6.1 Monte Carlo

We investigated again the design of Hansen (2000). In this case,

$$y_t = \beta^\top x_t + \delta^\top x_t 1(q_t \leq \psi) + \varepsilon_t,$$

where $x_t = (1, x_{2t})$, $q_t \sim N(2, 1)$, $\varepsilon_t \sim N(0, 1)$, $\delta = (\delta_1, \delta_2)^\top$, $\beta = 0$, $\delta_1 = 0$, and $\psi = 2$. In case I, $x_{2t} = q_t$ and in case II, $x_{2t} \sim N(0, 1)$. We compute ψ_n, ψ_n^+ using the kernel $K(x) = \Phi(x) + x\phi(x)$, where Φ and ϕ are the standard Gaussian c.d.f. and density functions respectively. The estimators

are computed by grid search over the sample of observed threshold values. We consider parameter values $\delta_2 \in \{0.25, 0.5, 1.0, 1.5, 2.0\}$ and sample sizes $n \in \{50, 100, 250, 500, 1000\}$ and do $ns = 1000$ replications for each experiment. In other work we have examined larger sample sizes, and we comment on these results.

6.1.1 Performance of the Estimator

In this section we describe the performance of the unsmoothed and smoothed threshold estimators. We take bandwidth parameter $\sigma_n = (\log n)n^{-1/2}$. In Table 1a we report results for the estimates of ψ, while in Table 1b we present the results for the estimates of δ_2. We present the interquartile range divided by 1.35, which is a robust estimate of the standard deviation of the estimates. The biases are very small in all cases and are not reported. There are several main results:

1. Results improve with sample size and with the value of δ_2

2. 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 with δ_2. This overprediction is also implicitly true for the unsmoothed least squares estimator.

3. The estimator ψ_n^+ is nearly always better than ψ_n

We have also examined the case with very large sample sizes and find that with $n = 10,000$ the mean squared errors are within 5% of the asymptotic predictions. Also in this case q-q plots reveal that normality is a good approximation.

6.1.2 Performance of the Confidence Intervals

We next compare our confidence intervals with those of Hansen (2000). We compute the estimators by the two different smoothing methods and we investigated three different t-statistic confidence intervals: those based on estimates of the asymptotic variance, those based on the percentile bootstrap, and those based on the pivotal bootstrap using the asymptotic standard errors to studentize. Hansen (2000) used the likelihood ratio, which can be expected to work particularly well in this design as it assumes normality and homoskedasticity. We report results for the parameter ψ and δ_2 for the fifty different combinations of sample sizes ($n \in \{50, 100, 250, 500, 1000\}$) and parameter values ($\delta_2 \in \{0.25, 0.5, 1.0, 1.5, 2.0\}$) for case I and II. We implemented the two methods as in the previous section.

The results of the simulations are shown in the tables. In Table 2abc we give the coverage rate for ψ_n intervals based on percentile bootstrap, pivotal bootstrap, and asymptotic method. In Table

3abc we give the same for intervals based on ψ_n^+. These tables correspond to Table II of Hansen (2000). Apart from the smallest value of δ_2, the bootstrap coverage rates are close to the nominal rate and because of the small number of replications are generally within 2 standard errors of the target value (0.02) except for the small δ_2 case. The coverage rates of the asymptotic intervals are less satisfactory for smaller samples sizes, but improve steadily with sample size and are competitive for $n = 1000$. There does not seem to be much difference between the intervals based on ψ_n and the intervals based on ψ_n^+. In Table 4ab and 5ab we give the bootstrap intervals for δ_2 based on the two estimators. These correspond to Table III of Hansen (2000).[3] The coverage rates of the bootstrap intervals are close to the nominal throughout.

We also investigated the bootstrap for the unsmoothed estimator. The coverage rates were very low (and not reported here) even in the largest sample sizes and we take this as evidence of inconsistency.

The results suggest that the small threshold case, $\delta_2 = 0.25$, is problematic. Indeed the asymptotic intervals are considerably undercovered for this case, although the bootstrap intervals are overcovered. This suggests that a combination of the two intervals may be useful in practice. Figure 1 shows a typical sample from this process - the threshold effect is indeed very small in this case. We investigated some different asymptotic confidence intervals for the special case $z = q, \delta_2 = 0.25, \sigma_n = 1.06 n^{-1/5}$. The results are reported in Tables 6 and Tables 7. We consider bandwidths $\sigma_n = 1.06 n^{-1/5}$ and $\sigma_n = (\log n) n^{-1/2}$. The results suggest that larger bandwidth gives better coverage. It also suggests that the Likelihood ratio intervals have the most accurate coverage, followed by the 'non-block diagonal' confidence intervals. We also report the median length of the confidence intervals; the smaller bandwidth procedures gives smaller length.

6.2 Application

6.2.1 Growth with multiple equilibria

We illustrate our methodology by examining the hypothesis that initial conditions may determine cross-section growth behavior using the Summers-Heston data set. Durlauf and Johnson (1995) studied it by a regression tree method due to Breiman et al. (1984) and Hansen (2000) by a threshold regression using the same data set. We specify the model similarly to the previous studies. Let $y_{i,t}$ be real GDP per member of the population aged 15-64 in year t, ι_i be investment to GDP ratio, n_i be growth rate of the working-age population, and S_i be the fraction of working-age population enrolled in secondary school. The variables other than $y_{i,t}$ are the annual averages over the period 1960-1985.

---

[3]In Table III of Hansen, the critical level of the table is 95%. And the confidence interval is constructed as union of confidence intervals based on a given set of threshold values that is a confidence interval for the threshold estimate.

Then, the log real GDP growth, $\ln y_{i,1985} - \ln y_{i,1960}$, is explained by $\ln y_{i,1960}, \ln \iota_i, \ln(n_i + 0.05)$, and $\ln S_i$.

Durlauf-Johnson proposed the initial output $y_{i,1960}$ and the literacy rate (lr_i) at the year 1960 as the possible threshold variables. Hansen examined each variable separately by the Lagrange multiplier test of Hansen (1996) and found some evidence for the presence of a threshold effect based on the initial output. Multiple threshold variables are not allowed in Hansen (1960) nor in Hansen (2000). Thus, he took a sequential approach in which he estimates the threshold with the initial output and test for the further threshold within the subsamples splitted by the threshold estimate. This procedure is repeated until we cannot find further evidence of threshold. He reports the first sample split at the output level of $863 and the second at the initial literacy rate of 45% within the subsample whose initial output is larger than $863.

We first estimate the model where the output is the threshold variable. The SLS estimate ψ_n^+ is $1781 with the standard error of $316. There are a couple of different methods to compute the standard error as explained in Section 4.1. The ones reported here are the most conservative and are robust to heteroskedasticity[4]. The 95% bootstrap confidence interval is $[0, 6675]$, much wider than the asymptotic interval. Figure 2 displays the smoothed sum of squares residuals as a function of the threshold in output. There are 47 of the 96 countries below the threshold. The second split is also based on the initial output of $777 with the standard error of $33. See Figure 3 for the sum of squared residuals.Unlike in Hansen, we could not find evidence for the threshold in the literacy rate in the subsamples generated by the initial output of $1781 using the LM test of Hansen (1996). We also estimate the model with the threshold in the linear combination of the initial output and the initial literacy rate. The coefficient of the output is normalized to 1, and the estimates are obtained by bivariate grid search. The estimated splitting line is

$$y_{1960} = 46 * lr - 294,$$

which reduces the sum of squared residuals by approximately 10% compared to that of one threshold in output above. The coefficient of lr appears significant as its standard error is 20 and its 95% bootstrap confidence interval is $[14.7, 390]$. On the contrary to the sequential approach, this indicates that both the initial output and the initial literacy rate may be related to the determination of the growth path.

The estimates for slope parameters are reported in Table 8. We observe that the initial output and the population growth have negative effect on the growth rate. In the subsample where the output is above $1781, 54% additional growth rate is expected while the average is 44%. And in the

---

[4]The heteroskedasticity appears clear. For example, the sample variance in the regime in which the output is below the threshold $1781 is 0.10 while that of the other regime is 0.07.

subsample where $y_{1960} > 46 * lr - 294$, the 57% increase is expected.

7   Conclusions

We have shown that the smoothed threshold estimator is asymptotically normal albeit at a slower rate than the corresponding unsmoothed estimator. This is born out by simulations. On the other hand, our simulations show that our confidence intervals can be more accurate than the confidence intervals of Hansen (2000) especially for larger thresholds. 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 results for smoothed LAD (SLAD) estimators, and perhaps faster than is the case for the unsmoothed estimator. Furthermore, we expect the smoothed estimation will enable the 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 the 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 purpose to which the estimation is put. For estimation itself, a small σ_n of the order $(\log n)n^{-1/2}$ seems to perform well. For testing problems bandwidth is likely to affect size and power in different ways so small is not necessarily best.

A   Proofs of Theorems

LEMMA 1. *Suppose that Assumption 1 holds. The following convergences hold almost surely uniformly over the parameter space:*

$$(i) \quad \frac{1}{n} \sum_{t=1}^{n} 1 \left\{ \left| q_{1t} + q_{2t}^{\top} \psi \right| < \eta \right\} \to \Pr \left\{ \left| q_{1t} + q_{2t}^{\top} \psi \right| < \eta \right\} \text{ for any } \eta > 0,$$

$$(ii) \quad \frac{1}{n} \sum_{t=1}^{n} x_t x_t^{\top} 1 \left\{ q_{1t} + q_{2t}^{\top} \psi > 0 \right\} \to E x_t x_t^{\top} 1 \left\{ q_{1t} + q_{2t}^{\top} \psi > 0 \right\}$$

$$(iii) \quad \frac{1}{n} \sum_{t=1}^{n} x_t \varepsilon_t 1 \left\{ q_{1t} + q_{2t}^{\top} \psi > 0 \right\} \to E x_t \varepsilon_t 1 \left\{ q_{1t} + q_{2t}^{\top} \psi > 0 \right\}$$

PROOF OF LEMMA 1. We apply the generic uniform law of large numbers by Andrews (1987, Corollary 1). Assumption A1 and B1 of that paper are also assumed here. Assumption B2 is trivially satisfied in (i) since the indicator function is bounded, and in (ii) and (iii), since:

$$E \left(\sup_{\psi \in \Theta_\psi} \left| x_t x_t^{\top} 1 \left\{ q_{1t} + q_{2t}^{\top} \psi > 0 \right\} \right| \right)^{\xi} \leq E \left| x_t x_t^{\top} \right|^{\xi} < \infty$$

$$E \left(\sup_{\psi \in \Theta_\psi} \left| x_t \varepsilon_t 1 \left\{ q_{1t} + q_{2t}^{\top} \psi > 0 \right\} \right| \right)^{\xi} \leq E \left| x_t \varepsilon_t \right|^{\xi} < \infty.$$

