UNIFIED INFERENCE ON CHANGE–POINT IN SEGMENTED REGRESSION MODELS

JAVIER HIDALGO, JUNGYOON LEE, AND MYUNG HWAN SEO

Abstract. This paper is concerned with unified inference on change-point regression models in two widely used frameworks when the threshold effect is caused by either time or a stochastic variable. More specifically, we examine estimators of the parameters of the model without resorting to prior knowledge on whether there is a kink or a jump in the regression function, where the jump size shrinks to zero as n increases. We also describe a test to distinguish between the jump and kink designs. One of our main and striking results is that the statistical properties of the estimator of the change-point is very different if the constraint of continuity is not employed in the estimation. A second aim of the paper is to establish an asymptotically valid bootstrap which does not require knowledge on whether the regression model is continuous at the change-point, enabling us to offer a unified treatment of change-point inference contrary to other works. Then, we propose to construct confidence intervals for the unknown change-point by bootstrap test inversion, also known as grid bootstrap. Finite sample performances of the bootstrap tests and the grid bootstrap confidence interval are examined and compared against tests based on the asymptotic critical value through Monte Carlo simulation. Finally, we implement our procedure to two empirical applications in economics.

Date: **11 Oct 2015**.

Key words and phrases. Structural Break. Threshold models. Grid Bootstrap. Test Inversion.

Thanks to Russell Davidson, Workshop on time series bootstrap, Anders Rahbek,

1. INTRODUCTION

Since the seminal work by Tong (1990), there has been a huge interest in nonlinear time series modelling, with the Self-Exciting-Threshold-Autoregressive (SETAR) model being one of the most popular models. One of the main reasons for success of these models is that they are parsimonious and yet able to capture many of the nonlinearities present in the data, see Chan and Tsay (1998) or Gonzalo and Wolf (2005) among others for some examples. The SETAR model, and its generalizations better known as Threshold models, were also brought to the attention of econometricians by Hansen (2000), who highlighted the flexibility that these models offer when modelling economic data in general. On the other hand, in a separate or parallel development, structural break regression models have attracted a huge attention since the work by Quandt (1960), as the survey by Perron (2006) demonstrates. Quite surprisingly these two avenues of research, despite their similarities, have mostly followed parallel paths without much reference to each other, although some exceptions exist, see Hansen (2000) or more implicitly Delgado and Hidalgo (2000) in a nonparametric setting.

Broadly speaking these models can be regarded to belong to the class of segmented regression models, which have been examined in the area of approximation theory in mathematics and are better known as splines. The latter methodology is frequently used in nonparametric estimation and routinely employed in functional regression models. However, one key difference between splines and thresholds/break point models is that the points of break or change (knots) between two different segments are chosen differently. In the former approach, the practitioner merely chooses the knots for convenience, whereas in the latter schemes the knots have to be estimated. It is precisely this distinction, e.g. the estimation of the breaks/knots, which yields theoretical challenges that otherwise will not be present. With this in mind, we shall refer to both threshold and structural break models as segmented (regression) models in this paper.

Segmented regression models have become even more relevant to economics due to a recent dramatic surge in theoretical and empirical literature of regression discontinuity (RD) and regression kink (RK) designs, see Lee and Lemieux (2010) for a comprehensive review of RD design, including an extensive list of empirical works and Ganong and Jäger (2015) for a list of empirical applications that uses the RK design. The main identification and estimation results can be found in Hahn, Todd and Van der Klaauw (2001) for RD design and Card, Lee, Pei and Weber (2015) for RK design. In most applications, researchers are interested in estimating the size of a discrete change in the level or slope of a conditional expectation function at a known change-point. The test on the location of the change-point offers a useful test of robustness and validity of the design, as advocated by Landais (2014), and we offer a unified test that works for both RD and RK setups and a bootstrap method that improves upon the finite sample performance of asymptotic test. In some applications, the estimation of unknown change-point is an important part of the analysis and the test of continuity at this change-point is also of interest, see e.g. Card, Mas and Rothstein (2008).

Thus the main aims of this paper are twofold. Firstly, this paper provides unified inferential procedures on segmented models that are either continuous or discontinuous at the point of change or knot. One motivation comes from the observation that the literature appears to examine separately these two frameworks, taking as given a prior knowledge of which framework the true model belongs to. See Chan (1993), Bai and Perron (1998) or Hansen (2000) who assume that the true segmented regression model is discontinuous and Chan and Tsay (1998) or

Feder ($1975a, b$) who assume continuity of the segmented model. However as we shall show the latter assumption is not as innocuous as one may think. So one of our aims is to enable inferences with no assumptions regarding the continuity of the segmented regression model. It is worth mentioning that, as in Bai and Perron (1998) or Hansen (2000), we shall look at the scenario where the size of the break shrinks to zero with the sample size if the segmented regression model is discontinuous. Recall that such a shrinkage is assumed to enable us to make inferences on the parameter of the change-point. However, we shall indicate situations where the conclusions of our results are not affected by the shrinkage, such as when testing for continuity of the regression model. So, we close the gap on inference for this type of models by looking at the properties of estimators and tests when the regression model is either continuous at the change-point or the jump size shrinks to zero as the sample size increases. We believe that this lack of knowledge is the rule rather than the exception with real data sets.

Some of the results of the paper are rather surprising and nonstandard. It is often the case, and belief, that the difference between estimating a regression model using or not using the true constraint among the parameters is related to the relative efficiency of the estimators, while the rates of convergence are the same and they are asymptotically Gaussian. However one of the consequences of not imposing the (true) constraint of continuity in the segmented regression model is a slower rate of convergence of the change/knot point estimator. More specifically, the rate becomes $n^{1/3}$ instead of $n^{1/2}$ obtained by Chan and Tsai (1998) and Feder ($1975a, b$) when they estimated the model imposing correctly the continuity on the segmented regression model. A second difference is that the asymptotic distribution of the change-point estimator is no longer normal, as was the case in the aforementioned work by Chan and Tsai or Feder, but the "*argmax*" of some Gaussian process. The latter casts some doubts on the findings in Gonzalo and Wolf (2005). Fortunately, we can show that a scaled quasi likelihood ratio test for the change-point has the same asymptotic distribution regardless of whether the true regression model is continuous or not, and the estimator of the slope parameters of the model are asymptotically independent of the change-point estimator. More specifically, we obtain the same asymptotic distribution as that obtained in Hansen (2000) up to a scale constant ξ which depends on the (dis-)continuity of the true regression model. Therefore, providing that we can consistently estimate the scale factor ξ without any prior knowledge on the (dis-)continuity of the regression model, we can still make robust inferences on the change/threshold point, see Theorem 2 and Proposition 2 below for details. To that end, we shall present an estimator $\widehat{\xi}$ for the scaling factor ξ based on the ratio of two kernel Nadaraya-Watson estimators, each of which converges to zero in probability under the constraint of continuity. However, similar to L'Hopital rule in real analysis, we show that the ratio still converges in probability to the proper scale factor ξ. In addition, rather surprisingly, using higher order kernels does not yield consistent estimation of ξ, contrary to many other cases in nonparametric estimation.

We now turn to our second main goal, that is to examine a bootstrap scheme. This is partly motivated by the fact that the Monte-Carlo simulation experiment suggests that the asymptotic distribution does not provide a good approximation to the finite sample one. Bootstrap methods have been shown to be highly successful in many standard inferential problems since it was introduced by Efron (1979). The main reason for its success is its ability to provide better approximations to the finite sampling distribution than their asymptotic counterparts. However, it has been noted that in nonregular scenarios when the asymptotic distribution is not Gaussian, the standard bootstrap schemes might no longer be valid. See for instance when testing for unit roots Basawa, Mallik, McCormick and Taylor (1991), or for inference of the threshold parameter (Seijo and Sen, 2011), or in the maximum score estimation (Abrevaya and Huang, 2005) among others. One of the possible reasons suggested for their lack of validity is that they are "non-regular/standard" inferential problems and the asymptotic distributions of the statistics are not Gaussian. Although correct resampling schemes have been proposed for all the frameworks just mentioned, they are by nature tailor made for each particular problem, suggesting that one should exercise some caution when applying the bootstrap to statistics in "non-regular" situations.

In our regression design, we shall discuss a bootstrap scheme, based on the wild bootstrap principle, which is valid to approximate the sampling distribution of the scaled quasi likelihood ratio statistic, see Section 2.2 below, for all the scenarios we have considered in this paper. More specifically, the bootstrap is asymptotically valid whether the true segmented regression model is continuous or not. The validity of the bootstrap is in contrast to Seijo and Sen's (2011a) results as they showed that the nonparametric bootstrap and the residual bootstrap for the threshold parameter are not valid under the non-shrinking asymptotics. This contrast is perhaps due to the fact that the nuisance parameter in the asymptotic distribution under the non-shrinking design is infinite-dimensional while the ones in our continuous or shrinking designs are finite-dimensional scaling terms.

We then present findings of a small Monte Carlo experiment to evaluate the finite sample performance of our bootstrap procedure for various frameworks covered by the paper. Specifically, we examine the Monte Carlo size of tests for a hypothesis concerning the change-point, based on asymptotic theory of Hansen (2000) and two bootstrap methods, one of which is proposed in this paper and the other that is standard. We then investigate coverage probabilities of confidence intervals, constructed from asymptotic theory of Hansen (2000), percentile-t method with the standard residual-based wild bootstrap, and test inversion based on the nonstandard bootstrap considered in this paper. Some interesting findings emerge. First, inference based on the asymptotic distribution of our statistics and the standard residual bootstrap method tends to have large size distortions and/or poor coverage probability in many cases. Second, the hypothesis testing and test inversion confidence interval based on the bootstrap proposed in this paper seem to work well across all the designs employed in the Monte-Carlo experiment. In particular, the bootstrap method proposed in this paper seems to be the only one that works in the kink design.

Finally, we apply inference methods based on the bootstrap scheme of this paper to data of Card, Mas, Rothstein (2008). They model a "tipping" behavior of neighborhood racial composition in the US cities over time, whereby a reduction in the share of white population occurs once the neighborhood's minority share exceeds a certain city-specific threshold. We present grid bootstrap confidence interval and test of continuity for three large cities. In two cities we find continuity is not rejected, suggesting the restricted estimation of change-point which yields a faster rate of convergence compared to unrestricted estimation would provide a better estimate of the change-point for those cities. We then implement a test of continuity on US unemployment data of Chan and Tsay (1998) and Gonzalo and Wolf (2005), which leads to a different conclusion from that obtained by Gonzalo and Wolf (2005) using a test based on subsampling. To gain some further insights on how our test compared to that in Gonzalo and Wolf (2005), we present a small Monte Carlo study that mirrors the data in order to compare relative size and power performance of various tests of continuity in the supplementary appendix.

The latter Monte Carlo experiment appears to corroborate the contrasting results we obtain.

The paper is organized as follows. Section 2 introduces the model and a quasi likelihood ratio test to make inferences on the threshold parameter. We also describe how to construct confidence intervals using asymptotic critical values and a bootstrap algorithm. Regularity assumptions and asymptotic distribution of the proposed quasi likelihood ratio statistic are given in Section 3, which also presents a test for the continuity of the segmented regression function. Section 4 gives formal theoretical justification for the bootstrap method. Section 5 examines finite sample property of the proposed methods via a Monte Carlo simulation experiment. Section 6 illustrates our findings in the two empirical examples mentioned above. Section 7 concludes, whereas all the proofs are delegated to the appendices in the supplement.

2. SEGMENTED REGRESSION MODEL

We shall consider the segmented regression model

$$\begin{aligned} y_t &= \beta' x_t + \delta' x_t \mathbf{1}\{q_t > \gamma\} + \varepsilon_t \\ &= \alpha' x_t(\gamma) + \varepsilon_t, \end{aligned} \tag{2.1}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function, x_t is a k-dimensional vector of regressors, q_t is the transition variable, $\alpha = (\beta', \delta')'$ is a $2k$-dimensional vector of coefficients, γ is the change-point parameter and ε_t is random error. The transition variable q_t may or may not be a part of the vector x_t. Abbreviating $\mathbf{1}\{q_t > \gamma\}$ by $\mathbf{1}_t(\gamma)$ and $\mathbf{1}\{q_t > \gamma_0\}$ by $\mathbf{1}_t$ in what follows, denote

$$x_t(\gamma) = \begin{pmatrix} x_t \\ x_t \mathbf{1}_t(\gamma) \end{pmatrix}; \quad \mathbf{x_t} = x_t(\gamma_0). \tag{2.2}$$

In addition, we shall abbreviate $\mathbf{1}\{\gamma < q_t < \widetilde{\gamma}\}$ by $\mathbf{1}_t(\gamma;\widetilde{\gamma})$, so that $\mathbf{1}_t(\gamma;\infty) =: \mathbf{1}_t(\gamma)$.

When q_t is a stochastic variable, the model (2.1) is known as the threshold regression, while it corresponds to structural break model when $q_t = t/n$. As indicated in the introduction, we shall refer to both setups as segmented regression models.

We estimate the unknown parameter $\theta_0 = (\alpha_0', \gamma_0)'$ by the least squares method. That is

$$\widehat{\theta} = (\widehat{\alpha}', \widehat{\gamma})' =: \arg \min_{\theta \in (\Theta_1, \Gamma)} \mathbb{S}_n(\theta),$$

where $\Theta = (\Theta_1, \Gamma)$ is a compact set in $\mathbb{R}^{2k+1}$ and

$$\mathbb{S}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - \alpha' x_t(\gamma)\right)^2. \tag{2.3}$$

The estimator $\widehat{\theta}$ is often obtained via a step-wise algorithm: define the concentrated (profiled) sum of squared residuals (with some abuse of notation)

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

where

$$\widehat{\alpha}(\gamma) = \arg\min_{\alpha\in\Theta_1} \frac{1}{n}\sum_{t=1}^{n}\left(y_t - \alpha' x_t(\gamma)\right)^2 \tag{2.4}$$

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

$$\widehat{\gamma} = \arg\min_{\gamma\in\Gamma_n} \mathbb{S}_n(\gamma) \tag{2.5}$$

and $\Gamma_n = \Gamma \cap \{q_1, ..., q_n\}$.

In the following subsections, we present and describe inferential procedures for θ_0, although our main interest lies in inferences on γ_0 since those for α_0 are the same as in the linear regression with known γ_0. The theoretical justification for the proposed procedures and some technical details will be postponed to Section 3 under quite general assumptions, which encompass conditions associated with the (dis-)continuity of the regression function.

2.1. **Unified Inference for Change-Point.** As an overview of one of our main findings of this paper, this section outlines how we will make inferences on the change-point γ_0 without knowing whether the regression function is continuous or not. For that purpose, suppose that we are concerned with the null hypothesis

$$\mathcal{H}_0 : \gamma = \gamma_0, \tag{2.6}$$

where γ_0 is an interior point of Γ, or with the construction of confidence intervals of γ_0. Following Hansen (2000), among others, one common and sensible approach is to employ the Gaussian quasi likelihood ratio (QLR_n) statistic, normalized by a suitable scaling term. Define the (unscaled) QLR_n test statistic as

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

Theorem 2 of Section 3.2 shows that QLR_n statistic is asymptotically pivotal up to an unknown scaling factor ξ for all the scenarios we consider. Furthermore, let

$$\widehat{\xi} = \frac{\sum_{t=1}^{n} \left(\widehat{\delta}' x_t\right)^2 \widehat{\varepsilon}_t^2 K\left(\frac{q_t - \widehat{\gamma}}{a}\right)}{\mathbb{S}_n\left(\widehat{\theta}\right) \sum_{t=1}^{n} \left(\widehat{\delta}' x_t\right)^2 K\left(\frac{q_t - \widehat{\gamma}}{a}\right)}, \tag{2.7}$$

where $K(\cdot)$ and a are, respectively, the kernel function and bandwidth parameter and $\widehat{\varepsilon}_t$ are the least squares residuals. See Assumption K below for specific assumptions on $K(\cdot)$ and a. As a consequence of Proposition 2 of Section 3.3, the above estimator $\widehat{\xi}$ consistently estimates the scale factor ξ irrespective of the (dis-)continuity of the regression function. In case of $q_t = t/n$, we may also set $K(\cdot) = 1$ and $q_t = \widehat{\gamma}$.

Interestingly, although the weak limit of QLR_n and the probability limit of $\widehat{\xi}$ differ depending on whether the model is continuous or not, we can still make inference without prior knowledge of the continuity of the segmented regression model since the rescaled QLR_n statistic $\widehat{\xi}^{-1}QLR_n$ has the same weak limit in all cases. Thus comes the title of this section. Then, we construct the s-level critical region for the rescaled QLR_n by

$$\left\{\widehat{\xi}^{-1}QLR_n \geq F^{-1}\left(1-s\right)\right\},$$

where $F^{-1}(\cdot)$ is the inverse function of $F(z) = \left(1 - e^{-z/2}\right)^2$.

2.2. **Bootstrap Test.** In this section we shall describe a bootstrap scheme for QLR_n that is valid regardless of whether the segmented regression model has a jump or a kink at γ_0. The motivation, as indicated in the introduction, is due to the observation that the asymptotic distribution appears not to be a good approximation of the finite sample distribution of QLR_n, as the Monte Carlo experiment in Section 5 suggests. As usual, the superscript "*" indicates the bootstrap quantities and convergences of bootstrap statistics conditional on the original sample. The bootstrap, denoted as **NB** in what follows, is described in the following *4 STEPS*. Some comments regarding the motivation to employ the bootstrap algorithm below can be seen at the end of the theoretical justification of the algorithm in Section 4.

***STEP 1*:** Obtain the estimators of $(\alpha_0', \gamma_0)'$, say $(\widehat{\alpha}', \widehat{\gamma})'$ by minimizing (2.3), using (2.4) – (2.5), and compute the residuals

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

and the restricted least squares estimator $\widetilde{\alpha} = \arg\min_{\alpha \in \Theta_1} \sum_{t=1}^{n} (y_t - \alpha' \mathbf{x_t})^2$, where $x_t(\gamma)$ and $\mathbf{x_t}$ are given in (2.2).

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

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

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

$$\widehat{\theta}^* = \arg\min_{\theta} \mathbb{S}_n^*(\theta) := \frac{1}{n} \sum_{t=1}^{n} \left(y_t^* - x_t(\gamma)' \alpha\right)^2. \tag{2.8}$$

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

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

and

$$\widehat{\xi}^* = \frac{\sum_{t=1}^{n} \left(\widehat{\delta}^{*\prime} x_t\right)^2 \widehat{\varepsilon}_t^{*2} K\left(\frac{q_t - \widehat{\gamma}^*}{a}\right)}{\mathbb{S}_n\left(\widehat{\theta}^*\right) \sum_{t=1}^{n} \left(\widehat{\delta}^{*\prime} x_t\right)^2 K\left(\frac{q_t - \widehat{\gamma}^*}{a}\right)}. \tag{2.9}$$

As usual to compute the bootstrap p-value, p^*, we repeat *STEPS 2-4* B times and obtain the proportion of times that $\widehat{\xi}^{*-1} QLR_n^*$ exceeds the sample statistic $\widehat{\xi}^{-1} QLR_n$. Then we reject the null hypothesis at a s level if $p^* < s$.

2.3. **Bootstrap Confidence Interval.** We propose using the bootstrap test inversion method (also known as the grid bootstrap) of Dümbgen (1991), see also Carpenter (1999) and Hansen (1999), to build confidence intervals for the parameter γ. For a given confidence level ζ, one can exploit the duality between hypothesis testing and confidence intervals by inverting tests to obtain

$$\left\{\gamma : \widehat{\xi}\left(\gamma\right)^{-1} QLR_n\left(\gamma\right) \leq q_n^*\left(\zeta|\gamma\right)\right\},$$

where $q_n^*\left(\zeta|\gamma\right)$ denotes the bootstrap quantile function. In practice, one would bootstrap $q_n^*\left(\zeta|\gamma\right)$ over a grid of γ and use some smoothing method such as linear interpolation or kernel averaging to obtain a smoothed bootstrap quantile function over a range of γ. Above interval can also be called ζ-level grid bootstrap confidence interval (BCI) of γ in the terminology of Hansen (1999). Such a test inversion BCI is known to have certain optimality property as in e.g. Brown, Casella and Hwang (1995) from the Bayesian perspective. Mikusheva (2007) showed that test inversion BCI attains correct coverage probability uniformly over the parameter space for sum of coefficients in autoregressive models, despite the behavior of the estimator being not uniform over the parameter space.

