Robust Inference for Threshold Regression Models[*]

Javier Hidalgo
London School of Economics

Jungyoon Lee
Royal Holloway, University London

Myung Hwan Seo[†]
Seoul National University

Abstract

This paper considers robust inference in threshold regression models when the practitioners do not know whether at the threshold point the true specification has a kink or a jump, nesting previous works that assume either continuity or discontinuity at the threshold. We find that the parameter values under the kink restriction are irregular points of the Hessian matrix, destroying the asymptotic normality and inducing the cube-root convergence rate for the threshold estimate. However, we are able to obtain the same asymptotic distribution as Hansen (2000) for the quasi-likelihood ratio statistic for the unknown threshold. We propose to construct confidence intervals for the threshold by bootstrap test inversion. Finite sample performances of the proposed procedures are examined through Monte Carlo simulations and an economic empirical application is given.

JEL Classification: C12, C13, C24.
Key words: Change Point, Kink, Grid Bootstrap, Cube Root.

1 INTRODUCTION

This paper examines robust inference in threshold models without a priori knowledge on whether the model is continuous or not at the threshold point. Since their introduction, threshold models have gained a lot of attention in econometrics, statistics and other fields, see Tong (1990) and Hansen (2000) among others. In the time series context, their popularity is due to the fact that they are capable of explaining nonlinear features present in many data such as chaos, cycles, irreversibility among others. In addition they have proved to have superior forecast performance in times of recession, see Tiao and Tsay (1994).

We nest previous works that assume either continuity or discontinuity at the threshold point and develop robust inference methods on the parameters of the model, which are valid under both specifications. The previous literature has explicitly assumed that either the threshold regression model has a kink or it is discontinuous at the threshold point. For instance, Chan (1993) and Hansen (2000) have focused on inference when the model is discontinuous at the threshold point, whereas Chan and Tsay (1998) , Hansen (2017) and Feder (1975a) have focused on inference in kink models. However, there is no a priori reason to believe that the model is or is not continuous.

We make an interesting finding that the estimator of the threshold parameter converges at the rate of $n^{1/3}$ if the model is continuous but the true restriction is not imposed in the estimation procedure. This is in clear contrast to the $n^{1/2}$ rate, which was first obtained by Feder (1975a) and in the time series context by Chan and Tsay (1998) by imposing the (true) constraint of a kink in its estimation. Furthermore, we show that the unconstrained estimator of the slope parameters is asymptotically independent of the estimator of the threshold point. The asymptotic independence also holds under the jump models of Chan (1993) or Hansen (2000) but not under the constrained estimation of Feder's (1975) or Chan and Tsay's (1998) kink models. These findings are interesting and new, when compared to standard results in regression models, where it is known that the consequence of not using the (true) restrictions is inefficiency but otherwise the asymptotic distribution is still Gaussian and the rate of convergence is the same.

Our preceding discussion motivates us to develop a robust inference in the threshold regression model. To that end, we first show that a quasi-likelihood ratio statistic for the threshold parameter has the same asymptotic distribution up to a scale constant that depends on whether

2

the true regression model has a kink or a jump. Second, we present an estimator for the scale factor based on the ratio of two kernel estimators. The consistency of this estimator is standard under the jump model but non-standard under the kink model because both its numerator and denominator converge to zero in probability. However, we prove that, similar to L'Hopital rule, the ratio of the two degenerating terms still converges in probability to the correct scale factor under the interesting requirement that higher-order kernels should not be used.

The last goal of this paper is to present valid bootstrap schemes as the asymptotic critical values appear to be a poor approximation to the finite-sample ones, as documented by Hansen (2000) and also in our Section 5. In addition, the validity of the bootstrap is of theoretical interest and it has not been established even under Hansen's (2000) shrinking jump design. The interest stems from two sets of findings in the literature regarding the failure of bootstrap for non-standard estimators: firstly with cube-root estimators such as the maximum score estimator, and secondly with super-consistent estimators such as the estimator of autoregressive coefficients of unit root processes and the threshold estimator under Chan's (1993) model, see Abrevaya and Huang (2005), Seijo and Sen (2011) and Yu (2014), just to name a few. Furthermore, we propose bootstrap test inversion confidence interval for the threshold, also known as the grid bootstrap in Hansen (1999), to enhance the finite-sample coverage probability.

A small Monte Carlo study reports good finite-sample performance of our bootstrap procedure for inference on the threshold location and the empirical application applies our robust inferential method to time series data on real GDP growth and debt-to-GDP ratio of a number of countries. Numerous works had fitted jump threshold models to a variety of datasets, see e.g. Caner, Grennes, and Koehler-Geib (2010), Cecchetti, Mohanty, and Zampolli (2011), and Lee et al. (2017), while Hansen (2017) had fitted kink threshold model to the US time series data. As there is little guidance from economic theory on suitability of jump or kink models, we advocate the use of our robust inference, and find substantial heterogeneity across countries in not just the estimated model parameters but also in the presence and location of threshold effect.

In Section 2 we introduce the model and present a set of regularity assumptions and describe how to estimate the parameters of the model. Section 3 then develops robust inferential methods for model parameters that are valid under both continuous and discontinuous settings, despite the slower rate of convergence for the estimate of the threshold under the kink specification. We then present in Section 4 a bootstrap algorithm for inference on the model

parameters, establishing their validity. Section 5 presents the results of a small Monte Carlo study, followed by Section 6, which contains the empirical application. Section 7 concludes. This paper has an appendix that contains some of the proofs and an online supplement that presents the remaining proofs, technical lemmas, and more numerical results for Section 5.

2   MODEL AND ESTIMATORS

We shall consider the following threshold regression model

$$y_t = \beta' x_t + \delta' x_t \mathbf{1}\{q_t > \gamma\} + \varepsilon_t, \tag{1}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function and x_t is a k-dimensional vector of regressors. The parameter γ is referred to as a threshold point, taking values in a compact parameter space Γ, which is a subset of the interior on the domain of the threshold variable q_t. It is worth mentioning that all our results hold true also when $q_t = t$, which is the case with structural break models. However, we have opted not to include this scenario for the sake of clarity and notational simplicity.

We assume that q_t is an element of the regressor vector x_t and denote

$$x_t = \left(1, x'_{t2}, q_t\right)'; \quad \delta = \left(\delta_1, \delta'_2, \delta_3\right)', \tag{2}$$

where δ is partitioned to match the dimensionality of x_t. Also we shall abbreviate $\mathbf{1}_t(\gamma) = \mathbf{1}\{q_t > \gamma\}$ and $x_t(\gamma) = (x'_t, x'_t \mathbf{1}_t(\gamma))'$, so that we can write (1) as

$$
\begin{aligned}
y_t &= \beta' x_t + \delta_1 \mathbf{1}_t(\gamma) + \delta'_2 x_{t2} \mathbf{1}_t(\gamma) + \delta_3 q_t \mathbf{1}_t(\gamma) + \varepsilon_t \\
&= \alpha' x_t(\gamma) + \varepsilon_t, \quad \text{where} \quad \alpha = (\beta', \delta')'.
\end{aligned}
\tag{3}
$$

Before stating some regularity assumptions on the model, we need to introduce some extra notation. Let $f(\cdot)$ denote the density function of q_t, which we assume to exist, and $\sigma^2(\gamma) = E\left(\varepsilon_t^2 \mid q_t = \gamma\right)$, the conditional variance function of error term, while $\sigma^2 = E(\varepsilon_t^2)$ denotes the unconditional variance. Denote $k \times k$ matrices $D(\gamma) = E\left(x_t x'_t | q_t = \gamma\right)$, $V(\gamma) = E\left(x_t x'_t \varepsilon_t^2 | q_t = \gamma\right)$ and let $D = D(\gamma_0)$ and $V = V(\gamma_0)$. As usual the "0" subscript on a parameter indicates its true unknown value. Finally, let $M = E(\mathbf{x_t x'_t})$ and $\Omega = E(\mathbf{x_t x'_t} \varepsilon_t^2)$ with $\mathbf{x_t} = x_t(\gamma_0)$.

4

Assumption Z. *Let $\{x_t, \varepsilon_t\}_{t \in \mathbb{Z}}$ be a strictly stationary, ergodic sequence of random variables such that their ρ-mixing coefficients satisfy $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$ and $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_t$ is the filtration up to time t. Furthermore, $M, \Omega > 0$, $E \|x_t\|^4 < \infty$, $E \|x_t \varepsilon_t\|^4 < \infty$ and $E |\varepsilon_t|^{4+\eta} < \infty$ for some $\eta > 0$.*

Assumption Q. *The functions $f(\gamma)$, $V(\gamma)$ and $D(\gamma)$ are continuous at $\gamma = \gamma_0$. For all $\gamma \in \Gamma$, the functions $f(\gamma)$, $E(x_t x_t' \mathbf{1}\{q_t \leq \gamma\})$ and $E(x_{t2} x_{t2}' | q_t = \gamma)$ are positive and continuous, and the functions $f(\gamma)$, $E(|x_t|^4 | q_t = \gamma)$ and $E(|x_t \varepsilon_t|^4 | q_t = \gamma)$ are bounded by some $C < \infty$.*

Assumptions **Z** and **Q** are commonly imposed on the distribution of $\{x_t, \varepsilon_t\}$, see e.g. Hansen (2000), so his comments apply here. As discussed therein, the self-exciting threshold autoregressive model of Tong (1990) satisfies Assumption **Z**. The condition for $E(x_{t2} x_{t2}' | q_t = \gamma)$ is written in terms of x_{t2} as the other elements in x_t are fixed given $q_t = \gamma$. While we allow conditional heteroscedasticity of a general form, Assumption **Q** requires continuity of the conditional variance function $\sigma^2(\cdot)$ at γ_0.

2.1 Estimators

We estimate $\theta_0 = (\alpha_0', \gamma_0)'$ by the (non-linear) least squares estimator (LSE), that is,

$$\widehat{\theta} = \left(\widehat{\alpha}', \widehat{\gamma}\right)' := \underset{\theta \in \Theta}{\operatorname{argmin}} \, \mathbb{S}_n(\theta), \tag{4}$$

where $\Theta = (\Lambda, \Gamma)$ is a compact set in $\mathbb{R}^{2k+1}$ and

$$\mathbb{S}_n(\theta) := \frac{1}{n} \sum_{t=1}^{n} \left(y_t - \alpha' x_t(\gamma)\right)^2, \tag{5}$$

which is a step function in γ at q_t's. For its computation, we shall employ a step-wise algorithm. To that end, one could employ the grid search algorithm on $\Gamma_n = \Gamma \cap \{q_1, ..., q_n\}$ to find $\widehat{\gamma}$. Define the concentrated sum of squared residuals

$$\widehat{\mathbb{S}}_n(\gamma) := \frac{1}{n} \sum_{t=1}^{n} \left(y_t - \widehat{\alpha}'(\gamma) x_t(\gamma)\right)^2, \tag{6}$$

where

$$\widehat{\alpha}(\gamma) := \underset{\alpha \in \Lambda}{\operatorname{argmin}} \, \frac{1}{n} \sum_{t=1}^{n} \left(y_t - \alpha' x_t(\gamma)\right)^2 \tag{7}$$

is the LSE of α for a given γ. Then, our estimator of α is $\widehat{\alpha} := \widehat{\alpha}(\widehat{\gamma})$ with

$$\widehat{\gamma} := \operatorname*{argmin}_{\gamma \in \Gamma_n} \widehat{\mathbb{S}}_n(\gamma). \tag{8}$$

This is the unconstrained LSE and for comparison we also describe the continuity constrained least squares estimator (CLSE), which minimizes (5) under Assumption **C** in the next section,

$$\widetilde{\theta} = (\widetilde{\alpha}', \widetilde{\gamma})' := \operatorname*{argmin}_{\theta \in \Theta: \delta_1 + \delta_3 \gamma = 0; \delta_2 = 0} \mathbb{S}_n(\theta). \tag{9}$$

This estimator was considered by Feder (1975a) and later by Chan and Tsay (1998) or Hansen (2017), who have established the asymptotic normality of $\widetilde{\theta}$ with the standard squared root consistency.

3 Robust Confidence Regions

This section presents our main results, namely how to perform robust inference in threshold models and in particular on the location of the threshold point. We begin with developing inference methods for the regression coefficients α_0 and the unknown threshold γ_0 based on the LSE $\widehat{\theta}$ when the true regression model has a kink. Then, they are compared with other inference methods that are developed under different sampling schemes such as Hansen (2000). In particular, we show that a judicious choice of statistics enables us to perform a robust inference in the sense that the same critical values can be employed for inference whether the model has a kink or a jump. That is, we do not need to know whether the model has a kink or a jump to make inference for the parameters α_0 and γ_0. As mentioned in the introduction the motivation comes from the rather surprising results given in Proposition 1 and Theorem 1 below.

First we state the kink model in terms of assumption.