Next, $1 \left\{ \left| q_{1t} + q_{2t}^{\top} \psi \right| < \eta \right\}$ and $1 \left\{ q_{1t} + q_{2t}^{\top} \psi > 0 \right\}$ satisfy Assumption A3 as shown in de Jong and Wootersen (2004, Lemma 4). Then, by Cauchy-Schwarz inequality, (ii) and (iii) satisfy Assumption A3, which completes the proof of Lemma. ∎

PROOF OF THEOREM 1. First, we show that $|S_n^* (\theta; \sigma_n) - S_n (\theta; \sigma_n)| \to 0$ almost surely uniformly over $\theta \in \Theta$. To do that, note that

$$|S_n^* (\theta) - S_n (\theta; \sigma_n)|$$

$$= \left| \frac{1}{n} \sum_{t=1}^{n} \left\{ (\tilde{x}_t^{\top} \delta)^2 - 2 \tilde{x}_t^{\top} \delta (y_t - x_t^{\top} \beta) \right\} \left[1 \left\{ q_{1t} + q_{2t}^{\top} \psi > 0 \right\} - \mathcal{K} \left(\frac{q_{1t} + q_{2t}^{\top} \psi}{\sigma_n} \right) \right] \right|$$

$$\leq \sqrt{ \frac{1}{n} \sum_{t=1}^{n} \left\{ (\tilde{x}_t^{\top} \delta)^2 - 2 \tilde{x}_t^{\top} \delta (y_t - x_t^{\top} \beta) \right\}^2 } \sqrt{ \frac{1}{n} \sum_{t=1}^{n} \left[1 \left\{ q_{1t} + q_{2t}^{\top} \psi > 0 \right\} - \mathcal{K} \left(\frac{q_{1t} + q_{2t}^{\top} \psi}{\sigma_n} \right) \right]^2 },$$

the first term of which almost surely converges to a finite number uniformly over $\left(\beta^{\top}, \delta^{\top} \right)^{\top} \in \Theta_\beta \times \Theta_\delta$ by Lemma 1. For the convergence of the second term, note that the same reasoning as in Lemma 4 of Horowitz (1992) applies. Then, it is sufficient to show that, for any $\eta > 0$, $(A4)$ in that paper, i.e.,

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

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. Next, we show that $\theta_n^* = \arg\min_\theta S_n^*(\theta)$ is consistent, which is sufficient for the consistency of θ_n. For a fixed ψ, the least squares estimator of β and δ are the OLS estimators, which are denoted as $\beta_n^*(\psi)$ and $\delta_n^*(\psi)$ respectively. Let $S_n^*(\psi) = S_n^*(\beta_n^*(\psi), \delta_n^*(\psi), \psi)$. Let $\varepsilon_t(\psi) = y_t - x_t^\top\beta(\psi) - \tilde{x}_t^\top\delta(\psi)\mathbf{1}\left\{q_{1t} + q_{2t}^\top\psi > 0\right\}$ such that $E(x_t\varepsilon_t(\psi)) = 0$ and $E\left(x_t\mathbf{1}\left\{q_{1t} + q_{2t}^\top\psi > 0\right\}\varepsilon_t(\psi)\right) = 0$. Let X_ψ be the matrix stacking $x_t^\top$ and $\tilde{x}_t^\top\mathbf{1}\left\{q_{1t} + q_{2t}^\top\psi > 0\right\}$ and ε_ψ with $\varepsilon_t(\psi)$. Then,

$$S_n^*(\psi) = \frac{1}{n}\varepsilon_\psi^\top\varepsilon_\psi - \frac{1}{n}\varepsilon_\psi^\top X_\psi\left(\frac{1}{n}X_\psi^\top X_\psi\right)^{-1}\frac{1}{n}X_\psi^\top\varepsilon_\psi \to E\varepsilon_t(\psi)^2,$$

almost surely uniformly over $\psi \in \Theta_\psi$, by Lemma 1. Note that $\beta = \beta(\psi_0)$ and $\delta = \delta(\psi_0)$, and that $E\varepsilon_t(\psi)^2$ is uniquely minimized at $\psi = \psi_0$, since θ_0 defines the conditional expectation, which minimizes MSE, and the threshold index $z_t(\psi)$ includes at least one continuous random variable. By the latter, $E\varepsilon_t(\psi)^2$ is continuous on $\psi \in \Theta_\psi$. Therefore, ψ_n^*, which also minimizes $S_n^*(\psi)$, converges to ψ_0 almost surely. Furthermore, it in turn implies that β_n and δ_n converge to β_0 and δ_0 almost surely by Lemma 1. ∎

PROOF OF THEOREM 2. The asymptotic distribution developed here is based on the Taylor series expansion of $T_n(\theta; \sigma_n)$:

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

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 $\sqrt{\sigma_n}$ and note that

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

The following is useful for the development below

$$\begin{aligned}\left\{\left(\tilde{x}_t^\top\delta\right)^2 - 2\tilde{x}_t^\top\delta\left(y_t - x_t^\top\beta\right)\right\} &= \left(\tilde{x}_t^\top\delta_0\right)^2 - 2\tilde{x}_t^\top\delta_0\left(y_t - x_t^\top\beta_0\right) + R_{nt}(\theta) \\ &= \left(\tilde{x}_t^\top\delta_0\right)^2 - 2\tilde{x}_t^\top\delta_0\left(\tilde{x}_t^\top\delta_0\mathbf{1}\{z_t > 0\} + \varepsilon_t\right) + R_{nt}(\theta) \\ &= \left(\tilde{x}_t^\top\delta_0\right)^2\left(1 - 2\cdot\mathbf{1}\{z_t > 0\}\right) - 2\tilde{x}_t^\top\delta_0\varepsilon_t + R_{nt}(\theta) \\ &= -\left(\left(\tilde{x}_t^\top\delta_0\right)^2\mathrm{sgn}(z_t) + 2\tilde{x}_t^\top\delta_0\varepsilon_t\right) + R_{nt}(\theta),\end{aligned} \tag{14}$$

where

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

23

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

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

$$
\begin{aligned}
T_n^\psi(\theta_0, \sigma_n) &= \frac{1}{n}\sum_{t=1}^n \left\{\left(\tilde{x}_t^\top \delta_0\right)^2 - 2\tilde{x}_t^\top \delta_0\left(y_t - x_t^\top \beta_0\right)\right\}\frac{q_{2t}}{\sigma_n}\mathcal{K}'\left(\frac{z_t}{\sigma_n}\right) \\
&= -\frac{1}{n}\sum_{t=1}^n \left(\left(\tilde{x}_t^\top \delta_0\right)^2 \operatorname{sgn}(z_t) + 2\tilde{x}_t^\top \delta_0 \varepsilon_t\right)\frac{q_{2t}}{\sigma_n}\mathcal{K}'\left(\frac{z_t}{\sigma_n}\right),
\end{aligned}
$$

and

$$
Q_n^s\left(\tilde{\theta}, \sigma_n\right) = 2\begin{pmatrix} \frac{1}{n}\sum_{t=1}^n x_t x_t^\top & \frac{1}{n}\sum_{t=1}^n x_t \tilde{x}_t^\top \mathcal{K}\left(\frac{q_{1t} + q_{2t}^\top \tilde{\psi}}{\sigma_n}\right) \\ \cdot & \frac{1}{n}\sum_{t=1}^n \tilde{x}_t \tilde{x}_t^\top \mathcal{K}\left(\frac{q_{1t} + q_{2t}^\top \tilde{\psi}}{\sigma_n}\right) \end{pmatrix}
$$

$$
Q_n^{s\psi}\left(\tilde{\theta}, \sigma_n\right) = 2\begin{pmatrix} \frac{1}{n}\sum_{t=1}^n x_t \tilde{x}_t^\top \tilde{\delta}\frac{q_{2t}}{\sigma_n}\mathcal{K}'\left(\frac{q_{1t} + q_{2t}^\top \tilde{\psi}}{\sigma_n}\right) \\ \frac{1}{n}\sum_{t=1}^n \left(\tilde{x}_t \tilde{x}_t^\top \tilde{\delta} + \tilde{x}_t x_t^\top \tilde{\beta} - \tilde{x}_t y_t\right)\frac{q_{2t}}{\sigma_n}\mathcal{K}'\left(\frac{q_{1t} + q_{2t}^\top \tilde{\psi}}{\sigma_n}\right) \end{pmatrix}
$$

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

where the last equality follows from (14). We show the convergences of T_n^s and Q_n^s and the others in the following sequence of Lemma's.

LEMMA 2 *Suppose* $\frac{\left(\tilde{\psi} - \psi_0\right)}{\sigma_n} = o(1)$. *Then,*

$$
\begin{aligned}
\sqrt{n}T_n^s(\theta_0, \sigma_n) &\Longrightarrow {}^d N\left(0, V^s\right), \\
Q_n^s\left(\tilde{\theta}, \sigma_n\right) &\to {}^p Q^s.
\end{aligned}
$$

PROOF OF LEMMA 2. Assumption 3 (d) implies that

$$
\left|1\{s > 0\} - \mathcal{K}\left(\frac{s}{\sigma_n}\right)\right| = o\left(n^{-1}\right) \tag{15}
$$

for all nonzero $s \in \mathbb{R}$. Therefore, it follows from the dominated convergence theorem that the followings are $o_p(1)$: $\frac{1}{\sqrt{n}}\sum_{t=1}^n x_t x_t^\top (1\{z_t > 0\} - \mathcal{K}\left(\frac{z_t}{\sigma_n}\right))$, $\frac{1}{\sqrt{n}}\sum_{t=1}^n x_t x_t^\top 1\{z_t \le 0\}\mathcal{K}\left(\frac{z_t}{\sigma_n}\right)$, and $\frac{1}{\sqrt{n}}\sum_{t=1}^n x_t \varepsilon_t(1\{z_t > 0\} - \mathcal{K}\left(\frac{z_t}{\sigma_n}\right))$. Then,