Figure 1 here (fig1.pdf)

Figure 1 illustrates how this confidence interval can be obtained in practice. The $QLR_n\left(\gamma\right)$ line is the linear interpolation of the rescaled $QLR_n\left(\gamma\right)$ statistic over the grid of γ at 50 points. The ACV line is the asymptotic critical value of Hansen (2000). We estimated bootstrap quantile function at 17 grid points and present the interpolated line as Grid quantile plot. The vertical arrow at intersections between $QLR_n\left(\gamma\right)$ and ACV yield the asymptotic confidence interval (ACI), while the vertical broken arrows indicate grid BCI based on the bootstrap.

3. REGULARITY CONDITIONS AND ASYMPTOTIC RESULTS

This section establishes the asymptotic validity of the proposed inferential procedures. We begin by stating some regularity assumptions on the distribution of the data, for which we introduce the following notations.

Let $f(\cdot)$ be the density function of the transition variable q_t. Denote $D(\gamma) = E(x_t x_t' | q_t = \gamma)$, $V(\gamma) = E(x_t x_t' \varepsilon_t^2 | q_t = \gamma)$, $\sigma^2(\gamma) = E(\varepsilon_t^2 \mid q_t = \gamma)$ and $D = D(\gamma_0)$, $V = V(\gamma_0)$, $\sigma^2 = E(\varepsilon_t^2)$. In case of $q_t = t/n$, we may regard q_t as a Uniform$[0,1]$ random variable that is independent of (other) elements of x_t and ε_t when constructing these quantities: that is, $f(\cdot) = 1$ and $\sigma^2(\cdot) = \sigma^2$ while $D(\gamma)$ and $V(\gamma)$ are unconditional expectations for all elements, apart from those involving q_t if q_t is part of x_t. Finally, let $M = \text{plim}_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbf{x_t x_t'}$, and $\Omega = \text{plim}_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbf{x_t x_t'}\varepsilon_t^2$.

Regarding the transition variable q_t and the parameter space Γ, we assume one of the following two conditions. Assumption 3 below is for the structural break model:

Assumption Q1. *In (2.1), $q_t = t/n$ and Γ is a proper compact subset of $(0,1)$.*

When $\{q_t\}_{t\in\mathbb{Z}}$ is a sequence of random variables, we make Assumption 3 instead:

Assumption Q2. *Let q_t have the probability density function $f(\cdot)$. Then, for all $\gamma \in \Gamma$, $f(\gamma)$, $E\left(|x_t|^4 | q_t = \gamma\right)$ and $E\left(|x_t \varepsilon_t|^4 | q_t = \gamma\right)$ are bounded by some $C < \infty$. The functions $f(\gamma)$, $V(\gamma)$ and $D(\gamma)$ are continuous at $\gamma = \gamma_0$ and $f(\gamma)$, $E\big(x_t x_t' \mathbf{1}\{q_t \le \gamma\}\big) > 0$ for all $\gamma \in \Gamma$.*

Assumption Z. *Let z_t be the stochastic component of $(x_t', \varepsilon_t)'$. Then, $\{z_t\}_{t\in\mathbb{Z}}$ is a strictly stationary, ergodic sequence of random variables such that $E\|x_t\|^4 < \infty$, $E\|x_t\varepsilon_t\|^4 < \infty$, and ρ-mixing satisfying that $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$. Also $E(\varepsilon_t | \mathcal{F}_{t-1}) =$*

0, where $\mathcal{F}_{t-1}$ is the information set up to time $t-1$ and $E\left|\varepsilon_t\right|^{4+\eta} < \infty$ for some $\eta > 0$. Furthermore, $M > 0$ and $\Omega > 0$.

Assumption Z is a set of fairly standard type of conditions on the dependence and moments as in e.g. Hansen (2000). Under the above assumptions, $M = E\mathbf{x_t}\mathbf{x_t}'$ and $\Omega = E\mathbf{x_t}\mathbf{x_t}'\varepsilon_t^2$ when q_t is stochastic. They are identical to the case where q_t is Uniform$[0,1]$ and independent of z_t, when $q_t = t/n$.

The next concerns conditions on θ_0, in particular, the size of the break δ_0. Here we use δ_n to indicate that the true value of δ can change over the sample size n.

Assumption J. *For some $0 < \varphi < 1/2$ and a vector $c \neq 0$, $\delta_0 = \delta_n = cn^{-\varphi}$ and $c'Dc$ and $c'Vc$ are positive.*

Assumption J states that the size of the break shrinks towards zero as the sample size increases while the regression function is discontinuous at the change-point for each n. Under these assumptions, the asymptotic distribution of $\widehat{\gamma}$ becomes the maximum of some Gaussian process, which was first obtained explicitly by Yao (1987). Otherwise, we may get very different asymptotic distribution of $\widehat{\gamma}$. When $\varphi = 0$, the asymptotic distribution of $\widehat{\gamma}$ is model-dependent and hence less useful for the purpose of making inferences. When φ is greater than or equal to $1/2$, δ_n is too small to point-identify γ_0.

To see the presence of a jump in more detail in contrast to the kink design, denote

$$x_t = (x_{t1}', q_t)'; \quad x_{t1} = (1, x_{t2}')', \tag{3.1}$$

noting that comparison between jump and kink setups is relevant when the transition variable q_t is a part of the vector x_t. We can partition δ accordingly, i.e.

$$\delta = (\delta_1, \delta_2', \delta_3)'.$$

Then, Assumption J implies that

$$\delta_n' D \delta_n = E\left[(x_t'\delta_n)^2 | q_t = \gamma_0 \right] = E\left[(\delta_{1n} + \delta_{2n}' x_{2t} + \delta_{3n}\gamma_0)^2 | q_t = \gamma_0 \right] > 0,$$

which implies that there is a nonzero jump in the regression function at $q_t = \gamma_0$ with positive probability. This is one of the reasons for the super consistent estimation of γ_0 in Hansen (2000). However when we allow for the continuous change with a kink, the statistical properties of $\widehat{\gamma}$ are quite different.

3.1. **Kink Design.** This section considers the case where the regression function $E(y_t|x_t)$ is continuous at γ_0, in the sense that $\delta_0' x_t = 0$ when $q_t = \gamma_0$ for all x_{2t}, and it is kinky at γ_0, noting the notation from (3.1). That is,

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

$$\delta_{10} + \delta_{30}\gamma_0 = 0; \quad \delta_{20} = 0. \tag{3.2}$$

Indeed, Assumption C implies that our regression model (2.1) becomes

$$y_t = x_t'\beta_0 + \delta_{30}(q_t - \gamma_0)\mathbf{1}_t(\gamma_0) + \varepsilon_t. \tag{3.3}$$

This kink design features prominently in the RK design literature of economics, and Landais (2014) advocates testing $H_0 : \gamma = \gamma_0$ to check for robustness and validity of the design. It is worth noting that this case was excluded by Hansen (2000), who assumes Assumption J. As we shall see in Proposition 1 and Theorem 1 this assumption is not innocuous and dropping it leads to some important consequences. This is a rather relevant result because with real data sets, it is difficult to know *a priori* whether the regression function has a kink or a jump.

Before we give our main results, it is worth providing a heuristic discussion on how imposing the continuity restriction (3.2) changes the asymptotic distribution of the least squares estimator and how the continuity of the segmented regression model affects the asymptotic distribution of the unrestricted estimator $\widehat{\theta}$, and in particular that of $\widehat{\gamma}$, i.e. when the constraints in (3.2) are not employed in the estimation. We will see that the consequences are rather striking and go beyond the standard loss of efficiency.

For simplicity of illustration, we begin with the model in (3.3) with $\beta_0 = 0$ and $x_t = (1, q_t)'$ and $\gamma_0 = 0$ since we can always rename the variable $q_t - \gamma_0$ as q_t. Let

$$\left(\widehat{\delta}, \widehat{\gamma}\right) = \arg\min_{\delta,\gamma} \ \mathbb{S}_n(\delta;\gamma) := \frac{1}{n}\sum_{t=1}^{n}\left(y_t - (\delta_1 + \delta_3 q_t)\,\mathbf{1}_t(\gamma)\right)^2.$$

On the other hand, when the (true) constraints given in (3.2) are imposed in the estimation procedure, the constrained least squares estimator is given by

$$\left(\widetilde{\delta}_3, \widetilde{\gamma}\right) = \arg\min_{\delta_3,\gamma} \ \widetilde{\mathbb{S}}_n(\delta_3;\gamma) := \frac{1}{n}\sum_{t=1}^{n}\left(y_t - \delta_3\left(q_t - \gamma\right)\mathbf{1}_t(\gamma)\right)^2.$$

Then, standard algebra yields that for $\gamma_0 = 0$,

$$\begin{aligned} (3.4)\quad & \mathbb{S}_n(\delta;\gamma) - \mathbb{S}_n(\delta_0;0) \\ = \quad & \frac{1}{n}\sum_{t=1}^{n}\left(\delta_{30} q_t \mathbf{1}_t - (\delta_1 + \delta_3 q_t)\,\mathbf{1}_t(\gamma)\right)^2 + \frac{2}{n}\sum_{t=1}^{n}\varepsilon_t\left(\delta_{30} q_t \mathbf{1}_t - (\delta_1 + \delta_3 q_t)\,\mathbf{1}_t(\gamma)\right), \end{aligned}$$

and

$$\begin{aligned} (3.5)\quad & \widetilde{\mathbb{S}}_n(\delta_3;\gamma) - \widetilde{\mathbb{S}}_n(\delta_{30};0) \\ = \quad & \frac{1}{n}\sum_{t=1}^{n}\left(\delta_{30} q_t \mathbf{1}_t - \delta_3\left(q_t - \gamma\right)\mathbf{1}_t(\gamma)\right)^2 + \frac{2}{n}\sum_{t=1}^{n}\varepsilon_t\left(\delta_{30} q_t \mathbf{1}_t - \delta_3\left(q_t - \gamma\right)\mathbf{1}_t(\gamma)\right). \end{aligned}$$

We first examine the behavior of (3.4). Assuming the consistency of the estimators, we have that the expectation of the right side of (3.4) is

$$\begin{aligned} & E\left(\delta_{30} q_t \mathbf{1}_t - \left(\delta_1 + \delta_3 q_t\right) \mathbf{1}_t(\gamma)\right)^2 \\ = \; & E\left(\delta_1 + \left(\delta_3 - \delta_{30}\right) q_t\right)^2 \mathbf{1}_t(\gamma) + \left(\delta_1 + \delta_3 q_t\right)^2 \mathbf{1}_t(0; \gamma) \\ \sim \; & \left\|\delta - \delta_0\right\|^2 + \gamma^3 \end{aligned}$$

since $E\left\{q_t^2 \mathbf{1}_t(0; \gamma)\right\} = |\gamma|^3 (f(0) + o(1))$. Whereas

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

Thus, the last two displayed expressions suggest that

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

since these rates of convergence balance the speed at which the bias and standard deviation of $\mathbb{S}_n(\delta; \gamma) - \mathbb{S}_n(\delta_0; \gamma_0 = 0)$ converge to zero.

Similarly for (3.5), we have that

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

so that we expect that

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

which is what Feder (1990) and Chan and Tsai (1975a, b) obtained.

An intuitive explanation of this phenomenon is that it is more difficult to find the minimizer of $\mathbb{S}_n$ when the limit of $\mathbb{S}_n$ is locally *cubic* in $|\gamma - \gamma_0|$ around

γ_0 than when it is locally quadratic or locally linear with a kink at γ : that is, the smoother the segmented regression model is around the kink, the more difficult is its estimation, see also Remark 1 below. This happens because we estimate a discontinuous function while the true regression function is continuous. This might be viewed as a mild form of model mis-specification. Note that in the standard M-estimation, the limit criterion function is quadratic around the minimum whether a restriction on the parameter is imposed or not.

We formalize the above heuristic arguments in the next proposition.

Proposition 1. *Under Assumptions Z, either Q1 or Q2, and C,*

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

As we see from the results of Proposition 1 the main consequence of not using the true constraints of the model is the slower rate of convergence of the estimator of the change-point parameter $\widehat{\gamma}$, as when Assumption C is employed in the estimation, the results of Chan and Tsay (1998) or Feder (1975a, b) imply root-n consistency of the change-point estimator $\widehat{\gamma}$ of γ_0. Then,

Theorem 1. *Under Assumptions Z, either Q1 or Q2, and C,*

$$n^{1/2}(\widehat{\alpha} - \alpha_0) \xrightarrow{d} \mathcal{N}\left(0, M^{-1}\Omega M^{-1}\right)$$
$$n^{1/3}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \arg\max_{g\in\mathbb{R}} \left(2\delta_{30}\sqrt{\frac{\sigma^2\left(\gamma_0\right) f\left(\gamma_0\right)}{3}} W\left(g^3\right) + \frac{\delta_{30}^2}{3} f\left(\gamma_0\right)\left|g\right|^3\right),$$

where the two limits are independent and $W(g) = B_1(-g)\,1\{g<0\} + B_2(g)\,1\{g>0\}$ *with two independent standard Brownian motions* $B_1(\cdot)$ *and* $B_2(\cdot)$.

The results in Theorem 1 corroborate the informal discussion given above regarding the rate of convergence of the estimators. This is a very interesting and surprising result because imposing restrictions on the parameter space improves the rate of convergence of the estimator, not just reducing the asymptotic variance of the estimator. We also note that Seo and Linton (2007) considered the smoothed least squares for the same setup. However, the convergence rate for their smoothed least squares estimator for γ was slower than our cube root under their assumptions for the smoothing parameter.

Another important consequence is that the estimators of the slope parameters and the change-point are mutually independent. Therefore, the inference on α_0 is the same as with the OLS of the linear regression model. This is not the case if the constraint on the parameters given in (3.2) were used, under which Chan and Tsay (1998) show the joint asymptotic normality of $\widehat{\alpha}$ and $\widehat{\gamma}$.

Remark 1. *The findings in Theorem 1 can be easily extended to more general settings such as those implicit in Feder* (1975a, b)*, where* $x_t'\delta\mathbf{1}_t(\gamma)$ *is replaced by* $g(x_t;\delta,\gamma)$ *and* g'*s first* $(m-1)$*-th order derivatives with respect to* q *degenerate. Then,* $\widehat{\gamma}-\gamma_0 = O_p\left(n^{-1/(2m+1)}\right)$ *whereas Feder obtained* $\widehat{\gamma}-\gamma_0 = O_p\left(n^{-1/2m}\right)$ *imposing the continuity constraint. See online-supplement for more details.*

Recalling the previous literature like Bai (1997) or Hansen (2000), we note that the inference for α_0 can be performed as in the linear regression (where γ is fixed at γ_0), whether the true parameter value satisfies Assumption J or Assumption C. However, it is not obvious if we can make inference for γ_0 without knowing the continuity of the regression function. This is answered in Section 3.2.

3.2. **Asymptotics for QLR$_n$.** This section establishes the validity of QLR_n based inference introduced in Section 2.1. Some of the results given below are reiteration of existing ones, while other results are new and quite surprising.

First, define the scaling factor that appears in the asymptotic distribution,

$$\xi=\lim_{n\to\infty}\frac{E\left(\delta_0'x_t\right)^2\varepsilon_t^2\phi\left(\left|q_t-\gamma_0\right|/a\right)}{\sigma^2E\left(\delta_0'x_t\right)^2\phi\left(\left|q_t-\gamma_0\right|/a\right)},$$

where ϕ is the standard normal density and $a\to0$. If $q_t=t/n$, as in Section 3, ξ is defined identically to the case where q_t is Uniform$[0,1]$ and independent of $z_t=(x_{t2}',\varepsilon_t)'$. This limit is well-defined by L'Hopital's rule. For instance, $\xi=\frac{c'Vc}{\sigma^2c'Dc}$ under Assumption J, $\xi=\frac{\sigma^2(\gamma_0)}{\sigma^2}$ under Assumption C, and $\xi=1$ under the conditional homoskedasticity or under Assumptions Q1 and C. Then,

Theorem 2. *Suppose that Assumptions Z and either Q1 or Q2 hold. Furthermore, suppose that either Assumption J or C hold. Then, as $n\to\infty$,*

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

Theorem 2 shows that, despite the different asymptotic properties of the change-point estimates under different scenarios, a properly scaled QLR_n statistic should yield a pivotal asymptotic distribution, regardless of whether the segmented regression model has a kink or it is discontinuous at γ_0. The distribution function of $Z:=\max_{g\in\mathbb{R}}\left(2W\left(g\right)-\left|g\right|\right)$ is $F\left(z\right)=\left(1-e^{-z/2}\right)^2$, see Section 2.1.

Generally, we would expect either having to estimate the unknown scaling factor presuming which case is the true data generating process, or resorting to some resampling scheme to make inference on γ_0. In Section 3.3 below we show that $\widehat{\xi}$ of (2.7) converges to a proper scaling term in every case. That is, the scaled QLR_n statistic $\widehat{\xi}^{-1}QLR_n$ converges to Z for all the cases. So, we obtain the important consequence that we do not need to know whether the data generating process is continuous or discontinuous in order to construct confidence intervals for the unknown change-point parameter γ_0.

3.3. **Consistency of the Scale Estimator.** This section gives the conditions under which $\widehat{\xi}$ defined in (2.7) consistently estimates the scale factor ξ in all the cases. When q_t is stochastic, dividing the numerator and denominator of (2.7) by $\sum_{t=1}^{n} K\left(\frac{q_t-\widehat{\gamma}}{a}\right)$, we notice that $\widehat{\xi}$ is actually the ratio of a Nadaraya-Watson kernel estimator for $E\left((\delta' x_t)^2 \varepsilon_t^2 \mid q_t = \gamma_0\right)$ and $E\left((\delta' x_t)^2 \mid q_t = \gamma_0\right)$. Under the setting of discontinuity of the segmented regression model, Hansen (2000) argued for the consistency of $\widehat{\xi}$. However when Assumption C holds true, we have that $\delta_0' x_t = \delta_{30}(q_t - \gamma_0) = 0$ at $q_t = \gamma_0$. So, both numerator and denominator of the kernel estimator $\widehat{\xi}$ converge to zero in probability. However as when obtaining the limit of the ratio of two sequences converging to zero, by L'Hopital rule, if both converge to zero at the same rate, the ratio may still possess a proper limit. This is precisely what Proposition 2 below shows and that the limit of the ratio is ξ. When $q_t = t/n$, it is sufficient to recall the standard uniform law of large numbers for heterogeneous sequences and the standard algebra for the sums involving t/n. We omit the details here.

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 $K'(u+w)/K'(u) \to 1$ for all u such that $|w/u| \leq C$ as $w \to 0$.
- **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 the higher order kernels by assuming $\kappa_2 \neq 0$. See Remark 2 following the statement of Proposition 2 for more discussion on this. Assumption K suffices to guarantee the consistency of $\widehat{\xi}$ under the discontinuity design by the standard consistency result of the Nadaraya-Watson estimator.

The following proposition establishes that $\widehat{\xi}$ is also consistent under the continuity constraint given in Assumption C.

Proposition 2. *Suppose Assumptions Z, Q2, C and K hold true. Then,*

$$\widehat{\xi} \xrightarrow{P} \frac{\sigma_{\varepsilon}^{2}\left(\gamma_{0}\right)}{\sigma^{2}}.$$

The main and important consequence of Proposition 2 is that $\widehat{\xi}$ is consistent regardless of whether the regression function is continuous, i.e. it has a kink at the change-point γ_0, or it is discontinuous. Thus, the latter implies that we have indeed been able to provide a robust estimator of the scalar constant ξ in Theorem 2 and we can make valid inferences without the need to pretest for the type of change we have at γ_0.

The consequence of dropping the assumption that $\kappa_2 \neq 0$ is discussed in detail in Remark 3 in the online supplement.