Assumption C. *Assume that $\delta_{30} \neq 0$ and*

$$\delta_{10} + \delta_{30}\gamma_0 = 0; \quad \delta_{20} = 0. \tag{10}$$

Under Assumption *C* the model (3) is written as

$$y_t = x_t'\beta_0 + \delta_{30}(q_t - \gamma_0)\mathbf{1}_t(\gamma_0) + \varepsilon_t. \tag{11}$$

6

Feder (1975), Chan and Tsay (1998), and Hansen (2017) considered the estimation of the model (11) along with an auxiliary condition of $\delta_{30} \neq 0$ to ensure the identification of the change-point γ_0. This is a model with a kink.

Then, the next proposition establishes the consistency and rates of convergence of the LSE $\widehat{\theta}$ defined in (4) under Assumption **C**.

Proposition 1. *Under Assumptions* **C**, **Z** *and* **Q**, *we have that*

$$\widehat{\alpha} - \alpha_0 = O_p(n^{-1/2}) \quad and \quad \widehat{\gamma} - \gamma_0 = O_p(n^{-1/3}).$$

The results of Proposition 1 are surprising because the convergence rate of $\widehat{\gamma}$ is slower than that of the CLSE $\widetilde{\gamma}$, which is known to be $n^{-1/2}$ as shown in the aforementioned works. That is, using the true restriction on the parameters leads to a faster rate of convergence of the estimator of γ_0, not just reducing its asymptotic variance as is often the case.

Next we present the asymptotic distribution of $\widehat{\theta}$.

Theorem 1. *Let Assumptions* **C**, **Z** *and* **Q** *hold and* $B_1(\cdot)$ *and* $B_2(\cdot)$ *be two independent standard Brownian motions. Define* $W(g) := B_1(-g)\mathbf{1}\{g < 0\} + B_2(g)\mathbf{1}\{g > 0\}$. *Then,*

$$n^{1/2}(\widehat{\alpha} - \alpha_0) \xrightarrow{d} \mathcal{N}\left(0, M^{-1}\Omega M^{-1}\right)$$

$$n^{1/3}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \underset{g \in \mathbb{R}}{\operatorname{argmax}}\left(2\delta_{30}\sqrt{\frac{\sigma^2(\gamma_0)f(\gamma_0)}{3}}W(g^3) + \frac{\delta_{30}^2}{3}f(\gamma_0)|g|^3\right),$$

where the two limit distributions are independent of each other.

The asymptotic independence is a consequence of the different convergence rates between the two sets of estimators $\widehat{\alpha}$ and $\widehat{\gamma}$ by similar arguments as in Chan (1993), albeit the rate for $\widehat{\gamma}$ being slower than that for $\widehat{\alpha}$ in our case. The asymptotic independence does not hold for the CLSE $\widetilde{\gamma}$ and $\widetilde{\alpha}$, which converge at the same rate as mentioned above and they are jointly asymptotically normal with a non-diagonal variance-covariance matrix.

Theorem 1 suggests that Gonzalo and Wolf's (2005) subsampling procedure would be correct if they had used the normalization $n^{1/3}$ instead of the incorrect one $n^{1/2}$. On the other hand, it is worth mentioning that Seo and Linton (2007) considered the smoothed least squares estimator for the same setup. The convergence rate for their smoothed least squares estimator for γ was slower than our cube-root rate under their assumptions for the smoothing parameter.

Remark 1. *We now present a heuristic discussion to illustrate why the constrained and unconstrained estimators of γ_0 have different rates of convergence and the unconstrained estimator belongs to the cube-root class explored by Kim and Pollard (1990) for the i.i.d. data and Seo and Otsu (2018) for more general setups. For simplicity of illustration, we begin with a simplified model, where $x_t = (1, q_t)'$, $\delta = (\delta_1, \delta_3)'$, β is fixed at $\beta_0 = 0$, and thus $\theta = (\delta', \gamma)'$. In addition we shall assume $\gamma_0 = 0$ and thus $\delta_{10} = 0$ by ([10](#)) without loss of generality since we can always rename the variable $q_t - \gamma_0$ as q_t. It is well known that the rates of convergence of an M-estimator is governed by the local behavior of its criterion function around the true value provided that the estimator is consistent. Then the convergence rate of LSE $\widehat{\theta} = \left(\widehat{\delta}', \widehat{\gamma}\right)'$ is determined by the stochastic expansion of*

$$\mathbb{S}_n(\theta) - \mathbb{S}_n(\theta_0) \tag{12}$$
$$= \frac{1}{n}\sum_{t=1}^n \left(\delta_{30}q_t \mathbf{1}_t(0) - (\delta_1 + \delta_3 q_t)\,\mathbf{1}_t(\gamma)\right)^2 + \frac{2}{n}\sum_{t=1}^n \varepsilon_t \left(\delta_{30}q_t \mathbf{1}_t(0) - (\delta_1 + \delta_3 q_t)\,\mathbf{1}_t(\gamma)\right),$$

in small neighborhoods of $\delta = \delta_0$ and $\gamma = \gamma_0 = 0$. Consider $\gamma > 0$. The case of $\gamma < 0$ is handled similarly. Then, as $\mathbf{1}_t(0) = \mathbf{1}_t(\gamma) + \mathbf{1}\{0 < q_t \le \gamma\}$ and $\mathbf{1}_t(\gamma)\mathbf{1}\{0 < q_t \le \gamma\} = 0$,

$$E\left(\delta_{30}q_t\mathbf{1}_t(0) - (\delta_1 + \delta_3 q_t)\,\mathbf{1}_t(\gamma)\right)^2$$
$$= E\left(\delta_1 + (\delta_3 - \delta_{30})\,q_t\right)^2 \mathbf{1}_t(\gamma) + E\left(\delta_{30}q_t\right)^2\mathbf{1}\{0 < q_t \le \gamma\}$$
$$\sim \|\delta - \delta_0\|^2 + \gamma^3,$$

because for some positive constant c,

$$E\left[q_t^2 \mathbf{1}\{0 < q_t \le \gamma\}\right] = \int_0^\gamma q^2 f(q)\, dq \sim \frac{c}{3}|\gamma|^3 \tag{13}$$

due to Assumption $\mathbf{Q}$. This cubic approximation at $\gamma = \gamma_0$ is non-standard and invalidates the asymptotic normality of $\widehat{\gamma}$, which builds on the quadratic approximation.[1] Similarly,

$$\mathrm{var}\left(\frac{1}{n}\sum_{t=1}^n \varepsilon_t \left(\delta_{30}q_t\mathbf{1}_t(0) - (\delta_1 + \delta_3 q_t)\,\mathbf{1}_t(\gamma)\right)\right) \sim \frac{\|\delta - \delta_0\|^2 + |\gamma|^3}{n}. \tag{14}$$

---

[1]This also shows that the asymptotic variance formula $U^{-1}VU^{-1}$ in Gonzalo and Wolf's (2005) Theorem A.1 and Remark A.1 is not properly defined due to the degeneracy of U, where U is the second derivative matrix of the expected criterion function that is evaluated under the continuity restriction.

Thus, the last two displayed expressions in (13) and (14) suggest that

$$\widehat{\delta} - \delta_0 = O_p\left(n^{-1/2}\right) \quad and \quad \widehat{\gamma} = O_p\left(n^{-1/3}\right),$$

as these rates of convergence balance the speeds at which the bias and standard deviation of $\mathbb{S}_n(\theta) - \mathbb{S}_n(\theta_0)$ *converge to zero. In comparison, the CLSE* $\left(\widetilde{\delta}_3, \widetilde{\gamma}\right)'$ *is governed by*

$$
\begin{aligned}
&\mathbb{S}_n(\theta) - \mathbb{S}_n(\theta_0) \\
= \ & \frac{1}{n} \sum_{t=1}^{n} \left(\delta_{30} q_t \mathbf{1}_t(0) - \delta_3 (q_t - \gamma) \mathbf{1}_t(\gamma)\right)^2 + \frac{2}{n} \sum_{t=1}^{n} \varepsilon_t \left(\delta_{30} q_t \mathbf{1}_t(0) - \delta_3 (q_t - \gamma) \mathbf{1}_t(\gamma)\right),
\end{aligned}
$$

due to the continuity constraint (10), for which we observe the quadratic expansion

$$
\begin{aligned}
E\left(\delta_{30} q_t \mathbf{1}_t(0) - \delta_3 (q_t - \gamma) \mathbf{1}_t(\gamma)\right)^2 &\sim |\delta_3 - \delta_{30}|^2 + \gamma^2 \\
\operatorname{var}\left(\frac{2}{n} \sum_{t=1}^{n} \varepsilon_t \left(\delta_{30} q_t \mathbf{1}_t(0) - \delta_3 (q_t - \gamma) \mathbf{1}_t(\gamma)\right)\right) &\sim \frac{|\delta_3 - \delta_{30}|^2 + |\gamma|^2}{n}.
\end{aligned}
$$

This yields that

$$\widetilde{\delta}_3 - \delta_{30} = O_p\left(n^{-1/2}\right) \quad and \quad \widetilde{\gamma} = O_p\left(n^{-1/2}\right),$$

which coincides with the rates of convergence that both Feder (1975a, b) and Chan and Tsay (1998) obtained.

An intuitive explanation for the preceding Proposition, Theorem and Remark is to appeal to "misspecification". Although the unconstrained model (1) encompasses both continuous and discontinuous models, the estimated regression function is almost surely discontinuous, since the probability that the LSE $\widehat{\theta}$ fulfills the continuity restriction is zero.

3.1 Inference on the Regression Coefficient α

Theorem 1 in Section 3.1, Lemma A.12 of Hansen (2000) and Theorem 2 of Chan (1993) report the same asymptotic distribution for $\widehat{\alpha}$, namely $\mathcal{N}\left(0, M^{-1}\Omega M^{-1}\right)$, which is asymptotically independent of $\widehat{\gamma}$. Thus, the inference for α_0 is uniform under any widely used sampling scheme with strongly identified γ_0, provided that the respective sample moments

$$\widehat{M} = \frac{1}{n} \sum_{t=1}^{n} x_t(\widehat{\gamma}) x_t(\widehat{\gamma})'; \quad \widehat{\Omega} = \frac{1}{n} \sum_{t=1}^{n} x_t(\widehat{\gamma}) x_t(\widehat{\gamma})' \widehat{\varepsilon}_t^2,$$

9

where $\widehat{\varepsilon}_t = y_t - x_t\,(\widehat{\gamma})'\,\widehat{\alpha}$, are consistent under each data generating process. This is the case due to the uniform law of large numbers, which only requires consistency of $\widehat{\gamma}$.

It is worthwhile to mention that this "oracle" property of $\widehat{\alpha}$ does not hold true for the CLSE $\widetilde{\alpha}$, whose asymptotic distribution is affected by that of $\widetilde{\gamma}$, as was first noticed and shown by Feder (1975a) and later extended to time series data by Chan and Tsay (1998).

3.2   Inference on the Threshold γ

The main purpose of this section is to develop a method to construct confidence regions for γ_0 that are valid regardless of whether the regression model has a kink or a jump at the true value of γ_0. Conventionally, inference on γ has been done after assuming either that the model has a kink or that it has a jump, i.e. the practitioner chooses between jump or kink models before estimating the threshold point. More specifically, if we decide that the model has a jump, then one follows e.g. Hansen (2000), whereas if one has chosen the kink model then one needs to employ the asymptotic normal inference as in Feder (1975a) and others. One of our findings is that Hansen (2000) results are not valid if the model had a kink and likewise Feder's results are not valid if the model had a jump.

Thus, this section develops robust confidence regions that are valid regardless which of the two models is the true specification. To ease reference, we recall Hansen's (2000) diminishing jump specification:

Assumption J. *For some* $0 < \varphi < 1/2$ *and* $d \neq 0$, $\delta_0 = d \cdot n^{-\varphi}$ *and* $d'Vd$ *and* $d'Dd$ *are positive for all* n.

When φ is greater than or equal to $1/2$, δ_0 is too small to consistently estimate γ_0, and such case is excluded. And we suppress the dependence of δ_0 on the sample size n to simplify the notation.

To develop robust confidence sets, we need to find a statistic whose asymptotic distribution is invariant to the true parameter value, that is, a statistic whose asymptotic distribution does not change suddenly under Assumption **C**. We begin by introducing a Gaussian quasi-likelihood ratio statistic based on the unconstrained model (1). Specifically, let

$$QLR_n = n\frac{\widehat{\mathbb{S}}_n\,(\gamma_0) - \widehat{\mathbb{S}}_n\,(\widehat{\gamma})}{\widehat{\mathbb{S}}_n\,(\widehat{\gamma})},$$

where $\widehat{\mathbb{S}}_n(\gamma)$ is defined in (6).

We now derive the following asymptotic distribution for QLR_n, which contrasts with the asymptotic distribution obtained by Hansen (2000) under Assumption **J**.