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

24

Furthermore, there is a $\vec{\psi}$ between ψ_0 and $\tilde{\psi}$ such that

$$
\frac{1}{n} \sum_{t=1}^{n} x_t x_t^\top \mathcal{K} \left(\frac{q_{1t} + q_{2t}^\top \tilde{\psi}}{\sigma_n} \right) = \frac{1}{n} \sum_{t=1}^{n} x_t x_t^\top \mathcal{K} \left(\frac{z_t}{\sigma_n} \right) + \frac{1}{n} \sum_{t=1}^{n} x_t x_t^\top \mathcal{K}' \left(\frac{q_{1t} + q_{2t}^\top \vec{\psi}}{\sigma_n} \right) \frac{q_{2t}^\top \left(\tilde{\psi} - \psi_0 \right)}{\sigma_n}
$$

$$
= \frac{1}{n} \sum_{t=1}^{n} x_t x_t^\top 1 \{ z_t > 0 \} + o_p (1),
$$

due to the dominated convergence theorem and (15). The LLN yields the desired results. $\blacksquare$

LEMMA 3. *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*

$$
\lim_n E \left[\sqrt{n \sigma_n} T_n^\psi (\theta_0, \sigma_n) \right] = 0
$$
$$
\lim_n \mathrm{var} \left[\sqrt{n \sigma_n} T_n^\psi (\theta_0, \sigma_n) \right] = V^\psi.
$$

PROOF OF LEMMA 3. Note that

$$
-E \left[\sigma_n^{-h} T_n^\psi (\theta, \sigma_n) \right]
$$
$$
= \sigma_n^{-h} \int \left\{ \left(\tilde{x}^\top \delta \right)^2 \mathrm{sgn} (z) \right\} \frac{q_2}{\sigma_n} \mathcal{K}' \left(\frac{z}{\sigma_n} \right) f_{z|X_2} (z|X_2) \, dz \, dF_{X_2}(X_2)
$$
$$
= \sigma_n^{-h} \int \left\{ \left(\sigma_n s \dot{\delta}_1 + X_2^\top \dot{\delta}_2 \right)^2 (\mathrm{sgn} (s)) \right\} q_2 \mathcal{K}' (s) f_{z|X_2} (\sigma_n s | X_2) \, ds \, dF_{X_2}(X_2),
$$

where $s = \frac{z}{\sigma_n}$ and $F_{X_2}(X_2)$ is the marginal distribution of X_{2t}. A Taylor series expansion about $\sigma_n s = 0$ yields, for ξ between 0 and $\sigma_n s$,

$$
f (\sigma_n s | X_2) = \sum_{j=0}^{h-3} (1/j!) f_{z|X_2}^{(j)} (0|X_2) (\sigma_n s)^j + (1/ (h-2)!) f_{z|X_2}^{(h-2)} (\xi|X_2) (\sigma_n s)^{h-2} .
$$

By Assumption 2, there is an $\eta > 0$ such that the derivatives exist and uniformly bounded for almost every X_2 if $|\sigma_n s| \leq \eta$. Let $C_n = \{ s : |\sigma_n s| \leq \eta \}$ and C_n^c denote the complement of C_n. Then $E \left[\sigma_n^{-h} T_n^\psi (\theta_0, \sigma_n) \right] = I_{n1} + I_{n2}$, where

$$
I_{n1} = \dot{\delta}_1^2 (1/ (h-2)!) \int_{C_n} s^h \mathrm{sgn} (s) q_2 \mathcal{K}' (s) f_{z|X_2}^{(h-2)} (\xi|X_2) \, ds \, dF_{X_2}(X_2)
$$
$$
\rightarrow \alpha_A \dot{\delta}_1^2 (1/ (h-2)!) \int f_{z|X_2}^{(h-2)} (0|X_2) q_2 \, dF_{X_2}(X_2),
$$

by the dominated convergence theorem, and

$$
I_{n2} \leq M \int_{C_n^c} \left(\sigma_n^{2-h} \left| s^2 \mathcal{K}' (s) \right| + \sigma_n^{1-h} |s \mathcal{K}' (s)| \left(X_2^\top \dot{\delta}_2 \right) + \sigma_n^{-h} |\mathcal{K}' (s)| \left(X_2^\top \dot{\delta}_2 \right)^2 \right) |q_2| \, ds \, dF_{X_2}(X_2) = o (1),
$$

by Assumption 3 (c), and 2.

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

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

and note that

$$
\sigma_n\operatorname{var}\left[\tau_{n,t}^\psi\right]
$$
$$
= \sigma_n E\left[\left(\left(\tilde{x}_t^\top\delta\right)^4 + 4\left(\tilde{x}_t^\top\delta\varepsilon_t\right)^2\right)\frac{q_{2t}q_{2t}^\top}{\sigma_n^2}\mathcal{K}'\left(\frac{z_t}{\sigma_n}\right)^2\right] + o\left(1\right).
$$

But, it follows from the same reasoning as above that

$$
\sigma_n E\left[\left(\tilde{x}_t^\top\delta\right)^4\frac{q_{2t}q_{2t}^\top}{\sigma_n^2}\mathcal{K}'\left(\frac{z_t}{\sigma_n}\right)^2\right]
$$
$$
= \int\left(\tilde{x}^\top\delta\right)^4\frac{q_2q_2^\top}{\sigma_n}\mathcal{K}'\left(\frac{z}{\sigma_n}\right)^2 f_{z|X_2}\left(z|X_2\right)dzdF_{X_2}(X_2)
$$
$$
= \int\left(\sigma_n s\dot{\delta}_1 + X_2^\top\dot{\delta}_2\right)^4 q_2q_2^\top\mathcal{K}'\left(s\right)^2 f_{z|X_2}\left(\sigma_n s|X_2\right)dsdF_{X_2}(X_2)
$$
$$
\rightarrow \|\mathcal{K}'\|_2^2\cdot E\left[\left(X_{2t}^\top\dot{\delta}_2\right)^4 q_{2t}q_{2t}^\top f_{z|X_2}\left(0|X_2\right)\right],
$$

where $s = z/\sigma_n$. [5]

Similarly,

$$
\sigma_n E\left[4\left(\tilde{x}_t^\top\delta\varepsilon_t\right)^2\frac{q_{2t}q_{2t}^\top}{\sigma_n^2}\mathcal{K}'\left(\frac{z_t}{\sigma_n}\right)^2\right]
$$
$$
= \int 4\left(\sigma_n s\dot{\delta}_1 + X_2^\top\dot{\delta}_2\right)^2 E\left(\varepsilon^2|z=\sigma_n s, X_2\right)q_2q_2^\top\mathcal{K}'\left(s\right)^2 f_{z|X_2}\left(\sigma_n s|X_2\right)dsdF_{X_2}(X_2)
$$
$$
= \int_{s>0} 4\left(\sigma_n s\dot{\delta}_1 + X_2^\top\dot{\delta}_2\right)^2 E\left(\varepsilon^2|z=\sigma_n s, X_2\right)q_2q_2^\top\mathcal{K}'\left(s\right)^2 f_{z|X_2}\left(\sigma_n s|X_2\right)dsdF_{X_2}(X_2)
$$
$$
+ \int_{s<0} 4\left(\sigma_n s\dot{\delta}_1 + X_2^\top\dot{\delta}_2\right)^2 E\left(\varepsilon^2|z=\sigma_n s, X_2\right)q_2q_2^\top\mathcal{K}'\left(s\right)^2 f_{z|X_2}\left(\sigma_n s|X_2\right)dsdF_{X_2}(X_2)
$$
$$
\rightarrow E\left[4X_{2t}^\top\dot{\delta}_2\overline{E}\left(\varepsilon_t^2|X_{2t}\right)q_{2t}q_{2t}^\top|z_t=0\right]f_z\left(0\right),
$$

where $\overline{E}\left(\varepsilon_t^2|X_{2t}\right)$ is defined in (6). This last step follows by the law of iterated expectation and reversing the order of expectations.

Next, by the mixing inequality (Davidson 1994, corollary 14.3), for $p\geq 2$,

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

---

[5] $E\left[\left(x_{2t}^\top\delta_2\right)^4 x_{2t}x_{2t}^\top\right] < \infty$. But it can be weakened to $E\left[\left(x_{2t}^\top\delta_2\right)^4\right] < \infty$ under the model where the threshold variable is a single variable instead of linear combination of x_t.

But, for any vector ξ s.t. $|\xi| = 1$,

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

Also note that

$$
\sigma_n \operatorname{cov}\left(\tau_{n,t}^\psi, \tau_{n,t-m}^\psi \right) = \sigma_n E \tau_{n,t-m}^{\psi'} \xi \tau_{n,t}^{\psi'} \xi + O\left(\sigma_n^{2h+1} \right) = O\left(\sigma_n \right),
$$

by the same reasoning as the convergence of $E\left[\sigma_n^{-h} T_n^\psi (\theta_0, \sigma_n) \right]$ provided that the boundedness of $E\left(\varepsilon_{t-m} | X_t, X_{t-m} \right)$ and $f\left(z_t, z_{t-m} | X_{2t}, X_{2t-m} \right)$. Then, for the same reason as Lemma 7 of de Jong and Wouterson $(2004, p.24)$,

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

Therefore, we conclude

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

By the same reasoning as for this, we can show that 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. ∎

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

PROOF OF LEMMA 4. See Lemma 6 of Horowitz (1992) and Theorem 2 of de Jong (1997).[6]

LEMMA 5. $\sigma_n^{-1} (\psi_n - \psi_0) = o_p(1)$.

PROOF OF LEMMA 5. The proof consists of two steps: First we show that

$$
\sup_{\theta \in \Theta_n} \left| T_n^\psi (\theta) - E T_n^\psi (\theta) \right| \to^p 0, \tag{16}
$$

where Θ_n is a neighborhood of θ_0, and then show that for any $\theta \to \theta_0$,

$$
E T_n^\psi (\theta) = \ddot{Q}_n + o(1), \tag{17}
$$

where

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

and $\ddot{\psi}_n = (\psi - \psi_0) / \sigma_n$. Since $T_n^\psi (\hat{\theta})$ is zero and $T_n^\psi (\theta)$ is uniformly continuous, $E T_n^\psi (\theta)$ converges uniformly to zero in Θ_n, which implies that $\ddot{Q}_n = o(1)$. We conclude $\ddot{\psi}_n = o(1)$ by contradiction as below.

---

[6] mixing condition: $\alpha_m \le C m^{-s/(s-2)-\eta}$.