3.4. **Testing for Continuity.** We conclude this section by describing and examining a simple test for the null hypothesis of continuity of the segmented regression model. More specifically, we describe a procedure, using the results of Theorem 1, to perform a very simple test for the null hypothesis of continuity in the framework of segmented regression models. The motivation of the test is similar to that for the Hausman-Durbin-Wu statistic, but while the latter exploits the fact that one of the estimators is efficient under the null hypothesis whereas only one of them is consistent under the alternative, in our present context we use the observation that the rates of convergence under the null and alternative are different. For that purpose, denote by $\widehat{\gamma}_{\mathsf{Res}}$ the estimator of γ_0 using (3.2) in

Assumption C and by $\widehat{\gamma}_{\mathsf{Unres}}$ when the latter restriction is not employed in the estimation of the parameters. Then, Theorem 1 and results in Chan and Tsay (1998) or Feder (1975a, b) suggest to test whether

$$\widehat{\gamma}_{\mathsf{Res}} - \widehat{\gamma}_{\mathsf{Unres}}$$

is significantly different from zero as the rate of convergence of $\widehat{\gamma}_{\mathsf{Res}}$ and $\widehat{\gamma}_{\mathsf{Unres}}$ to γ_0 are different under the null hypothesis of continuity/kink. More specifically, under the null hypothesis, i.e. (3.2), we have that

$$\begin{aligned} n^{1/3}\left(\widehat{\gamma}_{\mathsf{Unres}} - \widehat{\gamma}_{\mathsf{Res}}\right) &\overset{d}{=} n^{1/3}\left(\widehat{\gamma}_{\mathsf{Unres}} - \gamma_0\right) + n^{1/3}\left(\widehat{\gamma}_{\mathsf{Res}} - \gamma_0\right), \\ &\overset{d}{=} n^{1/3}\left(\widehat{\gamma}_{\mathsf{Unres}} - \gamma_0\right) + o_p(1) \end{aligned} \tag{3.6}$$

whereas under the alternative hypothesis of no kink but having a jump at γ_0, we expect that $\widehat{\gamma}_{\mathsf{Res}}$ does not converge to γ_0 while $\widehat{\gamma}_{\mathsf{Unres}}$ still does. Hence we have that, under the alternative hypothesis,

$$\begin{aligned} n^{1/3}\left(\widehat{\gamma}_{\mathsf{Res}} - \widehat{\gamma}_{\mathsf{Unres}}\right) &= n^{1/3}\left(\widehat{\gamma}_{\mathsf{Res}} - \gamma_0\right) + n^{1/3}\left(\widehat{\gamma}_{\mathsf{Unres}} - \gamma_0\right) \\ &= O_p\left(n^{1/3}\right) + o_p\left(n^{2\varphi - 2/3}\right), \end{aligned}$$

because under the alternative hypothesis, $\widehat{\gamma}_{\mathsf{Unres}} - \gamma_0 = o_p\left(n^{-1/2}\right)$. Indeed this is the case since we know that $\widehat{\gamma}_{\mathsf{Unres}} - \gamma_0 = O_p\left(n^{-1}\right)$ if there were no shrinkage, see Chan (1993) or Bai and Perron (1998), whereas $\widehat{\gamma}_{\mathsf{Unres}} - \gamma_0 = O_p\left(n^{-1+2\varphi}\right)$ if there were shrinkage, see Hansen (2000) or Bai and Perron (1998) with $0 < \varphi < 1/2$. So, $n^{1/3}\left|\widehat{\gamma}_{\mathsf{Res}} - \widehat{\gamma}_{\mathsf{Unres}}\right|$ diverges under the alternative hypothesis because $\varphi < 1/2$, which suffices for the consistency of the test. The following is an immediate consequence of (3.6) and Theorem 1.

Corollary 1. *Under the Assumptions of Theorem 1, we have that*

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

4. VALIDITY OF THE BOOTSTRAP

This section presents the consistency of the bootstrap statistic $\widehat{\xi}^{*-1}QLR_n^*$ that was introduced in Section 2.2. We begin with the asymptotic distribution of the least sqaures estimator $\widehat{\alpha}^*$ and $\widehat{\gamma}^*$.

Theorem 3. *Suppose that Assumptions Z and either Q1 or Q2 hold.*
(**a**) *Under Assumption C, $\widehat{\alpha}^*$ and $\widehat{\gamma}^*$ are asymptotically independent and*

$$n^{1/2}(\widehat{\alpha}^*-\widetilde{\alpha})\xrightarrow{d^*}\mathcal{N}\left(0,M^{-1}\Omega M^{-1}\right),\quad \textit{in probability,}$$
$$n^{1/3}(\widehat{\gamma}^*-\gamma_0)\xrightarrow{d^*}\arg\max_{g\in\mathbb{R}}\left(2\delta_{30}\sqrt{\frac{\sigma^{2}\left(\gamma_0\right)}{3}f\left(\gamma_0\right)}W\left(g^3\right)+\frac{\delta_{30}^{2}}{3}f\left(\gamma_0\right)\left|g\right|^3\right),\quad \textit{in probability.}$$

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

$$n^{1/2}(\widehat{\alpha}^*-\widetilde{\alpha})\xrightarrow{d^*}\mathcal{N}\left(0,M^{-1}\Omega M^{-1}\right),\quad \textit{in probability,}$$
$$n^{1-2\varphi}(\widehat{\gamma}^*-\gamma_0)\xrightarrow{d^*}\frac{2c'Vc}{\left(c'Dc\right)^2 f\left(\gamma_0\right)}Z,\quad \textit{in probability.}$$

Consistency of $\widehat{\xi}^*$ is established in the following proposition for the continuous regression case. For discontinuous case and structural break, consistency of $\widehat{\xi}^*$ to the corresponding scale factor follows naturally.

Proposition 3. *Suppose Assumptions Z, Q2, 3.1, and 3.3 hold. Then,*

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

We are ready to state the validity of bootstrap testing procedure.

Theorem 4. *Under the Assumptions of Theorem 3 and Assumption K,*

$$\widehat{\xi}^{*-1} QLR_n^* \xrightarrow{d^*} Z, \text{ in probability.}$$

Our results here can be contrasted with some existing ones regarding the validity of bootstrap for non-standard estimators. First, our consistency result seems to contradict Seijo and Sen's (2011), see also Yu (2014), result on the inconsistency of certain residual-based bootstrap and the nonparametric bootstrap (with independent and identically distributed data) for the case where $\varphi = 0$. The reason lies in the observation that our setup differs from theirs in an important and vital way, i.e. they consider the case of a fixed size of the break whereas as we consider the situation when $\delta = \delta_n$ decreases with the sample size. And the resulting asymptotic distributions differ significantly in that the limit depends on the complete conditional distribution of $\varepsilon_t \eta_t c' x_t$ given $q_t = \gamma_0$, when $\varphi = 0$, see e.g. Chan (1993) . Therefore, it is not possible to approximate the sampling distribution by the wild bootstrap scheme.

See Remark 4 in the online supplement for more discussion on our bootstrap resampling scheme in relation to the continuity of the regression function.

5. MONTE CARLO EXPERIMENT

We now present the results of a Monte Carlo study to shed some light on the finite sample performance of our bootstrap procedure in terms of the size of hypothesis testing and coverage probability of confidence intervals for γ_0. We compare our bootstrap procedures of Section 2.2 and 2.3 to asymptotic inference and a standard residual based wild bootstrap method, in which γ_0 in *STEP 2* of Section 2.2 is replaced with $\widehat{\gamma}$ so that $y_t^* = \widetilde{\alpha}' x_t(\widehat{\gamma}) + \widehat{\varepsilon}_t \eta_t$. We use '**NB**' to indicate the grid BCI as well as hypothesis testing based on **NB**, while we use

'**RB**', short for residual bootstrap, to indicate hypothesis testing and bootstrap percentile t confidence interval based on the standard bootstrap.

We consider the following four specifications:

$$A: \ y_t = 2 + 3x_t + \delta x_t 1\{q_t > 2\} + \varepsilon_t, \quad \varepsilon_t = |q_t| e_t, q_t \neq x_t$$

$$B: \ y_t = 2 + 3q_t + \delta q_t 1\{q_t > 2\} + \varepsilon_t, \quad \varepsilon_t = |q_t| e_t, q_t = x_t$$

$$C: \ y_t = 2 + 3q_t + \delta q_t 1\{q_t > 0\} + \varepsilon_t, \ \ \varepsilon_t = |q_t| e_t, \quad \text{continuous design}$$

$$D: \ y_t = 2 + 3x_t + \delta x_t 1\{t > n/2\} + \varepsilon_t, \ \ \varepsilon_t = |x_t| e_t, \ \ \text{structural break.}$$

In settings B and C, the regressor coincides with the transition variable, and specifying $\gamma_0 = 0$ in C leads to a continuous kink rather than a jump in the regression function. Random variables $\{\varepsilon_t\}_{t\geq 1}$, $\{q_t\}_{t\geq 1}$ and $\{x_t\}_{t\geq 1}$ were generated as mutually independent and independent and identically distributed, where $\varepsilon_t \sim N(0,1), q_t \sim N(2,1)$ and $x_t \sim N(2,1)$ in settings A, B and D, whereas $q_t \sim N(0,1)$ in setting C where $\gamma_0 = 0$.

We try two different rates of shrinkage in δ_n, at $\varphi = 1/4, 1/8$. Setting $\delta = \delta_n = \left(\sqrt{10}/4\right) n^{-1/4}$ yields values $\delta = 0.25, 0.199, 0.167$ for $n = 100, 250, 500$, respectively, whereas for $\delta = \left(\sqrt{10}/4\right) n^{-1/8}$ we have $\delta = 0.4446, 0.3965, 0.3636$. The models are such that the conditional heteroscedasticity is $E(\varepsilon_t^2|q_t) = q_t^2$ for settings A-C and $E(\varepsilon_t^2|x_t) = x_t^2$ for setting D. We estimate the scale factor ξ using Epanechnikov kernel and minimum-MSE bandwidth, as given in Härdle and Linton (1994).

5.1. **Size.** In this subsection, we report Monte Carlo size for test of hypothesis $H_0 : \gamma = \gamma_0$ based on the three methods, i.e. **NB**, **RB** and aysmptotic distribution given in Hansen (2000). We carried out 1000 iterations, with one bootstrap per each iteration. Columns 4-6 of Table 1 presents Monte Carlo sizes at significance levels $s = 0.01, 0.05, 0.1$ when $\varphi = 1/4$. Results for $\varphi = 1/8$ can be found in Table 3 in the supplement Appendix D.

The results are better when $\varphi = 1/8$ compared to $\varphi = 1/4$ in most cases. There are only 8 instances out of 108 results where the size result is better for $\varphi = 1/4$ than $\varphi = 1/8$. Across all settings, **NB** consistently reports the best size results. Asymptotic test delivers poor Monte Carlo sizes in settings A-C with substantial oversizing, although the results improve a little bit with larger n and smaller φ, with the results for setting A at $\varphi = 1/8$ and $n = 250, 500$ being somewhat satisfactory. For D, asymptotic test reports good size results but the two bootstrap methods perform even better when $\varphi = 1/4$ and the three methods do similarly when $\varphi = 1/8$. In general, the **RB** test performs better than the asymptotic test but worse than **NB** test in A-C, and the two bootstrap methods report similar results in D. For the kink design C, both asymptotic and **RB** tests seem not to work even at $n = 500$ while **NB** sizes are satisfactory for larger sample sizes $n = 250, 500$. This finding persists through the rest of Monte Carlo results and seems to suggest **NB** is the best method to use for the continuous design case.

5.2. **Coverage Probabilities.** In columns 8-13 of Table 1 we report Monte Carlo coverage probabilities for confidence interval of γ based on three methods, namely asymptotic confidence interval (ACI), percentile t method with **RB** and grid BCI with **NB** when $\varphi = 1/4$, and γ_0 is either at the 50-th or 75-th percentiles of q_t. Results for $\varphi = 1/8$ are given in Table 3 of Appendix D of the supplement. For the grid bootstrap method for **NB**, we divided the realized support of q_t to 10 equidistant points and used linear interpolation of bootstrap

TABLE 1. Monte Carlo size of test $H_0 : \gamma = \gamma_0$ and coverage probability of confidence intervals of γ_0, $\varphi = 1/4$, models A-D

			size				C.P.			C.P.		
		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
A	Asympt	0.01	0.095	0.059	0.044	0.9	0.733	0.77	0.774	0.811	0.834	0.844
		0.05	0.195	0.153	0.13	0.95	0.818	0.832	0.857	0.87	0.895	0.914
		0.1	0.29	0.242	0.2	0.99	0.916	0.938	0.95	0.953	0.971	0.98
	RB	0.01	0.004	0.023	0.026	0.9	0.664	0.719	0.741	0.725	0.792	0.814
		0.05	0.086	0.087	0.08	0.95	0.76	0.807	0.848	0.804	0.866	0.894
		0.1	0.155	0.183	0.147	0.99	0.877	0.918	0.941	0.92	0.965	0.975
	NB	0.01	0.003	0.015	0.009	0.9	0.756	0.81	0.84	0.783	0.826	0.852
		0.05	0.052	0.055	0.037	0.95	0.833	0.88	0.91	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	0.98
B	Asympt	0.01	0.185	0.145	0.155	0.9	0.608	0.612	0.658	0.74	0.73	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
	RB	0.01	0.044	0.034	0.039	0.9	0.605	0.663	0.704	0.725	0.777	0.786
		0.05	0.208	0.132	0.107	0.95	0.714	0.766	0.785	0.816	0.856	0.867
		0.1	0.36	0.239	0.194	0.99	0.849	0.884	0.906	0.904	0.936	0.947
	NB	0.01	0.022	0.013	0.021	0.9	0.77	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	0.98
C	Asympt	0.01	0.623	0.598	0.587	0.9	0.222	0.228	0.238	0.377	0.399	0.401
		0.05	0.738	0.704	0.682	0.95	0.267	0.283	0.281	0.443	0.475	0.474
		0.1	0.794	0.765	0.752	0.99	0.38	0.379	0.398	0.566	0.606	0.624
	RB	0.01	0.078	0.056	0.031	0.9	0.283	0.33	0.401	0.465	0.534	0.553
		0.05	0.281	0.205	0.193	0.95	0.348	0.417	0.484	0.565	0.617	0.673
		0.1	0.417	0.329	0.308	0.99	0.465	0.556	0.622	0.702	0.779	0.812
	NB	0.01	0.052	0.019	0.004	0.9	0.431	0.51	0.579	0.782	0.82	0.84
		0.05	0.137	0.081	0.074	0.95	0.529	0.592	0.671	0.849	0.886	0.895
		0.1	0.24	0.158	0.132	0.99	0.638	0.738	0.817	0.918	0.958	0.965
D	Asympt	0.01	0.006	0.003	0.004	0.9	0.846	0.878	0.883	0.849	0.88	0.892
		0.05	0.036	0.028	0.031	0.95	0.918	0.929	0.931	0.904	0.935	0.944
		0.1	0.074	0.067	0.07	0.99	0.968	0.99	0.985	0.974	0.982	0.992
	RB	0.01	0.006	0.001	0.006	0.9	0.775	0.825	0.855	0.767	0.814	0.846
		0.05	0.028	0.055	0.055	0.95	0.848	0.892	0.911	0.834	0.897	0.911
		0.1	0.068	0.106	0.103	0.99	0.943	0.975	0.981	0.943	0.964	0.983
	NB	0.01	0.006	0.004	0.011	0.9	0.794	0.84	0.869	0.786	0.84	0.865
		0.05	0.046	0.04	0.062	0.95	0.865	0.898	0.919	0.865	0.91	0.93
		0.1	0.085	0.092	0.12	0.99	0.944	0.973	0.978	0.952	0.972	0.983

Size results for test of $H_0 : \gamma = \gamma_0$ based on Hansen (2000)'s asymptotic distribution (Asympt), standard residual bootstrap (**RB**) and bootstrap of Section 2.2 (**NB**). Coverage probability (C.P.) results for γ_0 with asymptotic confidence interval based on Hansen (2000) (Asympt), percentile t confidence interval based on standard residual bootstrap (**RB**) and grid bootstrap confidence interval based on Section 2.3 (**NB**). Setup A: $q_t \neq x_t$, B: $q_t = x_t$ with a jump, C: $q_t = x_t$ with a kink, D: $q_t = t/n$. $\delta = \delta_n = \left(\sqrt{10}/4\right) n^{-1/4}$.

quantiles computed at these points. We carried out 1000 iterations with 399 bootstraps per iteration.

In settings A-C, results are better for $\varphi = 1/8$ compared to $\varphi = 1/4$ while for D they were similar. For $\varphi = 1/4$, in setting A, ACI and **NB** yielded similar results and outperformed **RB**, whereas **NB** preformed the best followed by **RB** then ACI in setting B. In setting C, **NB** substantially outperformed the other two while in setting D, ACI performed the best and the two bootstrap methods did similarly. For settings A-C, results are better when γ_0 is set at the 75-th percentile compared to 50-th percentile across all methods while for setting D, results are similar between the two different γ_0's.

For $\varphi = 1/8$, patterns similar to the case $\varphi = 1/4$ are observed, with the exception of setting B where ACI outperforms **RB** but not **NB**. Comparisons between two different values of γ_0 are similar to when $\varphi = 1/4$.

6. EMPIRICAL EXAMPLE

6.1. **Dynamics of racial segregation.** Card, Mas and Rothstein (2008) model the dynamics of racial segregation in US cities over time. They model a "tipping" behavior of racial mix in a neighborhood. We use a subset of Card, Mas and Rothstein (2008)'s US census tract level data, choosing three large cities for the period 1980-1990. The dependent variable y_i is the ten-year change in a neighborhood's white population while q_i is the minority share of the neighborhood in the base year 1980. There are also six control variables, see Card, Mas and Rothstein (2008) for details on the data. Each observation corresponds to a neighborhood within a city. There is ambiguity in the model of Card *et al.* as to whether there is a jump or a kink in the regression function of y_i at the change point. Due to this, it is not possible to get an interval estimation of γ_0 using the asymptotic theory of Hansen (2000) which only applies to the jump design. In

TABLE 2. Dynamics of racial segregation

City	n	$\widehat{\gamma}_{\mathsf{Unres}}$	$\widehat{\gamma}_{\mathsf{Res}}$	95% grid BCI	H_0: kink design
Chicago	1813	24.86	23.97	(19.5, 49.02)	not rejected
Los Angeles	2030	38.94	52.14	(35.86, 40.42)	rejected
Philadelphia	1299	72.61	77.63	(66.4, 87.75)	not rejected

contrast, our grid BCI based on **NB** is valid for both settings, and we also carry out the test of continuity of Section 3.4, by checking if $\widehat{\gamma}_{\mathsf{Res}}$ lies in the grid BCI, as shown in Table 2. With data of Chicago and Philadelphia we do not reject the null of kink design, while we reject the null for Los Angeles. This means that we may want to use the $\widehat{\gamma}_{\mathsf{Res}}$ which has a faster rate of convergence for Chicago and Philadelphia.

6.2. **Dynamics of US unemployment.** We apply the test of continuity proposed in Section 3.4 to first differences of seasonally adjusted quarterly US unemployment data from 1948-1993. Chan and Tsay (1998) had fitted both continuous and discontinuous SETAR(2) models, where $q_t = y_{t-2}$, to this data and commented that the two fitted models were similar. Gonzalo and Wolf (2005) carried out a test of continuity based on subsampling method and did not reject the null of continuity.

We obtained slightly different estimates from what the earlier authors report so we state our estimation results below. Fitting discontinuous SETAR(2) model to the data yields:

$$\widehat{y}_t = \begin{cases} 0.0182 + 0.6076 y_{t-1} + 0.0716 y_{t-2} & \text{if} \quad y_{t-2} \le 0.0333 \\ 0.2348 + 0.8792 y_{t-1} - 0.6964 y_{t-2} & \text{if} \quad y_{t-2} > 0.0333 \end{cases} \tag{6.1}$$

while fitting the continuous SETAR(2) model gives:

(6.2)

$$\widehat{y}_t = 0.0894 + 0.7874 y_{t-1} + \begin{cases} 0.1085(y_{t-2} - 0.1333) & \text{if} \quad y_{t-2} \leq 0.1333 \\ -0.5578(y_{t-2} - 0.1333) & \text{if} \quad y_{t-2} > 0.1333 \end{cases}$$

Given Corollary 1 in Section 3.4, our bootstrap confidence interval yields a test for the continuity as well. Figure 2 shows that 95% grid confidence interval based on our **NB** for γ_0 is given by $(-0.0202, 0.0796)$. This does not include the restricted estimate $\widehat{\gamma}_{\mathsf{Res}} = 0.1333$. Therefore, we reject the null of continuity at 5% significance level.

Figure 2 here (Fig2.pdf)

Given contrasting conclusion of our test compared to that of Gonzalo and Wolf (2005) and our doubts on their work as mentioned in Introduction, we further investigate relative performance of various tests of continuity. We first describe two new subsampling-based tests that take into account the cube root n rate of convergence of $\widehat{\gamma}_{\mathsf{Unres}}$ under the null hypothesis of continuity. We then carry out a small Monte Carlo study of performance of four tests in terms of the size and power for data generated to mirror our estimated equations (6.1) and (6.2). For size, we observe under rejection from all three subsampling tests, although one of the newly proposed subsampling tests report more satisfactory results than the other two. Our test based on **NB** of Section 3.4 reports the best size result. For power, our test based on **NB** dramatically outperforms all of the three subsampling-based tests. These findings corroborate the contrasting conclusions we obtained with the real data, since subsampling tests tend to have size smaller than the nominal one.

7. CONCLUSION

This paper has developed a unified framework for inference within the class of the segmented regression models. One of the main contributions is that we do not need a prior knowledge on certain qualitative features of the model, such as the continuity, which is often not available nor easily justifiable with real data sets. Also, we provided assumptions under which the bootstrap is valid for the segmented regression model, regardless of whether the model is continuous or not. Our Monte Carlo simulation further demonstrates that its finite sample performance is better than that using the asymptotic critical values. We have also presented and examined a test for continuity, which outperforms existing ones, such as that of Gonzalo and Wolf (2005). We expect that the success of the bootstrap together with the unified framework will enhance the practical application of the segmented regression models.