Proposition 2. *Suppose that Assumptions **C**, **Z** and **Q** hold. Then, as $n \to \infty$,*

$$QLR_n \xrightarrow{d} \zeta \max_{g \in \mathbb{R}} \left(2W(g) - |g| \right),$$

where

$$\zeta = \frac{\sigma^2(\gamma_0)}{\sigma^2}.$$

In comparison, we recall Hansen's (2000) results that

$$QLR_n \xrightarrow{d} \xi \max_{g \in \mathbb{R}} \left(2W(g) - |g| \right), \tag{15}$$

where

$$\xi = \frac{E\left(\left(x_t' d\varepsilon_t \right)^2 | q_t = \gamma_0 \right)}{\sigma^2 E\left(\left(x_t' d \right)^2 | q_t = \gamma_0 \right)},$$

and that the distribution function of $\max_{g \in \mathbb{R}} \left(2W(g) - |g| \right)$ is given by $F(z) = \left(1 - e^{-z/2} \right)^2$.

The results of our Proposition 2 and that in (15) indicate that the only difference between the limit distributions of QLR_n under the kink and jump specifications is the scaling factor. This is the case despite the fact the estimator $\widehat{\gamma}$ exhibits different rates of convergence across the two settings.

Next, we propose an estimator of the unknown scaling of QLR_n that converges in probability to ξ under Assumption **J**, while it converges to ζ under Assumption **C**, thus adapting to the unknown true scaling in each situation. We begin with a natural estimator of ξ, which is a ratio of two Nadaraya-Watson estimators of the conditional expectations. That is,

$$\widehat{\xi} = \frac{\frac{1}{n} \sum_{t=1}^n \left(\widehat{\delta}' x_t \right)^2 \widehat{\varepsilon}_t^2 K\left(\frac{q_t - \widehat{\gamma}}{a} \right)}{\mathbb{S}_n\left(\widehat{\theta} \right) \frac{1}{n} \sum_{t=1}^n \left(\widehat{\delta}' x_t \right)^2 K\left(\frac{q_t - \widehat{\gamma}}{a} \right)}, \tag{16}$$

where $K(\cdot)$ and a are, respectively, the kernel function and bandwidth parameter and $\widehat{\varepsilon}_t$'s are the least squares residuals. The consistency of $\widehat{\xi}$ to ξ is standard, as argued in Hansen (2000).

However, it is not trivial to establish that $\widehat{\xi} \xrightarrow{p} \zeta$ when the true model has a kink at γ_0 because both numerator and denominator degenerates asymptotically in Assumption **C**. It turns out that we need to impose some unconventional restrictions on the kernel function K and the bandwidth a. Specifically, we assume

Assumption K. *Assume the following for $K(\cdot)$ and a.*

K1 $K(\cdot)$ is symmetric and $\kappa_\ell = \int_{-\infty}^{\infty} u^\ell K(u)\,du < C$ for $\ell \leq 4$ and $\kappa_2 \neq 0$.

K2 $K(\cdot)$ is twice continuously differentiable with the first derivative $K'(\cdot)$ and for all u such that $|w/u| \leq C$ as $w \to 0$ $K'(u+w)/K'(u) \to 1$.

K3 $K(u) = \int \phi(v)e^{ivu}dv$, where the characteristic function $\phi(v)$ satisfies that $v\phi(v)$ is integrable.

K4 $a^{-3}n^{-1} + a \to 0$ as $n \to \infty$.

It is clear that the Epanechnikov and the Gaussian kernel functions satisfy **K1**, **K2** and **K3**. One important observation is that **K1** rules out higher-order kernels by assuming $\kappa_2 \neq 0$. The consequence of dropping the assumption that $\kappa_2 \neq 0$ is discussed in detail in Remark 2 that follows the next proposition.

Proposition 3. *Suppose Assumptions* **Z**, **Q** *and* **K** *hold true. Then, under Assumption* **C**

$$\widehat{\xi} \xrightarrow{P} \zeta,$$

while $\widehat{\widehat{\xi}} \xrightarrow{P} \xi$ *under Assumption* **J**.

Remark 2. *We now comment on the consequence of dropping the assumption that $\kappa_2 \neq 0$. If we allowed for higher-order kernels, that is $\kappa_2 = 0$ and $\kappa_3 = 0$ but $\kappa_4 \neq 0$, $\widehat{\xi}$ would not be consistent. Indeed, Proposition 3 and Lemma 2 in the Appendix indicate that, without loss of generality for $\gamma_0 = 0$ and $\sigma^2 = 1$, $\widehat{\xi}$ converges in probability to*

$$\frac{\frac{\partial^2}{\partial q^2} f(q)\, g_0(q)\,|_{q=0}}{\frac{\partial^2}{\partial q^2} f(q)\, g_0^*(q)\,|_{q=0}},$$

where $g_r(q) = E\left(x_{t2}^r \varepsilon_t^2 \mid q_t = q\right)$ and $g_r^*(q) = E\left(x_{t2}^r \mid q_t = q\right)$. This is the case because dropping in **K1** *the assumption of $\kappa_2 \neq 0$ and letting $\kappa_2 = \kappa_3 = 0$, the numerator in (16) will be*

$$\kappa_4 \delta_3^2 a^4 \frac{\partial^2}{\partial q^2} \left(f(0)\, g_0(0)\right)\left(1 + o_p(1)\right),$$

whereas the denominator in (16) becomes

$$\kappa_4 \delta_3^2 a^4 \frac{\partial^2}{\partial q^2} \left(f(0)\, g_0^*(0)\right)\left(1 + o_p(1)\right).$$

So that, unless $E(\varepsilon_t^2 \mid q_t = \gamma_0) = E(\varepsilon_t^2)$, we obtain that (similar to the L'Hopital rule):

$$\widehat{\xi} \xrightarrow{P} \frac{\frac{\partial^2}{\partial q^2} f(q) g_0(q) \mid_{q=0}}{\frac{\partial^2}{\partial q^2} f(q) g_0^*(q) \mid_{q=0}} = \frac{\frac{\partial^2}{\partial q^2} \left(f(q) E\left[\varepsilon_t^2 \mid q_t = q \right] \right) \mid_{q=0}}{\frac{\partial^2}{\partial q^2} f(q) \mid_{q=0}} \neq \zeta,$$

and hence $\widehat{\xi}$ would not be a consistent estimator of the scale factor ζ.

We can construct the $100s$ percent confidence set of γ_0 by

$$\widehat{\Gamma}_s = \left\{ \gamma \in \Gamma : \widehat{\xi}^{-1} QLR_n(\gamma) \leq F^{-1}(s) \right\}.$$

As we have already argued, this confidence set is valid under both scenarios, as the next theorem shows.

Theorem 2. *Let Assumption* **K**, **Z** *and* **Q** *hold true and suppose that either Assumption* **C** *or* **J** *hold. Then, for any* $s \in (0, 1)$,

$$P\{\gamma_0 \in \widehat{\Gamma}_s\} \to s.$$

4   BOOTSTRAP

This section develops a bootstrap-based test inversion confidence interval for the unknown threshold parameter γ_0, which is valid under Assumption **C** as well as under Assumption **J**. We do not discuss the bootstrap for α_0 in detail but note that the bootstrap for the linear regression can be employed,[2] see e.g. Shao and Tu (1995), since we can treat $\widehat{\gamma}$ as γ_0 for the inference on α_0 due to the arguments leading to the asymptotic independence between $\widehat{\alpha}$ and $\widehat{\gamma}$.

We propose using the bootstrap test inversion method, also known as the grid bootstrap, of Dümbgen (1991) to build confidence intervals for the parameter γ, see also Carpenter (1999) and Hansen (1999). Such a test inversion bootstrap confidence interval (BCI) is known to have certain optimality properties as in e.g. Brown, Casella and Hwang (1995) from the Bayesian

---

[2]This excludes the case where γ_0 is not strongly identified in the sense that $\delta_0 = d \cdot n^{-\varphi}$ with $\varphi \geq 1/2$. This case has not been explored except when $d = 0$, see e.g. Hansen (1996) and it is an interesting future research area.
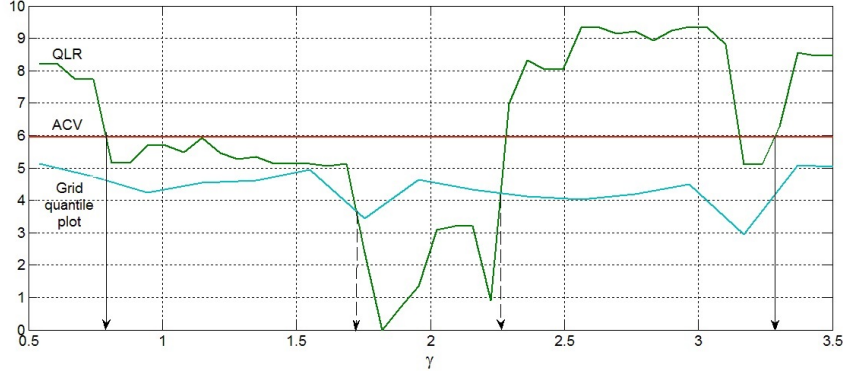
Figure 1: 90% asymptotic and grid bootstrap confidence intervals, $n = 100$, $\gamma_0 = 2$

perspective. Mikusheva (2007) showed that test inversion BCI attains correct coverage probability uniformly over the parameter space for the sum of coefficients in autoregressive models, despite the behavior of the estimator not being uniform over the parameter space.

For a given confidence level s, one can exploit the duality between hypothesis testing and confidence interval by inverting tests to obtain a confidence region

$$\widehat{\Gamma}_s^* = \left\{ \gamma \in \Gamma : \widehat{\xi}(\gamma)^{-1} QLR_n(\gamma) \leq F_n^*(s|\gamma) \right\} \ ,$$

where $F_n^*(s|\gamma)$ is the bootstrap estimate of the sth quantile of the statistic $\widehat{\xi}(\gamma)^{-1} QLR_n(\gamma)$ when $\gamma_0 = \gamma$. In other words, it denotes the bootstrap critical value of level $(1-s)$ testing for $\mathcal{H}_0 : \gamma_0 = \gamma$. In practice, one would estimate $F_n^*(s|\gamma)$ over a grid of $\gamma's$ and use some smoothing method such as linear interpolation or kernel averaging to obtain a smoothed bootstrap quantile function over a range of γ. The region $\widehat{\Gamma}_s^*$ is known as s-level grid bootstrap confidence interval (BCI) of γ in the terminology of Hansen (1999).

Figure 1 illustrates how this confidence interval can be obtained in practice. The $QLR_n(\gamma)$ line is the linear interpolation of the rescaled $QLR_n(\gamma)$ statistic over a grid of γ of 50 points. The ACV line is the asymptotic critical value of Hansen (2000). The true value of γ_0 was 2. We estimated bootstrap quantile function (described in the sequel) at 17 grid points and present the interpolated line. The vertical arrow at the intersections between $QLR_n(\gamma)$ and

14

ACV yield the asymptotic confidence interval (ACI), while the vertical broken arrows indicate grid BCI based on the bootstrap.

Now, we describe the bootstrap procedure for the grid bootstrap. We repeat the following procedure for each values of $\gamma_j \in \{\gamma_1, ..., \gamma_g\}$.

4.1   Bootstrap Algorithm for each γ_j

STEP 1 Obtain LSE $(\widehat{\alpha}', \widehat{\gamma})'$ by minimizing (5) and compute the LSE residuals

$$\widehat{\varepsilon}_t = y_t - \widehat{\alpha}' x_t(\widehat{\gamma}), \quad t = 1, ..., n.$$

STEP 2 Generate $\{\eta_t\}_{t=1}^n$ as *i.i.d.* zero mean random variables with unit variance and finite fourth moments, and compute

$$y_t^* = \widehat{\alpha}' x_t(\gamma_j) + \widehat{\varepsilon}_t \eta_t, \quad t = 1, ..., n.$$

STEP 3 Obtain the least squares estimate using $\{y_t^*\}_{t=1}^n$ and $\{x_t\}_{t=1}^n$,

$$\widehat{\theta}^* = \underset{\theta}{\mathrm{argmin}} \ \mathbb{S}_n^*(\theta) := \frac{1}{n} \sum_{t=1}^n \left(y_t^* - x_t(\gamma)' \alpha\right)^2. \tag{17}$$

STEP 4 Compute the bootstrap analogues of QLR_n and $\widehat{\xi}$ as

$$QLR_n^* = n \frac{\widehat{\mathbb{S}}_n^*(\gamma_j) - \widehat{\mathbb{S}}_n^*(\widehat{\gamma}^*)}{\widehat{\mathbb{S}}_n^*(\widehat{\gamma}^*)},$$

and

$$\widehat{\xi}^* = \frac{\sum_{t=1}^n (\widehat{\delta}^{*\prime} x_t)^2 \widehat{\varepsilon}_t^{*2} K\left(\frac{q_t - \widehat{\gamma}^*}{a}\right)}{\mathbb{S}_n(\widehat{\theta}^*) \sum_{t=1}^n (\widehat{\delta}^{*\prime} x_t)^2 K\left(\frac{q_t - \widehat{\gamma}^*}{a}\right)}, \tag{18}$$

where $\widehat{\mathbb{S}}_n^*(\gamma)$ is defined analogously as $\widehat{\mathbb{S}}_n(\gamma)$ in (6) by replacing y_t with y_t^*.