PROOF OF (16). Note that,

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

where R_{nt} is defined in (14). Define

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

and $g_{nt}^{C_n} (\theta) = g_{nt} (\theta) \mathbf{1} \{ |g_{nt} (\theta)| > C_n \}$, and decompose $\frac{1}{n\sigma_n} \sum_{t=1}^n (g_{nt} (\theta) - E g_{nt} (\theta))$ into

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

Let $nE |x_t|^r C_n^{-r/3} \to 0$. Then, for a constant $C > 0$,

$$\Pr \left\{ \sup_\theta \left| \frac{1}{n\sigma_n} \sum_{t=1}^n \left(g_{nt} (\theta) - g_{nt}^{C_n} (\theta) \right) \right| > 0 \right\} \leq n \Pr \{ |g_{nt} (\theta)| > C_n \}$$

$$\leq nCE |x_t|^r C_n^{-r/3} \to 0.$$

And, for a sequence m_n and σ_n as defined in Assumption 3(e),

$$\Pr \left\{ \left| \frac{1}{n\sigma_n} \sum_{t=1}^n \left(g_{nt}^{C_n} (\theta) - E g_{nt}^{C_n} (\theta) \right) \right| > \varepsilon \right\}$$

$$\leq O \left(\sigma_n^{-3k} \right) m_n \exp \left(-\varepsilon^2 n \sigma_n^2 C_n^{-2} m_n^{-2} \right) + O \left(\sigma_n^{-3k} \sigma_n^{-1} C_n \alpha_{m_n} \right)$$

$$= o (1),$$

by the same reasoning in the proof of Lemma 11 of De Jong and Woutersen (2004). Next, it is straightforward from the proof of (17) below that

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

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 (17). Since $E (\varepsilon_t | X_t) = 0$,

$$ET_n^\psi (\theta) = - \frac{1}{n} \sum_{t=1}^n E \left(\tilde{x}_t^\top \delta \right)^2 \operatorname{sgn} (z_t) \frac{q_{2t}}{\sigma_n} \mathcal{K}' \left(\frac{z_t + (\psi - \psi_0)^\top q_{2t}}{\sigma_n} \right) + o (1).$$

Let $\ddot{\psi}_n = (\psi - \psi_0)/\sigma_n$ and $s = \frac{z}{\sigma_n} + \ddot{\psi}_n^\top q_2$, and note that

$$-E\left\{ \left(\tilde{x}_t^\top \delta\right)^2 \operatorname{sgn}(z_t) \frac{q_{2t}}{\sigma_n} \mathcal{K}'\left(\frac{z_t + (\psi - \psi_0)^\top q_{2t}}{\sigma_n}\right)\right\}$$

$$= -\int \left(\sigma_n\left(s - \ddot{\psi}_n^\top q_2\right) + X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}\left(s - \ddot{\psi}_n^\top q_2\right) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s - \ddot{\psi}_n^\top q_2\right)|X_2\right) ds dF_{X_2}(X_2)$$

$$= -\int \left(X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}\left(s - \ddot{\psi}_n^\top q_2\right) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s - \ddot{\psi}_n^\top q_2\right)|X_2\right) ds dF_{X_2}(X_2)$$

$$\quad -\sigma_n^2 \int \left(s - \ddot{\psi}_n^\top q_2\right)^2 \operatorname{sgn}\left(s - \ddot{\psi}_n^\top q_2\right) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s - \ddot{\psi}_n^\top q_2\right)|X_2\right) ds dF_{X_2}(X_2)$$

$$\quad -2\sigma_n \int \left(s - \ddot{\psi}_n^\top q_2\right) X_2^\top \dot{\delta}_2 \operatorname{sgn}\left(s - \ddot{\psi}_n^\top q_2\right) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s - \ddot{\psi}_n^\top q_2\right)|X_2\right) ds dF_{X_2}(X_2)$$

$$= I_1 + I_2 + I_3.$$

Due to Assumption $3(b)$, for any $\ddot{\psi}_n$,

$$I_3 \leq o(1) + \sigma_n \left|\ddot{\psi}_n\right| C_1 = o(1).$$

Similarly,

$$I_2 \leq o(1) + \sigma_n^2 \left(\left|\ddot{\psi}_n\right| C_2 + \left|\ddot{\psi}_n\right|^2 C_3\right) = o(1).$$

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

$$J_1 = \int \left(X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}(s) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s - \ddot{\psi}_n^\top q_2\right)|X_2\right) ds dF_{X_2}(X_2)$$

$$= \int_{A_n^C} \left(X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}(s) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s - \ddot{\psi}_n^\top q_2\right)|X_2\right) ds dF_{X_2}(X_2)$$

$$\quad + \int_{A_n} \left(X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}(s) q_2 \mathcal{K}'(s) f_{z|X_2}(0|X_2) ds dF_{X_2}(X_2)$$

$$\quad + \int_{A_n} \sigma_n\left(s - \ddot{\psi}_n^\top q_2\right) \left(X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}(s) q_2 \mathcal{K}'(s) f_{z|X_2}(\xi|X_2) ds dF_{X_2}(X_2),$$

where ξ lies between zero and $\sigma_n\left(s - \ddot{\psi}_n^\top q_2\right)$. The first term is $o(1)$ due to Assumption $3(c)$, so are the second and third due to Assumption $3(b)$ and the dominated convergence theorem. Then, for a constant $C > 0$,

$$J_2 = \int \left(X_2^\top \dot{\delta}_2\right)^2 \left(\operatorname{sgn}\left(s - \ddot{\psi}_n^\top q_2\right) - \operatorname{sgn}(s)\right) q_2 \mathcal{K}'(s) f_{z|X_2}\left(\sigma_n\left(s - \ddot{\psi}_n^\top q_2\right)|X_2\right) ds dF_{X_2}(X_2)$$

$$= 2\int (X_2^\top \dot{\delta}_2)^2 (1\{0 < s < \ddot{\psi}_n^\top q_2\} - 1\{\ddot{\psi}_n^\top q_2 < s < 0\}) q_2 \mathcal{K}'(s) f_{z|X_2}(\sigma_n(s - \ddot{\psi}_n^\top q_2)|X_2) ds dF_{X_2}(X_2)$$

$$= 2\int \left(\mathcal{K}\left(\ddot{\psi}_n^\top q_2\right) - \mathcal{K}(0)\right) \left(X_2^\top \dot{\delta}_2\right)^2 q_2 f_{z|X_2}(0|X_2) dF_{X_2}(X_2) + R_J,$$

29

and

$$R_J \leq \sigma_n C \int \left(-\ddot{\psi}_n^\top q_2\right) \left(X_2^\top \dot{\delta}_2\right)^2 q_2 dF_{X_2}(X_2).$$

Therefore, we conclude that, for any $\ddot{\psi}_n$,

$$\ddot{Q}_n = 2 \int \left(\mathcal{K}\left(\ddot{\psi}_n^\top q_2\right) - \mathcal{K}(0)\right) \left(X_2^\top \dot{\delta}_2\right)^2 q_2 f_{z|X_2}(0|X_2) dF_{X_2}(X_2) = o(1). \tag{18}$$

Now consider $\ddot{\psi}_n^\top \ddot{Q}_n / \left|\ddot{\psi}_n\right|$. By Assumption $3(b)$,

$$\left(\mathcal{K}\left(\ddot{\psi}_n^\top q_2\right) - \mathcal{K}(0)\right) \ddot{\psi}_n^\top q_2 / \left|\ddot{\psi}_n\right| = \left|\left(\mathcal{K}\left(\ddot{\psi}_n^\top q_2\right) - \mathcal{K}(0)\right) \ddot{\psi}_n^\top q_2\right| / \left|\ddot{\psi}_n\right|$$

and

$$\ddot{\psi}_n^\top \ddot{Q}_n / \left|\ddot{\psi}_n\right| = \int \left(X_2^\top \dot{\delta}_2\right)^2 f_{z|X_2}(0|X_2) \frac{\left|\left(\mathcal{K}\left(\ddot{\psi}_n^\top q_2\right) - \mathcal{K}(0)\right) \ddot{\psi}_n^\top q_2\right|}{\left|\ddot{\psi}_n\right|} dF_{X_2}(X_2) + o(1) > 0,$$

which leads to contradiction unless $\left|\ddot{\psi}_n\right| \to 0$. If $\left|\ddot{\psi}_n\right| \to 0$, applying the dominated convergence theorem,

$$\begin{aligned}
\ddot{Q}_n &= 2 \int \frac{\mathcal{K}\left(\ddot{\psi}_n^\top q_2\right) - \mathcal{K}(0)}{\ddot{\psi}_n^\top q_2} \left(X_2^\top \dot{\delta}_2\right)^2 q_2 q_2^\top \ddot{\psi}_n f_{z|X_2}(0|X_2) dF_{X_2}(X_2) \\
&= 2 \int \mathcal{K}'(0) \left(X_2^\top \dot{\delta}_2\right)^2 q_2 q_2^\top f_{z|X_2}(0|X_2) dF_{X_2}(X_2) \ddot{\psi}_n + o(1) \\
&= Q\ddot{\psi}_n + o(1) = o(1).
\end{aligned}$$

No contradiction. ∎

LEMMA 7 *Let* $\{\tilde{\theta}\}$ *be any sequence in* Θ *such that* $(\tilde{\theta} - \theta_0)/\sigma_n \to 0$ *as* $n \to \infty$. *Then* $\sqrt{\sigma_n} Q_n^{s\psi}\left(\tilde{\theta}, \sigma_n\right) = o_p(1)$ *and*

$$\sigma_n Q_n^\psi\left(\tilde{\theta}, \sigma_n\right) \to^p Q.$$

PROOF OF LEMMA 7: Let $\ddot{\psi}_n = \left(\tilde{\psi} - \psi_0\right)/\sigma_n$ and $s = z/\sigma_n + q_2^\top \ddot{\psi}_n$. For $\sqrt{\sigma_n} Q_n^{s\psi}(\tilde{\theta}, \sigma_n) = o_p(1)$, it is sufficient to note that

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

Now we derive the limit of $\sigma_n Q_n^\psi$. Since $\left(\tilde{\theta} - \theta_0\right)/\sigma_n \to 0$, it follows from the boundedness of moments and $\mathcal{K}''$ that

$$\sigma_n \frac{1}{n} \sum_{t=1}^n R_{nt}\left(\tilde{\theta}\right) \frac{q_{2t}q_{2t}^\top}{\sigma_n^2} \mathcal{K}''\left(\frac{z_t + q_{2t}^\top\left(\tilde{\psi} - \psi_0\right)}{\sigma_n}\right) = o_p\left(1\right).$$

By Lemma 5,

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

Let $\{e_n\}$ be a sequence such that $e_n \to \infty$ and $e_n \ddot{\psi}_n \to 0$ as $n \to \infty$. Define $C_n = \{q_2 : |q_2| \le e_n\}$,