SUPPLEMENTAL MATERIALS

The supplement contains; (i) a generalization of Theorem 1 and further discussions on Assumption K and the bootstrap algorithm in Section 2.2, (ii) all the mathematical proofs of theorems, (iii) further Monte Carlo simulation results for the case $\varphi = 1/8$, and (iv) figures illustrating the test inversion BCI of Section 2.3 applied to the three cities in the empirical example of Section 6.1 as well as further discussion on relative performance of various tests of continuity for the US unemployment data of Section 6.2 including a small Monte Carlo study.

References

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

[2] Bai, J. (1997), "Estimation of a change point in multiple regression models" *The Review of Economics and Statistics*, 79, 551-563.

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

[4] Basawa, I.V., Mallik, A.K., McCormick, W.P., and Taylor, R.L. (1991), "Bootstrapping unstable first order autoregressive processes", *The Annals of Statistics*, 17, 1479-1486.

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

[6] Card, D., Lee, D., Pei, Z., and Weber, A. (2015), "Inference on causal effects in a generalized regression kink design", forthcoming in *Econometrica*

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

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

[9] 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.

[10] 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.

[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 behavior 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] Ganong, P., and Jäger, S.(2015), "A permutation test for the regression kink design", Working Paper 174531, Harvard University OpenScholar.

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

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

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

[18] Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, New York: Academic Press.

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

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

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

[23] Lee, D. S., and Lemieux, T. (2010), "Regression discontinuity designs in economics", *Journal of Economic Literature*, 48, 281-355.

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

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

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

[27] 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.

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

[29] Seijo, E., and Sen, B.(2011b), "Bootstrapping Manski's maximum score estimator", arXiv:submit/0245262

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

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

[32] Yao, Y.C.(1987), "Approximating the distribution of the ML estimate of the change-point in a sequence of independent r.v.'s", *The Annals of Statistics*, 3, 1321-3128.

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

Supplemental Materials to "Unified Inference on Change-Point in Segmented Regression models"

The supplement contains; (i) a generalization of Theorem 1 and further discussions on Assumption K and the bootstrap algorithm in Section 2.2, (ii) all the mathematical proofs of theorems, (iii) further Monte Carlo simulation results for the case $\varphi = 1/8$, and (iv) figures illustrating the test inversion BCI of Section 2.3 applied to the three cities in the empirical example of Section 6.1 as well as further discussion on relative performance of various tests of continuity for the US unemployment data of Section 6.2 including a small Monte Carlo study.

Appendix A. Some Generalization and More Discussions of Main Results

Remark 2. *(Continuation of Remark 1) The findings in Theorem 1 can be easily extended to more general settings such as those implicit in Feder* $(1975a, b)$*. For that purpose, consider the segmented regression model*

$$y_t = g(q_t; \delta_0, \gamma_0)\,\mathbf{1}_t + \varepsilon_t, \tag{A.1}$$

subject to

$$g(\gamma_0; \delta_0, \gamma_0) = 0. \tag{A.2}$$

That is, the true model is continuous. Of course, our main conclusions follow if we include the term $\beta' x_t$ *into the right side of* $(A.1)$*. However this does not add anything substantial to the discussion, so to keep it as simple as possible we examine the model in* $(A.1)$*. Proceeding as in Feder* $(1975a, b)$*, we can conclude* *that if* (*A*.2) *were used to estimate* $\theta_0 = (\delta_0, \gamma_0)'$, *we then have that*

$$n^{1/2m} (\widehat{\gamma} - \gamma_0) = O_p(1),$$

where, defining $g^{(\ell)}(q; \delta_0, \gamma_0) = \partial^\ell g(q; \delta_0, \gamma_0) / \partial q^\ell$,

$$m = \min\left\{\ell \geq 1 : g^{(\ell)}(\gamma_0; \delta_0, \gamma_0) \neq 0\right\}. \tag{A.3}$$

Observe that in the linear case, i.e. model (2.1), *we have* $m = 1$. *If we had* $g(q; \delta, \gamma) = \delta(q-\gamma)^2$ *then* $m = 2$, *so that our results would be similar to those in Dovonon and Renault* (2012). *On the other hand, if we do not use* (*A*.2) *in the estimation of the parameters, standard algebra yields that*

$$\begin{aligned} & \mathbb{S}_n(\delta;\gamma) - \mathbb{S}_n(\delta_0;\gamma_0) \\ = & \frac{1}{n}\sum_{t=1}^{n}\left(g(q_t;\delta,\gamma)\mathbf{1}_t(\gamma) - g(q_t;\delta_0,\gamma_0)\mathbf{1}_t\right)^2 \\ & + \frac{1}{n}\sum_{t=1}^{n}\left(g(q_t;\delta,\gamma)\mathbf{1}_t(\gamma) - g(q_t;\delta_0,\gamma_0)\mathbf{1}_t\right)\varepsilon_t. \end{aligned}$$

As before, we make comparable the "bias" with the "standard deviation". Consider the case $\gamma > \gamma_0$, *being the case* $\gamma < \gamma_0$ *similarly handled. Now, we have that*

$$\begin{aligned} & \left(g(q_t;\delta,\gamma)\mathbf{1}_t(\gamma) - g(q_t;\delta_0,\gamma_0)\mathbf{1}_t\right)^2 \\ = & \left(g(q_t;\delta,\gamma) - g(q_t;\delta_0,\gamma_0)\right)^2\mathbf{1}_t(\gamma) + g^2(q_t;\delta_0,\gamma_0)\mathbf{1}_t(\gamma_0;\gamma), \end{aligned}$$

so that (*A*.3) *implies that the first moment of the right side of the last displayed expression, that is, the "bias" of* $\mathbb{S}_n(\delta;\gamma) - \mathbb{S}_n(\delta_0;\gamma_0)$ *is of the following order*

$$E\left\|\frac{\partial}{\partial\theta}g(q_t;\delta,\gamma)\right\|^2\|\theta-\theta_0\|^2 + \left(g^{(m)}(\gamma_0;\delta_0,\gamma_0)\right)^2(\gamma-\gamma_0)^{2m+1},$$

while the "standard deviation" of $\mathbb{S}_n(\delta;\gamma)-\mathbb{S}_n(\delta_0;\gamma_0)$ *is governed by*

$$\frac{\|\theta-\theta_0\|}{n^{1/2}}\left\{E\left\|\frac{\partial}{\partial\theta}g\left(q_t;\delta,\gamma\right)\right\|^2\right\}^{1/2}+\frac{(\gamma-\gamma_0)^{m+1/2}}{n^{1/2}}\left|g^{(m)}\left(\gamma_0;\delta_0,\gamma_0\right)\right|.$$

So, the latter two displayed expressions indicate that to balance the rate of convergence to zero of both bias and standard deviation, the rate of convergence to zero of $(\theta-\theta_0)$ *and* $(\gamma-\gamma_0)$ *should be respectively* $O_p\left(n^{-1/2}\right)$ *and* $O_p\left(n^{-1/(2m+1)}\right)$. *Observe that in the segmented regression model* (2.1) *we have* $m=1$. *The reason is that the imposed linearity in the model implies that the "first derivative" is different from zero.*

Next is further remark on Assumption K in Section 3.3.

Remark 3. *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 2 and Lemma 2 in Appendix B (Section 9) below indicate that, without loss of generality, for* $\gamma_0=0$ *the probability limit of* $\widehat{\xi}$ *becomes*

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

where $g_r\left(q\right)=E\left(x_{t2}^r\varepsilon_t^2\mid q_t=q\right)$ *and* $g_r^*\left(q\right)=E\left(x_{t2}^r\mid q_t=q\right)$. *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* (2.7) *will be*

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

whereas the denominator in (2.7) *becomes*

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

So that, unless $E\left[\varepsilon_t^2 \mid q_t = \gamma_0\right] = E\left[\varepsilon_t^2\right]$*, we obtain that, similar to it happens with the L'Hopital rule,*

$$\widehat{\xi} \overset{P}{\to} \frac{\frac{\partial^2}{\partial q^2} f\left(q\right) g_0\left(q\right)|_{q=0}}{\sigma^2 \frac{\partial^2}{\partial q^2} f\left(q\right) g_0^*\left(q\right)|_{q=0}} = \frac{\frac{\partial^2}{\partial q^2}\left(f\left(q\right) E\left[\varepsilon_t^2 \mid q_t = q\right]\right)|_{q=0}}{\frac{\partial^2}{\partial q^2} f\left(q\right)|_{q=0}} \neq \xi,$$

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

Remark 4. *We offer some motivation for our bootstrap resampling scheme of Section 2.2. One important a priori aspect when computing the bootstrap algorithm is how to handle the continuity of the segmented regression model. Recall that one of our main interests is to describe and examine how we can make inferences without imposing the assumption. Now it is clear that when computing the bootstrap algorithm we imposed the continuity assumption of the segmented regression model, then our bootstrap method would not be valid if the true segmented regression model* (2.1) *were discontinuous. Similarly when computing the bootstrap if* (2.1) *were continuous, then one might expect the bootstrap to be invalid if we do not compute the bootstrap analogue of* (2.1) *imposing the continuity restriction. However, as the following heuristic argument suggests, this is in fact not the case. Again for simplicity, consider the model in* (2.1) *with* $\beta = 0$, $x_t = (1, q_t)'$ *and* $\delta = (\delta_1, \delta_3)'$*. That is, our true regression model is given by*

$$y_t = \delta_3\left(q_t - \gamma_0\right)\mathbf{1}_t + \varepsilon_t$$

and denote by $\widetilde{\delta}' = \left(\widetilde{\delta}_1, \widetilde{\delta}_3\right)$ *and* $\widetilde{\gamma}$ *the estimator of* δ *and* γ *in*

$$y_t = \left(\delta_1 + \delta_3 q_t\right)\mathbf{1}_t\left(\gamma\right) + \varepsilon_t$$

used to obtain the residuals in STEP 1 *below. Take* $\gamma_0 = 0$*, then we have that* $\widetilde{\delta} - (0, \delta_{30})' = O_p\left(n^{-1/2}\right)$ *and* $\widetilde{\gamma} = O_p\left(n^{-1/3}\right)$*. Now, the bootstrap analogue of*

$\mathbb{S}_n(\delta_3;\gamma) - \mathbb{S}_n(\delta_{30};\gamma_0)$ *becomes*

$$\begin{aligned}
& \mathbb{S}_n^*(\delta;\gamma) - \mathbb{S}_n^*(\widetilde{\delta};0) \\
= & \frac{1}{n}\sum_{t=1}^{n}\left\{\left(y_t^* - (\delta_1 + \delta_3 q_t)\,\mathbf{1}_t(\gamma)\right)^2 - \varepsilon_t^{*2}\right\} \\
= & \frac{1}{n}\sum_{t=1}^{n}\left(\left(\widetilde{\delta}_1 + \widetilde{\delta}_3 q_t\right)\mathbf{1}_t - (\delta_1 + \delta_3 q_t)\,\mathbf{1}_t(\gamma)\right)^2 \\
& + \frac{2}{n}\sum_{t=1}^{n}\varepsilon_t^*\left(\left(\widetilde{\delta}_1 + \widetilde{\delta}_3 q_t\right)\mathbf{1}_t - (\delta_1 + \delta_3 q_t)\,\mathbf{1}_t(\gamma)\right).
\end{aligned} \tag{A.4}$$

Next the bootstrap expectation of first term on the right of (A.4) *is*

$$\begin{aligned}
& \frac{1}{n}\sum_{t=1}^{n}E^*\left((\delta_1 + \delta_3 q_t)\,\mathbf{1}_t(\gamma) - \left(\widetilde{\delta}_1 + \widetilde{\delta}_3 q_t\right)\mathbf{1}_t\right)^2 \\
= & \frac{1}{n}\sum_{t=1}^{n}\left[(\delta_1 + \delta_3 q_t) - \left(\widetilde{\delta}_1 + \widetilde{\delta}_3 q_t\right)\right]^2\mathbf{1}_t(\gamma) \\
& + \frac{1}{n}\sum_{t=1}^{n}\left(\widetilde{\delta}_1 + \widetilde{\delta}_3 q_t\right)^2\mathbf{1}\{0 < q_t < \gamma\} \\
= & \left\|\widetilde{\delta} - \delta\right\|^2 + \widetilde{\delta}_1^2\gamma + \widetilde{\delta}_3^2\gamma^3 + 2\widetilde{\delta}_1\widetilde{\delta}_3\gamma^2 \\
= & h\left(\delta - \widetilde{\delta};\gamma\right)
\end{aligned}$$

whereas one would expect that the (bootstrap) variance of the second term on the right of (A.4) *is*

$$h\left(\delta - \widetilde{\delta};\gamma\right)O_p\left(n^{-1}\right).$$

So, when $\widetilde{\delta}^* - \widetilde{\delta} = O_{p^*}\left(n^{-1/2}\right)$ *and* $\widetilde{\gamma}^* = O_{p^*}\left(n^{-1/3}\right)$, *we obtain that the rates of convergence of the standard deviation and bias becomes of the same order of magnitude. Recall that* $\widetilde{\delta}_1 = O_p\left(n^{-1/2}\right)$ *when the segmented regression model is continuous.*

Appendix B. Proofs of Main Theorems

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

B.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

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

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

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

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

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

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

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

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

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

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

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

because Assumption Q2 implies that $E\left(x_t x_t' \mathbf{1}_t(\gamma)\right)$, $E\left(x_t x_t' \mathbf{1}_t(-\infty;0)\right)$ and $E\left(x_t x_t' \mathbf{1}_t(0;\gamma)\right)$ are positive definite matrices uniformly in $\gamma > \eta$ and $\left|\left|\bar{\beta}+\delta_0\right|\right| > \eta/3$ if $\left\|\overline{\beta}\right\| \leq \eta/3$ because we can always choose η such that $|\delta_0| \geq 2\eta/3$. we have that

$$C_1 \leq \frac{E\mathbb{A}_{n3}(\theta)}{\left(\overline{\tau}_1,\overline{\tau}_2'\right)E\left(x_{t1}x_{t1}'\mathbf{1}_t(0;\gamma)\right)\left(\overline{\tau}_1,\overline{\tau}_2'\right)' + \overline{\tau}_3^2 E\left(q_t^2\mathbf{1}_t(0;\gamma)\right)} \leq C_2, \tag{B.4}$$

where $\overline{\tau} = (\beta_0 - \beta) + \delta_0$. The motivation for the last displayed inequality comes from the fact that , say, implies that $E\left\{x_t x_t' \mathbf{1}_t(\gamma_1;\gamma_2)\right\}$ is a strictly positive and finite definite matrix which implies that for any vector $a' = (a_1', a_2)$,

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

for some finite constant $C > 0$. So, $(B.2)$ and $(B.3)$ implies that

$$\text{(B.5)} \qquad \inf_{\left\|\overline{\theta}\right\| > \eta} \sum_{\ell=1}^{3} E\left(\mathbb{A}_{n\ell}(\theta)\right) > C\eta^3.$$

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

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

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

$$\text{(B.6)} \qquad \sup_{\left\|\overline{\theta}\right\| > \eta/3} \left\| \sum_{\ell=1}^{3} \mathbb{D}_{n\ell}(\theta) \right\| = o_p(1).$$

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

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

using $(B.5)$ and $(B.6)$.

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

$$\text{(B.7)} \qquad \Pr\left\{ \inf_{\frac{C}{n^{1/2}} < \|\overline{v}\|, \left\|\overline{\beta}\right\| < \eta; \frac{C}{n^{1/3}} < \|\overline{\gamma}\| < \eta} \sum_{\ell=1}^{3} E\left(\mathbb{A}_{n\ell}(\theta)\right) + \mathbb{D}_{n\ell}(\theta) \leq 0 \right\} < \epsilon.$$

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

(B.8)
$$\Pr\left\{\inf_{\frac{C}{n^{1/2}}<\|\overline{v}\|,\left\|\overline{\beta}\right\|<\eta;\frac{C}{n^{1/3}}<\|\overline{\gamma}\|<\eta} E\left(\mathbb{A}_{n\ell}(\theta)\right)/2 + \left(\mathbb{A}_{n\ell}(\theta) - E\left(\mathbb{A}_{n\ell}(\theta)\right)\right) \le 0\right\} < \epsilon$$

(B.9)
$$\Pr\left\{\inf_{\frac{C}{n^{1/2}}<\|\overline{v}\|,\left\|\overline{\beta}\right\|<\eta;\frac{C}{n^{1/3}}<\|\overline{\gamma}\|<\eta} E\left(\mathbb{A}_{n\ell}(\theta)\right)/2 + \mathbb{B}_{n\ell}(\theta) \le 0\right\} < \epsilon.$$

To that end, we shall first examine

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

where

(B.10)
$$\begin{aligned}\Xi_j(\psi) &= \left\{\psi : \frac{C}{n^{1/2}}2^{j-1} < \left\|\overline{\psi}\right\| < \frac{C}{n^{1/2}}2^j\right\}; \qquad j = 1, ..., \log_2 \frac{\eta}{C}n^{1/2}\\ \Xi_k(\gamma) &= \left\{\gamma : \frac{C}{n^{1/3}}2^{k-1} < \overline{\gamma} < \frac{C}{n^{1/3}}2^k\right\}; \qquad k = 1, ..., \log_2 \frac{\eta}{C}n^{1/3}.\end{aligned}$$

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

First by standard arguments,

(B.11)
$$\begin{aligned}&\Pr\left\{\inf_{\Xi_j(v);\Xi_k(\gamma)} E\left(\mathbb{A}_{n1}(\theta)\right)/2 + \mathbb{B}_{n1}(\theta) \le 0\right\}\\ \le\ &\Pr\left\{\inf_{\Xi_j(v)} \|\overline{v}\| \lambda_{\min}\left(Ex_tx_t'\mathbf{1}_t(0)\right) \le \sup_{\Xi_k(\gamma)} \left\|\frac{4}{n^{1/2}}\sum_{t=1}^n x_t\varepsilon_t\mathbf{1}_t(\gamma)\right\|\right\}\\ \le\ &\Pr\left\{C2^{j-2} \le \sup_{\{\gamma:\|\overline{\gamma}\|<\eta\}} \left\|\frac{1}{n^{1/2}}\sum_{t=1}^n x_t\varepsilon_t\mathbf{1}_t(\gamma)\right\|\right\}\\ \le\ &C^{-1}2^{-j+2}\eta^{1/2}\end{aligned}$$

by Lemma 1 and the Markov's inequality. Observe that the latter inequality is independent of $\Xi_k(\gamma)$. Since $\sum_{j=1}^{\infty} 2^{-j} < \infty$, the probability in $(B.9)$ can be made arbitrary small for large C or small η, thus satisfying the condition $(B.9)$. $(B.8)$ follows similarly as is the case for $\ell = 2$ and thus it is omitted.

We next examine $(B.8)$ and $(B.9)$ for $\ell = 3$. Observing $(B.4)$ and the arguments that follow, defining

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

it suffices to show $(B.8)$ and $(B.9)$ for $\widetilde{\mathbb{A}}_{n3}(\theta)$ and $\widetilde{\mathbb{B}}_{n3}(\theta)$. To that end, because $\overline{\tau} > C_1$ as $|\delta_{30}| > C_1 > 0$, we obtain, since $Eq_t^2 \mathbf{1}_t(0;\eta) \geq C_1 \eta^3$

$$\begin{aligned}
& \Pr\left\{\inf_{\Xi_j(\upsilon);\Xi_k(\gamma)} E\left(\widetilde{\mathbb{A}}_{n3}(\theta)/2\right) + \widetilde{\mathbb{B}}_{n3}(\theta) \leq 0\right\} \\
\leq\ & \Pr\left\{\inf_{\Xi_k(\gamma)} \|\tau_0\| E\left(q_t^2 \mathbf{1}_t(0;\gamma)\right) \leq \sup_{\Xi_k(\gamma)} \left\|\frac{4}{n}\sum_{t=1}^{n} q_t \varepsilon_t \mathbf{1}_t(0;\gamma)\right\|\right\} \\
\leq\ & \Pr\left\{\frac{C}{n} 2^{3(k-2)} \leq \sup_{\Xi_k(\gamma)} \left\|\frac{1}{n}\sum_{t=1}^{n} q_t \varepsilon_t \mathbf{1}_t(0;\gamma)\right\|\right\} \qquad \text{(B.12)} \\
\leq\ & C^{-1} 2^{-3k/2},
\end{aligned}$$

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 $(B.9)$ holds true for $\ell = 3$ by choosing C large enough.