STEP 5 Compute the bootstrap $100s$-th quantile $F_n^*(s|\gamma_j)$ from the empirical distribution of $\widehat{\xi}^{*-1} QLR_n^*$ by repeating STEP 2-4.

Next, we derive the convergences of the bootstrap LSE $\widehat{\alpha}^*$ and $\widehat{\gamma}^*$ for both continuous and discontinuous setups and show the consistency of the bootstrap statistic $\widehat{\xi}^*$. These results then

yield the validity of the bootstrap test inversion confidence set following the same arguments in the proof of Theorem 2.

As usual, the superscript "*" indicates the bootstrap quantities and convergences of bootstrap statistics conditional on the original data. As in Shao and Tu (1995), the notation "$\xrightarrow{d^*}$, in Probability" signifies the the convergence in Probability of the random distribution functions of the bootstrap statistics in terms of the uniform metric and $A_n^* = o_{p^*}(1)$ means that $A_n^* \xrightarrow{d^*} 0$, in Probability.

Theorem 3. *Suppose that Assumptions **Z** and **Q** hold true.*

(**a**) *Under Assumption **C**, $\widehat{\alpha}^*$ and $\widehat{\gamma}^*$ are asymptotically independent and (in probability)*

$$n^{1/2}(\widehat{\alpha}^* - \widehat{\alpha}) \xrightarrow{d^*} \mathcal{N}\left(0, M^{-1}\Omega M^{-1}\right),$$

$$n^{1/3}(\widehat{\gamma}^* - \gamma_0) \xrightarrow{d^*} \arg\max_{g \in \mathbb{R}} \left(2\delta_{30} \sqrt{\frac{\sigma^2(\gamma_0)}{3} f(\gamma_0)} W\left(g^3\right) + \frac{\delta_{30}^2}{3} f(\gamma_0) |g|^3 \right).$$

(**b**) *Under Assumption **J**, $\widehat{\alpha}^*$ and $\widehat{\gamma}^*$ are asymptotically independent and (in probability)*

$$n^{1/2}(\widehat{\alpha}^* - \widehat{\alpha}) \xrightarrow{d^*} \mathcal{N}\left(0, M^{-1}\Omega M^{-1}\right),$$

$$n^{1-2\varphi}(\widehat{\gamma}^* - \gamma_0) \xrightarrow{d^*} \frac{2d'Vd}{(d'Dd)^2 f(\gamma_0)} \arg\max_{g \in \mathbb{R}} \left(2W(g) - |g| \right).$$

Our results can be compared with those already obtained in the literature regarding the validity of bootstrap for non-standard estimators. First, our consistency result seems to contradict Seijo and Sen's (2011) result on the inconsistency of a residual-based bootstrap and the nonparametric bootstrap (with *i.i.d.* data) for the case where $\varphi = 0$, see also Yu (2014). The reason behind such contradictory conclusions lies in the observation that our setup differs from theirs in an important and vital way: they consider the case of a fixed size of the break whereas we consider the situation that $\delta_0 = d \cdot n^{-\varphi}$ decreases with the sample size. Thus, their limiting distribution depends on the whole conditional distribution of $\varepsilon_t \eta_t d' x_t$ given $q_t = \gamma_0$ in a complicated manner, whereas ours contains only an unknown scaling factor.

It is worth mentioning that the centering term for $\widehat{\gamma}^*$ is γ_0, which reflects the fact that our resampling scheme imposes the hypothesized true value for the unknown threshold. This is important for the validity of our bootstrap since we do not impose the continuity restriction in our bootstrap resampling. By imposing the null value, our resampling scheme builds on $\sqrt{n}$-consistent estimates.

Next, the consistency of $\widehat{\xi}^*$ is established in the following proposition.

Proposition 4. *Suppose Assumptions* **Z**, **Q** *and* **K** *hold and either of Assumption* **J** *or Assumption* **C** *holds true. Then,*

$$\widehat{\xi}^* - \widehat{\xi} = o_{p^*}(1).$$

A direct consequence of Theorem 3 and Proposition 4 is the following theorem.

Theorem 4. *Now, suppose either Assumption* **J** *or Assumption* **C** *hold true in addition to Assumptions* **Z**, **Q** *and* **K**. *Then, (in probability)*

$$\widehat{\xi}^{*-1} QLR_n^* \xrightarrow{d^*} \max_{g \in \mathbb{R}} (2W(g) - |g|).$$

5 Monte Carlo Experiment

We generate data based on the following 3 specifications, with settings A and B being jump models akin to that considered in Hansen (2000, Section 4.2) and setting C representing the kink case.

$$A: y_t = 2 + 3x_t + \delta x_t 1\{q_t > \gamma_0\} + \varepsilon_t,$$
$$B: y_t = 2 + 3q_t + \delta q_t 1\{q_t > \gamma_0\} + \varepsilon_t,$$
$$C: y_t = 2 + 3q_t + \delta(q_t - \gamma_0)1\{q_t > \gamma_0\} + \varepsilon_t.$$

The main difference in our data generating process from that of Hansen (2000) is the conditional heteroscedasticity in ε_t: we set $\varepsilon_t = |q_t|e_t$ where $\{e_t\}_{t \geq 1}$ and $\{q_t\}_{t \geq 1}$ were generated as mutually independent and *i.i.d.* normal random variables with unit variance. This leads to conditional heteroscedasticity of the form $E(\varepsilon_t^2|q_t) = q_t^2$, in contrast to Hansen (2000) where ε_t was generated from $N(0,1)$. In setting A, we generated x_t as *i.i.d.* draws from $N(2,1)$, independent of $\{e_t\}_{t \geq 1}$ and $\{q_t\}_{t \geq 1}$, while setting $Eq_t = 2$. We generate $\{e_t\}_{t \geq 1}$ and $\{q_t\}_{t \geq 1}$ the same for setting B. For both settings A and B, we try $\gamma_0 = 2$ and 2.674, which correspond to the median and third quartile of q_t, respectively. In setting C, we set $\gamma_0 = 0$ and try $Eq_t = 0$ or -0.674 so that the threshold corresponds to the median or the third quartile of q_t, respectively. For the grid Γ_n used in estimation of γ_0, we discarded 10% of extreme values of realized q_t and used $n/2$ number of equidistant points.

Table 1: Monte Carlo size of test $H_0: \gamma = \gamma_0$ and coverage probability of confidence intervals of γ_0, model A: $q_t \neq x_t$, $\delta = n^{-\varphi}\sqrt{10}/4$

			Size				Coverage Probability					
		γ_0	median of $q_t(2)$			γ_0	median of $q_t(2)$			third quart. of $q_t(2.674)$		
φ		$s\backslash n$	100	250	500	$\zeta\backslash n$	100	250	500	100	250	500
1/4	Asym	0.01	0.095	0.059	0.044	0.9	0.733	0.770	0.774	0.811	0.834	0.844
		0.05	0.195	0.153	0.130	0.95	0.818	0.832	0.857	0.870	0.895	0.914
		0.1	0.290	0.242	0.200	0.99	0.916	0.938	0.950	0.953	0.971	0.980
	B/rap	0.01	0.003	0.015	0.009	0.9	0.756	0.810	0.840	0.783	0.826	0.852
		0.05	0.052	0.055	0.037	0.95	0.833	0.880	0.910	0.859	0.892	0.915
		0.1	0.106	0.095	0.083	0.99	0.928	0.959	0.969	0.935	0.965	0.980
1/8	Asym	0.01	0.068	0.037	0.029	0.9	0.79	0.837	0.897	0.817	0.835	0.872
		0.05	0.164	0.092	0.077	0.95	0.856	0.898	0.923	0.873	0.91	0.914
		0.1	0.214	0.15	0.129	0.99	0.933	0.961	0.975	0.949	0.964	0.972
	B/rap	0.01	0.006	0.009	0.008	0.9	0.791	0.846	0.881	0.792	0.827	0.871
		0.05	0.046	0.052	0.049	0.95	0.858	0.907	0.93	0.859	0.9	0.917
		0.1	0.099	0.095	0.105	0.99	0.936	0.968	0.98	0.938	0.963	0.972

Note: Size results for test of $H_0: \gamma = \gamma_0$ with nominal size s based on Hansen (2000)'s asymptotic distribution (Asym), and our bootstrap (B/rap). Coverage probability results for γ_0 with asymptotic confidence interval based on Hansen (2000) and our grid bootstrap confidence interval, with nominal confidence level ζ. $\delta = n^{-1/4}\sqrt{10}/4 = 0.25, 0.1988, 0.1672$, $\delta = n^{-1/8}\sqrt{10}/4 = 0.4446, 0.3965, 0.3636$ for $n = 100, 250, 500$

We investigate finite-sample performance of testing and confidence regions for γ given in Sections 3 and 4. We first compare the Monte Carlo size of tests for the correct location of the threshold, based on the asymptotic theory of Hansen (2000) which covers diminishing jump models, and our bootstrap method. We then investigate coverage probabilities of confidence intervals, constructed from either the asymptotic theory of Hansen (2000), or test-inversion based on our bootstrap. Our method has the virtue of robustness across different settings, and the objective is to see how it works across the jump settings of A and B and the kink setting of C. In A and B, we try two sets of δ with $\varphi = 1/4, 1/8$: $\delta = n^{-1/4}\sqrt{10}/4 = 0.25, 0.1988, 0.1672$, and $\delta = n^{-1/8}\sqrt{10}/4 = 0.4446, 0.3965, 0.3636$ for $n = 100, 250, 500$ reflecting Assumption **J**. In setting C, δ is fixed at $\delta = 2$ in line with Assumption **C**.[3] For the estimate $\widehat{\xi}$ of the scale factor for the QLR_n statistic, Epanechnikov kernel and minimum-MSE bandwidth choice, given in Härdle and Linton (1994), were deployed.

Columns 4-6 of Tables 1-3 present Monte Carlo size of test of $H_0 : \gamma = \gamma_0$ when γ_0 is the median of q_t for nominal sizes $s = 0.1, 0.05, 0.01$ for the three settings. We carried out 10,000 iterations, with one bootstrap per iteration, using the warp-speed method of Giacomini, Politis and White (2013). Using the asymptotic critical values delivers poor Monte Carlo sizes in settings A and B with substantial over-sizing, which is more severe in setting B. In contrast, the bootstrap test produces sizes that are close to the nominal ones, apart from $n = 100$ in B, for both φ. For the asymptotic test, the size results are somewhat better when $\varphi = 1/8$ compared to $\varphi = 1/4$ in settings A and B, although the over-sizing remains severe even for $\varphi = 1/8$ in setting B as shown in Table 2. For the kink setting C, asymptotic test based on Hansen's (2000) results produces sizes that become very small with increasing n, while the bootstrap test leads to good size results for $n = 250, 500$.

Columns 8-10 of Tables 1-3 report the coverage probabilities of confidence intervals for γ_0 in the three settings, when γ_0 is the median of q_t, and columns 11-13 present the case when γ_0 is the third quartile of q_t, for confidence levels $\zeta = 0.9, 0.95, 0.99$. Results are based on 1,000 iterations and in each iteration, we generated bootstrap quantile plots by interpolating bootstrap quantiles obtained at 10 equidistant points of the realized support of q_t from 399 bootstraps, and found intersections with the sample QLR_n plot formed by interpolating between $n/2$ number of equidistant points after discarding 10% of extreme values of realized q_t.

---

[3]Note that $\delta = 0.25, 2$ were the smallest and the largest values of δ tried in Hansen (2000), respectively.