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

and note that,

$$EQ_{nt}^\psi\left(\tilde{\theta}\right)$$
$$= -\sigma_n \int_{C_n} \left(\tilde{x}^\top \delta\right)^2 \operatorname{sgn}\left(z\right) \frac{q_2 q_2^\top}{\sigma_n^2} \mathcal{K}''\left(\frac{z}{\sigma_n} + q_2^\top \ddot{\psi}_n\right) f_{z|X_2}\left(z|X_2\right) dz dF_{X_2}(X_2)$$
$$\quad -\sigma_n \int_{C_n^c} \left(\tilde{x}^\top \delta\right)^2 \operatorname{sgn}\left(z\right) \frac{q_2 q_2^\top}{\sigma_n^2} \mathcal{K}''\left(\frac{z}{\sigma_n} + q_2^\top \ddot{\psi}_n\right) f_{z|X_2}\left(z|X_2\right) dz dF_{X_2}(X_2),$$
$$= -\int_{C_n/\{0\}} \left\{\left(\sigma_n s \dot{\delta}_1 + X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}\left(s - q_2^\top \ddot{\psi}_n\right)\right\} q_2 q_2^\top \mathcal{K}''\left(s\right) f_{z|X_2}\left(\left(s - q_2^\top \ddot{\psi}_n\right)\sigma_n|X_2\right) ds dF_{X_2}(X_2)$$
$$\quad -\int_{C_n^c} \left\{\left(\sigma_n s \dot{\delta}_1 + X_2^\top \dot{\delta}_2\right)^2 \operatorname{sgn}\left(s - q_2^\top \ddot{\psi}_n\right)\right\} q_2 q_2^\top \mathcal{K}''\left(s\right) f_{z|X_2}\left(\left(s - q_2^\top \ddot{\psi}_n\right)\sigma_n|X_2\right) ds dF_{X_2}(X_2)$$
$$= 2\mathcal{K}'\left(0\right) \int \left\{\left(X_2^\top \dot{\delta}_2\right)^2\right\} q_2 q_2^\top f_{z|X_2}\left(0|X_2\right) dF_{X_2}(X_2) + o\left(1\right),$$

where $o\left(1\right)$ follows from Assumption 3 (c) and the dominated convergence theorem. And for some $\eta > 0$,

$$\sup_{|c_n| < \eta} \left|\frac{1}{n} \sum_{t=1}^n Q_{nt}^\psi\left(\ddot{\psi}_n\right) - EQ_{nt}^\psi\left(\ddot{\psi}_n\right)\right| \to 0$$

by the same reasoning as for (16). ∎

Now, we prove the consistency of the variance estimators.

PROOF OF THEOREM 4. We first examine the convergence of $\hat{V}^s$. Let

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

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

$$
\tilde{\Gamma}_j = \begin{cases} \frac{1}{n} \sum_{t=j+1}^n \tau_t^s \tau_{t-j}^{s\top} & for\ j \geq 0, \\ \frac{1}{n} \sum_{t=-j+1}^n \tau_{t+j}^s \tau_t^{s\top} & for\ j < 0, \end{cases}
$$

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

It follows from Andrews (1991: Proposition 1 (c) and Theorem 1 (c)) that $\tilde{V}_n^s - V_n^s \to^p 0$ and $\hat{V}_n^s - \vec{V}_n^s \to^p 0$. Then, it remains to show that $\tilde{V}_n^s - \vec{V}_n^s \to^p 0$. 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 $\sum_{j=-n+1}^{n-1} w\left(\frac{j}{l_n}\right) < \infty$, it suffices to show that $\sup_j \left|\tilde{\Gamma}_j - \vec{\Gamma}_j\right| = o_p(1)$. But, for any nonzero a and β,

$$
\left| 1\{a > 0, \beta > 0\} - \mathcal{K}\left(\frac{a}{\sigma_n}\right) \mathcal{K}\left(\frac{\beta}{\sigma_n}\right) \right|
$$
$$
\leq 2\left(1 - \mathcal{K}\left(\frac{a}{\sigma_n}\right)\right) 1\{a > 0\} + 2\mathcal{K}\left(\frac{a}{\sigma_n}\right) 1\{a < 0\} + \mathcal{K}\left(\frac{\beta}{\sigma_n}\right) 1\{\beta < 0\}
$$
$$
\leq o\left(n^{-1}\right),
$$

which is sufficient for the purpose since, for any w_t s.t. $E|w_t|^2 < \infty$,

$$
\sup_j \left| n^{-1} \sum_{t=j+1}^n w_t w_{t-j} \left(1\{z_t > 0\} 1\{z_{t-j} > 0\} - \mathcal{K}\left(\frac{z_t}{\sigma_n}\right) \mathcal{K}\left(\frac{z_{t-j}}{\sigma_n}\right)\right) \right|
$$
$$
\leq \left[n^{-1} \sum_{t=1}^n |w_t|^2 \right]^{1/2} \left[n^{-1} \sum_{t=1}^n |w_t|^2 \sup_j \sum_{t=j+1}^n \left(\left| 1\{z_t > 0\} 1\{z_{t-j} > 0\} - \mathcal{K}\left(\frac{z_t}{\sigma_n}\right) \mathcal{K}\left(\frac{z_{t-j}}{\sigma_n}\right) \right| \right)^2 \right]^{1/2}
$$
$$
\leq \left[n^{-1} \sum_{t=1}^n |w_t|^2 \right] \sum_{t=1}^n \left(2\left| 1 - \mathcal{K}\left(\frac{z_t}{\sigma_n}\right) \right| 1\{z_t > 0\} + 3\left| \mathcal{K}\left(\frac{z_t}{\sigma_n}\right) \right| 1\{z_t < 0\} \right)
$$
$$
= o_p(1),
$$

by the dominated convergence theorem.

Next,

$$
\tau_{n,t}^\psi(\theta_n) = \left\{ \left(\tilde{x}_t^\top \delta_n\right)^2 - 2\tilde{x}_t^\top \delta_n \left(y_t - x_t^\top \beta_n\right) \right\} \frac{q_{2t}}{\sigma_n} \mathcal{K}'\left(\frac{q_{1t} + q_{2t}^\top \psi_n}{\sigma_n}\right)
$$
$$
= \left\{ \left(\tilde{x}_t^\top \delta + x_t^\top(\delta_n - \delta)\right)^2 - 2\left(\tilde{x}_t^\top \delta + x_t^\top(\delta_n - \delta)\right)\left(y_t - x_t^\top \beta_0 - x_t^\top(\beta_n - \beta_0)\right) \right\}
$$
$$
\times \frac{q_{2t}}{\sigma_n} \mathcal{K}'\left(\frac{z_t + q_{2t}^\top(\psi_n - \psi_0)}{\sigma_n}\right)
$$
$$
= -\left(\left(\tilde{x}_t^\top \delta\right)^2 \operatorname{sgn}(z_t) + 2\tilde{x}_t^\top \delta \varepsilon_t + R_{nt}(\theta_n) \right) \frac{q_{2t}}{\sigma_n} \mathcal{K}'\left(\frac{z_t + q_{2t}^\top(\psi_n - \psi_0)}{\sigma_n}\right).
$$

Then, since $\sqrt{n}\left(\theta_n - \theta_0\right) = O_p\left(1\right)$

$$
\begin{aligned}
\hat{V}^\psi &= \frac{\sigma_n}{n} \sum_{t=1}^n \tau_{n,t}^\psi\left(\theta_n\right) \tau_{n,t}^\psi\left(\theta_n\right)^\top \\
&= \frac{1}{n} \sum_{t=1}^n \left(\left(\tilde{x}_t^\top \delta\right)^2 \operatorname{sgn}\left(z_t\right) + 2\tilde{x}_t^\top \delta\varepsilon_t\right)^2 \frac{q_{2t}q_{2t}^\top}{\sigma_n} \mathcal{K}'\left(\frac{z_t + q_{2t}^\top\left(\psi_n - \psi_0\right)}{\sigma_n}\right)^2 \\
&\quad + \frac{2}{n} \sum_{t=1}^n \left(\left(\tilde{x}_t^\top \delta\right)^2 \operatorname{sgn}\left(z_t\right) + 2\tilde{x}_t^\top \delta\varepsilon_t\right) R_{nt}\left(\theta_n\right) \frac{q_{2t}q_{2t}^\top}{\sigma_n} \mathcal{K}'\left(\frac{z_t + q_{2t}^\top\left(\psi_n - \psi_0\right)}{\sigma_n}\right)^2 \\
&\quad + \frac{1}{n} \sum_{t=1}^n R_{nt}^2\left(\theta_n\right) \frac{q_{2t}q_{2t}^\top}{\sigma_n} \mathcal{K}'\left(\frac{z_t + q_{2t}^\top\left(\psi_n - \psi_0\right)}{\sigma_n}\right)^2 \\
&= \tilde{V}^\psi + o_p\left(1\right),
\end{aligned}
$$

where $\tilde{V}^\psi = \frac{1}{n} \sum_{t=1}^n \left(\left(\tilde{x}_t^\top \delta\right)^2 \operatorname{sgn}\left(z_t\right) + 2\tilde{x}_t^\top \delta\varepsilon_t\right)^2 \frac{q_{2t}q_{2t}^\top}{\sigma_n} \mathcal{K}'\left(\frac{z_t + q_{2t}^\top\left(\psi_n - \psi_0\right)}{\sigma_n}\right)^2$. And, for any ψ s.t. $\frac{\left(\psi - \psi_0\right)}{\sigma_n} \to 0$ as $n \to \infty$, by the same reasoning as the proof of Lemma 3,

$$
E\left(\left(\tilde{x}_t^\top \delta\right)^2 \operatorname{sgn}\left(z_t\right) + 2\tilde{x}_t^\top \delta\varepsilon_t\right)^2 \frac{q_{2t}q_{2t}^\top}{\sigma_n} \mathcal{K}'\left(\frac{z_t + q_{2t}^\top\left(\psi - \psi_0\right)}{\sigma_n}\right)^2 \to V^\psi.
$$

Therefore,

$$
E\tilde{V}^\psi \to V^\psi.
$$

For the variance of $\tilde{V}^\psi$, note that

$$
\left(n\sigma_n^2\right)^{-1} E\left(\left(\tilde{x}_t^\top \delta\right)^2 \operatorname{sgn}\left(z_t\right) + 2\tilde{x}_t^\top \delta\varepsilon_t\right)^4 \mathcal{K}'\left(\frac{z_t + q_{2t}^\top\left(\psi - \psi_0\right)}{\sigma_n}\right)^4 \operatorname{vec}\left(q_{2t}q_{2t}^\top\right) \operatorname{vec}\left(q_{2t}q_{2t}^\top\right)^\top \to 0,
$$

since the integral is bounded. And the cross product terms are negligible for the same reason as the proof of Lemma 3, which completes the proof that

$$
\tilde{V}^\psi \to^p V^\psi.
$$

$\blacksquare$

References

ANDREWS, D. W. K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817–858.