We now conclude the proof after we note that the left side of $(B.7)$ is bounded by

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

using $(B.11)-(B.12)$. ■

B.2. **Proof of Theorem 1.**

Because the "arg min" is a continuous mapping, see Kim and Pollard (1990), it suffices to examine the weak limit of

$$
\begin{aligned}
\mathbb{G}_n(h,g) &= n\left(\mathbb{S}_n\left(\alpha_0+\frac{h}{n^{1/2}},\gamma_0+\frac{g}{n^{1/3}}\right)-\mathbb{S}_n(\alpha_0,\gamma_0)\right) \\
&= \sum_{t=1}^{n}\left\{\left(\frac{h'}{n^{1/2}}x_t\left(\frac{g}{n^{1/3}}\right)+\delta_{20}q_t\mathbf{1}_t\left(0;\frac{g}{n^{1/3}}\right)+\varepsilon_t\right)^2-\varepsilon_t^2\right\},
\end{aligned}
$$

over any compact set, where we assume $\gamma_0=0$ as before for notational convenience. Let $\|h\|,|g|\leq C$. First, due to the uniform law of large numbers it follows 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 implies that

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

So, the latter implies that

$$\sup_{\|h\|,|g|\leq C}\left|\mathbb{G}_n(h,g)-\widetilde{\mathbb{G}}_n(h,g)\right|=o_p(1), \tag{B.13}$$

where

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

The consequence of $(B.13)$ 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

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

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

$$nEq_t^2\mathbf{1}_t\left(0;\frac{g}{n^{1/3}}\right)=n\int_0^{\frac{g}{n^{1/3}}}q^2f(q)\,dq=n\frac{2f\left(\frac{\widetilde{g}}{n^{1/3}}\right)}{3!}\left(\frac{g}{n^{1/3}}\right)^3\to 3^{-1}f(0)g^3, \tag{B.14}$$

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 $\widetilde{\mathbb{G}}_n^2(g)$ 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. (C.1), yields the tightness of the process as explained in Remark 5. 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 $u_{nt} = \sqrt{n} q_t \varepsilon_t \mathbf{1}_t \left(0; \frac{g}{n^{1/3}}\right)$,

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

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

For (i), note that $E n^{-2} \max_t |u_{nt}|^4 \le n^{-1} E |u_{nt}|^4 = n E q_t^4 \varepsilon_t^4 \mathbf{1}_t \left(0; \frac{g}{n^{1/3}}\right) \to 0$ 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 $(B.14)$. We now characterize the covariance kernel, 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 $(B.14)$, we have that

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

The cases for $g_1 > g_2 > 0$ or $g_2 < g_1 < 0$ are similar and thus omitted. Finally, the same argument shows that 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, yielding the independence between $\widetilde{h}$ and $\widetilde{g}$ and thus the asymptotic independence between $\widehat{\alpha}$ and the threshold estimator $\widehat{\gamma}$. ■

B.3. **Proof of Theorem 2.**

Theorem 2 part (**b**) was proved in Hansen (2000). For the structural break case in part (**a**), the proof of Proposition 3 in Bai (1997) established that

$$n\left(\mathbb{S}_n(\alpha_0; \gamma_0) - \mathbb{S}_n(\alpha_0; \widehat{\gamma})\right) \xrightarrow{d} \max_{v \in \mathbb{R}} \left(2\sqrt{\lambda} W(v) - \mu |v|\right),$$

where $\lambda := c' E\left(\mathbf{x_t} \mathbf{x_t}' \varepsilon_t^2\right) c$ and $\mu := c' E\left(\mathbf{x_t} \mathbf{x_t}'\right) c$. Now, using the change-of-variables $v = (\lambda/\mu^2) g$, the distributional equality $W(a^2 g) = a W(g)$ and definition

$\xi = \lambda/(\sigma^2\mu)$, we obtain that

$$\max_{v\in\mathbb{R}}\left(2\sqrt{\lambda}W(v)-\mu\left|v\right|\right)=\max_{g\in\mathbb{R}}\left(2\sqrt{\lambda}W\left(\frac{\lambda}{\mu^2}g\right)-\mu\left|\frac{\lambda}{\mu^2}g\right|\right)$$
$$=\frac{\lambda}{\mu}\max_{g\in\mathbb{R}}\left(2W\left(g\right)-\left|g\right|\right)=\xi\sigma^2\max_{g\in\mathbb{R}}\left(2W\left(g\right)-\left|g\right|\right).$$

Thus, we conclude that as $n\to\infty$,

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

Part (**c**) is similar. Due to the asymptotic independence between $\widehat{\alpha}$ and $\widehat{\gamma}$ and the results in Theorem 1, we have that

$$n\left(\mathbb{S}_n\left(\widehat{\alpha};\gamma_0\right)-\mathbb{S}_n\left(\widehat{\alpha};\widehat{\gamma}\right)\right)\overset{asy}{=}n\left(\mathbb{S}_n\left(\alpha_0;\gamma_0\right)-\mathbb{S}_n\left(\alpha_0;\widehat{\gamma}\right)\right)$$
$$\overset{asy}{=}f\left(\gamma_0\right)\max_{\phi\in\mathbb{R}}\left(2\delta_{30}\sqrt{3^{-1}f\left(\gamma_0\right)\sigma_\varepsilon^2\left(\gamma_0\right)}W\left(\phi^3\right)+3^{-1}\delta_{30}^2f\left(\gamma_0\right)\left|\phi\right|^3\right).$$

Then, applying the change-of-variable $\phi^3 = 3g\sigma_\varepsilon^2\left(\gamma_0\right)/\delta_{30}^2f\left(\gamma_0\right)$ yields the desired result. Observe that in showing Theorem 1, one of the steps in doing so is that $\mathbb{G}_n\left(h,g\right)$ converges weakly to a Gaussian process. ■

B.4. **Proof of Proposition 2.**

Recalling our notation in (3.1) and that $\delta_1+\delta_3\gamma_0=0$ and $\delta_2=0$ under Assumption C, we then have that

(B.15) $$\widehat{\delta}'x_t=\left(\widehat{\delta}_1-\delta_1\right)+\widehat{\delta}_2'x_{2t}+\left(\widehat{\delta}_3-\delta_3\right)q_t+\delta_3\left(q_t-\gamma_0\right).$$

Because we can rename $q_t-\gamma_0$ as q_t, we shall assume without loss of generality that $\gamma_0=0$ so that $\delta_1=0$.

Consider the case where $\widehat{\gamma} > 0$. The proof when $\widehat{\gamma} < 0$ is analogous and thus it is omitted. By construction, we have that

$$\widehat{\varepsilon}_t = \varepsilon_t + \left(\widehat{\beta} - \beta\right)' x_t + \left(\widehat{\delta} - \delta\right)' x_t \mathbf{1}_t\left(\widehat{\gamma}\right) + \delta_3 q_t \mathbf{1}_t\left(0; \widehat{\gamma}\right).$$

Because $(\delta_1, \delta_2') = 0$ and $\widehat{\beta} - \beta = O_p\left(n^{-1/2}\right)$, $\widehat{\delta} - \delta = O_p\left(n^{-1/2}\right)$ and $\widehat{\gamma} = O_p\left(n^{-1/3}\right)$, we obtain that

$$\begin{aligned}
\widehat{\varepsilon}_t^2 &= \varepsilon_t^2 + O_p\left(n^{-1}\right) + \left(\delta_3 q_t\right)^2 \mathbf{1}_t\left(0; \widehat{\gamma}\right) + 2\delta_3 \varepsilon_t q_t \mathbf{1}_t\left(0; \widehat{\gamma}\right) \\
&\quad + O_p\left(n^{-1/2}\right) \varepsilon_t x_t \left(1 + \mathbf{1}_t\left(\widehat{\gamma}\right)\right) + 2\delta_3 \left\|x_t\right\| q_t \mathbf{1}_t\left(0; \widehat{\gamma}\right) O_p\left(n^{-1/2}\right) \\
&= \varepsilon_t^2 + O_p\left(n^{-1/2}\right) \left\|x_t\right\| \varepsilon_t + 2\delta_3 \varepsilon_t q_t \mathbf{1}_t\left(0; \widehat{\gamma}\right) + \left\|x_t\right\| O_p\left(n^{-2/3}\right). \qquad \text{(B.16)}
\end{aligned}$$

Now $(B.15)$ implies that $\left(\widehat{\delta}' x_t\right)^2 = \delta_3^2 q_t^2 + O_p\left(n^{-1/2}\right) \delta_3 \left\|x_t\right\| q_t + O_p\left(n^{-1}\right)$. So, by Lemma 2 and 3 and by the standard arguments using $na^3 \to \infty$, we conclude that the behaviour of numerator of (2.7) is that of

$$\frac{1}{na^3} \sum_{t=1}^{n} \delta_3^2 q_t^2 \varepsilon_t^2 K\left(\frac{q_t - \widehat{\gamma}}{a}\right) = \kappa_2 \delta_3^2 a^2 \sigma^2(0) f(0) \left(1 + o_p(1)\right)$$

when $\kappa_2 \neq 0$, that is we do not assume higher order kernels. Observe that $g_0(q)$ in Lemma 2 corresponds to $\sigma^2(q)$. More specifically, the contribution due to other terms in $(B.16)$ are indeed negligible by Lemma 3.

Similarly, the leading term in the denominator in (2.7) is

$$\frac{1}{na^3} \sum_{t=1}^{n} \left(\widehat{\delta}' x_t\right)^2 K\left(\frac{q_t - \widehat{\gamma}}{a}\right) = \kappa_2 \delta_3^2 a^2 f(0) \left(1 + o_p(1)\right).$$

So, the convergence in (2.7) follows from the last two displayed expressions. Finally, it is standard to show that $\mathbb{S}_n(\widehat{\theta}) - \sigma^2 = o_p(1)$. This completes the proof of the proposition. ■

B.5. **Proof of Theorem 3.**

Recalling our definition of $\widehat{\alpha}^*$ and $\widehat{\gamma}^*$ in (2.8), we begin by showing their consistency and rate of convergence.

Proposition 4. *Suppose that Assumptions Z and either Q1 or Q2 hold. Then,*
(**a**) *Under Assumption C,*

$$\widehat{\alpha}^* - \widehat{\alpha} = O_{p^*}\left(n^{-1/2}\right) \quad \text{and} \quad \widehat{\gamma}^* - \gamma_0 = O_{p^*}\left(n^{-1/3}\right).$$

(**b**) *Under Assumption J,*

$$\widehat{\alpha}^* - \widehat{\alpha} = O_{p^*}\left(n^{-1/2}\right) \quad \text{and} \quad \widehat{\gamma}^* - \gamma_0 = O_{p^*}\left(n^{2\varphi-1}\right).$$

Proof of Proposition 4 Assuming without loss of generality that $\gamma \geq \widehat{\gamma} = \gamma_0$ and abbreviating $\widehat{\psi} - \psi$ by $\overline{\psi}$ for any parameter ψ, proceeding as in Proposition 1, we obtain that

$$\begin{aligned}
\mathbb{S}_n^*(\theta) - \mathbb{S}_n^*\left(\widehat{\theta}\right) &= \frac{1}{n}\sum_{t=1}^{n}\left\{\left(\overline{\beta}' x_t + \overline{\delta}' x_t \mathbf{1}_t(\gamma) + \widehat{\delta}' x_t \mathbf{1}_t(\widehat{\gamma};\gamma) + \varepsilon_t^*\right)^2 - \varepsilon_t^{*2}\right\} \\
&= \widehat{\mathbb{A}}_{n1}(\theta) + \widehat{\mathbb{A}}_{n2}(\theta) + \widehat{\mathbb{A}}_{n3}(\theta) + \mathbb{B}_{n1}^*(\theta) + \mathbb{B}_{n2}^*(\theta) + \mathbb{B}_{n3}^*(\theta),
\end{aligned}$$

where

$$\begin{aligned}
\widehat{\mathbb{A}}_{n1}(\theta) &= \overline{\upsilon}' M_n^x(\gamma)\overline{\upsilon}; \quad \widehat{\mathbb{A}}_{n2}(\theta) = \overline{\beta}' M_n^x(-\infty;\widehat{\gamma})\overline{\beta} \\
\widehat{\mathbb{A}}_{n3}(\theta) &= \left(\overline{\beta} + \widehat{\delta}\right)' M_n^x(\widehat{\gamma};\gamma)\left(\overline{\beta} + \widehat{\delta}\right) \\
\mathbb{B}_{n1}^*(\theta) &= \overline{\upsilon}' \frac{2}{n}\sum_{t=1}^{n} x_t \varepsilon_t^* \mathbf{1}_t(\gamma); \quad \mathbb{B}_{n2}^*(\theta) = \overline{\beta}' \frac{2}{n}\sum_{t=1}^{n} x_t \varepsilon_t^* \mathbf{1}_t(-\infty;\widehat{\gamma}) \\
\mathbb{B}_{n3}^*(\theta) &= \left(\overline{\beta} + \widehat{\delta}\right)' \frac{2}{n}\sum_{t=1}^{n} x_t \varepsilon_t^* \mathbf{1}_t(\widehat{\gamma};\gamma),
\end{aligned}$$

where, in what follows, for a generic sequence $\{z_t\}_{t\in\mathbb{Z}}$ we employ the notation $M_n^z(\gamma) = \frac{1}{n}\sum_{t=1}^{n} z_t z_t' \mathbf{1}_t(\gamma)$ and $M_n^z(\gamma_1;\gamma_2) = \frac{1}{n}\sum_{t=1}^{n} z_t z_t' \mathbf{1}_t(\gamma_1;\gamma_2)$. It is also worth recalling that for n large enough $0 < \sup_{\gamma\in\Gamma} \|M_n^x(\gamma)\| = H_n$ and $0 < \sup_{\gamma_1<\gamma_2} \|M_n^x(\gamma_1;\gamma_2)\| = H_n$, where in what follows H_n denotes a sequence of strictly positive $O_p(1)$ random variables. Finally as we have in the proof of Proposition 1, because $E(x_t x_t' \mathbf{1}_t(\gamma))$ and $E(x_t x_t' \mathbf{1}_t(0;\gamma))$ are strictly finite positive definite matrices, $M_n^x(-\infty;\gamma) - E(x_t x_t' \mathbf{1}_t(-\infty;\gamma)) = O_p\left(n^{-1/2}\right)$ and $M_n^x(\gamma) - E(x_t x_t' \mathbf{1}_t(\gamma)) = O_p\left(n^{-1/2}\right)$ uniformly in $\gamma \in \Gamma$, we have that

$$C_1 H_n \leq \frac{\widehat{\mathbb{A}}_{n2}(\theta)}{\left(\overline{\beta}_1, \overline{\beta}_2'\right) M_n^{x_1}(-\infty;0)\left(\overline{\beta}_1, \overline{\beta}_2'\right)' + \overline{\beta}_3^2 M_n^q \mathbf{1}_t(-\infty;0)} \;\leq\; C_2 H_n$$

$$C_1 H_n \leq \frac{\widehat{\mathbb{A}}_{n3}(\theta)}{\left(\overline{\tau}_1, \overline{\tau}_2'\right) M_n^{x_1}(0;\gamma)\left(\overline{\tau}_1, \overline{\tau}_2'\right)' + \overline{\tau}_3^2 M_n^q(0;\gamma)} \;\leq\; C_2 H_n, \tag{B.17}$$

where $\overline{\tau} = \left(\widehat{\beta} - \beta\right) + \widehat{\delta}$. The motivation is that we employ in the proof of Proposition 1, after observing that Proposition 1 implies that $\widehat{\gamma} - \gamma_0 = O_p\left(n^{-1/3}\right)$ and Lemma 1 that uniformly in $\gamma_1 < \gamma_2 \in \Gamma$,

$$M_n^x(\gamma_1;\gamma_2) - E x_t x_t' \,\mathbf{1}_t(\gamma_1;\gamma_2) = O_p\left(n^{-1/2}\right)$$

together with the fact that $M_n^x(-\infty;\widehat{\gamma}) = M_n^x(-\infty;\gamma_0) + M_n^x(\gamma_0;\widehat{\gamma})$.

Consistency. We begin with part (**a**). Arguing as in the proof of Proposition 1, it suffices to show that

$$\Pr{}^*\left\{\inf_{\|\overline{\theta}\|>\eta} \sum_{\ell=1}^{3} \widehat{\mathbb{A}}_{n\ell}(\theta) + \mathbb{B}_{n\ell}^*(\theta) \leq 0\right\} \leq \epsilon H_n. \tag{B.18}$$

First, when $\left\|\overline{\theta}\right\| > \eta$, it implies that either (**i**) $\|\overline{\gamma}\| > \eta/2$ or (**ii**) $\left\|\overline{\beta}\right\|, \|\overline{\upsilon}\| > \eta/2$. When (**ii**) holds true, it is clear that

$$\inf_{\|\overline{\upsilon}\|>\eta/2} \widehat{\mathbb{A}}_{n\ell}(\theta) > \eta^2 H_n \qquad \ell = 1, 2 \tag{B.19}$$

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

$$\inf_{\|\gamma\|>\eta/2} M_n^x\left(\widehat{\gamma};\gamma\right) > \eta H_n, \tag{B.20}$$

because $E\left(x_t x_t' \mathbf{1}_t\left(\gamma\right)\right)$ and $E\left(x_t x_t' \mathbf{1}_t\left(0;\gamma\right)\right)$ are strictly positive definite matrices, since say $E\left(x_t x_t' \mathbf{1}_t\left(0;\gamma\right)\right) - E\left(x_t x_t' \mathbf{1}_t\left(0;\eta/4\right)\right)$ is a positive definite matrix when $\|\overline{\gamma}\| > \eta/2$, $M_n^x\left(\widehat{\gamma};\gamma\right) = E\left(x_t x_t' \mathbf{1}_t\left(0;\gamma\right)\right)\left(1+o_p\left(1\right)\right)$ and $\widehat{\mathbb{A}}_{n\ell}\left(\theta\right) - E\left(\mathbb{A}_{n\ell}\left(\theta\right)\right) = o_p\left(1\right)$. Recall that $E\left(a' x_t \mathbf{1}_t\left(0;\eta\right)\right) > \eta \min_{q\in(0,\eta)} f\left(q\right) E\left(a' x_t\right)$. So, $(B.19)$ and $(B.20)$ implies that

$$\inf_{\|\overline{\theta}\|>\eta} \sum_{\ell=1}^{3} \widehat{\mathbb{A}}_{n\ell}\left(\theta\right) > \eta^2 H_n. \tag{B.21}$$

On the other hand, Lemma 4 implies that

(B.22)
$$E^*\left(\sup_{\gamma}\left\|\frac{1}{n^{1/2}}\sum_{t=1}^{n} x_t \varepsilon_t^* \mathbf{1}_t\left(\gamma\right)\right\|\right)^2 + E^*\left(\sup_{\gamma}\left\|\frac{1}{n^{1/2}}\sum_{t=1}^{n} x_t \varepsilon_t^* \mathbf{1}_t\left(-\infty;\gamma\right)\right\|\right)^2 = H_n,$$

so that

$$E^* \sup_{\|\overline{\theta}\|>\eta/2} \left\|\mathbb{B}_{n\ell}^*\left(\theta\right)\right\| = n^{-1/2} H_n \qquad \ell = 1,2,3. \tag{B.23}$$

Thus $(B.21)$ and $(B.23)$ yields that $\widehat{\theta}^* - \widehat{\theta} = o_{p^*}\left(1\right)$ because the left side of $(B.18)$ is bounded by

$$\Pr{}^*\left\{\inf_{\|\overline{\theta}\|>\eta}\sum_{\ell=1}^{3}\widehat{\mathbb{A}}_{n\ell}\left(\theta\right) \le \sup_{\|\overline{\theta}\|>\eta}\left\|\sum_{\ell=1}^{3}\mathbb{B}_{n\ell}^*\left(\theta\right)\right\|\right\}$$

and then Markov's inequality. This concludes the consistency proof.

Convergence rate. To that end, we shall show that for some $C > 0$ large enough and $\epsilon > 0$,

$$\Pr{}^{*}\left\{ \inf_{\frac{C}{n^{1/2}}<\|\overline{\upsilon}\|;\left\|\overline{\beta}\right\|<\eta;\frac{C}{n^{1/3}}<\|\gamma\|<\eta} \sum_{\ell=1}^{3} \widehat{\mathbb{A}}_{n\ell}(\theta) + \mathbb{B}^{*}_{n\ell}(\theta) \leq 0 \right\} < \epsilon H_n. \tag{B.24}$$

To that end, we shall first examine

$$\Pr{}^{*}\left\{ \inf_{\Xi_j(\upsilon);\Xi_j(\beta);\Xi_k(\gamma)} \sum_{\ell=1}^{3} \widehat{\mathbb{A}}_{n\ell}(\theta) + \mathbb{B}^{*}_{n\ell}(\theta) \leq 0 \right\}$$

where for some $j = 1, ..., \log_2 \frac{\eta}{C} n^{1/2}$ and $k = 1, ..., \log_2 \frac{\eta}{C} n^{1/3}$, and $\Xi_j(\upsilon)$ and $\Xi_k(\gamma)$ are defined similarly to $(B.10)$. Recall that we have assumed that $\gamma \geq 0$ since when $\gamma \leq 0$ the proof follows similarly.