Table 2: Monte Carlo size of test $H_0 : \gamma = \gamma_0$ and coverage probability of confidence intervals of γ_0, model B: $q_t = x_t$, $\delta = n^{-\varphi}\sqrt{10}/4$

			Size				Coverage Probability					
		γ_0	median of $q_t(2)$			γ_0	median of $q_t(2)$			third quart. of $q_t(2.674)$		
φ		$s\backslash n$	100	250	500	$\zeta\backslash n$	100	250	500	100	250	500
1/4	Asym	0.01	0.185	0.145	0.155	0.9	0.608	0.612	0.658	0.740	0.730	0.725
		0.05	0.344	0.293	0.268	0.95	0.687	0.707	0.742	0.813	0.817	0.827
		0.1	0.437	0.379	0.365	0.99	0.831	0.851	0.859	0.905	0.924	0.926
	B/rap	0.01	0.022	0.013	0.021	0.9	0.770	0.836	0.866	0.868	0.882	0.878
		0.05	0.101	0.066	0.071	0.95	0.853	0.894	0.924	0.932	0.943	0.943
		0.1	0.203	0.126	0.133	0.99	0.946	0.972	0.982	0.975	0.984	0.980
1/8	Asym	0.01	0.155	0.098	0.079	0.9	0.661	0.72	0.786	0.771	0.779	0.791
		0.05	0.285	0.207	0.158	0.95	0.745	0.802	0.852	0.852	0.844	0.855
		0.1	0.368	0.275	0.224	0.99	0.86	0.886	0.921	0.925	0.941	0.938
	B/rap	0.01	0.029	0.009	0.017	0.9	0.797	0.871	0.904	0.886	0.891	0.888
		0.05	0.093	0.073	0.065	0.95	0.878	0.917	0.945	0.936	0.946	0.943
		0.1	0.171	0.113	0.109	0.99	0.95	0.981	0.99	0.984	0.984	0.98

Note: Size results for test of $H_0 : \gamma = \gamma_0$ with nominal size s based on Hansen (2000)'s asymptotic distribution (Asym), and our bootstrap (B/rap). Coverage probability results for γ_0 with asymptotic confidence interval based on Hansen (2000) and our grid bootstrap confidence interval, with nominal confidence level ζ. $\delta = n^{-1/4}\sqrt{10}/4 = 0.25, 0.1988, 0.1672$, $\delta = n^{-1/8}\sqrt{10}/4 = 0.4446, 0.3965, 0.3636$ for $n = 100, 250, 500$.

In settings A and B reported in Tables 1 and 2, the coverage probability results are better when γ_0 is the third quartile of q_t for both methods when $\varphi = 1/4$. For $\varphi = 1/8$, this is still the case, with the exception of bootstrap coverage probabilities in setting A, which are similar between the two values of γ_0. In setting A as shown in Table 1, the asymptotic and bootstrap methods perform similarly, reporting lower-than-nominal coverage probabilities which improve with larger n. In setting B, the bootstrap method delivers substantially better coverage probabilities than the asymptotic confidence intervals based on Hansen (2000), which remain substantially lower than the nominal level even for $n = 500$ for $\varphi = 1/4$. Such under-coverage of asymptotic confidence intervals for small $\delta = 0.25$ was also reported in Hansen's (2000) Table 2, for homoskedastic error case. The coverage probability results are better when $\varphi = 1/8$ compared to $\varphi = 1/4$ for both methods in setting B, especially so for asymptotic confidence intervals. In Hansen's (2000) Table 2, coverage probability was also good for $\delta = 0.5$.

In setting C reported in Table 3, the asymptotic coverage probabilities becomes close to 1 for all values of ζ for $n = 250, 500$, while bootstrap coverage probabilities are satisfactory for $n = 250, 500$. The bootstrap coverage probability is better when γ_0 is the third quartile of q_t compared to when it is the median.[4]

6 EMPIRICAL APPLICATION: GROWTH AND DEBT

The so-called Reinhart-Rogoff hypothesis postulates that above some threshold (90% being their estimate of this threshold), higher debt-to-GDP ratio is associated with lower GDP growth rate. There have been numerous studies that utilize the threshold regression models to assess this hypothesis, including Hansen (2017) who fitted a kink model to a time series of US annual data, see Hansen (2017) for references on earlier studies which fitted jump models to various data sets. As there is little guidance from economic theory on the choice between kink and jump models in this setting, we advocate the use of our robust inference on the threshold and slope parameters of the model.

Hansen (2017) had fitted a kink model to US annual data on real GDP growth rate in year

---

[4]In Table 4 in Online Appendix, we report Monte Carlo size and coverage probability results for γ when $\varphi = 0$ with δ fixed at $\sqrt{10}/4 = 0.7906$ and 0.25 in setting A ($q_t \neq x_t$) with homoscedastic error. Fixed jump setup is not covered by Hansen (2000) or our bootstrap of Section 4, but nonetheless we investigate how the two methods perform in this setting for completeness.

Table 3: Monte Carlo size of test $H_0 : \gamma = \gamma_0$ and coverage probability of confidence intervals of γ_0, model C, kink, $\delta = 2$

		γ_0	median of q_t			γ_0	median of q_t			third quart. of q_t		
		$s\backslash n$	100	250	500	$\zeta\backslash n$	100	250	500	100	250	500
C	Asym	0.01	0.123	0.028	0.005	0.9	0.802	0.946	0.975	0.749	0.925	0.972
		0.05	0.168	0.043	0.015	0.95	0.84	0.965	0.983	0.784	0.945	0.98
		0.1	0.200	0.056	0.024	0.99	0.892	0.982	0.992	0.852	0.966	0.99
	B/rap	0.01	0.027	0.014	0.012	0.9	0.768	0.854	0.805	0.828	0.894	0.877
		0.05	0.091	0.054	0.052	0.95	0.817	0.918	0.889	0.88	0.949	0.943
		0.1	0.153	0.108	0.104	0.99	0.905	0.979	0.975	0.954	0.981	0.984

Note: Size results for test of $H_0 : \gamma = \gamma_0$ with nominal size s based on Hansen (2000)'s asymptotic distribution(Asym), and our bootstrap(B/rap). Coverage probability results for γ_0 with asymptotic confidence interval based on Hansen (2000) and our grid bootstrap confidence interval, with nominal confidence level ζ. When γ_0 is median, $q_t \sim N(0,1)$. When γ_0 is third quartile, $q_t \sim N(-0.674, 1)$ and $\gamma_0 = 0$.

t (y_t) and debt-to-GDP ratio from the *previous* year (q_t) for the period spanning 1792-2009 ($n = 218$), and estimated the threshold to be 43.8%, while the slope parameters of q_t were not significant. Before fitting the jump model to this data, we first tested for the presence of threshold effect using the testing procedure of Hansen (1996) with 1,000 bootstrap replications, and obtained p-value of 0.047, rejecting the null hypothesis of no threshold effect. This is in contrast to the p-value of 0.15 obtained by Hansen's (2017) test for presence of threshold effect when imposing the kink model. Hansen (2017) had remained inconclusive on the presence of kink threshold effect, since the bootstrap method used there did not account for the time series nature of data and the high p-value could have been due to modest power of the test.

The fitted jump model is given by:

$$
\widehat{y}_t = \begin{cases} \underset{(0.87)}{4.82} - \underset{(0.16)}{0.052} y_{t-1} - \underset{(0.049)}{0.114} q_t, & \text{if } q_t \leq 17.2 \\ \underset{(0.74)}{2.78} + \underset{(0.082)}{0.49} y_{t-1} - \underset{(0.012)}{0.017} q_t, & \text{if } q_t > 17.2 \end{cases}
$$

The sizes of the two regimes were 99 (below 17.2%) and 109 (above 17.2%). We obtained grid bootstrap confidence intervals for γ_0 to be (10.5, 39) for 95% confidence level and (10.8, 38.6) for 90%, based on 399 bootstrap iterations. Bootstrap quantiles were obtained at 38 grid points, which included $\widehat{\gamma}$, $\widetilde{\gamma}$ and equidistant points on the realized support of q_t after discarding
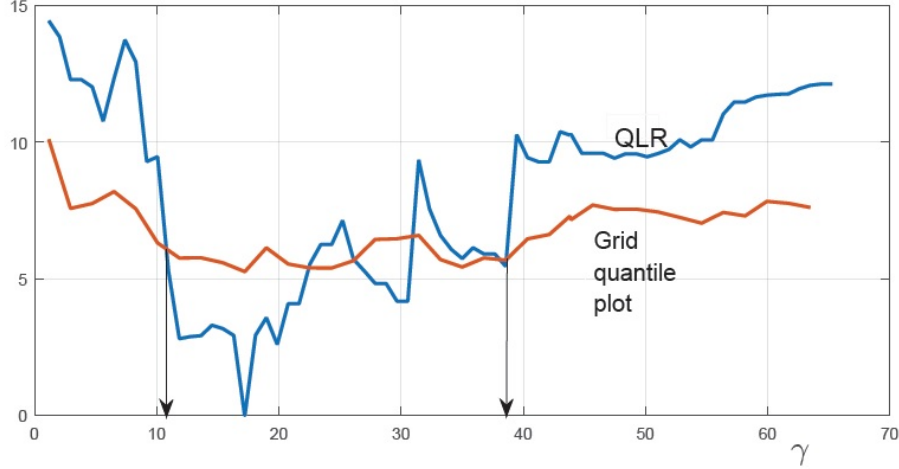
Figure 2: 90% grid bootstrap confidence interval for the US

7.5% of the largest and smallest values of q_t in the sample.[5] We find the points of intersection between the linearly interpolated bootstrap quantile line and the linear interpolation of sample $QLR_n(\gamma)$ test statistics for $H_0 : \gamma_0 = \gamma_j$ at grid points γ_j consisting of 73 equidistant points and $\widehat{\gamma}, \widetilde{\gamma}$, as shown in Figure 2 for 90% confidence level.

As the estimated threshold under the jump model is noticeably small at 17.2%, our estimated jump model which suggests insignificance of effect of q_t on y_t above the threshold does not necessarily contradict the Reinhart-Rogoff hypothesis. To see if this could be an indication of presence of further threshold points, we applied Hansen (1996)'s testing procedure for presence of threshold effect on the lower and upper subsamples with 1000 bootstraps and obtained p-values of 0.025 and 0.016, respectively. Hence, we conclude that the US time series data should be fitted to a threshold regression model with multiple threshold points.

To see if such conclusion holds across different countries, we proceeded by first applying Hansen (1996)'s test for the presence of threshold effect on Reinhart and Rogoff's (2010) data for countries with relatively long time spans without missing observations. For Australia($n =$

---

[5]There is currently no theoretical guide to the choice of the trimming parameter. Our choice of trimming out 7.5% was guided by Sweden's estimated $\tilde{\gamma}$ being the 12-th percentile of the q_t in the data. Sensitivity check on changing choices of the trimming value is recommended.

107) and the UK($n = 178$), the p-values with 1000 bootstraps were 0.795 and 0.98 so we conclude that there is no threshold effect for these countries in the relationship between the GDP growth and the debt-to-GDP ratio.

For data from Sweden for the period 1881-2009 ($n = 129$), the p-value for Hansen (1996)'s test of presence of threshold effect with 1000 bootstraps for the whole sample is 0.048, while for the lower and upper regimes, divided by $\widehat{\gamma}$, they were 0.979 and 0.131, respectively. The estimated jump model is:

$$\widehat{y}_t = \begin{cases} \underset{(2.17)}{1.12} - \underset{(0.24)}{0.2}\, y_{t-1} + \underset{(0.11)}{0.13}\, q_t, & \text{if } q_t \leq 21.3 \\ \underset{(0.58)}{1.86} + \underset{(0.11)}{0.48}\, y_{t-1} - \underset{(0.0082)}{0.004}\, q_t, & \text{if } q_t > 21.3 \end{cases}$$

with the lower regime having 61 observations and upper regime containing 68. The coefficient of debt-to-GDP ratio is not statistically significant.

The grid bootstrap confidence intervals for γ_0 were (15.3, ∞) and (16.4, ∞) for 95% and 90% confidence levels. Shown in Figures 3 are linear interpolation of 90% bootstrap quantiles at 27 grid points with 399 bootstraps and linear interpolation of QLR test statistic at each of 54 grid points.

We conclude that there is substantial heterogeneity across countries in the relationship between the GDP growth and the debt-to-GDP ratio, not only in the values of model parameters, but also in the type of models that are suitable.

7 CONCLUSION

This paper has developed unified inferential procedures for the threshold regression model. The unconstrained least squares estimator of the regression coefficient α turns out to enjoy a useful oracle property, which enables the standard asymptotic normal inference as in the linear regression model. On the other hand, we provide a judiciously constructed statistic, with which one can make inference of the unknown threshold without knowing the continuity of the threshold regression model. Asymptotically valid bootstrap inference is also proposed and shown to improve the finite sample performance of the asymptotic procedure.

An interesting future research area is extension to the nonparametric setting. For instance, see Card et al. (2008) and Pan (2015), who use the regression discontinuity methods [6] to

---

[6]Pan (2015, p.378) and a referee emphasize that this setting is not identical to the conventional regression
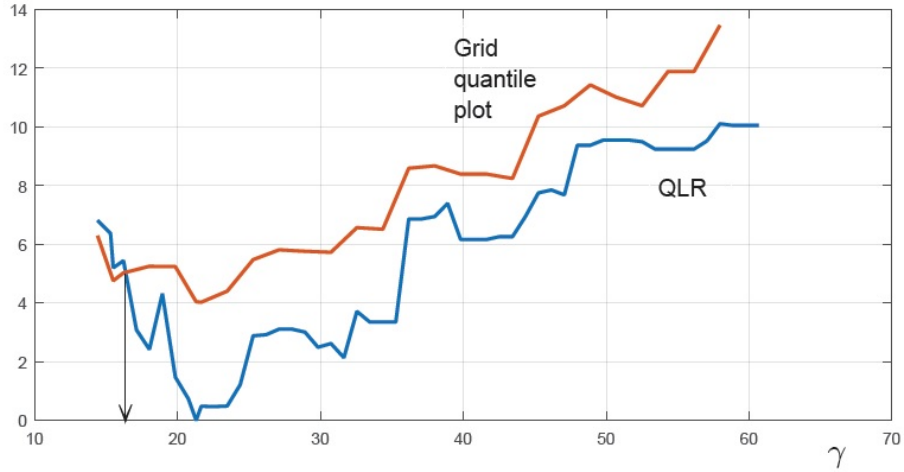
Figure 3: 90% grid bootstrap confidence interval for Sweden

test for the tipping phenomenon in racial segregation and gender segregation, respectively, or Landais (2014), who recommends testing for the location of the change-point as a validity check for the regression discontinuity design, even when the change-point is suggested by the institutional knowledge.