CARLSTEIN, E. (1986): "The use of subseries values for estimating the variance of a general statistic from a stationary sequence," *Ann. Statist.*, 14(3), 1171–1179.

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., AND R. S. TSAY (1998): "Limiting properties of the least squares estimator of a continuous threshold autoregressive model," *Biometrika*, 85(2), 413–426.

DE JONG, R., AND T. WOUTERSEN (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.

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

GONZALO, J., AND M. WOLF (2005): "Subsampling Inference in Threshold Autoregressive Models," *Journal of Econometrics*.

HALL, P. (1992): *The bootstrap and Edgeworth expansion*, Springer Series in Statistics. Springer-Verlag, New York.

HANSEN, B. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, 64, 413–430.

——— (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.

——— (1998): "Bootstrap methods for median regression models," *Econometrica*, 66(6), 1327–1351.

——— (2001): *The Bootstrap*vol. 5 of *Handbook of Econometrics*, chap. 52, pp. 3159–3228. Elsevier.

KIM, C.-J., AND C. R. NELSON (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.*, pp. 15–32. Inst. Math. Statist., Hayward, CA.

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," *J. Econometrics*, 3(3), 205–228.

POTTER, S. M. (1995): "A Nonlinear Approach to US GNP," *Journal of Applied Econometrics*, 10(2), 109–25.

SHAO, J., AND D. S. TU (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.

B Figures and Tables

B.1 Monte Carlo

				$z = q$					z is $N(0,1)$		
		0.25	0.50	1.0	1.5	2.0	0.25	0.50	1.0	1.5	2.0
$\widehat{\psi}_{LS}$	n=50	1.2514	0.8047	0.2335	0.1076	0.0793	1.0039	0.3917	0.0614	0.0441	0.0410
	n=100	1.1387	0.4255	0.1020	0.0539	0.0364	0.9337	0.1230	0.0296	0.0217	0.0205
	n=250	0.7991	0.1400	0.0371	0.0201	0.0155	0.2845	0.0300	0.0103	0.0083	0.0087
	n=500	0.2960	0.0590	0.0167	0.0099	0.0077	0.0785	0.0139	0.0054	0.0046	0.0046
	n=1000	0.1516	0.0272	0.0086	0.0051	0.0034	0.0278	0.0068	0.0026	0.0021	0.0021
$\widehat{\psi}^+$	n=50	1.4526	0.8655	0.2274	0.1241	0.1061	1.3251	0.5302	0.1012	0.0704	0.0646
	n=100	1.2934	0.3735	0.1159	0.0765	0.0607	1.1106	0.1498	0.0597	0.0402	0.0364
	n=250	0.8721	0.1287	0.0527	0.0380	0.0317	0.4307	0.0558	0.0271	0.0227	0.0185
	n=500	0.2895	0.0673	0.0307	0.0223	0.0198	0.0934	0.0306	0.0175	0.0119	0.0113
	n=1000	0.1257	0.0379	0.0191	0.0151	0.0131	0.0402	0.0176	0.0103	0.0074	0.0064
$\widehat{\psi}$	n=50	1.5129	1.1006	0.2645	0.1461	0.1423	1.4388	1.2621	0.1870	0.1282	0.1075
	n=100	1.5580	0.5378	0.1359	0.0934	0.0839	1.8743	0.8384	0.0892	0.0724	0.0580
	n=250	1.3445	0.1480	0.0633	0.0521	0.0444	1.9138	0.0766	0.0379	0.0309	0.0278
	n=500	0.3638	0.0742	0.0384	0.0291	0.0273	0.1721	0.0361	0.0228	0.0185	0.0183
	n=1000	0.1366	0.0372	0.0243	0.0196	0.0186	0.0475	0.0196	0.0126	0.0120	0.0105
as$\widehat{\psi}^+$	n=50	0.4984	0.2492	0.1246	0.0831	0.0623	0.2492	0.1246	0.0623	0.0415	0.0312
	n=100	0.3216	0.1608	0.0804	0.0536	0.0402	0.1608	0.0804	0.0402	0.0268	0.0201
	n=250	0.1771	0.0885	0.0443	0.0295	0.0221	0.0885	0.0443	0.0221	0.0148	0.0111
	n=500	0.1117	0.0559	0.0279	0.0186	0.0140	0.0559	0.0279	0.0140	0.0093	0.0070
	n=1000	0.0700	0.0350	0.0175	0.0117	0.0088	0.0350	0.0175	0.0088	0.0058	0.0044
as$\widehat{\psi}$	n=50	0.4958	0.2640	0.1603	0.1324	0.1212	0.2498	0.1355	0.0857	0.0728	0.0677
	n=100	0.3199	0.1703	0.1034	0.0854	0.0782	0.1611	0.0874	0.0553	0.0470	0.0437
	n=250	0.1762	0.0938	0.0569	0.0470	0.0430	0.0887	0.0481	0.0304	0.0259	0.0241
	n=500	0.1111	0.0592	0.0359	0.0297	0.0272	0.0560	0.0304	0.0192	0.0163	0.0152
	n=1000	0.0697	0.0371	0.0225	0.0186	0.0170	0.0351	0.0190	0.0120	0.0102	0.0095

Table 1a: std of estimates of ψ along with asymptotic predictions

		z = q					z is N(0,1)				
		0.25	0.50	1.0	1.5	2.0	0.25	0.50	1.0	1.5	2.0
$\widehat{\delta}_{2LS}$	n=50	0.8073	0.4849	0.2930	0.3083	0.2978	1.3119	0.8594	0.6130	0.5061	0.5341
	n=100	0.4971	0.2362	0.2020	0.1862	0.2072	0.8628	0.4635	0.3507	0.3775	0.3469
	n=250	0.2012	0.1163	0.1256	0.1281	0.1292	0.4133	0.2280	0.2154	0.2195	0.1984
	n=500	0.0959	0.0852	0.0876	0.0911	0.1017	0.1957	0.1587	0.1493	0.1472	0.1426
	n=1000	0.0616	0.0649	0.0630	0.0616	0.0677	0.1223	0.1073	0.1028	0.1138	0.1057
$\widehat{\delta}_2^+$	n=50	0.9516	0.5244	0.3132	0.2946	0.2986	1.9424	1.0380	0.6577	0.6658	0.7245
	n=100	0.5319	0.2433	0.2063	0.1883	0.2088	1.0608	0.5103	0.3957	0.4332	0.4330
	n=250	0.2024	0.1133	0.1231	0.1277	0.1327	0.4253	0.2335	0.2403	0.2447	0.2460
	n=500	0.0954	0.0866	0.0879	0.0879	0.1002	0.2113	0.1635	0.1619	0.1624	0.1583
	n=1000	0.0607	0.0642	0.0645	0.0607	0.0662	0.1178	0.1143	0.1037	0.1178	0.1163
$\widehat{\delta}_2$	n=50	1.1238	0.6689	0.3491	0.3505	0.3341	7.7863	6.1208	1.3128	1.2141	1.2464
	n=100	0.7613	0.2886	0.2303	0.2127	0.2235	8.2338	1.4964	0.6105	0.5978	0.6238
	n=250	0.2972	0.1292	0.1343	0.1338	0.1392	1.9394	0.3259	0.2766	0.2929	0.2934
	n=500	0.1091	0.0925	0.0881	0.0942	0.1014	0.3161	0.1806	0.1734	0.1797	0.1827
	n=1000	0.0671	0.0661	0.0639	0.0631	0.0722	0.1302	0.1131	0.1121	0.1228	0.1306
as	n=50	0.2000	0.2000	0.2000	0.2000	0.2000	0.4692	0.4692	0.4692	0.4692	0.4692
	n=100	0.1414	0.1414	0.1414	0.1414	0.1414	0.3318	0.3318	0.3318	0.3318	0.3318
	n=250	0.0894	0.0894	0.0894	0.0894	0.0894	0.2098	0.2098	0.2098	0.2098	0.2098
	n=500	0.0632	0.0632	0.0632	0.0632	0.0632	0.1484	0.1484	0.1484	0.1484	0.1484
	n=1000	0.0447	0.0447	0.0447	0.0447	0.0447	0.1049	0.1049	0.1049	0.1049	0.1049

Table 1b: std of estimates of δ_2 along with asymptotic predictions

δ_2			$z = q$					z is $N(0,1)$		
	0.25	0.50	1.0	1.5	2.0	0.25	0.50	1.0	1.5	2.0
n=50	0.9713	0.9952	0.9665	0.9426	0.9282	0.9856	0.9713	0.9426	0.9330	0.9282
n=100	0.9904	0.9904	0.9282	0.9187	0.9234	0.9856	0.9809	0.9426	0.9234	0.9282
n=250	0.9952	0.9617	0.9091	0.9139	0.8660	0.9904	0.9139	0.9139	0.9043	0.9474
n=500	0.9809	0.9569	0.9043	0.8469	0.8947	0.9713	0.9234	0.9139	0.9234	0.9378
n=1000	0.9617	0.9330	0.8947	0.8804	0.9091	0.9234	0.8852	0.9187	0.8995	0.8947

Table 2a. Percentile Bootstrap based on ψ_n

δ_2			$z = q$					z is $N(0,1)$		
	0.25	0.50	1.0	1.5	2.0	0.25	0.50	1.0	1.5	2.0
n=50	0.9282	0.9426	0.9569	0.9665	0.9187	0.9378	0.9713	0.9617	0.9474	0.9187
n=100	0.9330	0.9522	0.9426	0.9282	0.9234	0.9617	0.9713	0.9713	0.9522	0.9234
n=250	0.9569	0.9713	0.9282	0.8995	0.9091	0.9474	0.9522	0.9474	0.9330	0.9665
n=500	0.9713	0.9569	0.9330	0.8612	0.8900	0.9665	0.9474	0.9234	0.9569	0.9474
n=1000	0.9617	0.9474	0.9234	0.9043	0.9330	0.9234	0.8995	0.9187	0.9139	0.9091

Table 2b. Pivotal Bootstrap based on ψ_n

δ_2			$z = q$					z is $N(0,1)$		
	0.25	0.50	1.0	1.5	2.0	0.25	0.50	1.0	1.5	2.0
n=50	0.1483	0.3158	0.5167	0.6794	0.6746	0.1914	0.4163	0.6411	0.6986	0.7321
n=100	0.1340	0.3780	0.6603	0.7464	0.7608	0.2727	0.5933	0.7799	0.7703	0.7560
n=250	0.2727	0.6411	0.7943	0.8230	0.7560	0.4258	0.8182	0.8469	0.7847	0.8804
n=500	0.4019	0.7512	0.8852	0.8182	0.8373	0.6651	0.8900	0.8708	0.9139	0.8900
n=1000	0.5694	0.8612	0.8660	0.8660	0.8995	0.7943	0.8995	0.8756	0.8804	0.8708