Now Lemma 4 implies that

$$\begin{aligned}
&\Pr{}^{*}\left\{ \inf_{\Xi_j(\upsilon);\Xi_j(\beta);\Xi_k(\gamma)} \widehat{\mathbb{A}}_{n1}(\theta) + \mathbb{B}^{*}_{n1}(\theta) \leq 0 \right\} \\
\leq\ &\Pr{}^{*}\left\{ \inf_{\Xi_j(\upsilon);\Xi_j(\beta)} \|\overline{\upsilon}\| \|M^x_n(\gamma)\| \leq \sup_{\Xi_k(\gamma)} \left\| \frac{2}{n^{1/2}} \sum_{t=1}^{n} x_t \varepsilon^{*}_t \mathbf{1}_t(\gamma) \right\| \right\} \\
\leq\ &\Pr{}^{*}\left\{ \|M^x_n(\gamma)\| C 2^{j-1} \leq \sup_{\{\gamma:\|\gamma\|<\eta\}} \left\| \frac{1}{n^{1/2}} \sum_{t=1}^{n} x_t \varepsilon^{*}_t \mathbf{1}_t(\gamma) \right\| \right\} \qquad \text{(B.25)} \\
\leq\ &C^{-1} 2^{-2j} H_n.
\end{aligned}$$

Observe that the bound in $(B.25)$ is independent of k, i.e. the set $\Xi_k(\gamma)$. Defining

$$\begin{aligned}
\widetilde{\mathbb{A}}_{n2}(\theta) &= \left(\overline{\beta}_1, \overline{\beta}'_2\right) M^x_n(-\infty; 0) \left(\overline{\beta}_1, \overline{\beta}'_2\right)' \\
\widetilde{\mathbb{B}}^{*}_{n2}(\theta) &= \left(\overline{\beta}_1, \overline{\beta}'_2\right) \frac{2}{n} \sum_{t=1}^{n} x_{t1} \varepsilon^{*}_t \mathbf{1}_t(-\infty; \gamma),
\end{aligned}$$

$(B.17)$ yields that

$$
\begin{aligned}
& \Pr{}^{*}\left\{\inf_{\Xi_j(\upsilon);\Xi_k(\gamma)} \widetilde{\mathbb{A}}^{*}_{n2}(\theta)+\widetilde{\mathbb{B}}^{*}_{n2}(\theta)\leq 0\right\} \\
\leq\ & \Pr{}^{*}\left\{\inf_{\Xi_j(\upsilon)} \left\|\left(\overline{\beta}_1,\overline{\beta}_2'\right)\right\| M_n^{x_1}(-\infty;0)\leq \sup_{\Xi_k(\gamma)}\left\|\frac{2}{n}\sum_{t=1}^{n} x_{t1}\varepsilon_t^{*}\mathbf{1}_t(-\infty;\gamma)\right\|\right\} \\
\leq\ & \Pr{}^{*}\left\{\left\|M_n^{x_1}(-\infty;0)\right\| C2^{j-1}\leq \sup_{\{\gamma:\|\gamma\|<\eta\}}\left\|\frac{1}{n^{1/2}}\sum_{t=1}^{n} x_{t1}\varepsilon_t^{*}\mathbf{1}_t(-\infty;\gamma)\right\|\right\} \qquad \text{(B.26)} \\
\leq\ & C^{-1}2^{-2j}H_n,
\end{aligned}
$$

by Lemma 4, which once again the bound is independent of k.

Next, define

$$
\widetilde{\mathbb{A}}_{n3}(\theta)=\widehat{\tau}^2 q_t^2\mathbf{1}_t(0;\gamma)\,;\quad \widetilde{\mathbb{B}}^{*}_{n3}(\theta)=\widehat{\tau}\frac{2}{n}\sum_{t=1}^{n} q_t\varepsilon_t^{*}\mathbf{1}_t(0;\gamma)\,,
$$

then, because $\widehat{\tau}=H_n+C_1$,

$$
\begin{aligned}
& \Pr{}^{*}\left\{\inf_{\Xi_j(\upsilon);\Xi_k(\gamma)} \widetilde{\mathbb{A}}_{n3}(\theta)+\widehat{\tau}\widetilde{\mathbb{B}}^{*}_{n3}(\theta)\leq 0\right\} \\
\leq\ & \Pr{}^{*}\left\{\inf_{\Xi_j(\upsilon)} \|\widehat{\tau}\|\frac{1}{n}\sum_{t=1}^{n} q_t^2\mathbf{1}_t(0;\gamma)\leq \sup_{\Xi_k(\gamma)}\left\|\widetilde{\mathbb{B}}^{*}_{n3}(\theta)/\tau_0\right\|\right\} \\
\leq\ & \Pr{}^{*}\left\{\frac{C}{n}2^{3(k-1)}\leq \sup_{\Xi_k(\gamma)}\left\|\widetilde{\mathbb{B}}^{*}_{n3}(\theta)/\widehat{\tau}\right\|\right\} \qquad \text{(B.27)} \\
\leq\ & C^{-1}2^{-3k/2}H_n,
\end{aligned}
$$

by Lemma 4 and Markov's inequality. Observe that the latter displayed bound is independent of j, i.e. the set $\Xi_j(\upsilon)$.

So, the left side of $(B.24)$ is bounded by

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

using $(B.25)-(B.27)$. This concludes the proof of part (**a**).

The proof of part (**b**) is similarly handled after obvious changes, so it is omitted.

■

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 $\|h\|,|g|\le C$.

First, recall that $\widetilde{\delta}_1=O_p\left(n^{-1/2}\right)$ and $\widetilde{\delta}_2=O_p\left(n^{-1/2}\right)$ under Assumption C and note that Lemma 1 and Lemma 4 imply that, uniformly in $\|h\|,|g|<C$,

$$
\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\left\{x_t\left(\frac{g}{n^{1/3}}\right)x_t'\left(\frac{g}{n^{1/3}}\right)-\mathbf{x}_t\mathbf{x}_t'\right\} &= O_p\left(n^{-1/3}\right)\\
\frac{1}{n^{1/2}}\sum_{t=1}^{n}\left\{x_t\left(\frac{g}{n^{1/3}}\right)q_t\mathbf{1}_t\left(0;\frac{g}{n^{1/3}}\right)\right\} &= O_p\left(n^{-1/6}\right)\\
E^{*}\left\|\frac{1}{n^{1/2}}\sum_{t=1}^{n}\left(x_t\left(\frac{g}{n^{1/3}}\right)-\mathbf{x}_t\right)\varepsilon_t^{*}\right\|^2 &= O_p\left(n^{-1/3}\right).
\end{aligned}
$$

Thus, the latter implies that

$$E^{*}\sup_{h,g\in\mathbb{R}}\left|\mathbb{G}_{n}^{*}\left(h,g\right)-\widetilde{\mathbb{G}}_{n}^{*}\left(h,g\right)\right|=O_{p}\left(n^{-1/6}\right), \tag{B.28}$$

where

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

The consequence of $(B.28)$ is then that the minimizer of $\mathbb{G}_{n}^{*}\left(h,g\right)$ is asymptotically equivalent to that of $\widetilde{\mathbb{G}}_{n}^{*}\left(h,g\right)$. Thus, it suffices to show the weak convergence of $\widetilde{\mathbb{G}}_{1n}^{*}\left(h\right)$ and $\widetilde{\mathbb{G}}_{2n}^{*}\left(g\right)$ and that

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

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

The second term in the definition of $\widetilde{\mathbb{G}}_{2n}^{*}\left(g\right)$ converges weakly (in probability) to $2\delta_{30}\sqrt{3^{-1}f\left(0\right)\sigma_{\varepsilon}^{2}\left(0\right)}W\left(g^{3}\right)$. To see this note that Lemma 4's, and the Remark 4 that follows, yields the tightness of the process as explained in Remark 5. 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 $3^{-1}f\left(0\right)\sigma_{\varepsilon}^{2}\left(0\right)g^{3}$ and the Lindeberg's condition follows easily.

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

B.6. **Proof of Proposition 3.**

Let $\widehat{\gamma}^* > 0$. The case when $\widehat{\gamma}^* < 0$ is analogous and thus omitted. We shall examine the behaviour of the numerator of (2.9), that of its denominator being similarly handled. By construction,

$$\widehat{\varepsilon}_t^* = \varepsilon_t^* + \left(\widehat{\beta}^* - \widetilde{\beta}\right)' x_t + \left(\widehat{\delta}^* - \widetilde{\delta}\right)' x_t \mathbf{1}_t\left(\widehat{\gamma}^*\right) + \left(\widetilde{\delta}_1 + \widetilde{\delta}_3 q_t\right) \mathbf{1}_t\left(0; \widehat{\gamma}^*\right).$$

Recall that when the constraint given in (3.2) holds true $\widetilde{\delta}_2$ and $\widetilde{\delta}_1$ are both $O_p\left(n^{-1/2}\right)$. On the other hand Proposition 4 yields that $\widehat{\beta}^* - \widetilde{\beta} = O_{p^*}\left(n^{-1/2}\right)$, $\widehat{\delta}^* - \widetilde{\delta} = O_{p^*}\left(n^{-1/2}\right)$ and $\widehat{\gamma}^* = O_{p^*}\left(n^{-1/3}\right)$. Then, $\left(\widehat{\delta}^{*\prime} x_t\right)^2 = \widetilde{\delta}'^2 x_t^2 + O_{p^*}\left(n^{-1/2}\right) \widetilde{\delta}' x_t q_t + O_{p^*}\left(n^{-1}\right)$. And, proceeding as we did in the proof of Proposition 2, we easily deduce that

$$\widehat{\varepsilon}_t^{*2} = \varepsilon_t^{*2} + O_{p^*}\left(n^{-1/2}\right) x_t \varepsilon_t^* + 2\widetilde{\delta}_3 \varepsilon_t^* q_t \mathbf{1}_t\left(0; \widehat{\gamma}^*\right) + x_t O_{p^*}\left(n^{-2/3}\right). \tag{B.29}$$

By obvious arguments and those in $(C.23)$, it suffices to examine the behaviour of

$$\frac{1}{na} \sum_{t=1}^{n} \left(\widetilde{\delta}' x_t\right)^2 \varepsilon_t^{*2} K\left(\frac{q_t - \widehat{\gamma}^*}{a}\right).$$

Now, because $\widetilde{\delta}_2$ and $\widetilde{\delta}_1$ are both $O_p\left(n^{-1/2}\right)$ when (3.2) holds true the behaviour of the last displayed expression is governed by

$$\frac{1}{na} \sum_{t=1}^{n} \widetilde{\delta}_3^2 q_t^2 \varepsilon_t^{*2} K\left(\frac{q_t - \widehat{\gamma}^*}{a}\right)$$

which is $\kappa_2 \delta_{30}^2 a^2 E^*\left[\varepsilon_t^{*2} \mid q_t = \gamma_0\right] f(0)\left(1 + o_{p^*}(1)\right)$ by Lemma 5 when $\kappa_2 \neq 0$, that is we do not assume higher order kernels. Notice that, by standard results, the contribution due to other terms in $(B.29)$ are indeed negligible by Lemma 6.

Likewise the denominator in (2.9), is

$$\frac{1}{na}\sum_{t=1}^{n}\left(\widetilde{\delta}'x_t\right)^2 K\left(\frac{q_t-\widehat{\gamma}^*}{a}\right)=\kappa_2\delta_{30}^2a^2f\left(0\right)\left(1+o_{p^*}\left(1\right)\right).$$

So, the convergence in (2.9) follows from the last two displayed expressions. Finally, it is standard that $\mathbb{S}_n(\widehat{\theta}^*)-\sigma^2=o_{p^*}(1)$. This completes the proof of the proposition. ■

B.7. **Proof of Theorem 4.**

Theorem 2 was proved in Hansen (2000) under the discontinuity design. For the structural break case, the proof of Proposition 3 in Bai (1997) established that

$$n\left(\mathbb{S}_n^*\left(\alpha_0;\gamma_0\right)-\ \mathbb{S}_n^*\left(\alpha_0;\widehat{\gamma}^*\right)\right)\overset{d^*}{\to}\max_{v\in\mathbb{R}}\left(2\sqrt{\lambda}W(v)-\mu|v|\right),$$

where $\lambda:=c'E\left(\mathbf{x_t}\mathbf{x_t}'\varepsilon_t^2\right)c$ and $\mu:=c'E\left(\mathbf{x_t}\mathbf{x_t}'\right)c$. Then proceed as in Theorem 2. Therefore, we conclude that as $n\to\infty$,

$$QLR_n^*\overset{d^*}{\longrightarrow}\xi\max_{g\in\mathbb{R}}\left(2W\left(g\right)-|g|\right).$$

The continuous design case is similar. Due to the asymptotic independence and proceeding with analogues arguments to those in Theorem 1, we conclude that

$$\begin{aligned}&n\left(\mathbb{S}_n^*\left(\widehat{\alpha}^*;\gamma_0\right)-\mathbb{S}_n^*\left(\widehat{\alpha}^*;\widehat{\gamma}^*\right)\right)\overset{asy}{=}n\left(\mathbb{S}_n^*\left(\widehat{\alpha};\gamma_0\right)-\mathbb{S}_n\left(\widehat{\alpha}_0;\widehat{\gamma}^*\right)\right)\\&\overset{asy}{=}f\left(\gamma_0\right)\max_{\phi\in\mathbb{R}}\left(2\delta_{30}\sqrt{3^{-1}f\left(\gamma_0\right)\sigma_\varepsilon^2\left(\gamma_0\right)}W\left(\phi^3\right)+3^{-1}\delta_{30}^2f\left(\gamma_0\right)|\phi|^3\right).\end{aligned}$$

Then apply the change-of-variable $\phi^3=3g\sigma_\varepsilon^2\left(\gamma_0\right)/\delta_{30}^2f$ to yield the result. Observe again that in showing Theorem 3, one of the steps in doing so is that $\mathbb{G}_n^*\left(h,g\right)$ converges weakly (in probability) to a Gaussian process. ■

Appendix C. **Auxiliary Lemmas**

We begin with a set of maximal inequalities, which play a central role in deriving convergence rates and tightness of various empirical processes. Let

$$\begin{aligned} J_n(\gamma,\gamma') &= \frac{1}{n^{1/2}}\sum_{t=1}^{n}\varepsilon_t x_t \mathbf{1}_t(\gamma;\gamma') \\ J_{1n}(\gamma,\gamma') &= \frac{1}{n^{1/2}}\sum_{t=1}^{n}\varepsilon_t |q_t-\gamma|^j \mathbf{1}_t(\gamma;\gamma') \\ J_{2n}(\gamma) &= \frac{1}{n^{1/2}}\sum_{t=1}^{n}\Big\{|q_t-\gamma_0|^j \mathbf{1}_t(\gamma_0;\gamma) - E|q_t-\gamma_0|^j \mathbf{1}_t(\gamma_0;\gamma)\Big\} \end{aligned}$$

and for some sequence $\{z_t\}_{t=1}^n$,

$$J_{3n}(\gamma) = \frac{1}{n^{1/2}}\sum_{t=1}^{n}\left(z_t\mathbf{1}_t(\gamma_0;\gamma) - Ez_t\mathbf{1}_t(\gamma_0;\gamma)\right).$$

Lemma 1. *Suppose Assumptions Z and Q2 hold for the sequence* $\{x_{t1}, q_t, \varepsilon_t\}_{t=1}^n$. *In addition, for* $J_{3n}(\gamma)$, *assume that* $\{z_t, q_t\}_{t=1}^n$ *be a sequence of strictly stationary, ergodic, and* ρ *-mixing with* $\sum_{m=1}^{\infty}\rho_m^{1/2}<\infty$, $E|z_t|^4<\infty$ *and, for all* $\gamma\in\Gamma$, $E\left(|z_t|^4|q_t=\gamma\right)<C<\infty$. *Then, there exists* $n_0<\infty$ *such that for all* γ' *in a neighbourhood of* γ_0 *and for all* $n>n_0$ *and* $\epsilon\geq n_0^{-1}$,

$$(\mathbf{a})\ E\sup_{\gamma'<\gamma<\gamma'+\epsilon}|J_n(\gamma',\gamma)| \leq C\epsilon^{1/2}$$

$$(\mathbf{b})\ E\sup_{\gamma'<\gamma<\gamma'+\epsilon}|J_{1n}(\gamma',\gamma)| \leq C\epsilon^{1/2}(\epsilon+|\gamma_0-\gamma'|)^j \tag{C.1}$$

$$(\mathbf{c})\ E\sup_{\gamma_0<\gamma<\gamma_0+\epsilon}|J_{2n}(\gamma)| \leq C\epsilon^{j+1/2} \tag{C.2}$$

$$(\mathbf{d})\ E\sup_{\gamma_0<\gamma<\gamma_0+\epsilon}|J_{3n}(\gamma)| \leq C\epsilon^{1/2}. \tag{C.3}$$

Proof. Part (**a**) proceeds as in Hansen's (2000) Lemma A.3, so it is omitted.

Next part (**b**). This is almost identical to that of Hansen's (2000) Lemma A.3 once observing that if $|\gamma_1-\gamma'|\leq\epsilon$ and $|\gamma_2-\gamma'|\leq\epsilon$ and $h_t(\gamma_1,\gamma_2)=$

$|\varepsilon_t(q_t-\gamma_0)^j|\mathbf{1}_t(\gamma_1,\gamma_2)$, then the bound in his Lemma A.1 (12) should be updated to

$$Eh_i^r(\gamma_1,\gamma_2)\leq C\int_{\gamma_1}^{\gamma_2}|q-\gamma_0|^{jr}\,dq\leq C|\gamma_1-\gamma_2|\epsilon_1^{jr},$$

where $C<\infty$ and $\epsilon_1=(\epsilon+|\gamma_0-\gamma'|)$, since $E\left(|\varepsilon_t^r|\,|q_t\right)$ and the density $f(q)$ of q_t are bounded around $q_t=\gamma_0$. Hansen's bound in (13) should be changed to $|\gamma_1-\gamma_2|\epsilon_1^{jr}$ for the same reason. Then, these new bounds imply that the bounds (15) and (16) in his Lemma A.3 and the bounds (18) and (20) in the proof of his Lemma A.2 should change to $|\gamma_1-\gamma_2|^2\epsilon_1^{4j}$ and $n^{-1}|\gamma_1-\gamma_2|\epsilon_1^{4j}+|\gamma_1-\gamma_2|^2\epsilon_1^{4j}$, respectively, to yield the desired bound in $(C.1)$.

Part (**c**). For notational simplicity we assume that $\gamma_0=0$. Let $\gamma_k=k/n$, for $k=1,...,m$, where $m=[\epsilon n]+1$. By triangle inequality,

(C.4)

$$\sup_{\gamma_0<\gamma<\gamma_0+\epsilon}|J_{2n}(\gamma)|\leq\max_{k=1,\ldots,m-1}|J_{2n}(\gamma_k)|+\max_{k=1,\ldots,m}\sup_{\gamma_{k-1}\leq\gamma\leq\gamma_k}|J_{2n}(\gamma)-J_{2n}(\gamma_{k-1})|\,.$$

Now because $f(\cdot)$ is continuous differentiable at γ_0, standard algebra yields that

$$E\,|q_t|^j\;\mathbf{1}_t(\gamma_{k-1};\gamma_k)\leq C\gamma_k^j/n. \tag{C.5}$$

Next, using $(C.5)$

$$\begin{aligned}&\sup_{\gamma_{k-1}\leq\gamma\leq\gamma_k}\left|\frac{1}{n^{1/2}}\sum_{t=1}^{n}|q_t|^j\;\mathbf{1}_t(\gamma_{k-1};\gamma)\right|\\&\leq\left(J_{2n}(\gamma_k)-J_{2n}(\gamma_{k-1})\right)+n^{1/2}E\,|q_t|^j\;\mathbf{1}_t(\gamma_{k-1};\gamma_k)\\&=\left(J_{2n}(\gamma_k)-J_{2n}(\gamma_{k-1})\right)+C\gamma_k^j/n^{1/2}.\end{aligned}$$

Thus, using the inequality $\left(\sup_{j=1,\ldots,\ell}|c_j|\right)^4 \leq \sum_{j=1}^{\ell}|c_j|^4$, we conclude that second term on the right of $(C.4)$ has absolute moment bounded by

$$\left(\sum_{k=1}^{m} E\left|J_{2n}(\gamma_k) - J_{2n}(\gamma_{k-1})\right|^4\right)^{1/4} + C\gamma_m^j/n^{1/2}. \tag{C.6}$$

However, from Lemma 3.6 of Peligrad (1982), for any $k > j$,

$$E\left|J_{2n}(\gamma_k) - J_{2n}(\gamma_j)\right|^4 \leq C\left(n^{-1}E|q_t|^{4j}\ \mathbf{1}_t(\gamma_j;\gamma_k) + \left(E|q_t|^{2j}\ \mathbf{1}_t(\gamma_j;\gamma_k)\right)^2\right).$$

So, using again $(C.5)$ and that $m = [\varepsilon n] + 1$ and $n^{-1} < \varepsilon$, we conclude that the first moment of the second term on the right of $(C.4)$ is $C\epsilon^{j+1/2}$.