References

[1] Abrevaya, J., and Huang, J. (2005). "On the bootstrap of the maximum score estimator", *Econometrica*, 73, 1175-1204.

[2] Angrist J. D. and Lavy, V. (1999)."Using Maimonides' rule to estimate the effect of class size on scholastic achievement", *Quarterly Journal of Economics*, 114, 533-575.

[3] Bai, J., and Perron, P. (1998). "Estimating and testing linear models with multiple structural changes", *Econometrica*, 66, 47-78.

discontinuity method (e.g. Angrist and Lavy (1999); Hahn *et al.* (2001)) due to the lack of knowledge on the precise location of the discontinuity.

[4] Brown, L.D., Casella, G., and Hwang, J. T. G. (1995). "Optimal confidence sets, bioequivalence, and the Limaç on of Pascal", *Journal of the American Statistical Association*, 90, 880-889.

[5] Caner, M., Grennes, T., and Koehler-Geib, F. (2010). "Finding the tipping point-when sovereign debt turns bad", *Policy Research Working Paper Series* 5391, The World Bank.

[6] Card, D., Mas, A., and Rothstein, J. (2008), "Tipping and dynamics of segregation", *Quarterly Journal of Economics*, 123, 177-218.

[7] Carpenter, J. (1999). "Test inversion bootstrap confidence intervals", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 159-172.

[8] 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.

[9] Chan, K. S., and Tsay, R. S. (1998). "Limiting properties of the least squares estimator of a continuous threshold autoregressive model", *Biometrika*, 85, 413-426.

[10] Cecchetti, S. G., Mohanty, M. S., and Zampolli, F. (2011). "The real effects of debt", *Economic Symposium Conference Proceedings: Federal Reserve Bank of Kansas City* 145ï¿$\frac{1}{2}$196.

[11] Delgado, M. A., and Hidalgo, J. (2000). "Nonparametric inference on structural breaks", *Journal of Econometrics*, 96, 113-144.

[12] Dümbgen, L.(1991). "The asymptotic behaviour of some nonparametric changepoint estimators", *The Annals of Statistics*, 19, 1471-1495.

[13] Efron, B. (1979). "Bootstrap methods: another look at the jackknife", *The Annals of Statistics*, 7, 1-26.

[14] Feder, P. I. (1975a). "On asymptotic distribution theory in segmented regression problems-identified case", *The Annals of Statistics*, 3, 49-83.

[15] Feder, P. I. (1975b). "The log likelihood ratio in segmented regression", *The Annals of Statistics* , 3, 84-97.

26

[16] Giacomini, R., Politis, D. N. and White, H. (2013). "A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators ", *Econometric Theory*, 29, 567-589.

[17] Gonzalo, J., and Wolf, M. (2005). "Subsampling inference in threshold autoregressive models", *Journal of Econometrics*, 127, 201-224.

[18] Härdle, W., and Linton, O. (1994). "Chapter 38 Applied nonparametric methods", *Handbook of Econometrics*, 4, 2295-2339.

[19] Hahn, J., Todd, P., Van der Klaauw, W. (2001), "Identification and estimation of treatment effects with a regression-discontinuity design", *Econometrica*, 69, 201-209.

[20] Hansen, B. E. (1996). "Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.

[21] Hansen, B. E. (1999). "The grid bootstrap and the autoregressive model", *The Review of Economics and Statistics*, 81, 594-607.

[22] Hansen, B. E. (2000). "Sample splitting and threshold estimation", *Econometrica*, 68, 575-603.

[23] Hansen, B. E. (2017). "Regression kink with an unknown threshold", *Journal of Business and Economic Statistics*, 35, 228-240.

[24] Kim, J., and Pollard, D. (1990). Cube root asymptotics. The Annals of Statistics, 18, 191-219.

[25] Landais, C. (2014) "Assessing the welfare effects of unemployment benefits using the regression kink design", forthcoming in *American Economic Journal: Economic Policy*.

[26] Lee, S., Park, H., Seo, M. H., & Shin, Y. (2017). "Testing for a Debt-Threshold Effect on Output Growth." Fiscal studies, 38, 701-717.

[27] Lee, S., Seo, M.H. and Shin Y.(2011). "Testing for Threshold Effects in Regression Models," *Journal of the American Statistical Association* 106, 220-231.

[28] Li, D. and S. Q. Ling (2012). "On the least squares estimation of multiple-regime threshold autoregressive models," *Journal of Econometrics* 167, 240-253.

[29] Mikusheva, A. (2007). "Uniform inference in autoregressive models,", *Econometrica*, 75, 1411-1452.

[30] Pan, J. (2015). "Gender segregation in occupations: the role of tipping and social interactions", *Journal of Labor Economics*, 33, 365-408.

[31] Perron, P.(2006). "Dealing with structural breaks", *Pelgrave Handbook of Econometrics*, 278-352.

[32] Reinhart, C. M. and Rogoff, K. S.(2010). "Growth in a time of debt," *American Economic Review: Papers and Proceedings* 100, 573-578.

[33] Quandt, R.E.(1960). "Test of the hypothesis that a linear regression system obeys two separate regimes", *Journal of the American Statistical Association*, 55, 324-330.

[34] Seijo, E., and Sen, B.(2011). "Change-point in stochastic design regression and the bootstrap", *The Annals of Statistics*, 39, 1580-1607.

[35] Seo, M., and Linton, O.(2007). "A smoothed least squares estimator for the threshold regression", *Journal of Econometrics*, 141, 704-735.

[36] Seo, M. H., and Otsu, T. (2018). "Local M-estimation with discontinuous criterion for dependent and limited observations", *The Annals of Statistics*, 46, 344-369.

[37] Shao, J., and Tu, D. (1995). The jackknife and bootstrap. Springer.

[38] Tiao, G. C., & Tsay, R. S. (1994). "Some advances in non-linear and adaptive modelling in time-series", Journal of forecasting, 13(2), 109-131.

[39] Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*, New York: Oxford University Press.

[40] Yu, P.(2014). "The bootstrap in threshold regression", *Econometric Theory*, 30, 676-714.

A PROOFS OF MAIN THEOREMS

Let us introduce some notation first. In what follows $C, C_1,...$ denote generic positive finite constants, which may vary from line to line or expression to expression. Recall that $x_t = (1, x'_{t2}, q_t)'$, $x_{t1} = (1, x'_{t2})'$, and $\mathbf{1}_t(b) = \mathbf{1}\{q_t > b\}$, and introduce $\mathbf{1}_t(a; b) = \mathbf{1}\{a < q_t < b\}$. Finally, we abbreviate $\psi - \psi_0$ by $\overline{\psi}$ for any parameter ψ.

All the technical lemmas are given in the online supplement to this paper.

A.1 Proof of Proposition 1

Without loss of generality we assume that $\widehat{\gamma} \geq \gamma_0$ and $\gamma_0 = 0$, so that $\delta_{10} = 0$ and $\delta_{20} = 0$ under Assumption C. By definition, we have that

$$
\begin{aligned}
\mathbb{S}_n(\theta) - \mathbb{S}_n(\theta_0) &= \frac{1}{n} \sum_{t=1}^{n} \left\{ (y_t - \alpha' x_t(\gamma))^2 - \varepsilon_t^2 \right\} \\
&= \frac{1}{n} \sum_{t=1}^{n} \left\{ \left(\overline{\beta}' x_t + \overline{\delta}' x_t \mathbf{1}_t(\gamma) + \delta_0' x_t \mathbf{1}_t(0; \gamma) + \varepsilon_t \right)^2 - \varepsilon_t^2 \right\}.
\end{aligned}
$$

By standard algebra and denoting $\upsilon = \beta + \delta$,

$$
\begin{aligned}
&\overline{\beta}' x_t + \overline{\delta}' x_t \, \mathbf{1}_t(\gamma) + \delta_0' x_t \mathbf{1}_t(0; \gamma) \\
&= \overline{\upsilon}' x_t \mathbf{1}_t(\gamma) + \left(\overline{\beta} + \delta_0 \right)' x_t \mathbf{1}_t(0; \gamma) + \overline{\beta}' x_t \mathbf{1}_t(-\infty; 0),
\end{aligned}
$$

which implies, because of the orthogonality of the terms on the right of the last displayed expression, that

$$
\mathbb{S}_n(\theta) - \mathbb{S}_n(\theta_0) = \mathbb{A}_{n1}(\theta) + \mathbb{A}_{n2}(\theta) + \mathbb{A}_{n3}(\theta) + \mathbb{B}_{n1}(\theta) + \mathbb{B}_{n2}(\theta) + \mathbb{B}_{n3}(\theta),
$$

where

$$\mathbb{A}_{n1}(\theta) = \overline{v}'\frac{1}{n}\sum_{t=1}^{n}x_t x_t' \mathbf{1}_t(\gamma)\,\overline{v}; \quad \mathbb{A}_{n2}(\theta) = \overline{\beta}'\frac{1}{n}\sum_{t=1}^{n}x_t x_t' \mathbf{1}_t(-\infty;0)\,\overline{\beta}$$

$$\mathbb{A}_{n3}(\theta) = (\overline{\beta}+\delta_0)'\frac{1}{n}\sum_{t=1}^{n}x_t x_t' \mathbf{1}_t(0;\gamma)\,(\overline{\beta}+\delta_0)$$

$$\mathbb{B}_{n1}(\theta) = \overline{v}'\frac{2}{n}\sum_{t=1}^{n}x_t \varepsilon_t \mathbf{1}_t(\gamma); \quad \mathbb{B}_{n2}(\theta) = \overline{\beta}'\frac{2}{n}\sum_{t=1}^{n}x_t \varepsilon_t \mathbf{1}_t(-\infty;0)$$

$$\mathbb{B}_{n3}(\theta) = (\overline{\beta}+\delta_0)'\frac{2}{n}\sum_{t=1}^{n}x_t \varepsilon_t \,\mathbf{1}_t(0;\gamma).$$

Consistency. It suffices to show that for any $\epsilon > 0$, $\eta > 0$, there is n_0 such that for all $n > n_0$, $\Pr\left\{\left\|\widehat{\theta}-\theta_0\right\| > \eta\right\} < \epsilon$, which is implied by

$$\Pr\left\{\inf_{\|\overline{\theta}\|>\eta}\sum_{\ell=1}^{3}E\left(\mathbb{A}_{n\ell}(\theta)\right)+\mathbb{D}_{n\ell}(\theta)\leq 0\right\} < \epsilon, \tag{19}$$

where $\mathbb{D}_{n\ell}(\theta) = \mathbb{B}_{n\ell}(\theta) + (\mathbb{A}_{n\ell}(\theta)-E(\mathbb{A}_{n\ell}(\theta)))$ for $\ell = 1,2,3$.