Table 2c. Asymptotic interval based on ψ_n

δ_2		$z = q$					z is $N(0,1)$			
	0.25	0.50	1.0	1.5	2.0	0.25	0.50	1.0	1.5	2.0
n=50	0.9904	0.9904	0.9952	0.9474	0.9187	1.0000	0.9952	0.9904	0.9856	0.9904
n=100	1.0000	0.9904	0.9569	0.9378	0.9282	0.9952	0.9952	0.9761	0.9713	0.9856
n=250	0.9952	0.9761	0.9139	0.9234	0.8852	1.0000	0.9809	0.9522	0.9426	0.9139
n=500	0.9809	0.9713	0.9187	0.8660	0.8612	1.0000	0.9617	0.9091	0.9282	0.8947
n=1000	0.9761	0.9569	0.9091	0.8660	0.8947	0.9569	0.8852	0.9043	0.8900	0.8708

Table 3a. Percentile Bootstrap based on ψ_n^+

δ_2		$z = q$					z is $N(0,1)$			
	0.25	0.50	1.0	1.5	2.0	0.25	0.50	1.0	1.5	2.0
n=50	0.9856	0.9809	0.9856	0.9426	0.9187	0.9904	1.0000	0.9952	0.9713	0.9330
n=100	0.9809	0.9952	0.9761	0.8995	0.9043	0.9952	1.0000	0.9761	0.9474	0.9282
n=250	0.9856	0.9809	0.9043	0.8995	0.8565	1.0000	0.9809	0.9139	0.9043	0.8995
n=500	0.9904	0.9904	0.8995	0.8660	0.8612	1.0000	0.9378	0.8995	0.9378	0.9043
n=1000	0.9809	0.9569	0.8947	0.8708	0.9234	0.9665	0.8947	0.9043	0.8756	0.8708

Table 3b. Pivotal Bootstrap based on ψ_n^+

δ_2		$z = q$					z is $N(0,1)$			
	0.25	0.50	1.0	1.5	2.0	0.25	0.50	1.0	1.5	2.0
n=50	0.1340	0.2823	0.4354	0.5694	0.6124	0.1148	0.3014	0.5167	0.5502	0.6459
n=100	0.0718	0.2679	0.6124	0.6746	0.7177	0.1770	0.4163	0.6746	0.6842	0.7033
n=250	0.1818	0.5598	0.7847	0.8038	0.7943	0.3158	0.6507	0.7416	0.7656	0.8086
n=500	0.3062	0.7081	0.8278	0.7990	0.7895	0.5120	0.8565	0.8182	0.8469	0.8373
n=1000	0.4641	0.8325	0.8373	0.8421	0.8612	0.7033	0.8373	0.8469	0.8182	0.8325

Table 3c. Asymptotic interval based on ψ_n^+

δ_2	$z = q$					z is $N(0,1)$				
	0.25	0.50	1.0	1.5	2.0	0.25	0.50	1.0	1.5	2.0
n=50	0.9665	0.9426	0.9665	0.9378	0.9378	1.0000	0.9952	0.9665	0.9617	0.9426
n=100	0.9474	0.9617	0.9282	0.8995	0.8804	1.0000	0.9809	0.9426	0.9187	0.9282
n=250	0.9330	0.9617	0.9187	0.8947	0.8469	0.9952	0.9617	0.8756	0.9378	0.8995
n=500	0.9282	0.9234	0.8852	0.8565	0.8852	0.9952	0.8947	0.8565	0.9139	0.9139
n=1000	0.8756	0.9426	0.8995	0.8804	0.8565	0.9522	0.8995	0.9139	0.8947	0.9091

Table 4a. Percentile Bootstrap based on δ_{2n}

δ_2	$z = q$					z is $N(0,1)$				
	0.25	0.50	1.0	1.5	2.0	0.25	0.50	1.0	1.5	2.0
n=50	0.9665	0.9330	0.9665	0.9378	0.9187	0.9952	0.9856	0.9569	0.9569	0.9426
n=100	0.9378	0.9713	0.9330	0.8947	0.8947	1.0000	0.9665	0.9282	0.9234	0.9043
n=250	0.9330	0.9665	0.9234	0.8995	0.8421	0.9904	0.9426	0.8900	0.9474	0.9043
n=500	0.9378	0.9187	0.8804	0.8708	0.8852	0.9713	0.9043	0.8660	0.9234	0.9139
n=1000	0.8804	0.9474	0.8995	0.8756	0.8612	0.9426	0.8995	0.9139	0.9043	0.8947

Table 4b. Pivotal Bootstrap based on δ_{2n}

| | $z = q$ | | | | | z is $N(0,1)$ | | | | |
δ_2	0.25	0.50	1.0	1.5	2.0	0.25	0.50	1.0	1.5	2.0
n=50	0.9617	0.9569	0.9522	0.9282	0.8804	1.0000	0.9952	0.9856	0.9617	0.9856
n=100	0.9809	0.9617	0.9282	0.9091	0.9043	1.0000	1.0000	0.9904	0.9809	0.9426
n=250	0.9569	0.9330	0.9282	0.8900	0.8995	1.0000	0.9952	0.9330	0.9569	0.9282
n=500	0.9617	0.9378	0.9139	0.8852	0.9139	1.0000	0.9569	0.8756	0.9187	0.9043
n=1000	0.8852	0.9617	0.9426	0.9187	0.9091	0.9809	0.9139	0.9139	0.8804	0.8995

Table 5a. Percentile Bootstrap based on δ_{2n}^+

| | $z = q$ | | | | | z is $N(0,1)$ | | | | |
δ_2	0.25	0.50	1.0	1.5	2.0	0.25	0.50	1.0	1.5	2.0
n=50	0.8900	0.8565	0.9234	0.8900	0.9043	0.9091	0.8900	0.8804	0.8900	0.9282
n=100	0.8804	0.8852	0.9043	0.8804	0.8660	0.9522	0.8804	0.9330	0.9043	0.9234
n=250	0.9043	0.8756	0.8995	0.8947	0.8900	0.9569	0.8947	0.8469	0.9234	0.8995
n=500	0.8900	0.8900	0.8708	0.8852	0.9043	0.9426	0.8804	0.8421	0.9139	0.8995
n=1000	0.8469	0.9474	0.9426	0.9187	0.9091	0.9282	0.8804	0.9139	0.8660	0.8995

Table 5b. Pivotal Bootstrap based on δ_{2n}^+

	ci1	ci2	ci3	ci4	ci5	ci6	ci7	ci8
n=50	0.1790	0.2280	0.2330	0.1650	0.4190	0.2700	0.4920	0.7200
n=100	0.2150	0.2550	0.2670	0.1480	0.4160	0.3020	0.5310	0.7300
n=250	0.3090	0.3510	0.3550	0.1930	0.4710	0.3770	0.5740	0.7320
n=500	0.4560	0.5100	0.5160	0.2990	0.5980	0.5330	0.6820	0.7750
n=1000	0.6780	0.6910	0.6980	0.4720	0.7280	0.7120	0.8280	0.8500

Table 6a. Coverage of intervals for ψ

	ci1	ci2	ci3	ci4	ci5	ci6
n=50	0.4135	0.4975	0.5916	0.5974	1.0453	0.8201
n=100	0.4131	0.5384	0.6041	0.5336	0.9457	0.7365
n=250	0.3896	0.4991	0.5310	0.4410	0.8149	0.6022
n=500	0.3481	0.4467	0.4578	0.3650	0.6639	0.4982
n=1000	0.3070	0.3553	0.3574	0.3169	0.4768	0.3852

Table 6b. Length of Intervals for ψ

The intervals are all for the case $z = q, \delta_2 = 0.25, \sigma_n = 1.06n^{-1/5}$. c1 is using $\widehat{\psi}^+$ and Hessian; c2 is using ψ^+ and OPG; c3 is using ψ^+ and imposes homoskedasticity; c4 is using $\widehat{\psi}$; c5 is non-block diagonal using $\widehat{\psi}^+$ and Hessian; c6 is non-block diagonal using $\widehat{\psi}^+$ and OPG; c7 is the likelihood ratio based on $\widehat{\psi}^+$; c8 is the likelihood ratio based on $\widehat{\psi}$;

	ci1	ci2	ci3	ci4	ci5	ci6	ci7	ci8
n=50	0.5340	0.3820	0.6040	0.4640	0.5080	0.5090	0.6600	0.6100
n=100	0.5940	0.4620	0.6650	0.5260	0.4530	0.4550	0.6700	0.6440
n=250	0.6580	0.5920	0.7170	0.6620	0.4400	0.4550	0.7490	0.7250
n=500	0.7680	0.7410	0.8090	0.7830	0.5140	0.5130	0.8230	0.8110
n=1000	0.8620	0.8570	0.8760	0.8650	0.6000	0.5980	0.8760	0.8740

Table 6c. Coverage of Intervals for δ_2

	ci1	ci2	ci3	ci4	ci5	ci6
n=50	1.2045	0.8195	1.4895	1.1650	3.5939	4.5767
n=100	0.8245	0.6531	0.9664	0.7782	2.7405	3.8182
n=250	0.4676	0.4445	0.4774	0.4665	1.1578	1.2445
n=500	0.3062	0.3021	0.3010	0.3002	0.3800	0.3837
n=1000	0.2105	0.2095	0.2067	0.2057	0.2118	0.2139

Table 6d. Length of intervals for δ_2

The intervals are all for the case $z = q, \delta_2 = 0.25, \sigma_n = 1.06n^{-1/5}$. c1 is using $\widehat{\delta}_{2LS}$ imposing homoskedasticity; c1 is using $\widehat{\delta}_{2LS}$ not imposing homoskedasticity; c3 is using $\widehat{\delta}_2^+$ and imposing homoskedasticity; c4 is using $\widehat{\delta}_2^+$ and not imposing homoskedasticity; c5 is using $\widehat{\delta}_2$ and imposes homoskedasticity; c6 is using $\widehat{\delta}_2$ not imposing homoskedasticity; c7 is non-block diagonal using $\widehat{\delta}_2^+$ and Hessian; c8 is non-block diagonal using $\widehat{\delta}_2^+$ and OPG.

	ci1	ci2	ci3	ci4	ci5	ci6	ci7	ci8
n=50	0.1250	0.1940	0.2020	0.1450	0.3430	0.2420	0.4470	0.3830
n=100	0.1480	0.2120	0.2170	0.1280	0.3510	0.2400	0.4330	0.3600
n=250	0.2050	0.2710	0.2810	0.1860	0.3930	0.2920	0.4710	0.3990
n=500	0.2930	0.3690	0.3840	0.2770	0.4950	0.3760	0.5940	0.5330
n=1000	0.4630	0.5440	0.5510	0.4450	0.6460	0.5550	0.7110	0.6660