Next the first moment of the first term on the right of $(C.4)$ is also bounded by $C\epsilon^{j+1/2}$ by Billingsley's (1968) Theorem 12.2 using the last displayed inequality.

Finally part (**d**). This is similar to that of $(C.2)$. It is sufficient to note that, with $J_{3n}(\gamma)$, the bounds in $(C.5)$ and $(C.6)$ change to $C/n^{1/2}$ and $C\epsilon^2$, respectively. This yields the results as $n^{-1} < \epsilon$. ∎

Remark 5. *One of the consequences of the previous lemma* (**a**) *and* (**b**)*, which allows the maximal inequality to hold for any* γ' *in a neighbourhood of* γ_0*, is that*

$$nE\sup_{g_1<g<g_1+\epsilon}\left|J_n(\gamma_0 + g/r_n) - J_n(\gamma_0 + g_1/r_n)\right| \leq C(\epsilon + g_1)\,\epsilon^{1/2},$$

which can be made small by choosing small ϵ *and* $r_n \to \infty$*. This is used to verify the stochastic equicontinuity of the rescaled and reparameterized empirical processes in the proof of Theorem 1.*

The following two lemmas are used in the proof of Proposition 2. Before we state our next lemma, we need to introduce some notation. In what follows

$$\begin{aligned} g_r(q) &= E\left(x_{t2}^r \varepsilon_t^2 \mid q_t = q\right); \quad g_r^*(q) = E\left(x_{t2}^r \mid q_t = q\right) \\ h_{r,k}(q) &= \sum_{j=0}^{4-k} a^j \kappa_{j+k} \frac{\partial^j}{\partial q^j}\left(f(q) g_r(q)\right), \ k \leq 4 \\ h_{r,k}^*(q) &= \sum_{j=0}^{4-k} a^j \kappa_{j+k} \frac{\partial^j}{\partial q^j}\left(f(q) g_r^*(q)\right), \ k \leq 4. \end{aligned} \tag{C.7}$$

Note that we have implicitly assumed that $g_r(q)$ and $f(q)$ have four continuous derivatives. Also, without loss of generality, we assume $\gamma_0 = 0$ and x_{t2} is a scalar to ease notation.

Lemma 2. *Under* **K1**, **K2** *and* **K4**, *we have that for integers* $0 \leq \ell, r \leq 4$,

$$\begin{aligned} \frac{1}{na^{1+\ell}} \sum_{t=1}^{n} \varepsilon_t^2 x_{t2}^r q_t^\ell K\left(\frac{q_t - \widehat{\gamma}}{a}\right) - h_{r,\ell}(0) &= o_p(1) \\ \frac{1}{na^{1+\ell}} \sum_{t=1}^{n} x_{t2}^r q_t^\ell K\left(\frac{q_t - \widehat{\gamma}}{a}\right) - h_{r,\ell}^*(0) &= o_p(1). \end{aligned} \tag{C.8}$$

Proof. First, observe that we are using the normalization $\left(na^{1+\ell}\right)^{-1}$ instead of the standard $(na)^{-1}$. This is due to the factor q_t^ℓ. We shall consider only the first equality in $(C.8)$, the second one being similarly handled. Now abbreviating $K_t(\gamma) = K\left(\frac{q_t - \gamma}{a}\right)$, we have that standard kernel arguments imply

$$\frac{1}{na^{1+\ell}} \sum_{t=1}^{n} \varepsilon_t^2 x_{t2}^r q_t^\ell K_t(0) - h_{r,\ell}(0) = O_p\left((na)^{-1/2}\right) + o\left(a^{4-\ell}\right).$$

So, to complete the proof of the lemma, it suffices to show that

$$\frac{1}{na^{1+\ell}} \sum_{t=1}^{n} \varepsilon_t^2 x_{t2}^r q_t^\ell \left\{K_t(\widehat{\gamma}) - K_t(0)\right\} = o_p(1). \tag{C.9}$$

Proposition 1 implies that there exists C such that $\Pr\left\{|\widehat{\gamma}| > Cn^{-1/3}\right\} \leq \eta$, for any $\eta > 0$. So, we only need to show that $(C.9)$ holds true when $|\widehat{\gamma}| \leq Cn^{-1/3}$. In that case, we have that the left side of $(C.9)$ is bounded by

$$\begin{aligned}
& \sup_{|\gamma|\leq Cn^{-1/3}} \left| \frac{1}{na^{1+\ell}} \sum_{t=1}^{n} \varepsilon_t^2 x_{t2}^r q_t^\ell \left\{ K_t(\gamma) - K_t(0) \right\} \right| \\
\text{(C.10)} \quad \leq & \sup_{|\gamma|\leq Cn^{-1/3}} \left| \frac{1}{na^{1+\ell}} \sum_{t=1}^{n} \varepsilon_t^2 x_{t2}^r q_t^\ell \left\{ K_t(\gamma) - K_t(0) \right\} \mathbf{1}\left(|q_t| < a^{1/2}\right) \right| \\
& + \sup_{|\gamma|\leq Cn^{-1/3}} \left| \frac{1}{na^{1+\ell}} \sum_{t=1}^{n} \varepsilon_t^2 x_{t2}^r q_t^\ell \left\{ K_t(\gamma) - K_t(0) \right\} \mathbf{1}\left(|q_t| \geq a^{1/2}\right) \right|.
\end{aligned}$$

The expectation of second term on the right of $(C.10)$ is bounded by

$$\begin{aligned}
& \frac{C_1}{na} \sum_{t=1}^{n} E\left(\varepsilon_t^2 \left|x_{t2}\right|^r \left|\frac{q_t}{a}\right|^\ell K\left(\frac{q_t}{a}\right) \mathbf{1}\left(|q_t| \geq a^{1/2}\right) \right) \\
\leq \; & \frac{C_1}{a} \int_q \left|\frac{q}{a}\right|^\ell g_r(q) f(q) K\left(\frac{q}{a}\right) \mathbf{1}\left(|q_t| \geq a^{1/2}\right) dq \\
= \; & C_1 \int_{|q|\geq a^{-1/2}} |q|^\ell g_r(aq) f(aq) K(q)\, dq \\
= \; & o\left(a^{2-\ell/4}\right),
\end{aligned}$$

because by **K1**, $\kappa_\ell < C_1$, for $\ell \leq 4$.

For some $0 < \psi < 1$, the first term on the right of $(C.10)$ is bounded by

$$\begin{aligned}
& \frac{C}{n^{1/3}} \sup_{|\gamma|\leq Cn^{-1/3}} \left| \frac{1}{na^2} \sum_{t=1}^{n} \varepsilon_t^2 \left|x_{t2}\right|^r \left|\frac{q_t}{a}\right|^\ell K'\left(\frac{q_t - \psi\gamma}{a}\right) \mathbf{1}\left(|q_t| < a^{1/2}\right) \right| \\
\text{(C.11)} \leq \; & \frac{C}{n^{1/3}} \left| \frac{1}{na^2} \sum_{t=1}^{n} \varepsilon_t^2 \left|x_{t2}\right|^r \left|\frac{q_t}{a}\right|^\ell K'\left(\frac{q_t}{a}\right) \mathbf{1}\left(a^{3/2} < |q_t| < a^{1/2}\right) \right| \\
& + \frac{C}{n^{1/3}} \sup_{|\gamma|\leq Cn^{-1/3}} \left| \frac{1}{na^2} \sum_{t=1}^{n} \varepsilon_t^2 \left|x_{t2}\right|^r \left|\frac{q_t}{a}\right|^\ell K'\left(\frac{q_t - \phi\gamma}{a}\right) \mathbf{1}\left(|q_t| < a^{3/2}\right) \right|
\end{aligned}$$

because **K4** implies that $\gamma = o(a)$ when $|\gamma| \leq Cn^{-1/3}$, and hence if $a^{3/2} < |q_t| < a^{1/2}$ we have $\left| K'\left(\frac{q_t - \phi\gamma}{a}\right) / K'\left(\frac{q_t}{a}\right)\right| \leq C_1$ by **K2**. But, it is well known that the first moment of the first term on the right of $(C.11)$ is bounded, whereas that of the second term on the right is also bounded because $E\left|\frac{q_t}{a}\right|^{\ell} \mathbf{1}\left(|q_t| < a^{3/2}\right) < a^{(\ell+3)/2}$ and

$$\left| K'\left(\frac{q_t - \phi\gamma}{a}\right) - K_t'(0)\right| \mathbf{1}\left(|q_t| < a^{3/2}\right) \leq Ca^{1/2}. \tag{C.12}$$

So, the expectation of the first term on the right of $(C.10)$ is $O\left(n^{-1/3}\right)$. This concludes the proof of the lemma. ∎

Lemma 3. *Under* **K1** $-$ **K4***, we have that for integers* $0 \leq r, \ell \leq 4$*,*

$$\frac{1}{na}\sum_{t=1}^{n} x_{t2}^{r} q_t^{\ell} K_t(\widehat{\gamma})\,\varepsilon_t = o_p\left(a^{\ell} n^{1/2}\right). \tag{C.13}$$

Proof. To simplify the notation, we assume that $r = 0$. The left side of $(C.13)$ is

$$\frac{1}{na}\sum_{t=1}^{n} q_t^{\ell}\left\{K_t(\widehat{\gamma}) - K_t(0)\right\}\varepsilon_t + \frac{1}{na}\sum_{t=1}^{n} q_t^{\ell} K_t(0)\,\varepsilon_t.$$

The second term is easily shown to be $O_p\left(n^{-1/2}a^{\ell-1/2}\right)$. Next the first term of the last displayed expression is

$$\begin{aligned} &\frac{1}{na}\sum_{t=1}^{n} q_t^{\ell}\left\{K_t(\widehat{\gamma}) - K_t(0)\right\}\varepsilon_t \mathbf{1}\left(|q_t| < a^{\zeta}\right) \\ &+\frac{1}{na}\sum_{t=1}^{n} q_t^{\ell}\left\{K_t(\widehat{\gamma}) - K_t(0)\right\}\varepsilon_t \mathbf{1}\left(|q_t| \geq a^{\zeta}\right), \end{aligned} \tag{C.14}$$

where $\zeta = 1 - 2/\ell$, if $\ell > 2$, and $\zeta < 1$ if $\ell \leq 2$. The second term of $(C.14)$ is

$$a^{\ell}\frac{1}{na}\sum_{t=1}^{n}\left(\frac{q_t}{a}\right)^{\ell}\left\{K_t(\widehat{\gamma}) - K_t(0)\right\}\varepsilon_t \mathbf{1}\left(|q_t| \geq a^{\zeta}\right),$$

whose first absolute moment is bounded by

$$a^{\ell-1}\int_{|q|\geq a^{\zeta}}\left(\frac{q}{a}\right)^{\ell}K\left(\frac{q}{a}\right)f_q\left(q\right)dq\leq C_1a^{\ell}\int_{|q|\geq a^{\zeta-1}}q^{\ell}K\left(q\right)f_q\left(aq\right)dq=o\left(a^{\ell}\right)$$

because by **K1**, $\kappa_4<\infty$. So to complete the proof we need to examine the first term of $(C.14)$, which using the characteristic function of the kernel function is

$$\int\phi\left(av\right)\left(e^{iv\widehat{\gamma}}-1\right)\left\{\frac{1}{n}\sum_{t=1}^{n}q_t^{\ell}\varepsilon_te^{ivq_t}\mathbf{1}\left(|q_t|<a^{\zeta}\right)\right\}dv.$$

But its clear that the last displayed expression is bounded by

$$\begin{aligned}&\widehat{\gamma}\int v\left|\phi\left(av\right)\right|\left|\frac{1}{n}\sum_{t=1}^{n}q_t^{\ell}\varepsilon_te^{ivq_t}\mathbf{1}\left(|q_t|<a^{\zeta}\right)\right|dv=O_p\left(a^{\ell\zeta}n^{-1/2}\widehat{\gamma}\right)\int v\left|\phi\left(av\right)\right|dv\\ =\ &O_p\left(a^{\ell}\left(na^3\right)^{-4/3}n^{1/2}\right)\end{aligned}$$

using that $\zeta=1-2/\ell$, if $\ell\geq2$ and $\zeta<1$ when $0\leq\ell<2$, $\widehat{\gamma}=O_p\left(n^{-1/3}\right)$ and **K4** . This concludes the proof of the lemma. ∎

We now extend the maximal inequalities in Lemma 1 to its bootstrap analogues. Define $J_n^*\left(\gamma,\gamma'\right)$ and $J_{1n}^*\left(\gamma,\gamma'\right)$ by replacing ε_t in J_n and J_{1n} with $\widehat{e}_t\eta_t$, that is

$$\begin{aligned}J_n^*\left(\gamma,\gamma'\right)&=\frac{1}{n^{1/2}}\sum_{t=1}^{n}x_t\mathbf{1}_t\left(\gamma,\gamma'\right)\widehat{e}_t\eta_t\\ J_{1n}^*\left(\gamma,\gamma'\right)&=\frac{1}{n^{1/2}}\sum_{t=1}^{n}\left|q_t-\gamma\right|^{j}\mathbf{1}_t\left(\gamma;\gamma'\right)\widehat{e}_t\eta_t,\end{aligned}$$

and recall that H_n denotes a sequence of positive $O_p\left(1\right)$ random variables.

Lemma 4. *Under Assumption Z, we have that for all $\epsilon, \varsigma > 0$, there exists $\zeta > 0$ such that*

$$\Pr{}^{*}\left\{\sup_{\gamma'<\gamma<\gamma'+\epsilon}\left|J_n^*\left(\gamma',\gamma\right)\right|>\epsilon\right\} \le \zeta\varsigma H_n, \tag{C.15}$$

$$\Pr{}^{*}\left\{\sup_{\gamma'<\gamma<\gamma'+\epsilon}\left|J_{1n}^*\left(\gamma',\gamma\right)\right|>C\epsilon^{1/2}\left(\epsilon+\left|\gamma_0-\gamma'\right|\right)^{j}\right\} \le \zeta\varsigma H_n. \tag{C.16}$$

Proof. We shall assume for notational simplicity that $\gamma_0 < \widehat{\gamma}$, and that $\gamma_j = \gamma_1 + \frac{\zeta}{m}j$ and $n\zeta/2 < m < n\zeta$, as n can be chosen such that $n\zeta > 1$. By definition,

$$\begin{aligned} J_n^*\left(\gamma_k,\gamma_j\right) &= \frac{1}{n^{1/2}}\sum_{t=1}^{n} x_t\varepsilon_t\mathbf{1}_t\left(\gamma_j;\gamma_k\right)\eta_t \\ &+\frac{1}{n^{1/2}}\sum_{t=1}^{n} x_t x_t'\mathbf{1}_t\left(\gamma_j;\gamma_k\right)\eta_t\left(\widehat{\beta}-\beta\right) \\ &+\frac{1}{n^{1/2}}\sum_{t=1}^{n} x_t x_t'\mathbf{1}_t\left(\gamma_0\right)\mathbf{1}_t\left(\gamma_j;\gamma_k\right)\eta_t\left(\widehat{\delta}-\delta\right) \\ &+\frac{1}{n^{1/2}}\sum_{t=1}^{n} x_t x_t'\mathbf{1}_t\left(\gamma_0;\widehat{\gamma}\right)\mathbf{1}_t\left(\gamma_j;\gamma_k\right)\eta_t\widehat{\delta}. \end{aligned}$$

Now by standard inequalities and that $\eta_t \sim iid\left(0,1\right)$ with a finite fourth moments, the fourth (bootstrap) moment of the right side of last displayed equation is bounded by

$$\begin{aligned} &\left|\frac{1}{n}\sum_{t=1}^{n}\left\|x_t\right\|^2\varepsilon_t^2\mathbf{1}_t\left(\gamma_j;\gamma_k\right)\right|^2+\left\|\widehat{\beta}-\beta\right\|^4\left|\frac{1}{n}\sum_{t=1}^{n}\left\|x_t\right\|^4\mathbf{1}_t\left(\gamma_j;\gamma_k\right)\right|^2 \\ &+\left\|\widehat{\delta}-\delta\right\|^4\left|\frac{1}{n}\sum_{t=1}^{n}\left\|x_t\right\|^4\mathbf{1}_t\left(\gamma_j;\gamma_k\right)\mathbf{1}_t\left(\gamma_0\right)\right|^2 \\ &+\left\|\widehat{\delta}\right\|^4\left|\frac{1}{n}\sum_{t=1}^{n}\left\|x_t\right\|^4\mathbf{1}_t\left(\gamma_j;\gamma_k\right)\mathbf{1}_t\left(\gamma_0;\widehat{\gamma}\right)\right|^2. \end{aligned} \tag{C.17}$$

Because for fixed $\zeta > 0$, there exists n_0 such that for $n > n_0$, $Cn^{-1} < \zeta$, the expectation of the first term of $(C.17)$ is bounded by

$$C\left[(k-j)\zeta_m + \left(\frac{(k-j)\zeta_m}{n}\right)^{1/2}\right]^2 \leq C(k-j)^2\zeta_m^2,$$

arguing similarly as in Hansen's (2000) Lemma A.3 and $\zeta_m = \zeta/m$.

Next, recalling that $\widehat{\gamma} = \gamma_0 + D/n^{1/3}$, because $\mathbf{1}(\gamma_j < q_t < \gamma_k)\,\mathbf{1}(\gamma_0 < q_t < \widehat{\gamma}) \leq \mathbf{1}(\gamma_j < q_t < \gamma_k)$, the expectation of the fourth term of $(C.17)$ is bounded by

$$\begin{aligned} & \left|E\left\{\|x_t\|^4\,\mathbf{1}_t(\gamma_j;\gamma_k)\right\}\right|^2 + \left|\frac{1}{n}\sum_{t=1}^{n}\left\{\|x_t\|^4\,\mathbf{1}_t(\gamma_j;\gamma_k) - E\left\{\|x_t\|^4\,\mathbf{1}_t(\gamma_j;\gamma_k)\right\}\right\}\right|^2 \\ \leq\ & C(k-j)^2\zeta_m^2. \end{aligned}$$

Finally, the second and third terms of $(C.17)$ are

$$H_n\frac{1}{n^3}\sum_{t=1}^{n}E\left(\|x_t\|^8\,\mathbf{1}_t(\gamma_j;\gamma_k)\right) = H_n(k-j)^2\zeta_m^2.$$

From here we now conclude that $(C.15)$ holds true, so is the lemma proceeding as in Hansen's (2000) Lemma A.3 and in particular his expressions $(20)-(22)$ because if a sequence of random variables has finite first moments, it implies that it is $O_p(1)$. The proof of $(C.16)$ proceeds similarly and thus omitted. ∎

Remark 6. *One of the consequences of the previous lemma is that*

$$nE^*\sup_{g_1<g<g_1+\epsilon}\left|J_n^*(\gamma_0+g/r_n) - J_n^*(\gamma_0+g_1/r_n)\right| = (\epsilon+g_1)\,\epsilon^{1/2}H_n,$$

which can be made small by choosing small ϵ and $r_n \to \infty$.