First $\|\overline{\theta}\| > \eta$ implies that either **(i)** $\|\overline{\gamma}\| > \eta/3$ and $\|\overline{\beta}\| \leq \eta/3$, or **(ii)** $\|\overline{\beta}\| > \eta/3$ or $\|\overline{v}\| > \eta/3$. When (**ii**) holds true, it is clear that

$$\inf_{\|\overline{v}\|>\eta/3}E\left(\mathbb{A}_{n1}(\theta)\right) > C\eta^2 \quad \text{or} \quad \inf_{\|\overline{\beta}\|>\eta/3}E\left(\mathbb{A}_{n2}(\theta)\right) > C\eta^2 \tag{20}$$

whereas when **(i)** holds true, we have that

$$\inf_{\|\overline{\gamma}\|>\eta/3,\|\overline{\beta}\|\leq\eta/3}E\left(\frac{1}{n}\sum_{t=1}^{n}\left(x_t'(\overline{\beta}+\delta_0)\right)^2\mathbf{1}_t(0;\gamma)\right) > C\eta^3, \tag{21}$$

because Assumption Q implies that $E(x_t x_t' \mathbf{1}_t(\gamma))$, $E(x_t x_t' \mathbf{1}_t(-\infty;0))$ and $E(x_t x_t' \mathbf{1}_t(0;\gamma))$ are positive definite matrices uniformly in $\gamma > \eta$ and $\|\overline{\beta}+\delta_0\| > \eta/3$ if $\|\overline{\beta}\| \leq \eta/3$ because we can always choose η such that $|\delta_0| \geq 2\eta/3$. We have that

$$C_1 \leq \frac{E\mathbb{A}_{n3}(\theta)}{(\overline{\tau}_1,\overline{\tau}_2')\,E(x_{t1}x_{t1}'\mathbf{1}_t(0;\gamma))\,(\overline{\tau}_1,\overline{\tau}_2')' + \overline{\tau}_3^2 E(q_t^2\mathbf{1}_t(0;\gamma))} \leq C_2, \tag{22}$$

where $\overline{\tau} = (\beta_0 - \beta) + \delta_0$. The motivation for the last displayed inequality comes from the fact that , say, implies that $E\{x_t x_t' \mathbf{1}_t(\gamma_1;\gamma_2)\}$ is a strictly positive and finite definite matrix which

30

implies that for any vector $a' = (a'_1, a_2)$,

$$C^{-1} \leq \frac{a' E \{x_t x'_t \mathbf{1}_t (\gamma_1; \gamma_2)\} a}{a'_1 E \{x_{t1} x'_{t1} \mathbf{1}_t (\gamma_1; \gamma_2)\} a_1 + a_2^2 E \left(q_t^2 \mathbf{1}_t (\gamma_1; \gamma_2)\right)} \leq C.$$

So, (20) and (21) imply that

$$\inf_{\|\bar{\theta}\|>\eta} \sum_{\ell=1}^{3} E \left(\mathbb{A}_{n\ell} (\theta)\right) > C \eta^3. \tag{23}$$

On the other hand, Lemma 1 and the uniform law of large numbers, respectively, imply that

$$\sup_{\|\bar{\theta}\|>\eta} \|\mathbb{B}_{n\ell} (\theta)\| = O_p \left(n^{-1/2}\right) \quad \ell = 1, 2, 3; \qquad \sup_{\gamma_1, \gamma_2} \| \mathbb{F}_n (\gamma_1; \gamma_2)\| = o_p (1),$$

where $\mathbb{F}_n (\gamma_1; \gamma_2) = \frac{1}{n} \sum_{t=1}^{n} (x_t x'_t \mathbf{1}_t (\gamma_1; \gamma_2) - E (x_t x'_t \mathbf{1}_t (\gamma_1; \gamma_2)))$, and hence

$$\sup_{\|\bar{\theta}\|>\eta/3} \left\| \sum_{\ell=1}^{3} \mathbb{D}_{n\ell} (\theta) \right\| = o_p (1). \tag{24}$$

Thus $\widehat{\theta} - \theta_0 = o_p (1)$ because the left side of (19) is bounded by

$$\Pr \left\{ \inf_{\|\bar{\theta}\|>\eta} \sum_{\ell=1}^{3} E \left(\mathbb{A}_{n\ell} (\theta)\right) \leq \sup_{\|\bar{\theta}\|>\eta/3} \left\| \sum_{\ell=1}^{3} \mathbb{D}_{n\ell} (\theta) \right\| \right\} \to 0,$$

using (23) and (24).

Convergence Rate. We shall show next that for any $\epsilon > 0$ there exist $C > 0$, $\eta > 0$, n_0 such that for $n > n_0$ we have that

$$\Pr \left\{ \inf_{\frac{C}{n^{1/2}} < \|\bar{v}\|, \|\bar{\beta}\| < \eta; \frac{C}{n^{1/3}} < \|\bar{\gamma}\| < \eta} \sum_{\ell=1}^{3} E \left(\mathbb{A}_{n\ell} (\theta)\right) + \mathbb{D}_{n\ell} (\theta) \leq 0 \right\} < \epsilon. \tag{25}$$

Since $\Pr \{X_n + Y_n < 0\} \leq \Pr \{X_n < 0\} + \Pr \{Y_n < 0\}$ for any sequence X_n and Y_n and $\inf_x \{f (x) + g (x)\} \geq \inf_x f (x) + \inf_x g (x)$ for any functions f and g, it suffices to show that for each $\ell = 1, 2, 3$

$$\Pr \left\{ \inf_{\frac{C}{n^{1/2}} < \|\bar{v}\|, \|\bar{\beta}\| < \eta; \frac{C}{n^{1/3}} < \|\bar{\gamma}\| < \eta} E \left(\mathbb{A}_{n\ell} (\theta)\right) / 2 + \left(\mathbb{A}_{n\ell} (\theta) - E \left(\mathbb{A}_{n\ell} (\theta)\right)\right) \leq 0 \right\} < \epsilon \tag{26}$$

$$\Pr\left\{\inf_{\frac{C}{n^{1/2}}<\|\overline{v}\|, \|\overline{\beta}\|<\eta; \frac{C}{n^{1/3}}<\|\overline{\gamma}\|<\eta} E\left(\mathbb{A}_{n\ell}\left(\theta\right)\right)/2 + \mathbb{B}_{n\ell}\left(\theta\right) \le 0\right\} < \epsilon. \tag{27}$$

To that end, we shall first examine

$$\Pr\left\{\inf_{\Xi_j(v);\Xi_j(\beta);\Xi_k(\gamma)} E\left(\mathbb{A}_{n\ell}\left(\theta\right)\right)/2 + \mathbb{B}_{n\ell}\left(\theta\right) \le 0\right\}, \quad \ell = 1,2,3,$$

where

$$\Xi_j\left(\psi\right) = \left\{\psi : \frac{C}{n^{1/2}}2^{j-1} < \|\overline{\psi}\| < \frac{C}{n^{1/2}}2^j\right\}; \quad j = 1,...,\log_2\frac{\eta}{C}n^{1/2}$$

$$\Xi_k\left(\gamma\right) = \left\{\gamma : \frac{C}{n^{1/3}}2^{k-1} < \overline{\gamma} < \frac{C}{n^{1/3}}2^k\right\}; \quad k = 1,...,\log_2\frac{\eta}{C}n^{1/3}. \tag{28}$$

Recall that we have assumed that $\gamma \ge 0$, as the case $\gamma \le 0$ follows similarly.

First by standard arguments,

$$\Pr\left\{\inf_{\Xi_j(v);\Xi_k(\gamma)} E\left(\mathbb{A}_{n1}\left(\theta\right)\right)/2 + \mathbb{B}_{n1}\left(\theta\right) \le 0\right\}$$

$$\le \Pr\left\{\inf_{\Xi_j(v)} \|\overline{v}\| \lambda_{\min}\left(Ex_t x_t' \mathbf{1}_t\left(0\right)\right) \le \sup_{\Xi_k(\gamma)} \left\|\frac{4}{n^{1/2}}\sum_{t=1}^n x_t\varepsilon_t\mathbf{1}_t\left(\gamma\right)\right\|\right\}$$

$$\le \Pr\left\{C2^{j-2} \le \sup_{\{\gamma:\|\overline{\gamma}\|<\eta\}} \left\|\frac{1}{n^{1/2}}\sum_{t=1}^n x_t\varepsilon_t\mathbf{1}_t\left(\gamma\right)\right\|\right\} \tag{29}$$

$$\le C^{-1}2^{-j+2}\eta^{1/2}$$

by Lemma 1 and the Markov's inequality. Observe that the latter inequality is independent of $\Xi_k\left(\gamma\right)$. Since $\sum_{j=1}^\infty 2^{-j} < \infty$, the probability in (27) can be made arbitrary small for large C or small η, thus satisfying the condition (27). (26) follows similarly as is the case for $\ell = 2$ and thus it is omitted.

We next examine (26) and (27) for $\ell = 3$. Observing (22) and the arguments that follow, defining

$$\widetilde{\mathbb{A}}_{n3}\left(\theta\right) = \overline{\tau}^2 E\left(q_t^2\mathbf{1}_t\left(0;\gamma\right)\right); \qquad \widetilde{\mathbb{B}}_{n3}\left(\theta\right) = \overline{\tau}\frac{2}{n}\sum_{t=1}^n q_t\varepsilon_t\mathbf{1}_t\left(0;\gamma\right),$$

it suffices to show (26) and (27) for $\widetilde{\mathbb{A}}_{n3}\left(\theta\right)$ and $\widetilde{\mathbb{B}}_{n3}\left(\theta\right)$. To that end, because $\overline{\tau} > C_1$ as

$|\delta_{30}| > C_1 > 0$, we obtain, since $Eq_t^2 \mathbf{1}_t(0; \eta) \geq C_1 \eta^3$

$$
\begin{aligned}
&\Pr\left\{ \inf_{\Xi_j(\upsilon); \Xi_k(\gamma)} E\left(\widetilde{\mathbb{A}}_{n3}(\theta)/2\right) + \widetilde{\mathbb{B}}_{n3}(\theta) \leq 0 \right\} \\
&\leq \Pr\left\{ \inf_{\Xi_k(\gamma)} \|\tau_0\| E\left(q_t^2 \mathbf{1}_t(0; \gamma)\right) \leq \sup_{\Xi_k(\gamma)} \left\| \frac{4}{n} \sum_{t=1}^n q_t \varepsilon_t \mathbf{1}_t(0; \gamma) \right\| \right\} \\
&\leq \Pr\left\{ \frac{C}{n} 2^{3(k-2)} \leq \sup_{\Xi_k(\gamma)} \left\| \frac{1}{n} \sum_{t=1}^n q_t \varepsilon_t \mathbf{1}_t(0; \gamma) \right\| \right\} \\
&\leq C^{-1} 2^{-3k/2},
\end{aligned}
\tag{30}
$$

by Lemma 1 and Markov's inequality. Notice that this bound is independent of $\Xi_j(\upsilon)$. But by summability of $2^{-3k/2}$, we conclude that (27) holds true for $\ell = 3$ by choosing C large enough.

We now conclude the proof after we note that the left side of (25) is bounded by

$$
\begin{aligned}
&\Pr\left\{ \max_{j,k} \inf_{\Xi_j(\upsilon); \Xi_j(\beta); \Xi_k(\gamma)} \sum_{\ell=1}^3 \{E\mathbb{A}_{n\ell}(\theta) + \mathbb{B}_{n\ell}(\theta)\} \leq 0 \right\} \\
&\leq C^{-1}\left(\sum_{j=1}^{\log_2 \frac{n}{C} n^{1/2}} 2^{-2j} + \sum_{k=1}^{\log_2 \frac{n}{C} n^{1/3}} 2^{-3k/2} \right) < \epsilon
\end{aligned}
$$

using (29) − (30). ∎

A.2   Proof of Theorem 1

Because the "argmin" is a continuous mapping, see Kim and Pollard (1990), and the convergence rates of $\widehat{\alpha}$ and $\widehat{\gamma}$ are obtained in Proposition 1, it suffices to examine the weak limit of

$$
\begin{aligned}
\mathbb{G}_n(h, g) &= n\left(\mathbb{S}_n\left(\alpha_0 + \frac{h}{n^{1/2}}, \gamma_0 + \frac{g}{n^{1/3}} \right) - \mathbb{S}_n(\alpha_0, \gamma_0) \right) \\
&= \sum_{t=1}^n \left\{ \left(\varepsilon_t - \frac{h'}{n^{1/2}} x_t\left(\frac{g}{n^{1/3}}\right) - \delta_{30} q_t \mathbf{1}_t\left(0; \frac{g}{n^{1/3}}\right) \right)^2 - \varepsilon_t^2 \right\}
\end{aligned}
$$

over $\|h\|, |g| \leq C$, where we assume $\gamma_0 = 0$ as before for notational convenience and reparametrize $h = \sqrt{n}(\alpha - \alpha_0)$ and $g = n^{1/3}(\gamma - \gamma_0)$. First, due to the uniform law of large numbers it fol-

lows that

$$\sup_{|g|\leq C}\left|\frac{1}{n}\sum_{t=1}^{n}\left\{x_t\left(\frac{g}{n^{1/3}}\right)x_t'\left(\frac{g}{n^{1/3}}\right)-\mathbf{x_t}\mathbf{x_t'}\right\}\right|=o_p(1)$$

whereas Lemma 1 and the expansion of $E\left\{x_t\left(\frac{g}{n^{1/3}}\right)q_t\mathbf{1}_t\left(0;\frac{g}{n^{1/3}}\right)\right\}$ as in (32) imply that

$$\sup_{|g|\leq C}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{x_t\left(\frac{g}{n^{1/3}}\right)q_t\mathbf{1}_t\left(0;\frac{g}{n^{1/3}}\right)\right\}\right| = O_p\left(n^{-1/6}\right)$$

$$\sup_{|g|\leq C}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(x_t\left(\frac{g}{n^{1/3}}\right)-\mathbf{x_t}\right)\varepsilon_t\right| = O_p\left(n^{-1/6}\right).$$

Therefore

$$\sup_{\|h\|,|g|\leq C}\left|\mathbb{G}_n(h,g)-\widetilde{\mathbb{G}}_n(h,g)\right|=o_p(1),\tag{31}$$

where

$$\begin{aligned}\widetilde{\mathbb{G}}_n(h,g) &= \left\{h'\frac{1}{n}\sum_{t=1}^{n}\mathbf{x_t}\mathbf{x_t'}h-h'\frac{2}{n^{1/2}}\sum_{t=1}^{n}\mathbf{x_t}\varepsilon_t\right\}\\ &\quad+\delta_{30}\left\{\delta_{30}\sum_{t=1}^{n}q_t^2\mathbf{1}_t\left(0;\frac{g}{n^{1/3}}\right)-2\sum_{t=1}^{n}q_t\varepsilon_t\mathbf{1}_t\left(0;\frac{g}{n^{1/3}}\right)\right\}\\ &=: \widetilde{\mathbb{G}}_n^1(h)+\widetilde{\mathbb{G}}_n^2(g).\end{aligned}$$

The consequence of (31) is then that the minimizer of $\mathbb{G}_n(h,g)$ is asymptotically equivalent to that of $\widetilde{\mathbb{G}}_n(h,g)$. Thus, it suffices to show the weak convergence of $\widetilde{\mathbb{G}}_n^1(h)$ and $\widetilde{\mathbb{G}}_n^2(g)$ and that

$$\widetilde{h}=:\arg\min_{h\in\mathbb{R}}\widetilde{\mathbb{G}}_n^1(h);\qquad\widetilde{g}:=\operatorname*{argmin}_{g\in\mathbb{R}}\widetilde{\mathbb{G}}_n^2(g)$$

are $O_p(1)$. The convergence of $\widetilde{\mathbb{G}}_n^1(h)$ and its minimization is straightforward since it is a quadratic function of h.