Table 7a. Coverage of intervals for ψ

	ci1	ci2	ci3	ci4	ci5	ci6
n=50	0.2298	0.3350	0.4315	0.3389	0.5941	0.5048
n=100	0.2119	0.3463	0.4191	0.2752	0.5791	0.4387
n=250	0.1898	0.3227	0.3583	0.2111	0.4865	0.3629
n=500	0.1662	0.2636	0.2869	0.1775	0.4300	0.2807
n=1000	0.1399	0.2050	0.2130	0.1414	0.3172	0.2114

Table 7b. Length of Intervals for ψ

The intervals are all for the case $z = q, \delta_2 = 0.25, \sigma_n = (\log n)n^{-1/2}$. c1 is using $\widehat{\psi}^+$ and Hessian; c2 is using ψ^+ and OPG; c3 is using ψ^+ and imposes homoskedasticity; c4 is using $\widehat{\psi}$; c5 is non-block diagonal using $\widehat{\psi}^+$ and Hessian; c6 is non-block diagonal using $\widehat{\psi}^+$ and OPG; c7 is the likelihood ratio based on $\widehat{\psi}^+$; c8 is the likelihood ratio based on $\widehat{\psi}$;

	ci1	ci2	ci3	ci4	ci5	ci6	ci7	ci8
n=50	0.5760	0.4070	0.5950	0.4100	0.5460	0.4850	0.5910	0.5760
n=100	0.5790	0.4520	0.5980	0.4830	0.5360	0.4830	0.6270	0.5960
n=250	0.6710	0.6170	0.7210	0.6430	0.6090	0.5830	0.7230	0.7000
n=500	0.7770	0.7410	0.8070	0.7680	0.7170	0.6920	0.8120	0.7820
n=1000	0.8650	0.8540	0.8900	0.8860	0.8170	0.8140	0.8960	0.8920

Table 7c. Coverage of Intervals for δ_2

	ci1	ci2	ci3	ci4	ci5	ci6
n=50	1.1594	0.8138	1.3097	0.9753	1.7240	1.4020
n=100	0.8234	0.6513	0.8868	0.7384	1.1802	0.9662
n=250	0.4693	0.4424	0.4756	0.4555	0.5362	0.4997
n=500	0.3071	0.3021	0.3044	0.3016	0.3084	0.3076
n=1000	0.2111	0.2096	0.2091	0.2083	0.2098	0.2093

Table 7d. Length of intervals for δ_2

The intervals are all for the case $z = q, \delta_2 = 0.25, \sigma_n = (\log n)n^{-1/2}$. c1 is using $\widehat{\delta}_{2LS}$ imposing homoskedasticity; c1 is using $\widehat{\delta}_{2LS}$ not imposing homoskedasticity; c3 is using $\widehat{\delta}_2^+$ and imposing homoskedasticity; c4 is using $\widehat{\delta}_2^+$ and not imposing homoskedasticity; c5 is using $\widehat{\delta}_2$ and imposes homoskedasticity; c6 is using $\widehat{\delta}_2$ not imposing homoskedasticity; c7 is non-block diagonal using $\widehat{\delta}_2^+$ and Hessian; c8 is non-block diagonal using $\widehat{\delta}_2^+$ and OPG.
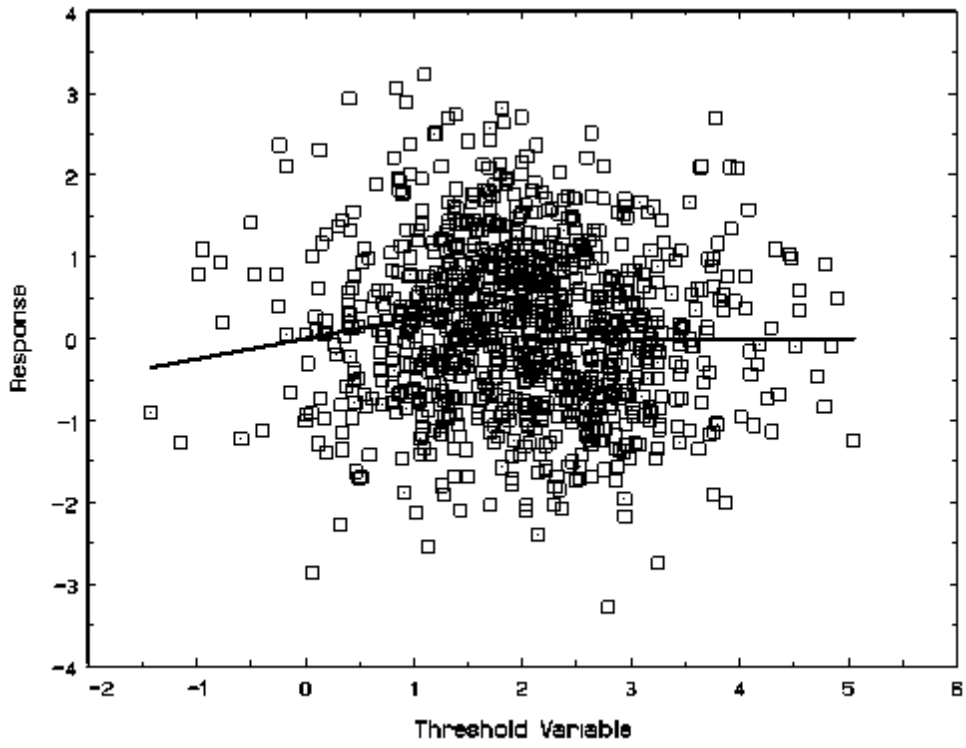
Figure 1.

B.2   Application

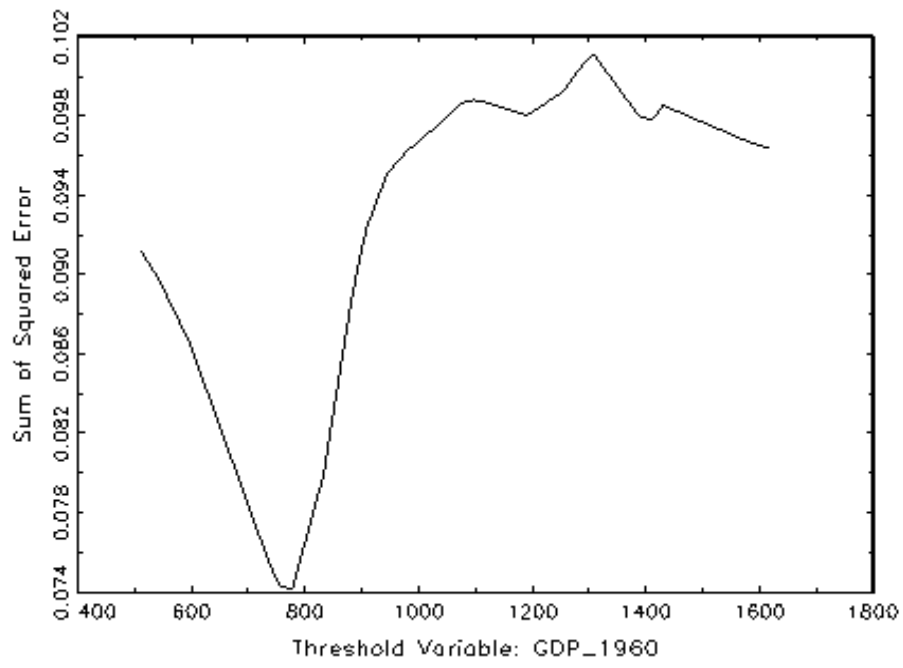| Variable	First Split		Second Split		Split with index	
	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
constant	5.09	1.85	3.55	2.42	2.30	5.13
GDP$_{1960}$	-0.53	0.20	-0.66	0.27	-0.11	0.34
INV/GDP	0.38	0.11	0.29	0.09	0.97	0.40
POP	-0.26	0.51	-0.48	0.51	-0.13	0.84
SCHOOL	0.25	0.08	-0.07	0.11	-0.19	0.15
δ						
constant	-1.25	2.28	-1.14	2.90	-0.08	5.24
GDP$_{1960}$	0.12	0.26	0.45	0.27	-0.18	0.36
INV/GDP	0.26	0.20	-0.06	0.24	-0.64	0.39
POP	-0.28	0.59	-0.01	0.72	-0.65	0.83
SCHOOL	-0.15	0.16	0.49	0.13	0.50	0.15

Table 8. Slope Estimates of Growth model
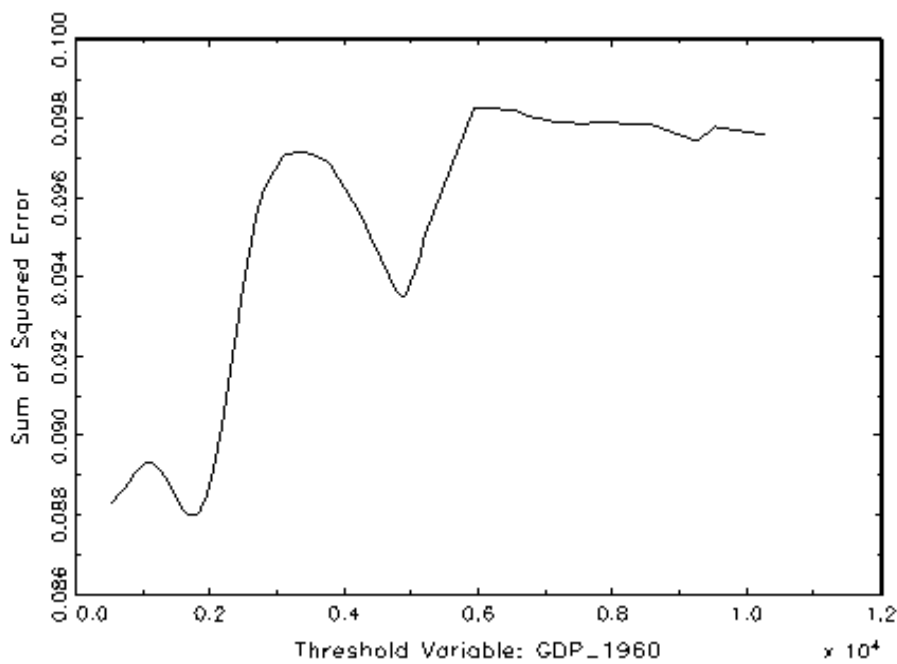
Figure 2. First Sample Split

47

Figure 3. Second Sample Split

48