Lemma 5. *Under* **K1**, **K2** *and* **K4** , *we have that for integers* $0 \leq \ell, r \leq 4$,

$$\frac{1}{na^{1+\ell}} \sum_{t=1}^{n} \varepsilon_t^{*2} x_{t2}^r q_t^\ell K\left(\frac{q_t - \widehat{\gamma}^*}{a}\right) - h_{r,\ell}(0) = o_{p^*}(1)$$

$$\frac{1}{na^{1+\ell}} \sum_{t=1}^{n} x_{t2}^r q_t^\ell K\left(\frac{q_t - \widehat{\gamma}^*}{a}\right) - h^*_{r,\ell}(0) = o_{p^*}(1). \tag{C.18}$$

Proof. We shall consider only the first equality in $(C.18)$, the second one being similarly handled. Now standard kernel arguments imply

$$\frac{1}{na^{1+\ell}} \sum_{t=1}^{n} \varepsilon_t^{*2} x_{t2}^r q_t^\ell K_t(0) - h_{r,\ell}(0) = O_{p^*}\left((na)^{-1/2}\right) + o_p\left(a^{4-\ell}\right).$$

So, to complete the proof of the lemma, it suffices to show that

$$\frac{1}{na^{1+\ell}} \sum_{t=1}^{n} \varepsilon_t^{*2} x_{t2}^r q_t^\ell \left\{K_t(\widehat{\gamma}^*) - K_t(0)\right\} = o_{p^*}(1). \tag{C.19}$$

Proposition 4 implies that there exists $C > 0$ such that $\Pr^*\left\{|\widehat{\gamma}^*| > Cn^{-1/3}\right\} \leq H_n$. So, we only need to show that $(C.9)$ holds true when $|\widehat{\gamma}^*| \leq Cn^{-1/3}$, so that we have that the left side of $(C.19)$ is bounded by

$$\sup_{|\gamma| \leq Cn^{-1/3}} \left| \frac{1}{na^{1+\ell}} \sum_{t=1}^{n} \varepsilon_t^{*2} x_{t2}^r q_t^\ell \left\{K_t(\gamma) - K_t(0)\right\} \right|$$

$$\leq \sup_{|\gamma| \leq Cn^{-1/3}} \left| \frac{1}{na^{1+\ell}} \sum_{t=1}^{n} \varepsilon_t^{*2} x_{t2}^r q_t^\ell \left\{K_t(\gamma) - K_t(0)\right\} \mathbf{1}\left(|q_t| < a^{1/2}\right) \right| \tag{C.20}$$

$$+ \sup_{|\gamma| \leq Cn^{-1/3}} \left| \frac{1}{na^{1+\ell}} \sum_{t=1}^{n} \varepsilon_t^{*2} x_{t2}^r q_t^\ell \left\{K_t(\gamma) - K_t(0)\right\} \mathbf{1}\left(|q_t| \geq a^{1/2}\right) \right|.$$

The expectation of second term on the right of $(C.20)$ is bounded by

$$\begin{aligned}
& \frac{C_1}{na}\sum_{t=1}^{n} E^*\left(\varepsilon_t^{*2}\left|x_{t2}\right|^r\left|\frac{q_t}{a}\right|^{\ell} K\left(\frac{q_t}{a}\right)\mathbf{1}\left(\left|q_t\right|\geq a^{1/2}\right)\right) \\
= & \frac{C_1}{na}\sum_{t=1}^{n}\left|x_{t2}\right|^r\left|\frac{q_t}{a}\right|^{\ell} K\left(\frac{q_t}{a}\right)\mathbf{1}\left(\left|q_t\right|\geq a^{1/2}\right)\frac{1}{n}\sum_{s=1}^{n}\widehat{\varepsilon}_t^2 \\
= & \frac{C_1}{na}\sum_{t=1}^{n}\left|x_{t2}\right|^r\left|\frac{q_t}{a}\right|^{\ell} K\left(\frac{q_t}{a}\right)\mathbf{1}\left(\left|q_t\right|\geq a^{1/2}\right) H_n,
\end{aligned}$$

where C_1 denotes a generic positive finite constant. Now,

$$E\frac{1}{na}\sum_{t=1}^{n}\left|x_{t2}\right|^r\left|\frac{q_t}{a}\right|^{\ell} K\left(\frac{q_t}{a}\right)\mathbf{1}\left(\left|q_t\right|\geq a^{1/2}\right) = o\left(a^{2-\ell/4}\right)$$

proceeding as we did in Lemma 2. So, we conclude that right of $(C.20)$ is $o\left(a^{2-\ell/4}\right)H_n$.

For some $0<\psi<1$, the first term on the right of $(C.20)$ is bounded by

$$\begin{aligned}
& \frac{C_1}{n^{1/3}}\sup_{|\gamma|\leq Cn^{-1/3}}\left|\frac{1}{na^2}\sum_{t=1}^{n}\varepsilon_t^{*2}\left|x_{t2}\right|^r\left|\frac{q_t}{a}\right|^{\ell} K'\left(\frac{q_t-\psi\gamma}{a}\right)\mathbf{1}\left(\left|q_t\right|<a^{1/2}\right)\right| \\
\text{(C.21)} \leq & \frac{C_1}{n^{1/3}}\left|\frac{1}{na^2}\sum_{t=1}^{n}\varepsilon_t^{*2}\left|x_{t2}\right|^r\left|\frac{q_t}{a}\right|^{\ell} K'\left(\frac{q_t}{a}\right)\mathbf{1}\left(a^{3/2}<\left|q_t\right|<a^{1/2}\right)\right| \\
& +\frac{C_1}{n^{1/3}}\sup_{|\gamma|\leq Cn^{-1/3}}\left|\frac{1}{na^2}\sum_{t=1}^{n}\varepsilon_t^{*2}\left|x_{t2}\right|^r\left|\frac{q_t}{a}\right|^{\ell} K'\left(\frac{q_t-\phi\gamma}{a}\right)\mathbf{1}\left(\left|q_t\right|<a^{3/2}\right)\right|
\end{aligned}$$

because **K4** implies that $\gamma=o(a)$ when $|\gamma|\leq Cn^{-1/3}$, and hence $\left|K'\left(\frac{q_t-\phi\gamma}{a}\right)/K'\left(\frac{q_t}{a}\right)\right|\leq C_1$ by **K2** if $a^{3/2}<\left|q_t\right|<a^{1/2}$. But, it is well known that the first moment of the first term on the right of $(C.21)$ is bounded, whereas that of the second term on the right is also bounded because $E\left|\frac{q_t}{a}\right|^{\ell}\mathbf{1}\left(\left|q_t\right|<a^{3/2}\right)<a^{(\ell+3)/2}$ and $(C.12)$. So, the expectation of the first term on the right of $(C.20)$ is $O_p\left(n^{-1/3}\right)$. This concludes the proof of the lemma. ∎

Lemma 6. *Under* $\mathbf{K1}-\mathbf{K4}$*, we have that for integers* $0 \leq r, \ell \leq 4$,

$$\frac{1}{na}\sum_{t=1}^{n} x_{t2}^{r} q_t^{\ell} K_t\left(\widehat{\gamma}^*\right)\varepsilon_t^* = o_{p^*}\left(a^{\ell} n^{1/2}\right). \tag{C.22}$$

Proof. To simplify the notation, we assume that $r=0$. The left side of $(C.22)$ is

$$\frac{1}{na}\sum_{t=1}^{n} q_t^{\ell}\left\{K_t\left(\widehat{\gamma}^*\right)-K_t\left(0\right)\right\}\varepsilon_t^* + \frac{1}{na}\sum_{t=1}^{n} q_t^{\ell} K_t\left(0\right)\varepsilon_t^*.$$

The second term is easily shown to be $O_{p^*}\left(n^{-1/2}a^{\ell-1/2}\right)$, whereas the first term is

$$\begin{aligned}\frac{1}{na}\sum_{t=1}^{n} q_t^{\ell}\left\{K_t\left(\widehat{\gamma}^*\right)-K_t\left(0\right)\right\}\varepsilon_t^* \mathbf{1}\left(\left|q_t\right| < a^{\zeta}\right) \\ +\frac{1}{na}\sum_{t=1}^{n} q_t^{\ell}\left\{K_t\left(\widehat{\gamma}^*\right)-K_t\left(0\right)\right\}\varepsilon_t^* \mathbf{1}\left(\left|q_t\right| \geq a^{\zeta}\right),\end{aligned} \tag{C.23}$$

where $\zeta = 1-2/\ell$ if $\ell > 2$ and $\zeta < 1$ if $\ell \leq 2$. The second term of $(C.23)$ is

$$a^{\ell}\frac{1}{na}\sum_{t=1}^{n}\left(\frac{q_t}{a}\right)^{\ell}\left\{K_t\left(\widehat{\gamma}^*\right)-K_t\left(0\right)\right\}\varepsilon_t^* \mathbf{1}\left(\left|q_t\right| \geq a^{\zeta}\right),$$

whose first absolute bootstrap moment is

$$\begin{aligned}&a^{\ell}\frac{1}{na}\sum_{t=1}^{n}\left|\frac{q_t}{a}\right|^{\ell}\left|K_t\left(\widehat{\gamma}^*\right)-K_t\left(0\right)\right| \mathbf{1}\left(\left|q_t\right| \geq a^{\zeta}\right)\frac{1}{n}\sum_{s=1}^{n}\left|\widehat{\varepsilon}_s\right| \\ &a^{\ell}\frac{1}{na}\sum_{t=1}^{n}\left|\frac{q_t}{a}\right|^{\ell}\left|K_t\left(\widehat{\gamma}^*\right)-K_t\left(0\right)\right| \mathbf{1}\left(\left|q_t\right| \geq a^{\zeta}\right) H_n.\end{aligned}$$

Now, proceed as in Lemma 5 to conclude that second term of $(C.23)$ is $O_{p^*}\left(a^{\ell}\right)$. So, to complete the proof we need to examine the first term of $(C.23)$ which, as we did with the first term of $(C.14)$, is

$$\int \phi\left(av\right)\left(e^{iv\widehat{\gamma}^*}-1\right)\left\{\frac{1}{n}\sum_{t=1}^{n} q_t^{\ell}\varepsilon_t^* e^{ivq_t}\mathbf{1}\left(\left|q_t\right| < a^{\zeta}\right)\right\}dv.$$

But it is clear that the last displayed expression is bounded by

$$\widehat{\gamma}^{*}\int v\left|\phi\left(av\right)\right|\left|\frac{1}{n}\sum_{t=1}^{n}q_{t}^{\ell}\varepsilon_{t}^{*}e^{ivq_{t}}\mathbf{1}\left(\left|q_{t}\right|<a^{\zeta}\right)\right|dv=O_{p^{*}}\left(a^{\ell\zeta}n^{-1/2}\widehat{\gamma}^{*}\right)\int v\left|\phi\left(av\right)\right|dv$$

$$= O_{p^{*}}\left(a^{\ell}\left(na^{3}\right)^{-4/3}n^{1/2}\right)$$

using **K4** and that $\zeta=1-2/\ell$ if $\ell\geq 2$ and $\zeta<1$ when $0\leq\ell<2$, $\widehat{\gamma}^{*}=O_{p^{*}}\left(n^{-1/3}\right)$ and that by standard arguments, it yields

$$E^{*}\left|\frac{1}{n}\sum_{t=1}^{n}q_{t}^{\ell}\varepsilon_{t}^{*}e^{ivq_{t}}\mathbf{1}\left(\left|q_{t}\right|<a^{\zeta}\right)\right|^{2}=O_{p}\left(a^{2\ell\zeta}n^{-1}\right).$$

This concludes the proof of the lemma. ∎

Appendix D. Further Monte Carlo results

Table 3 presents Monte Carlo size and coverage probabilities for the case $\varphi=1/8$.

Appendix E. Additional material for empirical examples.

E.1. **Figures for empirical example of Section 6.2.** The three figures demonstrate how we obtained the test inversion BCI for the three cities.

chicago.pdf, la.pdf, phily.pdf

E.2. **Comparison of tests of continuity for empirical example of Section 6.2.** In this subsection, we present a small Monte Carlo study that examine behavior of various tests of continuity.

In Section 6.2, it was noted that the test of continuity proposed in Section 3.4 of the present paper had rejected the null of continuity at 5 % significance level. Now, we present Gonzalo and Wolf's (2005) test of continuity based on

Table 3. Monte Carlo size of test $H_0 : \gamma = \gamma_0$ and coverage probability of confidence intervals of γ_0, $\varphi = 1/8$, models A-D

			size				C.P.			C.P.		
		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
A	Asympt	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
	RB	0.01	0.005	0.009	0.008	0.9	0.707	0.783	0.856	0.751	0.798	0.848
		0.05	0.062	0.066	0.06	0.95	0.794	0.863	0.914	0.826	0.882	0.91
		0.1	0.142	0.126	0.119	0.99	0.901	0.94	0.974	0.93	0.957	0.969
	NB	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
B	Asympt	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
	RB	0.01	0.055	0.025	0.023	0.9	0.624	0.684	0.739	0.731	0.788	0.822
		0.05	0.155	0.1	0.082	0.95	0.724	0.777	0.81	0.809	0.869	0.892
		0.1	0.308	0.184	0.145	0.99	0.842	0.884	0.897	0.907	0.941	0.955
	NB	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
C	Asympt	0.01	0.572	0.537	0.466	0.9	0.256	0.303	0.35	0.406	0.445	0.459
		0.05	0.678	0.628	0.552	0.95	0.316	0.352	0.405	0.468	0.502	0.537
		0.1	0.741	0.672	0.614	0.99	0.411	0.451	0.517	0.593	0.622	0.659
	RB	0.01	0.102	0.04	0.033	0.9	0.293	0.363	0.448	0.475	0.539	0.552
		0.05	0.255	0.175	0.16	0.95	0.371	0.442	0.532	0.561	0.617	0.674
		0.1	0.398	0.287	0.256	0.99	0.489	0.573	0.657	0.693	0.778	0.809
	NB	0.01	0.046	0.025	0.007	0.9	0.455	0.564	0.643	0.783	0.826	0.845
		0.05	0.116	0.081	0.061	0.95	0.546	0.644	0.726	0.855	0.891	0.9
		0.1	0.195	0.146	0.123	0.99	0.658	0.773	0.837	0.919	0.961	0.962
D	Asympt	0.01	0.01	0.008	0.006	0.9	0.88	0.889	0.898	0.876	0.894	0.91
		0.05	0.043	0.04	0.041	0.95	0.93	0.942	0.946	0.922	0.945	0.956
		0.1	0.088	0.093	0.082	0.99	0.972	0.988	0.987	0.975	0.984	0.991
	RB	0.01	0.007	0.007	0.005	0.9	0.789	0.835	0.845	0.776	0.841	0.874
		0.05	0.042	0.061	0.05	0.95	0.868	0.891	0.914	0.846	0.9	0.928
		0.1	0.091	0.122	0.118	0.99	0.951	0.98	0.976	0.929	0.969	0.983
	NB	0.01	0.007	0.013	0.007	0.9	0.816	0.853	0.864	0.813	0.853	0.881
		0.05	0.053	0.059	0.05	0.95	0.887	0.912	0.929	0.879	0.924	0.938
		0.1	0.099	0.129	0.116	0.99	0.961	0.978	0.983	0.951	0.974	0.986

Size results for test of $H_0 : \gamma = \gamma_0$ based on Hansen (2000)'s asymptotic distribution (Asympt), standard residual bootstrap (**RB**) and bootstrap of Section 2.2 (**NB**). Coverage probability (C.P.) results for γ_0 with asymptotic confidence interval based on Hansen (2000) (Asympt), percentile t confidence interval based on standard residual bootstrap (**RB**) and grid bootstrap confidence interval based on Section 2.3 (**NB**). Setup A: $q_t \neq x_t$, B: $q_t = x_t$ with a jump, C: $q_t = x_t$ with a kink, D: $q_t = t/n$. $\delta = \delta_n = \left(\sqrt{10}/4\right) n^{-1/8}$.

subsampling which uses the following test statistic,

$$h\left(\widehat{\delta},\widehat{\gamma}\right)=\left|\widehat{\delta}_1-\widehat{\delta}_3\widehat{\gamma}\right|+\left|\widehat{\delta}_3\right|.$$

Under the null hypothesis of continuity, $h(\delta_0,\gamma_0)=0$. For a block size b, one uses $(n-b+1)$ number of blocks of data in order to compute estimates of δ and γ. For the $a-th$ block, that is $[y_a,\cdots,y_{a+b-1}]$, denote the corresponding estimates as $\widehat{\delta}_{b,a}$ and $\widehat{\gamma}_{b,a}$. The subsampling p-value is then given by

$$PV_b^{(1)}=\frac{1}{n-b+1}\sum_{a=1}^{n-b+1}\mathbf{1}\left\{b^{1/2}h\left(\widehat{\delta}_{b,a},\widehat{\gamma}_{b,a}\right)\geq n^{1/2}h\left(\widehat{\delta},\widehat{\gamma}\right)\right\}.$$

For block sizes $b=30,40,50$ as used in Gonzalo and Wolf (2005), we find the subsampling p-values to be 0.1830, 0.0909 and 0.0977, respectively. So, we do not reject the null hypothesis of continuity at 5% significance level when using Gonzalo and Wolf's (2005) procedure. This is in contrast to the conclusion drawn from the test proposed in this paper.

One can also envisage to modify the subsampling algorithm to reflect the cube root n rate of convergence of $\widehat{\gamma}_{\mathsf{Unres}}$ under the null hypothesis of continuity. For that purpose, we first consider

$$PV_b^{(2)}=\frac{1}{n-b+1}\sum_{a=1}^{n-b+1}\mathbf{1}\left(b^{1/3}h\left(\widehat{\delta}_{b,a},\widehat{\gamma}_{b,a}\right)\geq n^{1/3}h\left(\widehat{\delta},\widehat{\gamma}\right)\right).$$

For block sizes $b=30,40,50$, the subsampling p-values are respectively 0.5294, 0.2797 and 0.2331, so we do not reject the null hypothesis of continuity. In addition, we also consider

$$PV_b^{(3)}=\frac{1}{n-b+1}\sum_{a=1}^{n-b+1}\mathbf{1}\left(b^{1/3}\left|\widehat{\gamma}_{\mathsf{Res},\ \mathrm{b,a}}-\widehat{\gamma}_{\mathsf{Unres},\ \mathrm{b,a}}\right|\geq n^{1/3}\left|\widehat{\gamma}_{\mathsf{Res}}-\widehat{\gamma}_{\mathsf{Unres}}\right|\right).$$

Table 4. Monte Carlo size and power of tests of continuity

	size			power		
s	0.1	0.05	0.01	0.1	0.05	0.01
Bootstrap	0.117	0.0536	0.015	0.4576	0.3006	0.0847
$PV\_b(1)$	0.0526	0.0268	0.0043	0.2231	0.1271	0.0393
$PV\_b(2)$	0.0086	0.0043	0	0.0527	0.0227	0.0041
$PV\_b(3)$	0.0805	0.0558	0.0365	0.0661	0.0372	0.0155

For block sizes $b = 30, 40, 50$, the subsampling p-values are 0.5621, 0.7063 and 0.8346. Therefore, for all three subsampling methods considered, we were not able to reject the null hypothesis of continuity at 5% significance level.

E.2.1. ***Monte Carlo for Continuity Test.*** Motivated by the contradictory results when testing for continuity, we decided to carry out a small Monte Carlo study to shed some light on the size performance in small samples of the different tests for continuity employed in the empirical example. Our Monte Carlo experiment mirrors the data set by simulating the estimated equation (6.2). For that purpose, we generate the innovations ε_t as independent draws from $N(0, 0.0938)$, where 0.0938 was the estimate of $Var(\varepsilon_t)$ from the data. Letting $y_{-99} = \varepsilon_{-99}$ and $y_{-98} = \varepsilon_{-98}$ and carrying on to simulate a time series of size 282 from (6.2), we then discard the first 100 time points and use the latter 182 observations. For 1000 iterations, 399 bootstraps per iteration were carried out and for subsampling, we used block size $b = 40$. We report Monte Carlo size of tests for the four tests of continuity for $10, 5$ and 1 percent significance levels in Table 4. f continuity for $10, 5$ and 1 percent significance levels in Table 4. We also investigate Monte Carlo power of test by generating data based on (6.1) and report results in Table 4.

Test based on (3.6) and **NB** of the current paper yielded the best size performance. Bootstrap test slightly over rejects while $PV_b^{(1)}$ and $PV_b^{(2)}$ under rejects, and $PV_b^{(3)}$ also under rejects at 10 percent nominal level. Importantly, the Monte

Carlo experiment may explain or shed some light into the reasons for the no rejection of the null hypothesis when using the test in Gonzalo and Wolf as the test tend to have size smaller than the nominal one. For power results, our test of continuity dramatically outperforms the three tests based on subsampling. The poor power performance of $PV_b^{(3)}$ is especially notable given its relatively satisfactory size performance.

References

[1] Bai, J. (1997), "Estimation of a change-point in multiple regression models" *The Review of Economics and Statistics*, 79, 551-563.

[2] Dovonon, P., and Renault, E.(2013), "Testing for common conditionally heteroskedastic factors", *Econometrica*, 81, 2561–2586.

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

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

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

[6] Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, New York: Academic Press.

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

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

[9] Peligrad, M. (1982), "Invariance principles for mixing sequences of random variables", *The Annals of Probability*, 10, 968-981.

London School of Economics, Houghton Street, WC2A 2AE, London, UK

E-mail address: m.seo@lse.ac.uk

URL: http://personal.lse.ac.uk/seo