Next, the first term of $\widetilde{\mathbb{G}}_n^2(g)$ converges to $3^{-1}\delta_{30}^2f(0)|g|^3$ uniformly in probability because Lemma 1, i.e. (49), implies the uniform law of large numbers and the Taylor series expansion up to the third order yields

$$nEq_t^2\mathbf{1}_t\left(0;\frac{g}{n^{1/3}}\right)=n\int_0^{\frac{g}{n^{1/3}}}q^2f(q)\,dq=n\frac{2f\left(\frac{\widetilde{g}}{n^{1/3}}\right)}{3!}\left(\frac{g}{n^{1/3}}\right)^3\to 3^{-1}f(0)g^3,\tag{32}$$

where $\widetilde{g} \in (0, g)$. When $g < 0$, it follows similarly as in this case the derivative should be multiplied by -1, so that the limit becomes $3^{-1} f(0) |g|^3$.

The second term in the definition of $\widetilde{\mathbb{G}}_n^2(g)$, that is $-2 \sum_{t=1}^{n} q_t \varepsilon_t \mathbf{1}_t \left(0; \frac{g}{n^{1/3}} \right)$ converges weakly to $2 \delta_{30} \sqrt{3^{-1} f(0) \sigma_\varepsilon^2(0)} W(g^3)$. To see this note that Lemma 1, i.e. (48), yields the tightness of the process as explained in Remark 3. For the finite dimensional convergence, we can verify the conditions for martingale difference sequence CLT (e.g. Hall and Heyde's (1980) Theorem 3.2). In particular, we need to show that for $u_{nt} = \sqrt{n} q_t \varepsilon_t \mathbf{1}_t \left(0; \frac{g}{n^{1/3}} \right)$,

$$(i) \quad n^{-1/2} \max_{1 \leq t \leq n} |u_{nt}| \xrightarrow{p} 0$$

$$(ii) \quad \frac{1}{n} \sum_{t=1}^{n} u_{nt}^2 \xrightarrow{p} \frac{1}{3} E(\varepsilon_t^2 | q_t = 0) f(0) g^3$$

For (i), note that $E n^{-2} \max_t |u_{nt}|^4 \leq n^{-1} E |u_{nt}|^4 = n E q_t^4 \varepsilon_t^4 \mathbf{1}_t \left(0; \frac{g}{n^{1/3}} \right) \to 0$ as $n \to \infty$. For (ii), apply the same argument for the first term in $\widetilde{\mathbb{G}}_n^2(g)$ and an expansion similar to that in (32). We now characterize the covariance kernel. To that end, we note that if g_1 and g_2 have different signs then the cross product becomes zero and for $g_2 > g_1 > 0$, similarly as with (32), we have that

$$n E \left(\varepsilon_t^2 (q_t - \gamma_0)^2 \mathbf{1} \left\{ \frac{g_1}{n^{1/3}} < q_t < \frac{g_2}{n^{1/3}} \right\} \right) = \frac{f(\gamma_0)}{3} \sigma_\varepsilon^2(\gamma_0) \left(g_2^3 - g_1^3 \right) + o(1).$$

The cases for $g_1 > g_2 > 0$ or $g_2 < g_1 < 0$ are similar and thus omitted.

Finally, the covariance between $n^{-1/2} \sum_{t=1}^{n} \mathbf{x_t} \varepsilon_t$ and $\sum_{t=1}^{n} q_t \varepsilon_t \mathbf{1}_t \left(0; g/n^{1/3} \right)$ vanishes for the same reasoning, yielding the independence between $\widetilde{h}$ and $\widetilde{g}$ and thus the asymptotic independence between $\widehat{\alpha}$ and the threshold estimator $\widehat{\gamma}$. ∎

A.3 Proof of Proposition 2

Due to the asymptotic independence between $\widehat{\alpha}$ and $\widehat{\gamma}$ in Theorem 1, see (31) in its proof, we have that

$$n \left(\mathbb{S}_n \left(\widehat{\alpha}(\gamma_0); \gamma_0 \right) - \mathbb{S}_n(\widehat{\alpha}; \widehat{\gamma}) \right) = n \left(\mathbb{S}_n(\alpha_0; \gamma_0) - \mathbb{S}_n(\alpha_0; \widehat{\gamma}) \right) + o_p(1),$$

which corresponds to $\min_g \widetilde{\mathbb{G}}_n^2 (g)$ in the proof of Theorem 1 due to the reparameterization $g = n^{1/3} (\gamma - \gamma_0)$. It also shows that

$$\min_g \widetilde{\mathbb{G}}_n^2 (g) \xrightarrow{d} f(\gamma_0) \min_{g \in \mathbb{R}} \left(2\delta_{30} \sqrt{3^{-1} f(\gamma_0) \sigma_\varepsilon^2 (\gamma_0)} W(g^3) + 3^{-1} \delta_{30}^2 f(\gamma_0) |g|^3 \right).$$

Finally, the desired result follows from applying the change of variables $g^3 = 3\phi \sigma_\varepsilon^2 (\gamma_0) / \delta_{30}^2 f(\gamma_0)$ because of the distributional equivalence $W(a^2 g) =^d aW(g)$ (and $W(s) =^d -W(s)$) and the fact that $\min_x g(x) = -\max_x -g(x)$ for any function g. ∎

A.4 Proof of Theorem 2

It is known that the distribution function of $\max_{g \in \mathbb{R}} (2W(g) - |g|)$ is F, as in Hansen (2000). Thus, under Assumption C, Propositions 2 and 3 yield the conclusion, while under Assumption J, Theorem 2 of Hansen (2000) verified the conclusion. ∎

A.5 Proof of Theorem 3

Recalling our definition of $\widehat{\alpha}^*$ and $\widehat{\gamma}^*$ in (17), we begin by showing their consistency and rate of convergence, which is given in Proposition 5.

We now discuss the asymptotic distribution of the bootstrap estimators. We begin with part (**a**). We assume $\gamma_0 = 0$ to simplify notation. Because the "arg max" is continuous as mentioned in Theorem 2, it suffices to examine the weak limit of

$$\begin{aligned}
\mathbb{G}_n^* (h, g) &= n \left(\mathbb{S}_n^* \left(\widetilde{\alpha} + \frac{h}{n^{1/2}}, \frac{g}{n^{1/3}} \right) - \mathbb{S}_n^* (\widetilde{\alpha}, 0) \right) \\
&= \sum_{t=1}^n \left\{ \left(\frac{h'}{n^{1/2}} x_t \left(\frac{g}{n^{1/3}} \right) + \widetilde{\delta}' q_t \mathbf{1}_t \left(0; \frac{g}{n^{1/3}} \right) + \varepsilon_t^* \right)^2 - \varepsilon_t^{*2} \right\},
\end{aligned}$$

where $\|h\|, |g| \leq C$.

First, recall that $\widetilde{\delta}_1 = O_p(n^{-1/2})$ and $\widetilde{\delta}_2 = O_p(n^{-1/2})$ under Assumption C and note that

36

Lemma 1 and Lemma 4 imply that, uniformly in $\|h\|, |g| < C$,

$$\frac{1}{n} \sum_{t=1}^{n} \left\{ x_t \left(\frac{g}{n^{1/3}} \right) x_t' \left(\frac{g}{n^{1/3}} \right) - \mathbf{x}_t \mathbf{x}_t' \right\} = O_p \left(n^{-1/3} \right)$$

$$\frac{1}{n^{1/2}} \sum_{t=1}^{n} \left\{ x_t \left(\frac{g}{n^{1/3}} \right) q_t \mathbf{1}_t \left(0; \frac{g}{n^{1/3}} \right) \right\} = O_p \left(n^{-1/6} \right)$$

$$E^* \left\| \frac{1}{n^{1/2}} \sum_{t=1}^{n} \left(x_t \left(\frac{g}{n^{1/3}} \right) - \mathbf{x}_t \right) \varepsilon_t^* \right\|^2 = O_p \left(n^{-1/3} \right).$$

Thus, the latter implies that

$$E^* \sup_{h,g \in \mathbb{R}} \left| \mathbb{G}_n^* (h, g) - \widetilde{\mathbb{G}}_n^* (h, g) \right| = O_p \left(n^{-1/6} \right), \tag{33}$$

where

$$\begin{aligned}
\widetilde{\mathbb{G}}_n^* (h, g) &= \left\{ h' \frac{1}{n} \sum_{t=1}^{n} \mathbf{x}_t \mathbf{x}_t' h + h' \frac{1}{n^{1/2}} \sum_{t=1}^{n} \mathbf{x}_t \varepsilon_t^* \right\} \\
&\quad + \widetilde{\delta}_3 \left\{ \widetilde{\delta}_3 \sum_{t=1}^{n} q_t^2 \mathbf{1}_t \left(0; \frac{g}{n^{1/3}} \right) + \sum_{t=1}^{n} q_t \varepsilon_t^* \mathbf{1}_t \left(0; \frac{g}{n^{1/3}} \right) \right\} \\
&= : \widetilde{\mathbb{G}}_{1n}^* (h) + \widetilde{\mathbb{G}}_{2n}^* (g).
\end{aligned}$$

The consequence of (33) is then that the minimizer of $\mathbb{G}_n^* (h, g)$ is asymptotically equivalent to that of $\widetilde{\mathbb{G}}_n^* (h, g)$. Thus, it suffices to show the weak convergence of $\widetilde{\mathbb{G}}_{1n}^* (h)$ and $\widetilde{\mathbb{G}}_{2n}^* (g)$ and that

$$\widetilde{h} =: \arg \max_{h \in \mathbb{R}} \widetilde{\mathbb{G}}_{1n}^* (h); \qquad \widetilde{g} =: \arg \max_{g \in \mathbb{R}} \widetilde{\mathbb{G}}_{2n}^* (g)$$

are $O_{p^*} (1)$. The convergence of $\widetilde{\mathbb{G}}_{1n}^* (h)$ and its minimization follows by standard arguments as it is a quadratic function of h so that it suffices to examine $\widetilde{\mathbb{G}}_{2n}^* (g)$ and it minimum.

Turning to the second term in the definition of $\widetilde{\mathbb{G}}_{2n}^* (g)$, we show that it converges to $2\delta_{30} \sqrt{3^{-1} f (0) \sigma_\varepsilon^2 (0)} W (g^3)$ weakly (in probability). To this end, note that Lemma 4's, and the Remark 4 that follows, yields the tightness of the process as explained in Remark 3. For the finite dimensional convergence, it follows by standard arguments as

$$E^* \left(\sum_{t=1}^{n} q_t \varepsilon_t^* \mathbf{1}_t \left(0; \frac{g}{n^{1/3}} \right) \right)^2 = \sum_{t=1}^{n} q_t^2 \widehat{\varepsilon}_t^2 \mathbf{1}_t \left(0; \frac{g}{n^{1/3}} \right)$$

which converges in probability to $3^{-1} f (0) \sigma_\varepsilon^2 (0) g^3$ and the Lindeberg's condition follows easily.

Part (**b**) is also proved similarly and thus omitted for the sake of space. ∎

A.6   Proof of Theorem 4

This is a direct consequence of Theorem 3 and Proposition 4 and the same arguments as the proof of Theorem 2. ∎